agencia española protección datos

lab

aepd Laboratorio de Privacidad

Revista de Privacidad, Innovación y Tecnología

Revista de Privacidad, Innovación y Tecnología (RPIT)

nº 1.
Abril 2026

Equipo editorial

Presidencia:

Lorenzo Cotino Hueso. Presidente de la AEPD.

Francisco Pérez Bes. Adjunto a la Presidencia de la AEPD.

Directores:

Jorge Castellanos Claramunt. Vocal Coordinador de la Unidad de Apoyo y Relaciones Institucionales de la AEPD y Profesor Titular de Derecho Constitucional de la Universitat de València.

Luis de Salvador Carrasco. Director de la División de Innovación Tecnológica (DIT) de la AEPD.

Comité Científico:

Unai Aberasturi Gorriño. Director de la Autoridad Vasca de Protección de Datos

Marcos Almeida Cerreda. Designado a Consejo consultivo AEPD por la CRUE

Belén Andreu Martínez. Universidad de Murcia

Mónica Arenas Ramiro. Universidad de Alcalá

Senén Barro Ameneiro. Universidade de Santiago de Compostela

Meritxell Borràs Solé. Directora de la Autoridad Catalana de Protección de Datos

María Belén Cardona Rubert. Universitat de València

Ginevra Cerrina Feroni. Vice Presidente del Garante per la protezione dei dati personali / Università di Firenze

Juan Gustavo Corvalán. Universidad de Buenos Aires

Iñigo de Miguel Beriain. Universidad del País Vasco

Juan Manuel Fernández López. Exdirector AEPD

Diana Urania Galetta. Università degli Studi di Milano

Yolanda Gómez Sánchez. UNED

Jesús María González García. Presidente del Consejo de Transparencia y Protección de Datos de la Comunidad de Madrid

Jesús Jiménez López. Director del Consejo de Transparencia y Protección de Datos de Andalucía

Pablo Lucas Murillo de la Cueva. Universidad de Córdoba / Magistrado Tribunal Supremo

Alessandro Mantelero. Politecnico di Torino

Cristina Pauner Chulvi. Universitat Jaume I de Castellón

José Luis Piñar Mañas. Exdirector AEPD

Miguel Recio Gayo. Designado a Consejo consultivo AEPD por la APEPIA

Nelson Remolina Angarita. Universidad de los Andes

José Luis Rodríguez Álvarez. Exdirector AEPD

Jesús Rubí Navarrete. Écija. Exadjunto a la dirección de la AEPD

Carlos Alberto Saiz Peña. Designado a Consejo consultivo AEPD por ISMS

Julián Valero Torrijos. Universidad de Murcia

Consejo de Redacción:

Blanca Álvarez Yaque. Consejo de Transparencia y Protección de datos de Andalucía

Eduard Chaveli Donet. Head of Consulting Strategy, Govertis

Horacio Roberto Granero. Universidad Católica Argentina

Estrella Gutiérrez David. Universidad Complutense de Madrid

Joana Marí Cardona. Autoridad Catalana de Protección de Datos
Juan Antonio Martín Zubiaur. Autoridad Vasca de Protección de Datos
Domingos Soares Farinho. Universidad de Lisboa
María de Lourdes Zamudio Salinas. Universidad de Lima

Listado de evaluadores:

Ana María Aba Catoira. Universidade da Coruña
Miguel J. Agudo Zamora. Universidad de Córdoba
Arminda Mª Teresa Álamo Bolaño. Universidad de Las Palmas de Gran Canaria
Carlos Arce Jiménez. Universidad de Córdoba
Noel Armas Castilla. Universidad de Sevilla
David Aviñó Belenguer. Universitat de València
Anna Buchardó Parra. Universitat de València
L. Javier Cabeza Ramírez. Universidad de Córdoba
Agustí Cerrillo Martínez. UOC
Nuria Cuadrado Gamarra. Universidad Complutense de Madrid
Valentina Faggiani. Universidad de Granada
Ana Galdámez Morales. UNED
Pablo Gallego Rodríguez. Universidad de Córdoba
Mª Teresa García-Berrio Hernández. Universidad Complutense de Madrid
Joaquín Garrido Martín. Universidad de Sevilla
Ana Garriga Domínguez. Universidad de Vigo
Luis Miguel González de la Garza. UNED
Emilio Guichot Reina. Universidad de Sevilla
Ángel Guillén Pajuelo. Universitat de València
Mª Guadalupe Imormino De Haro. Universidad Autónoma de Coahuila
Ferdinando A. Insanguine Mingarro. Università degli Studi di Bari
José Miguel Iturmendi Rubia. CUNEF Universidad
Daniel Jove Villares. Universidad de Oviedo
Mónica Martínez López-Sáez. Universitat de València
Ciro Milione. Universidad de Córdoba
Bernardo Olivares Olivares. Universidad Complutense de Madrid
Adrián Palma Ortigosa. Universitat de València
David Parra Gómez. Universidad de Murcia
Nuria Reche Tello. Universidad Miguel Hernández
José María Ruz López. Universidad de Córdoba
Mohamed Saad Bentaouet. Universidad de Sevilla
Marco Emilio Sánchez Acevedo. Universidad Católica de Colombia
Sandra Sánchez Cañizares. Universidad de Córdoba
Javier Sierra Rodríguez. UNED
Pere Simón Castellano. UNIR
Alba Soriano Arnanz. Universitat de València
Bartolomé Torralbo Muñoz. Universidad de Córdoba
Raquel Valle Escolano. Universitat de València
Pilar Vargas Martínez. Universidad Complutense de Madrid
Gabriele Vestri. Universidad Pablo de Olavide

Índice

Presentación de la Revista de Privacidad, Innovación y Tecnología

Presidente de la AEPD

Lorenzo Cotino Hueso

Poco después de un año en la Presidencia de la Agencia Española de Protección de Datos es un privilegio presentar el primer número de la Revista PIT (Privacidad, Innovación y Tecnología) con enorme ilusión.

No es frecuente que una iniciativa editorial nazca con una vocación tan claramente estratégica como la que aquí se presenta. No estamos ante una iniciativa meramente editorial, sino ante la materialización de una línea estratégica del reciente Plan Estratégico 2025–2030 de la AEPD ("Innovación responsable y defensa de la dignidad en la era digital"). La transformación digital es un elemento estructural de nuestras sociedades. En este nuevo contexto, la privacidad y la protección de datos personales no son en modo alguno ámbitos especializados o marginales, sino condiciones de posibilidad del ejercicio de los derechos fundamentales en entornos tecnológicos complejos. Buena parte de las decisiones que afectan a las personas —desde el acceso a servicios hasta la configuración de oportunidades vitales— se encuentran hoy mediadas por sistemas tecnológicos basados en datos. Por ello, la articulación entre desarrollo tecnológico y garantía de derechos no es una opción política o técnica entre otras, sino una exigencia inherente al propio Estado de Derecho y, en consecuencia, un criterio ineludible de legitimidad de cualquier proceso de innovación tecnológica.

Es más, de forma particular hay que situar la revista en el Eje 2 del Plan ("Por una innovación tecnológica con garantías") y, en concreto, la Revista PIT se configura como una de las líneas del Laboratorio de Privacidad, orientado a la observación, análisis e investigación colaborativa de las tecnologías emergentes y sus implicaciones en la protección de datos.

El ecosistema académico y profesional cuenta, sin duda, con publicaciones de gran calidad. Aquí se pretende hacer confluir, de forma sistemática y con vocación de transferencia, la reflexión académica rigurosa, la experiencia práctica y la perspectiva institucional de una autoridad independiente. La Revista se sitúa deliberadamente en ese punto de intersección, no como una mera agregación de enfoques, sino como un espacio de integración y articulación orientado a generar conocimiento útil, contrastado y proyectable, especialmente en aquellos ámbitos en los que la distancia entre la regulación, la práctica y el desarrollo tecnológico resulta más acusada.

Este proyecto ve la luz poco después del primer año de Presidencia de la AEPD. Este primer número exterioriza y presenta el resultado de un intenso proceso de trabajo desde julio de 2025, en el que se ha definido con cuidado tanto su orientación como su arquitectura editorial. En este sentido, la Revista nace con vocación de permanencia más allá de una determinada presidencia, como una apuesta estratégica y estructural que pretende consolidarse en la Agencia como instrumento estable al servicio del interés general.

La posición de la AEPD resulta particularmente idónea para impulsar una iniciativa de estas características. Como autoridad independiente, la Agencia no solo ejerce funciones de supervisión y control, sino que dispone de una visión transversal derivada de su contacto directo con los problemas reales que plantea la aplicación del Derecho en entornos tecnológicos. Esta posición permite identificar tendencias, anticipar riesgos y detectar necesidades regulatorias con una inmediatez difícilmente alcanzable desde otros ámbitos. Al mismo tiempo, la Agencia ha asumido como propia una lógica de innovación institucional que no se limita a la reacción frente a los problemas, sino que busca anticiparlos y comprenderlos de forma sistemática. Afortunadamente, se cuenta desde hace años con una prestigiosa y envidiable División de Innovación Tecnológica.

Y es precisamente en este marco donde se inserta el Laboratorio de Privacidad de la Agencia, lanzado en otoño de 2025, del que la Revista PIT forma parte. El Laboratorio no se concibe como un espacio experimental en sentido débil ni como un conjunto aislado de iniciativas, sino como

una estructura flexible orientada a —si se me permite— *trenzar* de manera estable la interacción entre los planos institucional, académico y profesional. Su función no se agota en la generación de conocimiento, sino que se extiende a su contraste, validación y, de forma especialmente relevante, a su transferencia. La Revista constituye, en este contexto, uno de los instrumentos del Laboratorio, en la medida en que permite ordenar, sistematizar y proyectar ese conocimiento hacia la comunidad.

Desde esta misma lógica, la Revista se dirige a una comunidad de la privacidad: investigadores, profesionales, responsables públicos, expertos del ámbito tecnológico y, en un sentido más amplio, la comunidad interesada en la gobernanza de la innovación digital, están llamados a participar en la construcción de este espacio. La Revista aspira a configurarse como un lugar de encuentro en el que el conocimiento no solo se difunde, sino que se elabora colectivamente, mediante el contraste de enfoques y la integración de perspectivas diversas.

La Revista nace desde el diseño con los estándares de calidad propios de las publicaciones científicas de referencia. No se trata solo de la mera indexación, sino de la convicción de que solo desde el rigor metodológico puede contribuirse de manera efectiva al debate jurídico y técnico. La evaluación por pares ciega, la transparencia en los procesos editoriales, el uso de plataformas estandarizadas y el acceso abierto son garantías de calidad, independencia y credibilidad de los contenidos.

En cuanto a su objeto, la Revista PIT se configura como un espacio de reflexión, análisis y difusión del conocimiento en torno a la privacidad, la protección de datos y su interacción con las tecnologías digitales. Si bien presta una atención singular a las tecnologías emergentes —y, en particular, a la inteligencia artificial—. La relevancia de la inteligencia artificial deriva de la naturaleza de los retos que plantea: sistemas con capacidad de inferencia, con grados crecientes de autonomía y con un impacto potencialmente masivo sobre derechos e intereses protegidos. Ahora bien, estos desarrollos no pueden analizarse al margen del acervo conceptual y garantista propio del Derecho de protección de datos, que sin duda se proyecta a la IA. Categorías como la calidad de los datos, la minimización, la transparencia, la explicabilidad o la seguridad, así como instrumentos como las evaluaciones de impacto o las garantías desde el diseño, constituyen hoy referencias imprescindibles para abordar los riesgos asociados a estos sistemas. La interacción entre ambos ámbitos es estructural, en la medida en que contribuye a configurar un modelo de innovación orientado a la responsabilidad y a la efectiva garantía de los derechos fundamentales.

Este enfoque se traduce en una delimitación temática de la Revista que, sin pretender ser exhaustiva, abarca un conjunto amplio y articulado de ámbitos. Junto al análisis de la evolución normativa y jurisprudencial, la Revista aborda la formación y revisión de criterios regulatorios y las propuestas de desarrollo normativo, así como las buenas prácticas de cumplimiento y el desarrollo de tecnologías orientadas al propio cumplimiento normativo. A ello se suman tanto el estudio de los avances tecnológicos y de la investigación aplicada como el análisis de riesgos emergentes en ámbitos como la ciberseguridad, los sistemas de perfilado o la desinformación. Asimismo, la Revista incorpora dimensiones transversales como la educación y la cultura de la privacidad, la ética digital y los derechos humanos, junto con la perspectiva internacional y comparada, cada vez más necesaria en este ámbito. Finalmente, la atención a sectores específicos —como el sanitario, el financiero, el educativo o el laboral— y a cuestiones estructurales como la gobernanza de los datos y la economía digital permite situar los problemas en contextos materiales concretos, evitando formulaciones excesivamente abstractas o descontextualizadas y reforzando su proyección práctica.

Con todo, la Revista parte de una premisa que conviene subrayar: la regulación, siendo imprescindible, no resulta suficiente por sí sola para afrontar la complejidad de los fenómenos tecnológicos. Se requieren espacios intermedios —de carácter necesariamente híbrido— en los que se genere, contraste y transfiera conocimiento entre quienes diseñan, desarrollan, aplican y supervisan las tecnologías. La Revista PIT aspira a desempeñar esa función, facilitando un diálogo estructurado y exigente entre los distintos actores implicados, en un contexto en el que no siempre resulta sencillo alinear intereses, lenguajes y tiempos entre los distintos sectores.

Este planteamiento se alinea con los principios que orientan la actuación de la Agencia: defensa de la dignidad y los derechos en los entornos digitales, independencia, cooperación, excelencia técnica, innovación y apertura institucional. La Revista no se limita a enunciarlos, sino que aspira a materializarlos en su práctica editorial. La apertura se concreta en el acceso abierto y en la invitación a la participación; la cooperación, en la interacción interdisciplinar; la excelencia, en la exigencia metodológica; y la innovación, en su propia inserción en el Laboratorio como instrumento de comprensión y anticipación de los cambios tecnológicos.

No puede desconocerse, en todo caso, la dificultad inherente a la consolidación de una nueva revista en un ecosistema académico amplio, competitivo y, en ocasiones, fragmentado. Precisamente por ello, la identificación de un espacio propio —basado en la articulación entre academia, práctica

e institución— no solo es posible, sino necesaria. La Revista nace con la ambición de ocupar ese espacio, consciente de que su consolidación exigirá tiempo, calidad y compromiso continuado, y de que dicha consolidación no está asegurada, sino que deberá construirse progresivamente mediante la calidad y utilidad de sus contribuciones y, sobre todo, con el apoyo e interés de la comunidad investigadora y académica y de los expertos y profesionales.

En este punto, resulta obligado reconocer el papel de quienes han hecho posible que este primer número vea la luz. Contamos con la amable e ilusionada entrega de un comité científico y asesor y un comité de redacción de auténtico lujo. Ellas y ellos han aportado no solo su conocimiento y experiencia, sino también su compromiso con la construcción de un espacio de reflexión exigente. De manera muy especial, hay que agradecer la codirección de la revista con Jorge Castellanos, que ha asumido un esfuerzo particularmente intenso y continuado, determinante para la puesta en marcha y culminación de este primer número, afrontando con rigor y dedicación las exigencias propias de un proyecto de esta naturaleza. Asimismo, hay que destacar el papel de la División de Innovación Tecnológica de la Agencia y, especialmente, mencionar a Luis de Salvador, al frente de la misma.

Para quien lleva más de treinta años en el servicio público, en el ámbito académico e investigador, sorprende realmente apreciar que contemos en el sector público con auténticas joyas del conocimiento y su aplicación práctica, con una ilimitada vocación e ilusión por la innovación y la garantía de la persona. Como se ha adelantado, y hay que agradecer especialmente, la División de Innovación Tecnológica ha asumido desde el inicio la convicción de que este proyecto del Laboratorio no es accesorio, sino estratégico para la AEPD, y lo impulsan con una ilusión y exigencia nada habitual en el ámbito institucional. A todos ellos corresponde un reconocimiento expreso. Y, por supuesto, a los atrevidos autores y autoras de este primer número, que han arriesgado y puesto en esta Revista su magnífico conocimiento. Son los primeros de muchos que han de venir.

Presentación nº1 de la Revista de Privacidad, Innovación y Tecnología

El **monográfico** que abre este primer número de la *Revista de Privacidad, Innovación y Tecnología*, titulado ***IA, gobernanza algorítmica y riesgos para los derechos fundamentales***, se articula en torno a una cuestión central en la evolución contemporánea del Derecho: el impacto de la inteligencia artificial y de la gobernanza de datos en la configuración, ejercicio y garantía de los derechos fundamentales. Lejos de abordar la inteligencia artificial desde una perspectiva meramente tecnológica o instrumental, los trabajos aquí reunidos ponen de relieve su profunda dimensión jurídica, institucional y social, así como los riesgos estructurales que su despliegue comporta cuando se inserta en procesos de decisión que afectan directamente a las personas.

El eje vertebrador del monográfico reside en una idea compartida: la inteligencia artificial no es neutral. Los sistemas algorítmicos se nutren de datos, incorporan decisiones humanas y operan sobre realidades complejas, lo que implica que pueden reproducir —e incluso intensificar— sesgos, desigualdades y dinámicas de exclusión. Desde esta premisa, los tres trabajos seleccionados ofrecen una mirada complementaria que permite comprender tanto los riesgos asociados a la automatización como las posibles respuestas normativas y organizativas para su adecuada gobernanza.

El primer estudio, de **Diana-Urania Galetta**, se sitúa en el núcleo del problema al analizar la decisión administrativa algorítmica y su relación con los datos. A partir de la conocida máxima *garbage in, garbage out*, la autora muestra cómo la calidad, representatividad y configuración de los datos condicionan de forma decisiva el resultado de los procesos automatizados, pudiendo generar efectos discriminatorios especialmente intensos en ámbitos como la asignación de recursos públicos o prestaciones sociales. El trabajo destaca, asimismo, el papel esencial de la transparencia algorítmica, la motivación de las decisiones y el Derecho de la Unión Europea como pilares para garantizar la legalidad, el control y la legitimidad de la actuación administrativa automatizada.

Desde una perspectiva más amplia, el trabajo de **Ginevra Cerrina Feroni** aborda los sistemas de puntuación social y de reputación algorítmica como una de las manifestaciones más avanzadas —y preocupantes— de la gobernanza automatizada. A través del análisis del Sistema de Crédito Social chino, del *Toeslagenaffaire* neerlandés y de la experiencia italiana, la autora pone de manifiesto cómo las tecnologías de clasificación pueden convertirse en instrumentos de vigilancia, condicionamiento de oportunidades y restricción de derechos. Frente a esta realidad, el artículo propone un giro conceptual de especial relevancia: el tránsito desde la algocracia hacia una algorética basada en los derechos fundamentales, la proporcionalidad, la transparencia y el control humano.

Cierra el monográfico la aportación de **Eduard Chaveli Donet**, que introduce una perspectiva sistemática orientada a la gestión de los riesgos derivados de los sesgos en la inteligencia artificial. El autor ofrece una precisa delimitación conceptual del sesgo y analiza su posible aparición a lo

largo de todo el ciclo de vida de los sistemas de IA. Sobre esta base, el trabajo destaca el papel de la gobernanza de datos y de la gobernanza de la IA como instrumentos clave para articular mecanismos de control, supervisión y mejora continua. En particular, se subraya la relevancia del marco regulatorio europeo y de estándares como ISO 42001 como herramientas para generar confianza y garantizar un desarrollo responsable de estas tecnologías.

En su conjunto, los trabajos que integran este monográfico ofrecen una aproximación rigurosa y plural a uno de los grandes desafíos del Derecho contemporáneo. La inteligencia artificial, en cuanto tecnología de decisión, obliga a repensar categorías clásicas del Derecho administrativo, constitucional y de la protección de datos, al tiempo que exige reforzar los mecanismos de garantía frente a nuevas formas de poder algorítmico. Este monográfico pretende contribuir a ese debate, poniendo de relieve que la innovación tecnológica solo puede considerarse legítima cuando se desarrolla en coherencia con los principios del Estado de Derecho y con el respeto efectivo de los derechos fundamentales.

Junto al bloque monográfico, este número inaugural incorpora una sección de **Estudios de privacidad** que amplía y diversifica el análisis, abordando distintas manifestaciones de la relación entre tecnología, datos y derechos fundamentales desde perspectivas complementarias. Los trabajos reunidos en esta sección comparten una misma preocupación de fondo: comprender cómo las transformaciones tecnológicas reconfiguran los equilibrios jurídicos existentes y qué respuestas normativas, institucionales y organizativas resultan necesarias para garantizar una protección efectiva de la persona en entornos cada vez más mediados por datos y sistemas inteligentes.

Desde una perspectiva especialmente innovadora, el trabajo de **Áurea Rodríguez López** introduce el concepto de inteligencia artificial agéntica y plantea la necesidad de desarrollar nuevas capacidades —cognitivas, socio-técnicas y de gobernanza— para interactuar con sistemas capaces de planificar, decidir y ejecutar acciones de manera autónoma. El artículo desplaza el foco desde la mera eficiencia tecnológica hacia la construcción de una verdadera co-inteligencia humano-máquina, subrayando la importancia de la trazabilidad, la supervisión y la rendición de cuentas como elementos estructurales de un uso responsable de la IA.

En una línea distinta pero convergente, el estudio de **Thiago Guimaraes Moraes** examina los *regulatory sandboxes* como instrumentos de experimentación normativa en contextos de innovación tecnológica. A través de un análisis sistemático de la literatura y de experiencias comparadas, el autor identifica los elementos definitorios de estos entornos controlados —flexibilidad regulatoria, carácter temporal, aprendizaje institucional o gobernanza colaborativa— y pone de relieve su potencial para articular modelos de regulación más adaptativos sin renunciar a la protección de intereses públicos y derechos fundamentales.

La contribución de **Nelson Remolina Angarita** sitúa el debate en una frontera especialmente sensible: la de las neurotecnologías y los neurodatos. El trabajo advierte sobre los riesgos que el tratamiento de información cerebral puede suponer para la dignidad humana y la autonomía personal, planteando interrogantes de gran calado sobre el tipo de sociedad que se configura a partir de estas tecnologías. Desde esta perspectiva, se subraya la necesidad de anticipar respuestas jurídicas capaces de preservar los derechos humanos ante escenarios tecnológicos que desbordan los marcos tradicionales de protección.

Por su parte, el estudio de **Mikel Recuero** ofrece un análisis crítico de los datos genéticos como categoría especialmente sensible dentro del Derecho de protección de datos. A través de un recorrido que abarca su evolución histórica, su configuración actual en el Reglamento General de Protección de Datos y sus posibles desarrollos futuros, el autor pone de relieve tanto las certezas como las incertidumbres que rodean a estos datos, cuya creciente relevancia científica y económica plantea desafíos regulatorios de primer orden.

En el ámbito de la gobernanza global del dato, el trabajo de **Pilar Vargas Martínez** propone una sugerente lectura comparada de los modelos regulatorios de la Unión Europea, los Estados Unidos y China a través de tres metáforas —el Camino de Santiago, la Ruta 66 y la Ruta de la Seda— que reflejan distintas concepciones de la privacidad: como derecho fundamental, como bien de mercado o como instrumento de control estatal. Este enfoque permite comprender las tensiones estructurales que caracterizan el actual escenario internacional y sus implicaciones para la regulación de los flujos de datos.

La dimensión social de la digitalización es abordada por **Manuel Jesús Garnica Corbacho**, quien analiza la brecha digital como una nueva forma de desigualdad en el ámbito laboral. El trabajo examina sus manifestaciones en el acceso al empleo, la segmentación del mercado de trabajo y el teletrabajo, destacando su impacto diferencial en colectivos vulnerables y proponiendo mecanismos de tutela jurídica orientados a garantizar la igualdad de oportunidades digitales y la transparencia en el uso de sistemas algorítmicos.

Desde el ámbito educativo, **Carolina López Medina** examina la incorporación de la inteligencia artificial en la educación superior, poniendo de relieve la tensión entre innovación pedagógica y protección de datos personales. El artículo defiende la necesidad de una integración ética y jurídicamente garantista de estas tecnologías, apoyada en la alfabetización digital, la supervisión humana y la aplicación de principios como la privacidad desde el diseño y por defecto.

El estudio de **Darío Echeverría Muñoz** aporta una perspectiva comparada sobre los modelos regulatorios en materia de protección de datos, analizando el caso ecuatoriano frente al paradigma europeo del RGPD. Su trabajo muestra la transición hacia modelos de metarregulación basados en la tecnificación del cumplimiento y la gestión cuantitativa del riesgo, evidenciando tanto sus ventajas en términos de seguridad jurídica como los riesgos de rigidez y exclusión que pueden derivarse de este enfoque.

Finalmente, **Patricio Monreal Vilanova** aborda las implicaciones del uso de la inteligencia artificial en la Administración local desde la perspectiva de la protección de datos personales. Su contribución ofrece un análisis práctico de las principales obligaciones jurídicas asociadas al tratamiento de datos en sistemas de IA —licitud, transparencia, responsabilidad o evaluación de impacto—, poniendo de relieve la necesidad de integrar las exigencias del Derecho de protección de datos en los procesos de modernización administrativa.

En su conjunto, los trabajos que integran esta sección muestran la amplitud y complejidad de los desafíos que plantea la sociedad digital. Desde ámbitos tan diversos como la organización del trabajo, la educación, la biotecnología o la administración pública, todos ellos coinciden en subrayar que la protección de datos y la privacidad no constituyen un obstáculo para la innovación, sino una condición necesaria para su legitimidad y sostenibilidad en el marco del Estado de Derecho.

El número se cierra con una **sección de recensiones** dedicada a dos obras recientes que abordan, desde perspectivas complementarias, la relación entre inteligencia artificial, poder público y garantías jurídicas:

- Castellanos Claramunt, Jorge, *DemocracIA. Un análisis en clave constitucional*, Dykinson, Madrid, 2025.

- Olivares Olivares, Bernardo, *Sistemas de Inteligencia Artificial en la Agencia Estatal de Administración Tributaria: despliegue tecnológico y control jurídico*, Atelier, Madrid, 2026.

Monográfico: IA, gobernanza algorítmica y riesgos para los derechos fundamentales

Decisión administrativa algorítmica discriminatoria y datos: reflexiones y propuestas*

Discriminatory Algorithmic Administrative Decision-Making and Data: Reflections and Proposals

Diana-Urania Galetta
Catedrática de Derecho Administrativo y Derecho Administrativo Europeo
Universidad de Milán
Directora del CERIDAP https://ceridap.eu/

Resumen:
El uso de algoritmos en la actividad de las administraciones públicas plantea importantes desafíos jurídicos, especialmente en relación con el riesgo de discriminación derivado del tratamiento de datos. Este trabajo analiza la relación estructural entre algoritmos y datos, subrayando cómo los sesgos presentes en los conjuntos de datos pueden reproducirse y amplificarse en los procesos de toma de decisiones automatizados. A partir del análisis del caso italiano del algoritmo de la «Buena Escuela» y de otros ejemplos de decisiones administrativas basadas en sistemas automatizados, el artículo muestra cómo la automatización puede generar efectos discriminatorios, especialmente en ámbitos sensibles como la asignación de recursos públicos o prestaciones sociales. El estudio examina el papel de la transparencia algorítmica y de la obligación de motivación como instrumentos esenciales para garantizar la legalidad, la controlabilidad y la legitimidad de las decisiones administrativas automatizadas. Asimismo, se destaca la contribución del Derecho de la Unión Europea, en particular la Carta de los Derechos Fundamentales y el Reglamento de IA, en la construcción de un marco normativo destinado a prevenir la discriminación algorítmica. El artículo concluye subrayando la función central del Derecho administrativo en la regulación y orientación del uso de algoritmos en la administración pública, con el objetivo de garantizar el respeto de los derechos fundamentales y el derecho a una buena administración.

Palabras clave:
Algoritmos; Discriminación; Transparencia; Datos; Administración.

Abstract:
The increasing use of algorithms in public administration raises significant legal challenges, particularly regarding the risk of discrimination resulting from data-driven decision-making. This article explores the structural relationship between algorithms and data, highlighting how biases embedded in datasets may be reproduced and amplified through automated administrative decisions. Drawing on the Italian case of the "Good School" algorithm and other examples of automated decision-making in the allocation of public benefits, the study shows how algorithmic governance may generate discriminatory outcomes, especially in sensitive areas such as access to social assistance. The paper examines the role of algorithmic transparency and the duty to give reasons as key legal safeguards to ensure the legality, accountability, and reviewability of automated administrative decisions. It also analyses the contribution of European Union law—particularly the Charter of Fundamental Rights and the EU Artificial Intelligence Regulation—in establishing a regulatory framework aimed at preventing algorithmic discrimination. The article concludes by emphasising the central role of administrative law in guiding and regulating the use of algorithms within public administration in order to safeguard fundamental rights and uphold the right to good administration.

Keywords:
Algorithms; Discrimination; Transparency; Data; Administration.

* El presente artículo es la versión en castellano (adaptada y complementada) de una contribución mía que se publicará en italiano el volumen de G. P. Ruotolo y A. Stiano, eds., *Discriminazioni automatiche. Come il diritto può aiutare l'intelligenza artificiale e vice versa* (Turín: Giappichelli, 2026).

1. Notas introductorias

Mi participación en el debate académico sobre cuestiones relacionadas con el uso de las tecnologías de la información y la comunicación (TIC) en el contexto de la actividad de las administraciones públicas comenzó hace más de una década, como consecuencia de mi participación, en calidad de miembro, en el comité directivo de la «Red de Investigación sobre Derecho Administrativo de la UE» (*Research Network on EU Administrative Law – ReNEUAL*). En ese contexto específico fui, durante más de una década, *jefa de equipo* (junto con H.C.H. Hofmann y J.P. Schneider) del grupo de investigación responsable de la redacción de los Libros V y, en particular, VI del código ReNEUAL, centrado en las cuestiones jurídicas relacionadas con la cooperación dentro de la Unión Europea para la gestión de la información administrativa[1].

La creciente importancia de las TIC en este contexto me resultó evidente desde el principio, al igual que los retos que esto suponía y supone, y cuyo análisis ha sido objeto de muchas de mis publicaciones científicas de los últimos años, algunas de las cuales también he compartido con otros miembros de la red ReNEUAL[2].

Otro aspecto que me gustaría destacar es que profundizar en estas cuestiones me ha exigido una considerable inversión de tiempo para poder comprender los problemas técnicos subyacentes. Imaginar el posible impacto de las TIC en la actividad de las administraciones públicas y sus implicaciones jurídicas presupone, de hecho, una comprensión de qué son y en qué consisten las TIC. Abordar las cuestiones jurídicas que plantea o plantearía la gestión de un proceso de toma de decisiones administrativas algorítmicas requiere, como mínimo, una comprensión básica de lo que es un algoritmo, de los distintos tipos de algoritmos y de sus características distintivas, así como de los retos (técnicos) relacionados con su desarrollo y aplicación. Si se habla de algoritmos de inteligencia artificial (IA), es imprescindible comprender qué se entiende por IA y, para ello, es necesario comprender al menos los rudimentos básicos sobre la naturaleza y el funcionamiento de la inteligencia humana[3].

[1] Oriol Mir, Herwig C. H. Hofmann, Jens-Peter Schneider y Jacques Ziller, dirs., *Código ReNEUAL de procedimiento administrativo de la Unión Europea* (Madrid: Instituto Nacional de Administración Pública, 2015).

[2] Diana-Urania Galetta, Herwig C. H. Hofmann y Jens-Peter Schneider, "Information Exchange in the European Administrative Union: An Introduction," *European Public Law* 20 (2014): 65 y ss; Diana-Urania Galetta, "Informal Information Processing in Dispute Resolution Networks: Informality versus the Protection of Individual's Rights?," *European Public Law* 20 (2014): 71 y ss; Diana-Urania Galetta, "Decision-Making and Information Management," en *The Model Rules on EU Administrative Procedures: Adjudication*, ed. Matthias Ruffert (Europa Law Publishing, 2016), 185 y ss; Diana-Urania Galetta, "Le Model Rules di ReNEUAL e gli aspetti più innovativi della collaborazione fra amministrazioni nell'UE: procedimento amministrativo, scambio dei dati e gestione delle banche dati," *Rivista Italiana di Diritto Pubblico Comunitario* (2018), no. 2: 347 y ss. Véase también Diana-Urania Galetta y Jacques Ziller, eds., *Information and Communication Technologies Challenging Public Law: Beyond Data Protection* (Nomos Verlagsgesellschaft, 2018); Diana-Urania Galetta, "Information and Communication Technology and Public Administration: Through the Looking-Glass," en *Information and Communication Technologies Challenging Public Law: Beyond Data Protection*, ed. Diana-Urania Galetta y Jacques Ziller (Nomos Verlagsgesellschaft, 2018), 119 y ss.

[3] Diana-Urania Galetta, *Artificial Intelligence and Public Administration. A Journey* (Editoriale Scientifica, 2025), esp. caps. I y IX

Esto significa, en esencia (y para concluir esta nota introductoria), que el jurista que desee abordar con conocimiento de causa el tema de la discriminación que puede surgir contra las personas en un universo de toma de decisiones administrativas algorítmicas debe poseer una base de partida lo suficientemente sólida como para poder reconstruir, desde un punto de vista puramente técnico e informático, como se producen estas «discriminaciones algorítmicas» y cómo y por qué se diferencian de las discriminaciones que pueden producirse en el contexto de una decisión administrativa no algorítmica. Partiendo de esta base, es posible (y necesario) preguntarse cómo se puede definir un sistema de normas que permita abordar adecuadamente las cuestiones y los problemas planteados por la innovación tecnológica relacionada con la automatización de las decisiones administrativas, o al menos de parte del proceso de toma de decisiones administrativas, mediante el uso de algoritmos.

2. *Garbage in Garbage out*: cómo la decisión administrativa algorítmica está determinada por los datos disponibles

La cuestión está estrechamente relacionada con el problema de los datos. De hecho, los *data scientists* utilizan el acrónimo GIGO, que significa «*garbage in garbage* out» (basura entra, basura sale)[4]. Evidentemente, un algoritmo no puede sino reflejar la calidad de los datos sobre los que se construye. Y esto se aplica, por supuesto, a las hipótesis predictivas que es capaz de producir, si se trata un algoritmo de «*machine learning*» (aprendizaje automático)[5]. Los algoritmos, especialmente los de aprendizaje automático, aprenden de los datos que se les proporcionan durante la fase de entrenamiento. Si los datos son de mala calidad, incompletos, distorsionados o no representativos de la realidad que se quiere modelar, el algoritmo tendrá dificultades para hacer predicciones precisas, ya que no podrá captar las relaciones correctas entre las variables. Por ejemplo, si los datos de entrenamiento contienen errores sistemáticos, el algoritmo tenderá a reproducir y amplificar estos errores en sus predicciones[6].

El problema es, obviamente, aún mayor si los datos contienen sesgos: el algoritmo tenderá a reproducirlos, lo que dará lugar a decisiones erróneas o injustas. Un ejemplo clásico es el de los sistemas de reconocimiento facial entrenados con conjuntos de datos que no representan adecuadamente a todas las etnias, resultando más precisos para algunos grupos y menos para otros. Esto provoca un efecto dominó en las hipótesis predictivas que formulan.

En el contexto del aprendizaje automático, de hecho, el algoritmo no hace más que construir modelos predictivos basados en datos. Las «hipótesis predictivas» del algoritmo son esencialmente las predicciones que este produce a partir de la información aprendida. Si los datos están distorsionados, son incompletos o no son representativos, las hipótesis predictivas se verán afectadas de manera similar.

El algoritmo, en el contexto del aprendizaje automático, depende en gran medida de la calidad de los datos. Una buena calidad de los datos permite construir modelos predictivos más precisos y útiles, mientras que los datos de mala calidad dan lugar a modelos que pueden ser imprecisos, distorsionados o ineficaces. Por eso, una parte fundamental del proceso de aprendizaje automático es la preparación de los datos (preprocesamiento de datos), que incluye actividades como la limpieza, la normalización, la gestión de valores faltantes y la eliminación de anomalías.

A pesar de los posibles métodos de control de la calidad de los datos que se pueden implementar en su interior, los algoritmos de aprendizaje automático no son capaces de reconocer la calidad intrínseca de los datos y, por lo tanto, tienen en cuenta en sus predicciones todos los datos proporcionados, ya sean erróneos o «corruptos». Por esta razón, es esencial que los datos introducidos sean correctos y de alta calidad antes de entrenar un modelo de aprendizaje automático.

En el contexto de los algoritmos simples o condicionales, los datos siguen siendo muy relevantes, pero su influencia es diferente a la de los sistemas de aprendizaje automático. Los algoritmos simples, como los basados en estructuras condicionales (por ejemplo, instrucciones *if-else*), no «aprenden» de los datos en el sentido tradicional. En cambio, operan con datos ya definidos y

[4] Traducible como «entra basura, sale basura».

[5] Véase al respecto lo que ya he observado en Diana-Urania Galetta, "Human-stupidity-in-the-loop? *Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale," Federalismi.it*, fasc. 5/2023, 22 de febrero de 2023, 1 y ss., http://www.federalismi.it

[6] Véase también Cosimo Accoto, *Il mondo dato: Cinque brevi lezioni di filosofia digitale* (Milán, 2017).

pueden verse influidos por el tipo y la calidad de los datos con los que se alimentan, pero de una manera más directa y predecible.

En los algoritmos condicionales, los datos influyen explícitamente en el flujo de decisiones. El algoritmo ejecuta una secuencia de verificaciones y acciones basadas en los valores específicos de los datos. La diferencia principal es que, mientras que en los algoritmos complejos de aprendizaje automático el algoritmo construye modelos basados en patrones en los datos, en los algoritmos simples los datos determinan directamente el resultado de las operaciones, y una buena calidad de los datos es esencial para evitar errores o comportamientos inesperados.

3. El caso (italiano) del algoritmo de la buena escuela

Un ejemplo claro de lo que acabamos de describir es el caso, bien conocido por los estudiosos del derecho administrativo italiano, relativo al uso del algoritmo de la llamada «buena escuela»[7], utilizado para asignar profesores a las escuelas en el marco de la «reforma de la Buena Escuela», introducida en 2015[8].

En concreto, el algoritmo (de tipo simple/condicional) debía asignar a los profesores a los centros educativos en función de las preferencias expresadas por los propios profesores y de la disponibilidad de los centros.

El sistema se basaba en un algoritmo que tenía en cuenta diversos factores, como la ubicación geográfica, las preferencias de los profesores, las necesidades de las escuelas y otras variables. Sin embargo, los resultados obtenidos con la aplicación de este algoritmo han suscitado mucha polémica y han conducido a una gran cantidad de apelaciones y, por lo tanto, han sido objeto de muchas decisiones judiciales de nuestros jueces administrativos[9].

Los problemas que se han puesto de manifiesto en el caso mencionado son típicamente problemas relacionados con la calidad de los datos, así como con el diseño del propio algoritmo[10]. En particular, los datos recopilados resultaron ser incompletos, distorsionados o, en algunos casos, incluso totalmente erróneos. Por ejemplo, los datos relativos a las escuelas (como la disponibilidad de plazas) no siempre eran precisos o estaban actualizados, lo que creaba discrepancias entre las expectativas de los solicitantes y la realidad. La presencia de errores en los datos relativos a los currículos o a la formación de los profesores dificultó que el algoritmo realizara asignaciones correctas. En términos más generales, dado que el algoritmo se basaba en datos erróneos o inexactos, las asignaciones de plazas realizadas como resultado de su uso no eran correctas. Por ejemplo, algunos profesores fueron asignados a escuelas muy alejadas de sus preferencias o ubicación geográfica; las necesidades de algunas escuelas no se satisfacían, ya que el algoritmo no tenía acceso a datos precisos sobre los recursos efectivos necesarios o las necesidades educativas reales, etc.

Otro aspecto crítico es que el algoritmo no era transparente. Los usuarios (los profesores y las escuelas) no tenían una comprensión clara del funcionamiento del algoritmo y de cómo se ponderaban las preferencias. Esto contribuyó a la percepción de que las decisiones que tomaba eran ineficaces e injustas y estaban mal tomadas.

Así pues, y esta es la paradoja y la razón por la que este ejemplo resulta especialmente pertinente en

[7] Véase, al respecto, por último, en Diana-Urania Galetta, "El derecho a una buena administración en un entorno de Administración pública digital. Reflexiones a partir del ejemplo de Italia," en *De Luis XIV al Estado inteligente*, ed. A. A. Martino (Buenos Aires: Editorial Astrea, 2022), 48 y ss. Véase también, entre otros, Alfonso Celotto, "Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto," *Analisi giuridica dell'economia* (2019): 47 y ss.; Ida Angela Nicotra y Vincenzo Varone, "L'algoritmo, intelligente ma non troppo," *Rivista AIC* 4 (2019): 86 y ss.; Silvia Sassi, "Gli algoritmi nelle decisioni pubbliche tra trasparenza e responsabilità," *Analisi giuridica dell'economia* (2019): 109 y ss.; Andrea Simoncini, "L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà," *BioLaw Journal – Rivista di BioDiritto* (2019): 63 y ss.

[8] Véase al respecto https://www.istruzione.it/allegati/2017/La_Buona_Scuola_Approfondimenti.pdf.

[9] Véase la doctrina citada en la nota n.º 7.

[10] Véase Lorenzo Cotino Hueso, "Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público," en *Inteligencia artificial y sector público. Retos, límites y medios*, ed. Eduardo Gamero Casado y Francisco Luis Pérez Guerrero (Valencia: Tirant lo Blanch, 2023), 257 y ss.; Agustí Cerrillo i Martínez, "Preventing Biased Results and Discriminatory Effects of AI Use in Public Administration," en *The EU Artificial Intelligence Act and the Public Sector. Humans and AI Systems in Public Administration in the Light of the European Regulation on Artificial Intelligence of 2024 (with Foreword by C. Coglianese)*, ed. Juli Ponce y Agustí Cerrillo i Martínez (Athens: EPLS, 2025), 175 y ss.

este contexto, un algoritmo, diseñado con la intención de automatizar el proceso de asignación para generar decisiones más equitativas e imparciales, se encontró con la complejidad y la variabilidad de la situación real. Las condiciones locales de las escuelas y las situaciones particulares de los profesores no siempre han estado bien representadas en los datos, y el algoritmo ha tomado decisiones que han dado lugar a situaciones problemáticas, como la movilidad forzosa de profesores o asignaciones a escuelas que no se ajustaban a las necesidades.

Por último, hay un elemento técnico adicional que vale la pena destacar aquí y que se refiere a una característica específica de la que el algoritmo de la buena escuela carecía por completo: la capacidad de adaptación dinámica. Es decir, el algoritmo no fue capaz de adaptarse en tiempo real a los cambios en los datos. Por ejemplo, durante el año escolar, la disponibilidad de profesores o la demanda de las escuelas podían cambiar, pero el algoritmo no era lo suficientemente flexible como para responder a estos cambios sin generar inconvenientes.

En conclusión, el caso del algoritmo de la Buena Escuela demuestra claramente cómo un algoritmo, aunque esté bien diseñado en términos teóricos, puede producir resultados problemáticos si los datos en los que se basa son incompletos, imprecisos o mal gestionados. Pero, más aún, pone de manifiesto la importancia de un diseño transparente y una comunicación clara sobre el funcionamiento de los algoritmos, especialmente cuando estos tienen un impacto significativo en las personas y sus derechos. Esto suele ocurrir cuando nos enfrentamos a una decisión administrativa, que tiene en sí misma la fuerza jurídica para modificar, de forma totalmente unilateral, la esfera jurídica de su destinatario.

En este sentido, la «discriminación algorítmica» que se ha producido en el caso que nos ocupa es totalmente peculiar, ya que es consecuencia de la inexactitud de los datos, pero también del «efecto multiplicador» que puede producir la decisión administrativa algorítmica (y sobre el que volveré en los próximos párrafos).

4. Ejemplos de decisiones administrativas basadas en algoritmos discriminatorios

El caso (italiano) del algoritmo de la buena escuela no es más que un ejemplo de los efectos distorsionadores y discriminatorios que pueden producir los algoritmos que, por el contrario, tendrían la ambición de hacer que las decisiones de las administraciones públicas fueran más imparciales, al estar desvinculadas del decisor humano, con su bagaje de preferencias y prejuicios, y más acordes con el canon de buena administración codificado en el artículo 41 de la Carta de los Derechos Fundamentales de la Unión Europea[11].

De hecho, en la práctica se han puesto de manifiesto muchos otros ejemplos negativos de los resultados distorsionadores a los que puede conducir la administración por algoritmos.

El problema siempre ha sido generado por la mala calidad de los conjuntos de datos utilizados: insuficientemente amplios y representativos, o basados en prejuicios sociales y culturales preexistentes (los llamados sesgos)[12].

Un ejemplo especialmente impactante, precisamente por la delicadeza de las situaciones que se crean como consecuencia, es el de los casos en los que se han utilizado algoritmos para apoyar procesos de toma de decisiones (administrativas) destinadas a la asignación de subvenciones públicas y han producido efectos discriminatorios en detrimento de los sectores más vulnerables de la población: aquellos que, por el contrario, deberían ser los beneficiarios del servicio, prestado con arreglo a la lógica del objetivo de la denominada «igualdad sustantiva» que la intervención pública asistencial pretende alcanzar[13].

El resultado paradójico es que los sujetos más discriminados en las decisiones algorítmicas o en la investigación algorítmica son precisamente aquellos que ya se enfrentan a la marginación o los prejuicios.

[11] Véase Diana-Urania Galetta, "Digitalizzazione e diritto ad una buona amministrazione (Il procedimento amministrativo, fra diritto UE e tecnologie ICT)," en *Il Diritto dell'Amministrazione Pubblica digitale*, ed. Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. (Turín, 2025), 77 y ss.

[12] Véase Kate Crawford, "Artificial Intelligence's White Guy Problem," *New York Times*, 25 de junio de 2016.

[13] Véase Jaime Rodríguez-Arana, "La buena administración como principio y como derecho fundamental en Europa," *Misión Jurídica. Revista de Derecho y Ciencias Sociales* (2013): 23 y ss. Sobre este punto, véanse las reflexiones y la doctrina mencionadas en el siguiente apartado 8, dedicado a las conclusiones.

Este fenómeno se ha observado en diversos contextos y, obviamente, me limitaré aquí a algunos ejemplos significativos.

En Estados Unidos, los sistemas automatizados que gestionan la asignación de subsidios de asistencia social, como los del programa SNAP (*Supplemental Nutrition Assistance Program*) y otros subsidios federales, han suscitado preocupación por los sesgos y la discriminación hacia determinadas comunidades, en particular las de bajos ingresos o las minorías étnicas[14]. Algunos estudios han puesto de manifiesto que los sistemas informáticos que gestionaban la asignación de subsidios alimentarios (y de otro tipo) estaban influenciados por discriminaciones históricas. Por ejemplo, el algoritmo tenía el efecto de penalizar o dificultar el acceso a las ayudas a las familias negras o a las personas que vivían en determinadas zonas geográficas. Esto se debía a que los sistemas automáticos utilizaban información que reflejaba los prejuicios raciales presentes en las bases de datos, como la residencia en barrios de bajos ingresos o el hecho de proceder de familias monoparentales.

La introducción de algoritmos para automatizar el proceso de verificación de la idoneidad para recibir subsidios también ha dado lugar a dificultades para identificar correctamente a las familias con necesidades reales, que a menudo no estaban bien representadas en los datos tradicionales o en las métricas utilizadas por los algoritmos. Por ejemplo, algunos algoritmos dieron mayor peso a criterios relacionados con la estabilidad financiera a largo plazo, penalizando a quienes viven en contextos económicos inestables, sin tener en cuenta otras formas de vulnerabilidad social.

Lo mismo ocurrió en el Reino Unido, donde el sistema «automatizado» de asignación de subsidios de desempleo fue criticado por los casos en los que el algoritmo excluyó erróneamente a muchos desempleados, especialmente entre las personas con discapacidad o las minorías étnicas. En el caso del sistema *Universal Credit*[15], un programa que combina diferentes formas de asistencia social, entre ellas las prestaciones por desempleo y las ayudas a las rentas más bajas, el algoritmo que gestionaba la asignación de recursos dio lugar a errores significativos, en parte porque no era capaz de gestionar correctamente la información relativa a personas con discapacidad o con situaciones familiares complejas. De hecho, algunas personas vulnerables quedaron excluidas del sistema debido a la rigidez de los criterios utilizados para determinar la elegibilidad, mientras que otras fueron objeto de investigaciones injustificadas o sufrieron retrasos en los pagos[16].

El algoritmo que gestionaba las solicitudes del *Universal Credit* también ha sido acusado de dar menos acceso a las prestaciones a las personas que pertenecían a determinados grupos étnicos o que vivían en condiciones socioeconómicas precarias, lo que ha creado desigualdades en el trato de las personas. También en este caso, los sesgos implícitos en los datos históricos sobre el empleo, la familia y los ingresos han tenido un impacto directo en las decisiones tomadas por el algoritmo.

Se han registrado discriminaciones similares en Italia, por ejemplo, en lo que respecta a la asignación de la «renta básica» (*reddito di cittadinanza*)[17], en relación con la cual el sistema de «discriminación institucional»[18] ya existente se ha agravado aún más con la automatización, mediante el uso de algoritmos destinados a determinar la idoneidad para obtener la prestación. Los criterios basados en los datos sobre ingresos y patrimonio han creado dificultades a algunos grupos, en particular a aquellos que no disponían de la documentación fiscal adecuada o vivían en situaciones de pobreza denominada «invisible». El sistema ha tenido dificultades para identificar a todas las familias necesitadas, penalizando las situaciones de vulnerabilidad que no eran fácilmente detectables mediante el sistema automatizado, como las personas que vivían en condiciones de pobreza extrema pero no tenían los medios para documentarlo oficialmente.

Otro problema que surgió fue la fiabilidad de los datos en los que se basaba el propio algoritmo. Algunos errores en los datos de residencia o en las verificaciones de ingresos acabaron excluyendo

[14] Louis DiPietro, "Google Algorithm Found to Overlook Spanish Speakers in Online SNAP Ads," *TechXplore*, 10 de agosto de 2023, https://techxplore.com/news/2023-08-google-algorithm-overlook-spanish-speakers.html

[15] Véase Simon Harwood, "The Parrot Is Not Dead, Just Resting: The UK Universal Credit System – An Empirical Narrative," *University of Edinburgh Business School Working Paper Series* (2018): 1 y ss.

[16] Véase Michele Gilman, "AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead," *The Conversation*, 14 de febrero de 2020, https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625

[17] https://www.redditodicittadinanza.gov.it/

[18] Maurizio Ambrosini, Deborah Erminio y Alberto Guariso, "*L'accesso degli stranieri al Reddito di Cittadinanza: possibile discriminazione istituzionale*," *Welforum.it*, 23 de octubre de 2023, https://www.welforum.it/un-caso-di-persistente-discriminazione-istituzionale-il-reddito-di-cittadinanza/

del beneficio a algunas familias que deberían haber recibido la ayuda, o provocaron retrasos en los pagos.

Un caso especialmente inquietante, que de hecho saltó a los titulares, es el del algoritmo de aprendizaje automático que utilizó la ciudad de Róterdam para identificar posibles casos de fraude al sistema de seguridad social. Cada año, miles de personas en toda Rotterdam eran investigadas por funcionarios antifraude, que basaban el inicio de sus investigaciones en los resultados producidos por un algoritmo entrenado en investigaciones anteriores y que debía ayudarles a determinar qué personas eran más susceptibles de cometer fraudes en materia de seguridad social. En 2021, el uso de este modelo de evaluación automática del riesgo se suspendió después de que auditores externos, nombrados por el gobierno central, constataran la falta de transparencia de los resultados producidos por el algoritmo y la presencia de sesgos en su propio diseño[19] .

Los ejemplos podrían continuar. Pero los que se han mencionado hasta ahora me parecen más que suficientes para confirmar la situación tal y como se describió inicialmente: en todos estos casos el uso de algoritmos para la asignación de recursos públicos ha puesto de manifiesto que la calidad de los datos en los que se basan los algoritmos es fundamental para evitar la discriminación. Si los datos están distorsionados o son incompletos, el algoritmo puede perpetuar o amplificar las desigualdades existentes, en lugar de resolverlas. Además, la falta de transparencia en los procesos algorítmicos, junto con la dificultad de corregir los errores, puede empeorar aún más la situación, causando daños reales a las personas vulnerables a las que el sistema de asistencia pública debería favorecer/proteger.

5. ¿La transparencia como solución?

Desde la última perspectiva mencionada, uno de los principales problemas de la decisión algorítmica (o respaldada por algoritmos) es evitar que los errores o prejuicios ocultos en los datos y/o en la lógica decisional programada y gestionada por el algoritmo puedan afectar negativamente a la decisión final de la administración pública, alejándola, en lugar de acercarla, a los cánones de imparcialidad y buen funcionamiento que, para Italia, están codificados en el artículo 97 de la Constitución y, para la Unión Europea, se explicitan en el artículo 41 de la Carta de los Derechos Fundamentales de la Unión Europea[20].

La transparencia puede desempeñar aquí un papel crucial[21]. Puede contribuir a construir un sistema de toma de decisiones administrativas basado en algoritmos que sea capaz de garantizar decisiones administrativas no solo más rápidas, sino también más equitativas e imparciales, ya que sus resultados finales son verificables y pueden mejorarse y adaptarse con el tiempo[22].

Desde este punto de vista, la transparencia puede referirse tanto a la disponibilidad de la información relativa al código de programación del algoritmo como al conocimiento de los datos utilizados para el entrenamiento de los algoritmos. En ambos casos, se trata de información que puede permitir a expertos externos, como auditores independientes u organizaciones no gubernamentales, examinar los sistemas de decisión algorítmica en busca de problemas que puedan conducir a decisiones discriminatorias.

La posibilidad de rastrear los procesos de toma de decisiones también posibilita cuestionar cualquier decisión equivocada. Si el algoritmo es transparente, cualquier error o resultado injusto puede identificarse y corregirse más fácilmente, antes de que cause daños a gran escala. Si, por

[19] Véase el informe de Matt Burgess, Ellen Schot y Georgina Geiger, "This Algorithm Could Ruin Your Life," *Wired*, 6 de junio de 2023, https://www.wired.com/story/welfare-algorithms-discrimination/. Véase también Andrea Alù, "L'algoritmo dei sussidi sociali discrimina e fa cadere il governo: il caso olandese," *Agenda Digitale*, 19 de enero de 2021, https://www.agendadigitale.eu/cultura-digitale/algoritmi-troppo-invasivi-contro-le-frodi-fiscali-la-lezione-delle-dimissioni-del-governo-olandese/

[20] Véase Diana-Urania Galetta, "Digitalizzazione e diritto ad una buona amministrazione," 77 y ss.

[21] Más en general sobre el tema de la transparencia de algoritmos e IA véase Lorenzo Cotino Hueso, "Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida," *Revista Española de la Transparencia*, núm. 16 (2023): 17 y ss.; y el reciente volumen: Lorenzo Cotino Hueso y Jorge Castellanos Claramunt, eds., *Transparencia y explicabilidad de la inteligencia artificial* (Valencia: Tirant lo Blanch, 2024).

[22] La bibliografía sobre este tema es muy amplia. Entre otros: Deven R. Desai y Joshua A. Kroll, "Trust but Verify: A Guide to Algorithms and the Law," *Harvard Journal of Law & Technology* (2017): 31 y ss.; Mike Ananny y Kate Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability," *New Media & Society* (2018): 973 y ss.; Tobias D. Krafft, Katharina A. Zweig y Philipp D. König, "How to Regulate Algorithmic Decision-Making: A Framework of Regulatory Requirements for Different Applications," *Regulation & Governance* (2020): 1 y ss.

ejemplo, un algoritmo de asignación de subvenciones crea desigualdades entre los beneficiarios, los *feedback* pueden dar lugar a ajustes rápidos, como la reelaboración de los datos y/o la actualización de los modelos. La transparencia también permite adaptar el sistema para responder a situaciones nuevas o imprevistas, reduciendo el riesgo de consecuencias inesperadas.

Por último, este tipo de transparencia también puede favorecer la promoción de un diálogo público sobre el diseño del algoritmo, lo que induce a reflexionar sobre las implicaciones éticas de las decisiones, como el uso de determinados datos o el diseño de los modelos de predicción[23]. Por ejemplo, si un algoritmo destinado a la adopción de decisiones para distribuir subsidios públicos se basa en datos que reflejan desigualdades históricas (como los ingresos o la raza), la transparencia permite identificar estos problemas. La conciencia de estos posibles sesgos puede ayudar a mejorar el algoritmo, modificando los datos o cambiando los criterios de asignación para evitar la reproducción de las desigualdades. De lo contrario, el algoritmo, si es de tipo determinista, es capaz de elevar el error a sistema, reproduciéndolo un número infinito de veces, de forma totalmente automática. Si es de aprendizaje automático, es capaz de adquirir el error en ese patrimonio personal de conocimientos en el que basa su «entrenamiento», produciendo así una especie de «efecto multiplicador» del propio error, con resultados finales en gran medida impredecibles, por otra parte.

Desde esta perspectiva, también se pueden recopilar y utilizar de forma útil los *feedback* de los usuarios finales (ciudadanos o beneficiarios de subsidios) para mejorar continuamente el algoritmo. Así que deben ser recibidas con satisfacción todas las propuestas legislativas que fomentan esta modalidad de «transparencia algorítmica». Esta se realiza a través de la implementación de regulaciones que impongan a las administraciones públicas la obligación de divulgar los procesos de diseño y empleo de algoritmos, así como su potencial impacto en los individuos, con el objetivo de prevenir prácticas discriminatorias.

El Reglamento de la UE sobre inteligencia artificial[24] hace referencia, en este sentido, tanto a la circunstancia de que debe darse a conocer que se ha recurrido al uso de herramientas algorítmicas, como a que debe facilitarse «información significativa» sobre el funcionamiento de dichos sistemas. En particular, el artículo 53 del Reglamento sobre IA especifica, entre las obligaciones de los proveedores de modelos de IA para fines generales, que estos «elaborarán y pondrán a disposición del público un resumen suficientemente detallado del contenido utilizado para el entrenamiento del modelo de IA de uso general, con arreglo al modelo facilitado por la Oficina de IA» (art. 53, apartado 1, letra d).

En este sentido, me parece muy relevante también la disposición del artículo 14 de la Ley italiana sobre Inteligencia Artificial[25]. En consonancia con lo dispuesto en el Reglamento de la UE sobre IA, en su apartado 1 esta nueva ley establece que, cuando las Administraciones Públicas «utilicen la inteligencia artificial», será necesario garantizar «a los interesados el conocimiento de su funcionamiento y la trazabilidad de su uso». La disposición retoma, de hecho, lo establecido en el artículo 30 del Código italiano de Contratos Públicos de 2023, que, en su apartado 2, establece que, en caso de utilizar procedimientos automatizados en ese contexto específico, es necesario garantizar la disponibilidad del código fuente, de la documentación correspondiente y de cualquier otro elemento útil para comprender su lógica de funcionamiento; y, en el apartado 3, especifica que las decisiones adoptadas mediante automatización respetarán los principios de conocimiento y comprensibilidad, por lo que todo operador económico tiene derecho a conocer la existencia de procesos de decisión automatizados que le afecten y, en tal caso, a recibir información significativa sobre la lógica utilizada.

Las normativas mencionadas hasta ahora intentan, de hecho, hacer frente a la conocida dificultad que existe, a nivel de sistema, para garantizar una transparencia efectiva. Dado que a menudo es difícil comprender cómo se toman las decisiones, muchos algoritmos se perciben como una «caja negra»[26]. La complejidad del código informático (que a menudo se oculta en el *software* propietario)

[23] Véase Helena Webb, Ansgar Koene, Menisha Patel y Elvira Perez Vallejos, "Multi-Stakeholder Dialogue for Policy Recommendations on Algorithmic Fairness," en *Proceedings of the 9th International Conference on Social Media and Society* (Nueva York, 2018), 395 y ss.

[24] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

[25] Ley 132/2025, Disposiciones y delegación al Gobierno en materia de inteligencia artificial, que entró en vigor el 10 de octubre de 2025.

[26] Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (Cambridge, 2015).

y la posibilidad de que este código pueda ser comprendido por personas que no son expertas en áreas específicas de la informática dificultan su comprensión[27]. En los denominados «sistemas expertos», esto podría ser un verdadero problema, pero lo es aún más cuando el código de los sistemas de decisión algorítmicos está compuesto por tecnologías de aprendizaje automático que dificultan incluso a los expertos predecir la variedad de resultados que pueden producir los procedimientos de decisión automatizados[28]. En muchos casos, la posibilidad de «deducción lineal» no existe en realidad, lo que significa que los propios programadores del código podrían no comprender completamente cómo el código informático utilizado para ayudar en el proceso de toma de decisiones ha llevado realmente al contenido de la propia decisión[29].

6. Transparencia y obligación de motivación, entre el derecho interno y el derecho de la UE

A la luz de lo observado al final del párrafo anterior, existe, de hecho, otro aspecto de la transparencia que ha sido mucho menos estudiado pero que, en mi opinión, es aún más relevante desde la perspectiva de evitar la denominada discriminación algorítmica. Se trata de esa forma particular de transparencia relacionada con la obligación de las administraciones públicas de motivar sus decisiones. Esta obligación está presente en la mayoría de las legislaciones nacionales (de los Estados miembros de la UE y no solo[30]) y figura expresamente en el artículo 3 de la Ley italiana n.º 241/1990 sobre el procedimiento administrativo[31].

A este respecto, el juez administrativo nacional ha subrayado, no por casualidad, en la jurisprudencia relativa al denominado algoritmo de la buena escuela, del que ya se ha hablado ampliamente en el apartado 3 anterior, que «no es conforme con el conjunto normativo vigente y con lo dispuesto en el artículo 97 de la Constitución, con los principios que lo sustentan y con los institutos de participación en el procedimiento definidos en los artículos 7, 8, 10 y 10-*bis* de la Ley n.º 241, de 7 de agosto de 1990, a la obligación de motivar las medidas administrativas consagrada en el artículo 3 de la misma ley (...) confiar el desarrollo de los procedimientos administrativos a la activación de mecanismos y sistemas informáticos y a su consiguiente funcionamiento impersonal»[32]. Si esto ocurriera, el resultado sería «una frustración también de las garantías procesales relacionadas que se declinan en el ámbito del derecho de acción y defensa en juicio a que se refiere el artículo 24 de la Constitución, derecho que se ve comprometido cada vez que la ausencia de motivación no permite inicialmente al interesado y, posteriormente, a instancias de este, al juez, percibir el proceso lógico-jurídico seguido por la administración para llegar a una determinada resolución» [33].

Este aspecto de la compleja relación entre la decisión respaldada por el algoritmo y la obligación de motivación[34] se puso aún más de relieve en una sentencia anterior, de marzo de 2017, también de la

[27] Véanse también las reflexiones críticas de Cary Coglianese y David Lehr, "Transparency and Algorithmic Governance," *Administrative Law Review* 71 (2019): 1–60; Cary Coglianese y David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era," *Georgetown Law Journal* (2017): 1147 y ss.

[28] Como ya se ha señalado en Diana-Urania Galetta, Herwig C. H. Hofmann y Jacques Ziller, "Automazione, Intelligenza Artificiale e Pubblica Amministrazione, fra diritto interno e diritto UE," en *Il Diritto dell'Amministrazione Pubblica digitale*, ed. Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. (Turín, 2025), 109 y ss., esp. 113 y ss.

[29] Véase al respecto Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland y Patrick Vinck, "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes," *Philosophy & Technology* (2018): 611 y ss.

[30] Véase, por ejemplo: Mark C. Elliott, "Has the Common Law Duty to Give Reasons Come of Age Yet?," *University of Cambridge Faculty of Law Research Paper*, n.º 7 (2012), https://ssrn.com/abstract=2041362; Jerry L. Mashaw, "Reasoned Administration: The European Union, the United States, and the Project of Democratic Governance," en *The Accountability of Expertise: Making the Un-Elected Safe for Democracy*, ed. Erik O. Eriksen (Londres, 2022), 34 y ss.; Sosten Wekesa y Nixon Otieno, "The Duty to Give Reasons under Kenya's Fair Administrative Action Act, 2015 in Kenya: Seven Years Later," *Kabarak Journal of Law and Ethics* (2022): 99 y ss.

[31] Sobre esto, véase Antonio Cassatella, *Il dovere di motivazione nell'attività amministrativa* (Padua: CEDAM, 2013). Véase también Tommaso Autieri, *La motivazione del provvedimento amministrativo: raccolta di dottrina, giurisprudenza e legislazione* (Padua: CEDAM, 2002); Riccardo Villata y Margherita Ramajoli, *Il provvedimento amministrativo*, 2.ª ed. (Turín: Giappichelli, 2017), esp. 269 y ss.

[32] TAR. Lazio, sec. *III-bis*, sentencia n.º 9230, de 10 de septiembre de 2018. La traducción del italiano es mía.

[33] TAR Lazio, citado en la nota anterior. La traducción del italiano es mía.

[34] Véase, por último, Andrea Simoncini, "Amministrazione digitale algoritmica. Il quadro costituzionale," en *Il Diritto dell'Amministrazione Pubblica digitale*, ed. Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. (Turín: Giappichelli, 2025), 1 y ss., esp. 26 y ss.

sección III-bis del Tribunal administrativo de Lacio. En ella se afirma, de hecho, que «la admisibilidad de la elaboración electrónica del acto administrativo no está vinculada a la naturaleza discrecional o vinculante del acto, sino a la posibilidad, que sin embargo es científica y no jurídica, de reconstruir el proceso lógico sobre cuya base el acto puede ser promulgado mediante procedimientos automatizados en lo que respecta a su contenido dispositivo»[35]. Por lo tanto, en el derecho italiano se pone de manifiesto la importancia de la motivación como elemento esencial de transparencia: sirve para explicar el proceso lógico sobre cuya base se ha promulgado el acto. Así pues, para que el uso de algoritmos en el contexto del procedimiento administrativo sea (jurídicamente) aceptable, siempre debe ser posible proporcionar esta motivación. Esta posición ha sido confirmada también por el Consejo de Estado en su jurisprudencia posterior, que ha subrayado que, en caso de automatización del procedimiento administrativo, es aún más importante cumplir con la obligación de motivación, ya que «la regla algorítmica no solo debe ser conocida en sí misma, sino también estar sujeta al pleno conocimiento y al pleno control del juez administrativo»[36]. De hecho, se destaca que «la "caracterización multidisciplinar" del algoritmo (una construcción que sin duda requiere no solo competencias jurídicas, sino también técnicas, informáticas, estadísticas y administrativas) no exime de la necesidad de que la "fórmula técnica", que de hecho representa el algoritmo, vaya acompañada de explicaciones que la traduzcan en la "regla jurídica" subyacente y la hagan legible y comprensible, tanto para los ciudadanos como para el juez»[37].

La obligación de motivación siempre ha sido fundamental también en el Derecho de la UE. Ya el Tratado constitutivo de la Comunidad Europea del Carbón y del Acero, firmado en París el 18 de abril de 1951, establecía que «*Les décisions, recommandations et avis de la Haute Autorité sont motivés*» (art. 15 del Tratado CECA). Y según el artículo 190 del Tratado CEE de 1957 los reglamentos, las directivas y las decisiones del Consejo y de la Comisión deberán estar motivados. Disposición que se retomó y reformuló en el artículo 296, apartado 2, del TFUE, que establece que «los actos jurídicos deberán estar motivados». La Carta de los Derechos Fundamentales de la Unión Europea codificó posteriormente la obligación específica «que incumbe a la administración de motivar sus decisiones» en el artículo 41 sobre el derecho a una buena administración (apartado 2, letra c)[38].

7. En general, sobre la contribución del derecho de la UE a la lucha contra la discriminación algorítmica

En lo que respecta más específicamente a la delicada cuestión de la protección contra la discriminación y la contribución del Derecho de la UE al respecto, cabe recordar aquí que el artículo 21, apartado 1, de la Carta de los Derechos Fundamentales de la Unión Europea «prohíbe toda discriminación, y en particular la ejercida por razón de sexo, raza, color, orígenes étnicos o sociales, características genéticas, lengua, religión o convicciones, opiniones políticas o de cualquier otro tipo, pertenencia a una minoría nacional, patrimonio, nacimiento, discapacidad, edad u orientación sexual».

A este último respecto, el Tribunal de Justicia de la UE ha precisado que «los Estados miembros no pueden utilizar, como *criterios predeterminados*, criterios basados en características (…) cuya utilización puede dar lugar a discriminaciones»; por lo que «los *criterios predeterminados* deben establecerse de forma que, aunque se formulen de un modo neutro, su aplicación no pueda resultar particularmente desventajosa para las personas que tengan las características protegidas»[39] .

Esta jurisprudencia relativa a los «criterios predeterminados» es evidentemente relevante también en el contexto de las decisiones basadas en algoritmos. E implica que las administraciones públicas deben ser capaces de garantizar y documentar que los algoritmos utilizados en su proceso de toma de decisiones no infrinjan en modo alguno la prohibición del artículo 21 de la Carta de los Derechos Fundamentales de la Unión Europea, ni directa ni indirectamente, y con total independencia de que se trate de algoritmos predeterminados en sistemas expertos, de algoritmos basados en el

[35] TAR Lazio, sec. *III-bis*, sentencia n.º 3769, de 22 de marzo de 2017. La traducción del italiano es mía.

[36] Consejo de Estado, sección VI, sentencia n.º 2270, de 8 de abril de 2019, punto 8.4., sobre la que remito a todos a Diana-Urania Galetta, "Algoritmi, procedimento amministrativo e garanzie: brevi riflessioni, anche alla luce degli ultimi arresti giurisprudenziali in materia," *Rivista Italiana di Diritto Pubblico Comunitario*, n.º 2 (2020): 501 y ss.

[37] Consejo de Estado, sentencia citada en la nota anterior. La traducción del italiano es mía.

[38] Sobre este punto, véase nuevamente Diana-Urania Galetta, "Digitalizzazione e diritto ad una buona amministrazione," 103 y ss.

[39] Sentencia del Tribunal de Justicia de 21 de junio de 2022, 817/19, *Ligue des droits humains*, apartado 197. El resaltado en cursiva es mío.

aprendizaje automático (ML) o basados en IA generativa[40].

Desde esta perspectiva, la transparencia relacionada con la obligación de motivación es sin duda un instrumento poderoso para garantizar que los algoritmos, que son herramientas utilizadas para lograr una asignación más eficiente y mejor de los recursos públicos, no se conviertan en instrumentos para operar discriminaciones sistemáticas. Y también lo es desde la perspectiva de la cognoscibilidad del código fuente y de los conjuntos de datos utilizados, ya que permite, al menos potencialmente, identificar y corregir posibles sesgos y errores en los datos utilizados y responsabiliza a quienes diseñan e implementan los algoritmos, garantizando que estos sean equilibrados, inclusivos y adaptables a las necesidades reales.

En esta lógica, no es casualidad que el Reglamento de la UE sobre inteligencia artificial[41] prevea una serie de disposiciones en materia de transparencia cuando la IA se utilice para adoptar decisiones que puedan afectar a los derechos de los ciudadanos, como en el caso de la asignación de subvenciones públicas y la prestación de servicios públicos, por ejemplo, en los ámbitos de la educación, la asistencia sanitaria, los servicios sociales, la vivienda y la administración de justicia. En estos casos, el Reglamento de la UE sobre IA establece que las administraciones públicas (y los gestores/proveedores privados) deben garantizar que los ciudadanos estén informados sobre el uso de la IA y comprendan cómo se toman las decisiones. Y el artículo 77 especifica que «las autoridades u organismos públicos nacionales encargados de supervisar o hacer respetar las obligaciones contempladas en el Derecho de la Unión en materia de protección de los derechos fundamentales, incluido el derecho a la no discriminación, con respecto al uso de sistemas de IA de alto riesgo (…) tendrán la facultad de solicitar cualquier documentación creada o conservada con arreglo al presente Reglamento y de acceder a ella, en un lenguaje y formato accesibles, ». Por su parte, el artículo 27 del Reglamento establece que «antes de desplegar uno de los sistemas de IA de alto riesgo (…), los responsables del despliegue que sean organismos de Derecho público, o entidades privadas que prestan servicios públicos (…) llevarán a cabo una evaluación del impacto que la utilización de dichos sistemas puede tener en los derechos fundamentales».

Mientras que en el anexo III del Reglamento, relativo a los «Sistemas de IA de alto riesgo a que se refiere el artículo 6, apartado 2», están incluidos los «sistemas de IA destinados a ser utilizados por las autoridades públicas o en su nombre para evaluar la admisibilidad de las personas físicas para beneficiarse de servicios y prestaciones esenciales de asistencia pública, incluidos los servicios de asistencia sanitaria, así como para conceder, reducir o retirar dichos servicios y prestaciones o reclamar su devolución» (art. 1, apartado 5, letra a), del anexo III). Esto se debe a que, como se especifica en el considerando n.º 5 del Reglamento, «dependiendo de las circunstancias relativas a su aplicación, utilización y nivel de desarrollo tecnológico concretos, la IA puede generar riesgos y menoscabar los intereses públicos y los derechos fundamentales que protege el Derecho de la Unión». Por lo tanto, el Reglamento de la UE establece que las administraciones públicas que utilicen IA de alto riesgo deben supervisar continuamente el uso de estos sistemas para garantizar que cumplan con la normativa y no generen efectos discriminatorios o perjudiciales, y este seguimiento incluye tanto controles periódicos del rendimiento de los algoritmos y su capacidad para adaptarse a los cambios en los datos, como sistemas de auditoría externa para garantizar que los sistemas sean equitativos y transparentes, con la obligación de hacer públicos los resultados de estas verificaciones. Por último, se prevén evaluaciones periódicas del impacto sobre la protección de datos y los posibles riesgos para los derechos fundamentales.

En resumen, puede decirse que el Reglamento de la UE sobre IA crea un marco normativo que garantiza el uso de la IA en las administraciones públicas con transparencia, responsabilidad y atención a los derechos fundamentales[42]. Estos principios son fundamentales para evitar la discriminación y garantizar que la IA se utilice de forma correcta y transparente, especialmente en los casos en que las decisiones respaldadas por algoritmos tienen un impacto directo en las personas, como cuando se trata de asignar subvenciones públicas o conceder acceso a servicios públicos. La transparencia, junto con la auditoría y la gestión de riesgos, ayuda a evitar que los algoritmos se utilicen para crear o perpetuar las desigualdades.

[40] Para más información sobre este punto, véase, en particular, Diana-Urania Galetta y Herwig C. H. Hofmann, "Evolving AI-Based Automation – The Continuing Relevance of Good Administration," *European Law Review* (2023): 617 y ss.

[41] Reglamento 1689/2024, cit.

[42] Véase el comentario de Lorenzo Cotino Hueso y Pere Simón Castellano, eds., *Tratado sobre el Reglamento Europeo de Inteligencia Artificial* (Madrid: Editorial Aranzadi, 2024), que también ha sido traducido al inglés: Lorenzo Cotino Hueso y Diana-Urania Galetta, eds., *The European Union Artificial Intelligence Act. A Systematic Commentary* (Napoli: Editoriale Scientifica, 2025) y la doctrina allí mencionada.

8. El papel de la ciencia del derecho (administrativo) en la lucha contra la discriminación algorítmica: conclusiones

En conclusión, cabe destacar que, en primer lugar, hay que ser consciente de la «*liaison très dangereuse*» entre los algoritmos y los datos: si se utiliza un algoritmo de tipo determinista, porque este es capaz de elevar el error a sistema, reproduciendo el mismo error de forma totalmente automática un número infinito de veces; si se utiliza un algoritmo de aprendizaje automático, porque este es capaz de incorporar el error en el patrimonio de conocimientos personales en el que se basa su «entrenamiento», generando una especie de «efecto multiplicador» del propio error. Por lo tanto, en lugar de aumentar la inteligencia humana[43], es posible que los algoritmos actúen, de hecho, como multiplicadores de las imperfecciones de la inteligencia humana, creando un escenario que ya he definido irónicamente como de «*Human-stupidity-in-the-loop*»[44]. La decisión administrativa algorítmica discriminatoria podría, por lo tanto, revelarse como una consecuencia inevitable del acto de trasladar los mecanismos de decisión y las capacidades de toma de decisiones del hombre a la «máquina»: si no se adoptan las precauciones necesarias y no se tienen en cuenta las cuestiones y problemas a los que se ha hecho referencia en los párrafos anteriores.

Desde este punto de vista, el derecho, y el derecho público/administrativo en particular, está llamado a desempeñar su papel esencial de ciencia social que se ocupa del estudio y la reflexión sobre las normas (jurídicas) destinadas a garantizar la convivencia social pacífica, regulando las relaciones entre los componentes de la sociedad, así como entre la sociedad y sus componentes[45]. En concreto, el derecho está llamado a «justificar, controlar y orientar los caminos»[46] del uso de algoritmos como herramienta para apoyar y ayudar a la adopción de decisiones que afectan a los distintos ámbitos de la convivencia entre individuos[47] y que también pueden tener el efecto de crear o perpetuar patrones discriminatorios existentes.

Es decir, no necesitamos principios éticos, sino normas jurídicas adecuadas: porque el hecho de que los seres humanos seamos siempre «moralmente imperfectos»[48] implica que también nuestros procesos de toma de decisiones estén viciados desde el principio por esta imperfección nuestra, que podría transferirse simplemente a procedimientos de toma de decisiones automatizados o semiautomatizados mediante algoritmos capaces de producir un «efecto multiplicador». Este es el verdadero problema de los sesgos en las decisiones asistidas por algoritmos: la multiplicación exponencial del número de decisiones erróneas porque se basan en errores de evaluación, de juicio o cognitivos incorporados en el algoritmo. Lo cual, por supuesto, es aún más grave cuando se trata de decisiones de las administraciones públicas, como las relativas a la asignación de subvenciones públicas[49], cuyo objetivo sería alcanzar la igualdad sustantiva consagrada, por ejemplo, en el artículo 3 de la Constitución italiana[50] y no crear nuevas desigualdades[51] .

[43] Como ha hipotetizado, entre otros, Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (Nueva York: Viking Press, 1999).

[44] Diana-Urania Galetta, "Human-stupidity-in-the-loop? Reflexiones (de un jurista) sobre el potencial y los riesgos de la inteligencia artificial," 1 y ss.

[45] Véase Francesco Viola, "Il diritto come arte della convivenza civile," *Rivista di filosofia del diritto*, n.º 1 (2015): 57 y ss.

[46] Ibidem, p. 57 y ss. La traducción del italiano es mía.

[47] Como ya he señalado en Diana-Urania Galetta, "Decidere con l'IA: un problema comune a tutte le aree della scienza," *CERIDAP*, n.º 2 (2024): 374 y ss., https://ceridap.eu

[48] «Si los hombres fueran ángeles, no sería necesario ningún gobierno», escribía uno de los padres fundadores de la Constitución estadounidense, James Madison, "Federalist No. 51: The Structure of the Government Must Furnish the Proper Checks and Balances Between the Different Departments," *The Federalist Papers*, publicado en *The New York Packet*, 8 de febrero de 1788, https://guides.loc.gov/federalist-papers/text-51-60. La traducción del inglés es mía.

[49] Véase Massimo Airoldi, "Lo spettro dell'algoritmo e le scienze sociali. Prospettive critiche su macchine intelligenti e automazione delle disuguaglianze," *Polis* 34, n.º 1 (2020): 111 y ss.; Biagio Aragona, "Sistemi di decisione algoritmica e disuguaglianze sociali," *La Rivista delle Politiche Sociali*, n.º 2 (2020): 1 y ss.

[50] Art. 3 de la Constitución: «Todos los ciudadanos tienen la misma dignidad social y son iguales ante la ley, sin distinción de sexo, raza, lengua, religión, opiniones políticas, condiciones personales y sociales. Es tarea de la República eliminar los obstáculos de orden económico y social que, limitando de hecho la libertad y la igualdad de los ciudadanos, impiden el pleno desarrollo de la persona humana y la participación efectiva de todos los trabajadores en la organización política, económica y social del país».

[51] Paradigmático a este respecto es el informe de Matt Burgess, Ellen Schot y Georgina Geiger, "This Algorithm Could Ruin Your Life." Véase también: Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Nueva York: Crown, 2016); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Nueva York: St. Martin's Press, 2019).

En esta perspectiva, me parece más necesario que nunca promover lo que ya he definido en otra ocasión como la «cultura del derecho a una buena administración»[52]: ese derecho consagrado en el artículo 41 de la Carta de los Derechos Fundamentales de la Unión Europea y que, en mi opinión, adquiere aún más importancia en un contexto de decisiones que podrían «automatizarse» gracias al uso de algoritmos.

A este último respecto, es necesario precisar, a modo de conclusión, que, si bien sin duda debe permitirse a las administraciones públicas recurrir a las herramientas que hoy en día ponen a su disposición las TIC, cuando estas resulten adecuadas para garantizar una instrucción del expediente más completa y más acorde con los principios de imparcialidad y buen funcionamiento, también es necesario que ello se haga respetando el principio de transparencia, entendido ante todo en términos de plena cognoscibilidad y de la existencia de posibles procesos de decisión automatizados y del algoritmo. Esto es aún más cierto, por Italia, tras la introducción, mediante la Ley n.º 120/2020[53] de conversión del denominado decreto de simplificación, de un apartado 2-*bis* en el artículo 1 de la Ley n.º 241/1990 sobre el procedimiento administrativo, según el cual «Las relaciones entre el ciudadano y la administración pública se basan en los principios de colaboración y buena fe»[54].

Bibliografía

Accoto, Cosimo. *Il mondo dato: Cinque brevi lezioni di filosofia digitale*. Milán, 2017.

Alù, Andrea. "L'algoritmo dei sussidi sociali discrimina e fa cadere il governo: il caso olandese." *Agenda Digitale*. 19 de enero de 2021. https://www.agendadigitale.eu/cultura-digitale/algoritmi-troppo-invasivi-contro-le-frodi-fiscali-la-lezione-delle-dimissioni-del-governo-olandese/

Ambrosini, Maurizio, Deborah Erminio y Alberto Guariso. "L'accesso degli stranieri al Reddito di Cittadinanza: possibile discriminazione istituzionale." *Welforum.it*. 23 de octubre de 2023. https://www.welforum.it/un-caso-di-persistente-discriminazione-istituzionale-il-reddito-di-cittadinanza/

Ananny, Mike, y Kate Crawford. "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society* (2018).

Airoldi, Massimo. "Lo spettro dell'algoritmo e le scienze sociali. Prospettive critiche su macchine intelligenti e automazione delle disuguaglianze." *Polis* 34, n.º 1 (2020): 111–136.

Aragona, Biagio. "Sistemi di decisione algoritmica e disuguaglianze sociali." *La Rivista delle Politiche Sociali*, n.º 2 (2020).

Autieri, Tommaso. *La motivazione del provvedimento amministrativo: raccolta di dottrina, giurisprudenza e legislazione*. Padua: CEDAM, 2002.

Burgess, Matt, Ellen Schot y Georgina Geiger. "This Algorithm Could Ruin Your Life." *Wired*. 6 de junio de 2023. https://www.wired.com/story/welfare-algorithms-discrimination/

Cassatella, Antonio. *Il dovere di motivazione nell'attività amministrativa*. Padua: CEDAM, 2013.

Celotto, Alfonso. "Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto." *Analisi giuridica dell'economia* (2019).

Cerrillo i Martínez, Agustí. "Preventing Biased Results and Discriminatory Effects of AI Use in Public Administration." En *The EU Artificial Intelligence Act and the Public Sector. Humans and AI Systems in Public Administration in the Light of the European Regulation on Artificial Intelligence of 2024 (with Foreword by C. Coglianese)*, editado por Juli Ponce y Agustí Cerrillo i Martínez. Athens: EPLS, 2025.

Coglianese, Cary, y David Lehr. "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era." *Georgetown Law Journal* (2017).

"Transparency and Algorithmic Governance." *Administrative Law Review* 71 (2019): 1–60.

[52] Así ya en Diana-Urania Galetta, "Decidere con l'IA: un problema comune a tutte le aree della scienza."

[53] Ley n.º 120, de 11 de septiembre de 2020.

[54] La traducción del italiano es mía.

Cotino Hueso, Lorenzo. "Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público." En *Inteligencia artificial y sector público. Retos, límites y medios*, editado por Eduardo Gamero Casado y Francisco Luis Pérez Guerrero. Valencia: Tirant lo Blanch, 2023.

"Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida." *Revista Española de la Transparencia*, núm. 16 (2023): 17–63.

Cotino Hueso, Lorenzo, y Diana-Urania Galetta, eds. *The European Union Artificial Intelligence Act. A Systematic Commentary*. Napoli: Editoriale Scientifica, 2025.

Cotino Hueso, Lorenzo, y Jorge Castellanos Claramunt, eds. *Transparencia y explicabilidad de la inteligencia artificial*. Valencia: Tirant lo Blanch, 2024.

Cotino Hueso, Lorenzo, y Pere Simón Castellano, eds. *Tratado sobre el Reglamento Europeo de Inteligencia Artificial*. Madrid: Editorial Aranzadi, 2024.

Crawford, Kate. "Artificial Intelligence's White Guy Problem." *New York Times*. 25 de junio de 2016.

Desai, Deven R., y Joshua A. Kroll. "Trust but Verify: A Guide to Algorithms and the Law." *Harvard Journal of Law & Technology* (2017).

DiPietro, Louis. "Google Algorithm Found to Overlook Spanish Speakers in Online SNAP Ads." *TechXplore*. 10 de agosto de 2023. https://techxplore.com/news/2023-08-google-algorithm-overlook-spanish-speakers.html

Elliott, Mark C. "Has the Common Law Duty to Give Reasons Come of Age Yet?" *University of Cambridge Faculty of Law Research Paper*, n.º 7 (2012). https://ssrn.com/abstract=2041362

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Nueva York: St. Martin's Press, 2019.

Galetta, Diana-Urania. "Decision-Making and Information Management." En *The Model Rules on EU Administrative Procedures: Adjudication*, editado por Matthias Ruffert, 185–210. Europa Law Publishing, 2016.

"Le Model Rules di ReNEUAL e gli aspetti più innovativi della collaborazione fra amministrazioni nell'UE: procedimento amministrativo, scambio dei dati e gestione delle banche dati." *Rivista Italiana di Diritto Pubblico Comunitario* (2018), no. 2.

"Information and Communication Technology and Public Administration: Through the Looking-Glass." En *Information and Communication Technologies Challenging Public Law: Beyond Data Protection*, editado por Diana-Urania Galetta y Jacques Ziller. Nomos Verlagsgesellschaft, 2018.

"Algoritmi, procedimento amministrativo e garanzie: brevi riflessioni, anche alla luce degli ultimi arresti giurisprudenziali in materia." *Rivista Italiana di Diritto Pubblico Comunitario*, n.º 2 (2020): 501–526.

"Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale." *Federalismi.it*. Fasc. 5/2023. 22 de febrero de 2023. http://www.federalismi.it

"El derecho a una buena administración en un entorno de Administración pública digital. Reflexiones a partir del ejemplo de Italia." En *De Luis XIV al Estado inteligente*, editado por A. A. Martino. Buenos Aires: Editorial Astrea, 2022.

"Decidere con l'IA: un problema comune a tutte le aree della scienza." *CERIDAP*, n.º 2 (2024): 374–386. https://ceridap.eu

Artificial Intelligence and Public Administration. A Journey. Editoriale Scientifica, 2025.

"Digitalizzazione e diritto ad una buona amministrazione (Il procedimento amministrativo, fra diritto UE e tecnologie ICT)." En *Il Diritto dell'Amministrazione Pubblica digitale*, editado por Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. Turín, 2025.

Galetta, Diana-Urania, y Herwig C. H. Hofmann. "Evolving AI-Based Automation – The Continuing Relevance of Good Administration." *European Law Review* (2023).

Galetta, Diana-Urania, Herwig C. H. Hofmann y Jens-Peter Schneider. "Information Exchange in the European Administrative Union: An Introduction." *European Public Law* 20 (2014): 65–70.

"Informal Information Processing in Dispute Resolution Networks: Informality versus the Protection of Individual's Rights?" *European Public Law* 20 (2014): 71–88.

Galetta, Diana-Urania, Herwig C. H. Hofmann y Jacques Ziller. "Automazione, Intelligenza Artificiale e

Pubblica Amministrazione, fra diritto interno e diritto UE." En *Il Diritto dell'Amministrazione Pubblica digitale*, editado por Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. Turín, 2025.

Galetta, Diana-Urania, y Jacques Ziller, eds. *Information and Communication Technologies Challenging Public Law: Beyond Data Protection*. Nomos Verlagsgesellschaft, 2018.

Gilman, Michele. "AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead." *The Conversation*. 14 de febrero de 2020. https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625

Harwood, Simon. "The Parrot Is Not Dead, Just Resting: The UK Universal Credit System – An Empirical Narrative." *University of Edinburgh Business School Working Paper Series*. 2018.

Krafft, Tobias D., Katharina A. Zweig y Philipp D. König. "How to Regulate Algorithmic Decision-Making: A Framework of Regulatory Requirements for Different Applications." *Regulation & Governance* (2020).

Kurzweil, Ray. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Nueva York: Viking Press, 1999.

Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland y Patrick Vinck. "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes." *Philosophy & Technology* (2018).

Madison, James. "Federalist No. 51: The Structure of the Government Must Furnish the Proper Checks and Balances Between the Different Departments." *The Federalist Papers*. Publicado en *The New York Packet*, 8 de febrero de 1788. https://guides.loc.gov/federalist-papers/text-51-60

Mashaw, Jerry L. "Reasoned Administration: The European Union, the United States, and the Project of Democratic Governance." En *The Accountability of Expertise: Making the Un-Elected Safe for Democracy*, editado por Erik O. Eriksen. Londres, 2022.

Mir, Oriol, Herwig C. H. Hofmann, Jens-Peter Schneider y Jacques Ziller, dirs. *Código ReNEUAL de procedimiento administrativo de la Unión Europea*. Madrid: Instituto Nacional de Administración Pública, 2015.

Nicotra, Ida Angela, y Vincenzo Varone. "L'algoritmo, intelligente ma non troppo." *Rivista AIC* 4 (2019).

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Nueva York: Crown, 2016.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, 2015.

Rodríguez-Arana, Jaime. "La buena administración como principio y como derecho fundamental en Europa." *Misión Jurídica. Revista de Derecho y Ciencias Sociales* (2013).

Sassi, Silvia. "Gli algoritmi nelle decisioni pubbliche tra trasparenza e responsabilità." *Analisi giuridica dell'economia* (2019).

Simoncini, Andrea. "L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà." *BioLaw Journal – Rivista di BioDiritto* (2019).

"Amministrazione digitale algoritmica. Il quadro costituzionale." En *Il Diritto dell'Amministrazione Pubblica digitale*, editado por Roberto Cavallo Perin y Diana-Urania Galetta, 2.ª ed. Turín: Giappichelli, 2025.

Villata, Riccardo, y Margherita Ramajoli. *Il provvedimento amministrativo*. 2.ª ed. Turín: Giappichelli, 2017.

Viola, Francesco. "Il diritto come arte della convivenza civile." *Rivista di filosofia del diritto*, n.º 1 (2015): 57–72.

Webb, Helena, Ansgar Koene, Menisha Patel y Elvira Perez Vallejos. "Multi-Stakeholder Dialogue for Policy Recommendations on Algorithmic Fairness." En *Proceedings of the 9th International Conference on Social Media and Society*. Nueva York, 2018.

Wekesa, Sosten, y Nixon Otieno. "The Duty to Give Reasons under Kenya's Fair Administrative Action Act, 2015 in Kenya: Seven Years Later." *Kabarak Journal of Law and Ethics* (2022).

Artificial Intelligence and Social Scoring Systems: Between Dystopia and Reality

Inteligencia artificial y sistemas de puntuación social: entre la distopía y la realidad

Ginevra Cerrina Feroni
Catedrática de Derecho Constitucional Italiano y Comparado
Universidad de Florencia
Vicepresidenta del Garante per la protezione dei dati personali

Resumen:

Este trabajo examina los sistemas de puntuación social y de reputación algorítmica como una de las manifestaciones más intensas de la gobernanza automatizada contemporánea. A partir del caso paradigmático de la República Popular China y de su Sistema de Crédito Social, el artículo analiza la lógica de clasificación y vigilancia que subyace a estos mecanismos, así como sus efectos sobre las oportunidades, libertades y condiciones de vida de las personas. El estudio aborda también el caso neerlandés del *Toeslagenaffaire*, en el que el uso de algoritmos en la detección de fraude produjo efectos discriminatorios y decisiones gravemente lesivas, y la experiencia italiana en materia de rating reputacional, especialmente a la luz de la jurisprudencia y de las decisiones de la Autoridad de Protección de Datos. Sobre esta base, el trabajo sostiene que la expansión de la toma de decisiones automatizada exige reforzar las garantías de transparencia, control humano, proporcionalidad y protección de la dignidad de la persona. La reflexión final propone desplazar la atención desde la algocracia hacia una algorética orientada por los derechos fundamentales.

Palabras clave:

Inteligencia artificial; puntuación social; crédito social; rating reputacional; protección de datos.

Abstract:

This article examines social scoring systems and algorithmic reputational rating as one of the most far-reaching manifestations of contemporary automated governance. Taking the People's Republic of China and its Social Credit System as a paradigmatic case, the paper analyses the logic of classification and surveillance underlying these mechanisms, as well as their effects on individuals' opportunities, freedoms, and living conditions. It also addresses the Dutch Toeslagenaffaire, where the use of algorithms in fraud detection generated discriminatory outcomes and seriously harmful decisions, and the Italian experience with reputational rating, especially in light of case law and decisions of the Data Protection Authority. On that basis, the article argues that the expansion of automated decision-making requires stronger guarantees of transparency, human oversight, proportionality, and protection of human dignity. The final section advocates a shift from algocracy to an algoretics grounded in fundamental rights.

Keywords:

Artificial intelligence; social scoring; social credit; reputational rating; data protection.

Sumario:

1. Una distopía que ya es realidad. **2.** La República Popular China: el Sistema de Crédito Social. **2.1.** Planificación e implementación. **2.2.** Factores para evaluar el comportamiento de los ciudadanos. **2.3.** Efectos del Sistema de Crédito Social. **3.** Los Países Bajos: el *Toeslagenaffaire*. **3.1.** Los hechos. **3.2.** Aspectos jurídicos relevantes. **4.** La experiencia italiana. La calificación reputacional en la jurisprudencia de la Autoridad de Protección de Datos. **4.1.** El caso *Mevaluate*. **4.2.** El contexto escolar. **5.** Reflexiones finales. De la algocracia a la algorética.

Summary:

1. *A dystopia that is already reality.* ***2.*** *The People's Republic of China: the Social Credit System.* ***2.1.*** *Planning and implementation.* ***2.2.*** *Factors for evaluating citizens' behavior.* ***2.3.*** *Effects of the Social Credit System.* ***3.*** *The Netherlands: the Toeslagenaffaire.* ***3.1.*** *The facts. 3.2. Relevant legal aspects.* ***4.*** *The Italian experience. The reputational rating in the case law of the Data Protection Authority.* ***4.1.*** *The Mevaluate case.* ***4.2.*** *The school context.* ***5.*** *Concluding reflections. From algocracy to algoretics.*

1. A dystopia that is already reality

Let us imagine if every one of our actions, interactions, movements were reduced to an assessment on a five-point scale. A higher score opens the doors to fabulous opportunities and special advantages, while a low score can, essentially, isolate us from the rest of society. Thus, for example, by reason of how we answered the waiter at the café where we ordered a coffee, it could be forbidden for us to get on public transport, or depending on the evaluations of our previous neighborhood, the possibility could be precluded of moving to another district populated by subjects whose score is higher than a certain threshold. The subject of one of the most watched episodes of the science-fiction series "Black Mirror", this, however simplified, is how a system of social score or social credit works, that is, of social scoring or reputational rating.

But what still sounds like the dystopian plot of a science-fiction narrative is a reality that already surrounds us and directly concerns the citizens of some regions of the world.

One of the best-known and most evident social credit systems is the one introduced by the Chinese Government, although China is not the only legal system that is adopting solutions for social monitoring. In January 2022, the government of the United Kingdom revealed that it would use real-time facial recognition software on the streets of London to find suspects wanted by the police, and similar systems are already used by law-enforcement agencies and intelligence agencies around the world. In addition to public authorities, other entities, such as some insurance companies, have announced or already implemented systems that would allow their operators to effectively collect information about people's behavior to be used in important decisions. For example, companies specializing in life insurance, in New York, are authorized to make decisions about their clients using information found on social networks. After all, the automobile insurance sector already uses black boxes installed on cars as part of the contract, capable of calculating the malus on the basis of the driving style that they perceive.

There are programs that allow owners of restaurant or entertainment businesses to create lists of people not authorized to enter because of previous improper behavior. And all sharing-economy services, such as Airbnb, Uber or home delivery services, have a kind of scoring system that evaluates those who participate in the service both as workers and as consumers. With the Covid-19 health emergency, the issue of monitoring citizens imposed itself in an even more pervasive manner, concentrating upon itself the radical cultural divergences, before political and legal ones, regarding the possibility that regulatory authorities around the world could or could not hypothesize solutions to trace infected or unvaccinated citizens and, on the basis of their health data, allow or prohibit for them access, opportunities and even the very exercise and enjoyment of constitutionally guaranteed rights and freedoms.

Hannah Arendt opens Part Three of her fundamental work on the origins of totalitarianism with a quotation from David Rousset: "ordinary men do not know that everything is possible"[1]. The real risk, therefore, is not that of failing to notice that certain scenarios that we did not even know we could imagine are today reality and thus failing to protect ourselves from them, but precisely of failing to notice them because we are getting used to them. To address the topic it is worth dwelling, albeit in broad strokes, on the to-date most developed experience of a social credit system, namely the case of the People's Republic of China. We will then go on to recall a social credit affair that concerned the Netherlands and was the subject of a broad public debate, and then analyze the Italian case, also in the light of the most significant case law of the Data Protection Authority.

[1] Hannah Arendt, *The Origins of Totalitarianism* (New York: Schocken Books, 1951), 421.

2. The People's Republic of China: the Social Credit System

The Social Credit System (SCS) is an initiative developed by the People's Republic of China since the early two-thousands with the aim of developing a national system to classify the reputation of its citizens. Through the SCS each citizen is assigned a score representing his "social credit," on the basis of information held by the Government, such as those regarding his economic and social condition, his preferences and his habits, thus operating essentially as a mass-surveillance mechanism entrusted to AI technologies for the analysis of big data[2]. In addition to individual citizens, as will be seen, the SCS has also been extended to the activities of all enterprises operating in the Chinese market.

2.1. Planning and implementation

The Social Credit System is placed within the Chinese so-called "top-level design" approach and is coordinated by the "Central Leading Group for Comprehensively Deepening Re-forms". According to the "Planning for the Establishment of the Social Credit System (2014-2020)" published by the State Council, integrated with the recent "Social Credit Action Plan 2024-2025"[3], the Social Credit System focuses on four areas: "honesty in government af-fairs," "commercial integrity," "social integrity," and "judicial credibility".

The objective of the initiative is twofold: on the one hand there is the individual plan that per-tains to the ethical sphere of the actions of the individual: in this dimension the SCS is aimed at identifying and punishing deviance and at incentivizing sincerity, honesty, and integrity of individual Chinese citizens[4]. On the other hand - and perhaps this is the most neglected aspect of the phenomenon - the intent is expressly declared to create prodromal conditions for the perfection and regulation of the economy of a socialist market[5]. The strengthening of the State's dirigiste capacity and its ability to steer the behavior and actions of citizens are in fact aimed not only at needs of social control, but also at the need for uniformity of the Chinese market. From this latter point of view the SCS project pursues two fundamental economic objectives: on a small scale, to eliminate market pathologies, such as fraud and counterfeiting; on a large scale, to solve the problem of financial reliability within and by the Chinese market and, in this way, to contribute to realizing the union between social harmony and the harmo-nious economic development of the Country[6].

The Social Credit System, for now, appears confined to mainland China, and not also to Hong Kong and Macao. Moreover, the plans for the enlargement and development of the system do not speak of a distinction between Chinese enterprises and foreign enterprises op-erating in the Chinese market, implying the possibility that foreign enterprises operating in China are also subjected to classification within the system.

The system has already been involved in quite a few controversies, particularly with regard to the way in which it will be applied to people and to enterprises, and yet neither the cases won before the courts nor the errors made by the algorithm and recognized as such by the authori-ties themselves have led to a modification of the process of functioning of the same[7].

As regards citizens, examples are already detectable of punishments caused by the violation of social protocols: the system has already denied nine million people with "low scores" the right to

[2] Among the most recent works that have attempted to systematically describe and define the topic: C. Liu and A. Rona Tas, "Trusting by Numbers: An Analysis of a Chinese Social Credit System Governance Infrastructure," *Critical Sociology* 51, no. 6 (September 2025): 1247–1265; C. Loefflad, M. Chen, and H. Grossklags, "Reputational Discrimination and Fairness in China's Social Credit System," *Digital Government* 5, no. 4 (December 2024): 1–27; see also X. Dai, "Toward a Reputation State: A Comprehensive View of China's Social Credit System Project," in *Social Credit Rating*, ed. O. Everling (Wiesbaden: Springer, 2020), 139–165; B. Ahl, L. Catá Backer, and Y. Chen, "Law and Social Credit in China," *China Review* 24, no. 3 (2024): 1–15; S. Hofmann, *Social Credit: Technology-Enhanced Authoritarian Control with Global Consequences*, International Cyber Policy Center, Policy Brief Report no. 6 (2018).

[3] National Development and Reform Commission, *Social Credit Action Plan 2024–2025* (June 2024), testo in inglese su https://www.chinalawtranslate.com/en/2024-2025social-credit-plan/

[4] Not by chance, R. Creemers, *China's Social Credit System: An Evolving Practice of Control*, SSRN, 2018, 2, notes that in Mandarin the word "credit" (xinyong) is semantically identical to the term "integrity".

[5] B. Ahl, L. Catá Backer, Y. Chen, "Law and Social Credit in China", cited work, esp. pp. 7 ff.

[6] L. Yu-Hsin Lin, C. Milhaupt, *China's Corporate Social Credit System: The Dawn of Surveillance State Capitalism?*, The China Quarterly, 256, 2023, 835–853, esp. 836; see also D. Mac Sithing, M. Siems, *The Chinese Social Credit System: a Model for other Countries*, Modern Law Review, 82(6), 2019, 1034-1071, esp. pp. 1048 ff.

[7] M. van Blomberg, *The Social Credit System in China's Rule of Law*, Mapping China Journal, no. 2, 2018, pp. 78-162, esp. p. 100.

buy domestic flights. The system is also used to exclude the possibility for parents to enroll their children in certain schools, to prohibit those who have a low score from renting hotel rooms, from using credit cards, and to insert certain individuals on a blacklist with the aim of precluding for them the obtaining of work[8]. The system has also been used to evaluate people's internet browsing habits (too much time spent playing video games reduces the score for example), purchasing habits and a variety of personal and absolutely harmless ac-tions that have no impact on the community[9].

The "Notice of the State Council regarding the issuance of the 'Planning for the Establish-ment of the Social Credit System (2014-2020)'" was issued by the Chinese State Council on June 14, 2014. In 2015 a license was granted to eight companies so that they could begin to develop prototypes of credit systems. The selected enterprises are Ant Financial of the Aliba-ba Group, the software developer Tencent and six others. The credit systems created by these enterprises used databases such as Sesame Credit at the beginning of the experimentation phase[10].

Until now, there exists no comprehensive social credit system extended to the whole nation, but multiple experimental projects have been implemented that are testing the system on a local scale and in some sectors of the economy. A program of this type was implemented in Shanghai by means of the Honest Shanghai app, which uses facial recognition software to search within government archives and classify users in this way. The scores can also be based on information extrapolated from the online search habits of Chinese citizens.

The infrastructure of the Social Credit System was completed in 2020 and strengthened in recent years, but the restrictions on citizens and enterprises with "low scores" began to take effect starting already in 2018.

With regard specifically to enterprises, instead, as already said, the Social Credit System was designed as a mechanism of market regulation. The objective is to establish a structure of self-imposed regulation, powered by big data, within which enterprises control each other. The basic idea is that with a functional credit system active, companies will endeavor to comply with Government policies and provisions in order to avoid their score going down. As it was conceived, enterprises with high scores will have advantages such as better conditions on loans, lower taxes and more investment opportunities. Enterprises with low scores will re-ceive disadvantageous conditions for new loans, higher taxes, restrictions on investments and fewer opportunities to participate in projects financed by the public sector. The Government's plans also provide for real-time monitoring of the activities of enterprises. In that case, in-fringements by an enterprise could take shape in an instantaneous lowering of the score[11].

2.2. Factors for evaluating citizens' behavior

The Chinese social credit system is based on three variables:

1. A first factor, not dissimilar indeed from that taken into consideration in "private" social credit systems (e.g., Schufa in Germany, or FICO in the USA) is that linked to the ability of each citizen to regularly fulfill his own financial obligations. A person who pays his bills punctually can be considered reliable and solvent and for this reason will receive a higher score.

2. A second factor concerns the ability of each individual to fulfill his own duties as defined in the contracts entered into. For example, an individual with a substantial level of savings will receive a higher score than a person who struggles to make it to the end of the month.

3. A third factor, the one that presents the greatest criticalities, concerns social behavior extensively considered: purchasing orientations on online portals such as Alibaba, the trends of offline purchases paid with Alipay, social interactions on WeChat, the online history of Baidu, the areas of frequentation (identified through geolocation of the cell phone), that is in general those elements that can give an all-encompassing idea of the personality of each citizen. For example, a user who buys diapers is very likely a parent and, therefore, also a more responsible citizen. On the contrary, a citizen who spends large sums on online games is considered lazier and less productive for society. It is easy to understand how such results can be obtained only by an algorithmic evaluation system

[8] Among the many who reported these examples, D. Mac Sithing, M. Siems, *The Chinese Social Credit System: a Model for other Countries*, cited; and B. Ahl, L. Catá Backer, Y. Chen, *Law and Social Credit in China*, cited.

[9] A. Fenwick Elliott, *China is banning people with bad 'social credit' from using planes and trains*, The Telegraph, 19 March 2018.

[10] M. van Blomberg, *The Social Credit System in China's Rule of Law*, cited, p. 93 (Sesame Credit description).

[11] See B. Ahl, L. Catá Backer, Y. Chen, *Law and Social Credit in China*, cited, esp. pp. 7 ff.

whose source code is highly influenced by social values of collective interest[12].

2.3. Effects of the Social Credit System

The Chinese system does not limit itself to analyzing the behavior of the individual[13]. It is in fact provided that each individual's score is also influenced by the behavior of the people closest to him. For example, a subject who violates one or more laws of the Chinese Communist Party will also negatively influence partners and persons connected to him[14].

The effects of the Social Credit System can concern a series of activities.

a) Flight bans: currently, the flight ban for persons considered unreliable is already common practice in China[15].

b) Exclusion from private schools: if the parents' score were to be below a certain threshold, their children would be excluded from the best schools in the region.

c) Slowing of the internet connection: citizens considered unreliable could see their internet connection slowed down or even be completely excluded from access to specific websites.

d) Exclusion from high-prestige jobs: the score of the Chinese SCS could in the future be part of a person's curriculum vitae and thus provide essential information to potential employers. In this way for people with a low score, it would almost surely be [impossible] to reach the most sought-after and qualified work positions.

e) Registration on a public blacklist: a prototype of a blacklist currently exists in China and all those who are registered on such a list could expect one or more of the penalties mentioned previously.

Another crucial step in the process of constructing the SCS is also represented by the parallel reward system that is envisaged for citizens aimed at incentivizing the latter to behave according to the rules in order to be able to enjoy facilitations with reference to the same type of activities that are forbidden to others. The "virtuous circle" that follows consists above all in the advantage that the Chinese State is not required to provide any direct intervention either of control or corrective on this type of activities[16]. Among the reward measures, one can recall facilitated access to financing (in the Sesame Credit pilot project, those who reach a score of 600 can request a loan up to an amount of 5,000 Yuan) and to rentals or leases (including exemption from the obligation of a security deposit), facilitation of travel (depending on the score, fewer documents are required to justify the request for travel visas) and the increase in social status that follows its official recognition (one's own score can be used as a status symbol on social platforms and dating apps)[17].

3. The Netherlands: the *Toeslagenaffaire*

Although the Chinese case of profiling on a reputational basis certainly remains the most extreme, nonetheless it constitutes only one part of a broader phenomenon. One cannot fail to note, in fact, that, favored by the growing power of the algorithm, the use of social scoring on the public-law level has found its first applications on a global scale and, indeed, in Europe. The most striking case, which in January 2021 led to the resignation of the then Prime Minister Mark Rutte, concerns the so-called *Toeslagenaffaire*, by reason of which it was brought to light that the Dutch tax administration and, in particular, its *Toeslagen* unit (the welfare department) had unlawfully requested the retroactive repayment of subsidy allowances received by about 35,000 parents. It was precisely the use of the unlimited automation of machine-learning algorithms that played a significant role in the scandal which, moreover, caused the Dutch public finances damage of about 500 million euros, thus revealing the uneconomical nature of the entire operation.

[12] Hence the co-ordinate value that Chinese civilization attributes to social norms and to law. See Z. Zuo, *Governance by Algorithm: China's Social Credit System*, University of Cambridge Working Paper, 16 June 2020.

[13] J. Chin and L. Lin, *Surveillance State*, St. Martin's Press, 2022, pp. 219 ff.

[14] *China's Chilling Social Credit' Blacklist*, Human Rights Watch, 12 December 2017.

[15] *China Releases Investigative Journalist After Almost Year in Jail*, Newsweek, 3 August 2014.

[16] M. van Blomberg, *The Social Credit System in China's Rule of Law*, cited, esp. p. 93 ff.

[17] As reported by S. Johnson, How China's 'social credit score' will punish and reward citizens, Big Think, 26 April 2018, the Baihe website, for example, already allows its users to publish their own score.

3.1. The facts

Towards the end of 2011 the Dutch press revealed a series of cases of fraud against social security concerning payments made by the tax administration in favor of people who resided in other EU Member States[18]. This series of cases, christened by the press *Bulgarenfraude* (Bulgarian fraud)[19], thus represented the incentive to reform the welfare system, in order to increase the efficiency of the process of identifying fraud and recovering undue emoluments. This streamlining was pursued through various means including the implementation of machine-learning algorithms to automate the identification of culpable errors or fraud in the processing of social-security forms submitted by parents. In practice, however, the same sanctions, namely the suspension of allowances and the request for restitution of what had been unduly received, were imposed not only for intentional ideologically false crimes, but also for material errors and simple formal oversights, such as, for example, the erroneous completion of a box or the omission of a signature. This was due to the fact that Art. 26 of the Income Schemes Act (AWIR) did not contemplate the obligation for the administration to apply the principle of proportionality[20].

The use of this AI algorithm for "zero-tolerance" risk detection therefore obliged the full repayment of the debt with no possibility of its temporal deferral and often with the addition of the payment of a surcharge, even for individuals to whom only material errors in completing the application forms for the welfare contribution could be imputed.

The welfare department, moreover, had used a machine-learning algorithm also in the risk-assessment process for the selection of beneficiaries of childcare allowances to be subjected to specific checks. As is typical of machine-learning algorithms, the system inferred the existence of risk factors on the basis of the analysis of historical data, that is, of previous known cases of fraud. This is, as has been known for some time, a method that is not free from risks of producing biases and discriminations, which for this purpose requires a twofold protocol of safeguards: first of all the data processed both during and after the automatic-learning process must accurately represent the target population, that is, the recipient of the welfare measure; secondly the risk factors (the so-called "weights" in the terminology of machine learning) must be as free as possible from any kind of influence that concerns the conduct, generally considered, held by minorities or by individuals of lower socio-economic status. It is no coincidence that, when it is said that algorithms (and also predictive algorithms) are based on decisions already taken, this means that it is not the algorithm that contains the bias, since what is discriminatory is the historical sequence of decisions that lies upstream. For these reasons, circumventing such distortion requires vigilance, constant monitoring, spot checks and also a certain degree of skepticism with regard to the conclusions of the algorithm[21].

In the *Toeslagenaffaire* the Dutch Data Protection Authority (AP) and the National Audit Service (ADR) showed how the attitude held by the tax administration was in reality of a completely opposite sign. The *Toeslagen* unit would not have exercised prudent oversight nor would it have verified whether the algorithm was free of bias, indirectly inducing the algorithm to return distorted results. Proof of this is the so-called "twin test"[22] carried out by the AP and from which it emerged that, all other conditions being equal, the algorithm flagged a higher risk of fraud on the part of individuals who were not of Dutch nationality compared to Dutch citizens who presented similar characteristics[23].

[18] According to calculations by the Dutch Data Protection Authority (Autoriteit Persoonsgegevens), "Werkwijze Belastingdienst in strijd met de wet en discriminerend," July 17, 2020, the events led to net revenue losses of about €10 million for the Belastingdienst, while the Dutch government had to step in by creating a €500 million fund to compensate the victims.

[19] The label was linked to the fact that some of these fraud cases involved families of Bulgarian nationality: "*Uitspraak Bulgarenfraude: hoe zat het ook alweer?*", on RTL Nieuws, 19 May 2015.

[20] See also Venice Commission, Council of Europe, Opinion no. 1030/2021, p. 5. The legitimacy of that law had been upheld on several occasions by the Dutch Supreme Administrative Court, which deemed lawful the failure to apply the principle of proportionality to the seriousness of the offence.

[21] Parlementaire ondervragingscommissie Kinderopvangtoeslag, *Tweede Kamer der Staten-Generaal* 11 July 2020.

[22] A "twin test" is a simple procedure in which two identical fictitious profiles are created, differing only by one characteristic; in this case, the characteristic was "Dutch/non-Dutch."

[23] Thus the Autoriteit Persoonsgegevens, "*Belastingdienst beloofde ambtenaren niet te straffen om toeslagenaffaire*", 20 May 2019, according to which, in substance, the algorithm discriminated against foreign residents. In its report, the AP describes a particularly troubling case in which, based on alerts concerning a specific childcare institution where about 100 parents of Ghanaian origin were suspected of fraud, the *Toeslagen* unit arbitrarily decided to investigate all 6,047 parents of Ghanaian origin living in the Netherlands. Under such conditions - where supposedly "objective," "data-driven" decisions are in fact arbitrary, biased, and discriminatory - it is hard to deny that the machine adopted a discriminatory decision, much as human tax officials might have done. Moreover, the case of the Ghanaian parents is not the only instance in which foreigners were subjected to an arbitrary decision by the leadership of the *Belastingdienst*.

Adding up little by little, these arbitrary manual targetings of foreigners in the machine-learning system, which learns and constantly updates the risk factors on the basis of the historical data entered by the Tax Administration officials, induced a veritable distortion in the algorithm, which heuristically concluded that welfare beneficiaries of foreign origin and citizenship were more inclined to fraud. Even more problematic is the fact that welfare beneficiaries, considered potential fraudsters, suffered the interruption of the benefit without even any *ex post* assessment by human operators.

3.2. Relevant legal aspects

The first lesson to be drawn from the Toeslagenaffaire is that one cannot automate a process that has original flaws, since automation will only amplify its scope. Machine-learning algorithms were used in the Toeslagenaffaire to reduce the serious backlog in the process of recovering the childcare allowance. However, since the recovery process was in the first place gravely flawed - particularly due to the lack of application of the principle of proportionality in the case of material errors by welfare beneficiaries - the algorithms determined a result completely opposite to what was envisaged: that is, they further increased the backlog at the *Belastingdienst* and the number of appeals brought by welfare beneficiaries against the Toeslagen unit[24].

The second conclusion is that States should use machine learning and automation by arranging ex ante the legal and technological conditions so that these algorithms can be used lawfully and ethically. Well before the Toeslagenaffaire, it had been repeatedly pointed out how automatic learning was inclined to create serious risks of biases and discriminations, as has repeatedly emerged over the last years, for example with reference to the right to a fair trial, to good administration and to other principles derivable from constitutional or interposed norms[25]. Starting from this, the adoption of machine-learning algorithms, in the absence of strengthened safeguards and the protection of these rights, is inevitably the harbinger of perverse effects.

Which leads to the third lesson to be gleaned from the Dutch case: the necessary control by the human operator. The systematic adoption of machine-learning tools roughly coincides with the global financial crisis of 2008-2011, in which austerity measures led to a reduction in the workforce of tax administrations throughout the European Union. This should in itself be a cause for concern, since it is likely that a reduction in staff affects the number of officials in charge of taxpayers' ex post complaints, and therefore their review capacity. By doing so, we are not creating the conditions for an anthropocentric AI, but rather a decision-making system centered on uncontrolled artificial intelligence, completely antithetical to the objectives that one intends to promote with algorithmic-governance projects of a Community stamp.

The reported affair is, therefore, of great interest in that it is easy to imagine that given the prerequisites that determined it, it is not an isolated phenomenon, but that, on the contrary, it may be repeated in other EU Member States.

4. The Italian experience. The reputational rating in the case law of the Data Protection Authority

The proliferation of applications or private databases that indiscriminately collect and store personal data on the net in order to make decisions also with negative effects and consequences on the private, social and economic life of individuals has also reached our country. These projects aggregate information coming from different sources, not always of an objective nature - e.g. also posts from social profiles - and extend the evaluation model, now consolidated and applied to accommodation or restaurant facilities, directly also to the evaluation of people.

These are systems that aim, in substance, to make measurable the reputation of the registered subjects, raising more than one concern in terms of personal-data protection, notwithstanding that the declared purposes are almost always justified by the pursuit of significant objectives such as, for example, the fight against corruption, against false identities, against fraud, etc. But the very idea of entrusting the "review" of a person to an algorithm risk truly opening a very dangerous drift. Hence a reflection is required right from the start that goes beyond the technicist application of the

[24] M. Fenger, R. Simonse, *The implosion of the Dutch surveillance welfare state, in Social Policy & Administration*, vol. 58, Issue 2, 2024, pp. 264-276.

[25] [27] C. Castro, *What's Wrong with Machine Bias*, Ergo, 5(15), 2019–2020; J. Angwin et al., Machine Bias, ProPublica, 23 May 2016; E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in *Consulta Online*, 20 march 2020, online su www.giurcost.org, in part. pp. 9 ff.; A. Gatti, *L'algoritmo tra volontà e rappresentazione*, in *DPCE online*, n. 3, 2020, pp. 3457-3461.

GDPR rules (and in particular of Art. 22, on automated decisions including profiling and subsequent ones on security), to dwell rather on the general principles enshrined by it.

Reputation is a complex element and proper to an entire personality, corresponding to several aspects that concern multiple personal data that are also significant and sensitive; therefore, as such it is protected by the GDPR and by the Privacy Code and is guaranteed by an independent Authority. The prejudice to reputation is expressly listed among possible damages consequent to a personal-data violation of the data subject, pursuant to Recital 85 of the GDPR. An emblematic case can occur precisely in the hypothesis of publication of personal data processed for profiling purposes.

It is evident that the reputation of individuals that one would like to quantify in order to assign reliability scores is closely linked to the person considered in his social projection and is intimately connected with dignity, a keystone of the discipline on personal-data protection, which looks precisely at the impact of inconsiderate processing, especially if automated and of profiling, on the rights and freedoms of individuals, paying attention so that the data subject can always exercise his own control over his data processed by others[26].

Many doubts concern the actual quality of the data collected - destined precisely to be processed by specific algorithms - or the genuineness of any reviews entered by third parties that can easily lend themselves to distorted, defamatory or harmful uses. Judgments that risk seriously compromising the right to the integrity of personal identity and whose prejudicial effects are irreparably amplified by users' free access through search engines. And crystallized in the hypermnesia of the net for an indefinite time.

Further evaluations are imposed both by the concrete methods of managing such rating systems and by the discretion of the measurement criteria identified by the companies that intend to offer such services on the market. It is no secret that in the past managers, in some cases, even knowingly altered the reference market by inserting fake reviews. These are projects and practices that, in the face of the various critical issues raised, have a considerable impact on the personal sphere of users, on their right to informational self-determination. This dimension, whose centrality has been reaffirmed by the case law of the Court of Justice of the European Union, which since 2014 has recognized the exercise of a new right to digital oblivion, is therefore crucial[27].

A human being is not a company name, a commercial asset, a service and cannot and must not become one[28]. The Court of Luxembourg has established that, in the event of conflict, the mere economic interest succumbs, when certain conditions occur, in relation to the category of fundamental rights that are an expression of the principles of dignity contained in the Charter of Rights[29].

4.1. The Mevaluate case

With measure no. 488/2016, the Privacy Authority judged as unlawful the platform of the "Mevaluate Onlus" Association which collects numerous categories of personal data (pertaining to the following spheres: criminal, tax, civil, work and civic engagement, studies and training) for the calculation of a reputational rating of the applicant (natural or legal person), in order to "make socio-economic relations more efficient, transparent and secure". Such information is collected on the basis of the consent of the data subjects and is provided by sources other than the data subjects, possibly supplemented by "press articles, radio/TV" and other supporting documents optionally produced by the data subjects.

As a premise the Authority observes that the rating processed by the system "could have a heavy impact on the (also private) life of the registered individuals, influencing their choices and prospects and conditioning their very admission to (or exclusion from) specific services, services or benefits";

[26] On data profiling from a constitutional point of view, see the very recent contribution by A. Gatti, *Profilazione e diritti fondamentali*, ES, Napoli, 2025.

[28] As early as 2015, President Antonello Soro argued: "On the basis of these premises, it becomes difficult to implement databases or services aimed at profiling a person - even in moral and relational terms - in order to assess his or her alleged reliability. The European legal order protects privacy, recognized as a fundamental human right, and forbids 'rating' any person in the absence of a public interest, which certainly cannot be invoked when one enters the world of personal and sentimental relationships of anyone who has a social-network account." Antonello Soro, *La reputazione online tra principi di dignità e diritto all'oblio*, Huffington Post, 6 October 2015. On the connection between reputational rating and dignity, see E. di Carpegna Brivio, *Pari dignità sociale e Reputation scoring. Per una lettura costituzionale della società digitale*, Turin, Giappichelli, 2024.

[29] CJEU, Kranemann v Land Nordrhein-Westfalen, 17 March 2005 (ECLI:EU:C:2005:187).

and that the matter must be approached with extreme caution, since "the reputation that [the platform] would like to measure, [...] closely correlated to the consideration of people and to their very projection, social, is intimately connected with their dignity, a keystone of the discipline of personal-data protection".

With reference to the lawfulness of the processing the Authority underlines that the processing was not based on an appropriate source of regulation, since the legal system recognizes and regulates exclusively the "business rating"[30] and not that relating to natural persons. Moreover, the consent of the data subjects, in order to be in accordance with the law, should have been expressed freely, a condition that does not occur in the case of autonomous processing of documents freely usable (e.g. judgments) by an automated system, both with reference to the profiles relating to users who have expressly consented to the processing, and with reference to the "profiles against third parties" relating to subjects not registered on the platform; not to mention that the processing was based on inadequate information.

The processing in question then appeared in violation of the principles of data minimization, since it was based on a massive collection of data and documents, and of proportionality of processing, in that the relevance and pertinence of the data and documents collected appeared doubtful and, in any case, unproven and the period of data retention was not adequately justified. Finally, the processing, involving a potentially very large number of subjects, would have had reliable significant repercussions for the rights of the data subjects in the event of a breach of security measures and the latter were not endowed with sufficient reliability.

The Mevaluate Association challenged the Authority's measure before the Court of Rome which, by judgment no. 5715/2018, held that the lack of regulatory discipline on reputational rating did not prevent the development of infrastructures capable of assigning it. Of the opposite view compared to the Authority, which had considered that the complex system of collection and processing of personal data at issue would have been able to affect both the economic and social representation of a broad category of subjects and the private life of the registered individuals, the ordinary judge noted how private evaluation and certification bodies, recognized also for the purposes of certification of quality and/or conformity to technical standards, are now widely used.

The Court, establishing that "one cannot deny to private autonomy the power to organize accreditation systems of subjects, providing broadly 'evaluative' services, in view of their entry into the market, for the conclusion of contracts and for the management of economic relations," thus considered the processing to be in accordance with the requirement of lawfulness of processing, starting from the assumption that "the activities of uploading information and of validation and certification of the documents are subject to the consent of the data subject and to the voluntariness of his action". On the basis of such arguments, therefore, it annulled the Authority's measure, subject to the prohibition of processing the personal data of subjects not registered on the platform, even if taken from documents that are freely knowable.

With reference, instead, to the profiles against third parties and to the other issues raised by the Authority, the Court confirmed the unlawfulness of the processing carried out by Mevaluate.

On the question of the validity of the consent given by the data subject, the Court of Cassation ultimately ruled, which, in judgment 14381/2021, upheld the Authority's appeal and expressed the following principle of law: "in the matter of processing of personal data, consent is validly given only if it is expressed freely and specifically with reference to clearly identified processing; it follows that in the case of a web platform (with annexed IT archive) preordained to the processing of reputational profiles of individual natural or legal persons, centered on a calculation system with an algorithm at its base aimed at establishing reliability scores, the requirement of awareness cannot be considered satisfied where the executive scheme of the algorithm and the elements of which it is composed remain unknown or not knowable by the data subjects"[31].

In fact, the Supreme Court underlined that for the lawfulness of processing based on consent, Art. 23 of Legislative Decree No. 196 of 2003 (i.e. the pre-GDPR privacy code, in force at the time of the facts as the reference legislation) presupposed not only consent tout court, but also that the consent was validly given[32]. Specifically, indeed, Art. 23 provided that "(a) the processing of personal data by private parties or public economic bodies is permitted only with the express consent of

[30] Managed by the National Anti-Corruption Authority (ANAC) on the basis of Article 83(10) of Legislative Decree No. 50 of 18 April 2016.

[31] Decision of the Italian Data Protection Authority (Garante per la protezione dei dati personali) No. 488 of 24 November 2016 [web doc No. 5796783]

[32] Corte di Cassazione, decisions No. 17278/2018 and No. 16358/2018

the data subject; (b) consent may concern the entire processing or one or more operations of the same; (c) consent is validly given only if it is expressed freely and specifically with reference to 'clearly identified' processing, if it is documented in writing, and if the information referred to in Art. 13 has been given to the data subject; (d) consent is expressed in written form when the processing concerns sensitive data".

In the present case, therefore, the processing was (and is) functional to determining the reputational profile of the subjects and, consequently, the assessment of the lawfulness of such processing, based on consent, could not be envisaged by the court without a prior consideration of the elements likely to affect the seriousness of the manifestation, including the elements implicated and considered in the relevant algorithm, the functioning of which is essential to the calculation of the rating[33].

In particular, the Court of Cassation highlights that the poor transparency of the algorithm with respect to the specific purpose to which it was preordained had not been considered decisive by the appealed judgment. This approach is not shared by the Supreme Court which maintains that the prerequisite of the lawfulness of processing is precisely constituted by the validity of the consent that is assumed to have been given at the time of accession. And it cannot logically be affirmed that the accession to a platform by the members of society also includes the acceptance of an automated system, which makes use of an algorithm, for the objective evaluation of personal data, where the executive scheme in which the algorithm is expressed and the elements considered for the purpose are not made knowable.

As is well known, the observations of the Supreme Court, pertinent and precise, are amply confirmed by the GDPR where Art. 4, paragraph 11, establishes that the consent of the data subject, in an unequivocal manner, must first of all be free because if the data subject does not make a real choice and feels obliged to give his consent also to avoid negative consequences in the event that he refused, then the consent cannot be considered valid. If, for example, consent is inserted in a non-negotiable part of terms and conditions it is presumed that it has not been given freely. Moreover, consent must be specific, since if it is used to justify multiple processing operations it must be freely given for each one. Data subjects must be able to choose for which purposes they consent to processing. Finally, consent must be informed: without accessible information, data subjects cannot make informed decisions and therefore, clarifies the WP29 (in its 2017 guidelines) "user control would become illusory and the consent invalid for the processing".

With reference then to forms of automatic profiling, it must be pointed out that in the field of decisions based solely on automated processing, as provided by the already mentioned Art. 22, the Regulation introduces the need to provide the data subject with more information on the methods of creation and use of these processes. In this regard, indeed, the GDPR finds significant risks for the rights and freedoms of individuals connected to the tendency toward opacity of automated processes and mechanisms, which often leads the individual, the subject of profiling, not to be aware of it and to the creation, by the controller, of new data, additional to the original ones, which could reduce the data subject to a category to which he does not recognize himself, thus conditioning his choices and, in some cases, also leading to forms of discrimination.

Therefore, the GDPR, to correct this informational imbalance between controller and data subject and to avoid prejudice to the legal sphere of the latter, identifies a series of requirements on which to focus in order to make these automated processing operations compliant with the legislation: specific prescriptions on transparency and fairness; greater obligations of accountability; specific legal bases for the legitimization of processing; safeguards for individuals in terms of the right to object to profiling and, in particular, to profiling for marketing purposes; the conduct of a data-protection impact assessment where certain conditions are not met.

In particular, Art. 13, para. 2, letter f) and Art. 15, para. 1, letter h) establish the right of the data subject to know of the existence of automated decision-making and, in particular, to obtain meaningful information about the logic used (the criteria assumed to reach the decision, without thereby necessarily having to provide a complex explanation of the algorithms used) and about the envisaged consequences of such processing (through examples one will have to provide information on how the automated process could in future affect the person concerned).

Taking into account the significant risks to the rights and freedoms of the data subject for these types of processing, the Regulation, on the one hand, obliges the controller to implement appropriate and strengthened protective measures (it will also be important to provide methods that regularly verify the correctness of the processes to limit classification or evaluation errors with a negative impact on the profiled subjects), on the other hand, recognizes the power of the data subject to obtain human intervention on the part of the controller, to express his opinion and to contest the

[33] F. Galli, *Rating reputazionale e diritto alla spiegazione,* in *Labour & Law Issues*, vol. 9, No. 2, 2023, pp. 62-97.

decision, in cases where such decision is provided for by contract or consented to by the data subject (Art. 22, para. 3). Among other things, an automated decision-making process involving special categories of data, as per Art. 9, para. 1, is permitted only in the presence of the explicit consent of the data subject or for reasons of important public interest on the basis of Union or Member State law.

It is worth recalling, in fact, what is established by Art. 22, para. 1 GDPR: "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

For the Group of European Data Protection Authorities[34], therefore, Art. 22, para. 1, establishes a prohibition for that individual decision-making process that is completely automated, including profiling, which has a legal or similar effect on the data subject. By "decision based solely on automated processing" one must understand a decision taken without the involvement of a human being who can influence and possibly change the result through his authority or competence.

For the right of the data subject not to be subject to a decision based solely on automated processing to be recognized, it is necessary that such decision "produces legal effects or similarly significantly affects his person". The reference to "legal effects" concerns, obviously, the impact that an automated decision can produce on the legal sphere of the individual (e.g. penalizing the right of association, of voting, of contractual freedom, of free movement, etc.). In addition to this, the rule also opens up to circumstances that "in a similar way" can potentially and significantly influence the behaviors and choices of the individuals concerned.

Recital 71 of the GDPR cites, as examples of automated decisions that can significantly affect the rights and freedoms of individuals, the automatic refusal of an online credit application or electronic hiring practices without human interventions.

Art. 22 in para. 2 provides three exceptions to the general prohibition of a completely automated decision-making process that brings effects in the legal sphere of the individual: when the decision is necessary for the conclusion or performance of a contract between the data subject and a controller; when the decision is authorized by the law of the Union or of the Member State to which the controller is subject; when the decision is based on the explicit consent of the data subject.

With reference to the first point, the European Authorities clarify that the necessity to use automated decisions for the performance or conclusion of a contract must be interpreted restrictively: this means that the controller must be able to demonstrate that the profiling is necessary (principle of necessity) and that no less invasive alternative means are available (principle of proportionality), along the lines of what has already been established by the ECtHR case law, with particular reference to Art. 8 of the Convention[35]. With reference to the second point, the legislation of the Member States can, in specific cases, authorize the use of an automated decision-making process for the monitoring and prevention of fraud and tax evasion or to guarantee the security and reliability of a service provided by the controller. Finally, as regards the third derogation, the Regulation requires the explicit consent of the data subject, that is, confirmed by an express declaration and not inferred from conclusive conduct.

On the administrative front, one must finally not forget that the Council of State[36], too, in assessing the legitimacy of the adoption of an algorithm for carrying out an administrative activity, has noted that three fundamental principles must be duly taken into consideration in the examination and use of IT tools. First, the principle of knowability, by which everyone has the right to know of the existence of automated decision-making processes that concern him and, in this case, to receive meaningful information about the logic used. The second principle can be defined as the principle of non-exclusivity of the algorithmic decision. In the case in which an automated decision "produces legal effects that concern or significantly affect a person," that person has the right that such decision is not based solely on such automated process. The third principle is that of algorithmic non-discrimination, according to which it is appropriate that the controller use appropriate mathematical or statistical procedures for profiling, implementing adequate technical and organizational measures in order to ensure, in particular, that the factors that entail inaccuracies in the data are corrected and that the risk of errors is minimized.

[34] EDPB, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 22 August 2018.

[35] S. & Marper v United Kingdom, ECtHR, apps. nos. 30562 and 30566, 4 December 2008, §§ 35, 119.

[36] Consiglio di Stato, decision No. 8472/2019.

4.2. The school context

Also to be recalled on the subject is the question of reputational rating in the school context.

The Data Protection Authority sent a request for information to the association Crop News (where "Crop" is an acronym for Personalized Objective Reputational Chronicles, which operates within the Mevaluate group). Press reports revealed that the association, within the framework of the "Virtute 4 Students" Project, was experimenting on students the reputational rating processed on the basis of algorithms by the Mevaluate Platform itself and that a higher education institution had joined the project.

The Garante, considering the delicacy of the project which is addressed to particularly vulnerable subjects (students and minors), asked the association for clarifications, in particular the functioning of the platform and of the connected database in order to allow the Authority to evaluate the impact of the use of algorithms and the effects that they can determine on students, as well as the measures possibly adopted for their protection[37].

The Association declared that the purpose of the initiative was to stimulate students to build a real reputation that coincides with the virtual reputation through their own digitized reputational rating in order to promote a transparent behavioral economic model. Virtute calculates the individual reputations published by the online periodical Crop News, thus promoting lawful behaviors founded on the fear of appearing in the newspaper. The reduction of unlawful acts and contractual defaults would guarantee both economic operators and consumers and users, but by reason of the moral "blackmail" of an automated reputational rating, which moreover assigns variable remuneration in relation to the contribution provided by each person to qualify the documented and traceable reputation of oneself and of the counterparties in obligatory relationships.

On the basis of the information provided and of further investigations, the Authority considered the processing unlawful, in violation of the mandatory safeguards of Art. 22 of the GDPR and, above all, contrary to the principles of lawfulness (since it was carried out in the absence of a sufficient legal basis), fairness (since it is capable of engendering unreasonable and disproportionate discriminations and prejudices), transparency (in that it was carried out in the absence of adequate information notice), minimization (processing quantities of data that are indeterminate and not determinable a priori), limitation (since it produces new data that can be processed for the pursuit of purposes that are different and not necessarily lawful), and storage limitation (entrusting the data to an online newspaper). Crop News was therefore admonished, while a new investigation was opened against Mevaluate, collector of all the data collected.

5. Concluding reflections. From algocracy to algoretics

The experiences and examples shown in the course of the contribution allow us an overview of the practical realizations, mostly experimental, that to date concern reputational profiling. However, there is no need to say that, because of its implications, the subject needs to be addressed at a higher level, that is, at the level of its theoretical and philosophical foundations that lead us to come to terms with the phenomenon of the so-called algocracy, that is, the increasing dependence on algorithms and automated decision-making processes in various aspects of society. This tendency has been driven by the progress of artificial intelligence and machine learning, which have made it possible to automate activities that previously were considered the exclusive domain of human judgment and skills. However, since the use of algorithms has become more widespread, many concerns have arisen about the ethical implications of these systems and their impact on individuals and on society as a whole.

In response to these questions, we have witnessed a shift of attention to algoretics, that is, the study of the ethical implications of algorithms and automated decision-making processes. Algoretics seeks to address issues such as the potential for algorithms to perpetuate and amplify existing inequalities and prejudices, as well as the lack of transparency and accountability in these systems. Overall, the shift from algocracy to algoretics reflects a growing recognition of the need to consider the ethical implications of algorithms and automated decision-making systems. Although these technologies have the potential to improve efficiency and objectivity in decision-making, it is important to ensure that they are used in a way that promotes fairness and transparency. This will require constant attention to the ethical and social implications of algocracy and a commitment to addressing any negative consequences that might arise.

[37] "*Scuola: rating reputazionale sotto la lente del Garante privacy. Avviata istruttoria su una piattaforma rivolta agli studenti*", Garante press release, 3 May 2022.

Machine learning, AI and Big Data are undoubtedly changing the way of reading and managing reality. These technologies not only help man by increasing his capabilities, but in an ever greater number of situations give rise to systems, bots or robots, completely autonomous, which therefore replace him.

In particular, artificial intelligences are technological artifacts different from all the artifacts produced up to today. All the tools that we have contrived have enabled man to perform certain tasks by enhancing his physical capacities: from primitive clubs up to large industrial machines, all these tools have served to perform specific tasks better, faster and more effectively. AIs, both in bots and in robots, go beyond the concept known up to now. All the automatic mechanisms built during the industrial revolution were designed according to what their purpose was and, therefore, performed exclusively that specific task.

Today AIs are not designed like this: they are not programmed software, but trained systems, surpassing the classic "if this then that" model in which an engineer first foresaw all the possible occurrences. AIs respond autonomously to a problem that is posed to them. These artifacts are a new species among machines. Of machine sapiens that share the world with homo sapiens, however reasoning differently.

Data scientists underline how the problem is linked to the quality and to the quantity of the data that the machines have at their disposal to make their decisions: when we have a perfect database on which to run AI systems - so they say - the machine will make "perfect" choices.

But is it really so? Already in the past we had this impression. Laplace maintained that if we had known the position at the same instant of all the particles that contain the universe we would have been able to predict all the future and to know all the past of the universe. But even if we were able to create a map that is the exact copy of reality, including within it everything, this map would still be useless, since it would prove to be as complex as reality and, therefore, too complex to make decisions more effective than those that we already make.

We would thus find ourselves before the well-known paradox recounted by Jorge Luis Borges in a fragment On Exactitude in Science, the last of A Universal History of Infamy first published in 1935. As is his habit, the Argentine author attributes the quotation to a book that in reality does not exist: "... In that Empire, the art of cartography attained such perfection that the map of a single Province occupied an entire city, and the map of the empire an entire Province. Over time, these unwieldy maps were no longer sufficient. The colleges of cartographers made a map of the Empire that had the vastness of the Empire and coincided perfectly with it (Suárez Miranda, *Viajes de varones prudentes*, book IV, chap. XIV, Lérida, 1658)".

Data are a map of reality, they represent a reduction of reality and for this they are useful for making decisions. Moreover AIs, besides databases, work on sensors which in turn are not able to read all reality: they learn only a part of it transforming it into data. This is the key point of the question. Since artificial intelligences base their decisions on data, and since these are not a perfect copy of reality, it cannot be thought a priori that the machine endowed with artificial intelligence can make a choice devoid of errors. The machine sapiens will always and constitutively be fallible.

Since artificial intelligences can make mistakes, it is necessary to understand how to manage such errors, assigning to AIs an ethics. That is, it is necessary to find a shared ethical system so that the use of these systems does not produce injustices, does not harm people and does not create strong global imbalances.

The existence of machine sapiens calls for setting up a new universal language that knows how to translate these ethical guidelines into directives executable by the machine and since the dimension of the digital era is regulated by algorithms, here algocracy imposes itself, the dominion of the algorithm, in the face of which it is urgent to develop the common language of algoretics.

To the extent that we want to entrust human skills of understanding, judgment and autonomy of action to AI software systems we must understand the value, in terms of knowledge and capacity for action, of these systems that claim to be intelligent and cognitive, and it is for this reason that the problem is first and foremost philosophical and epistemological. AIs "work" according to schemes that connect data; but what kind of knowledge is this? What value does it have? How should it be treated and considered? It is first of all necessary to clarify in what sense one speaks of value. In fact algorithms work on values of a numerical nature, while ethics weighs moral value. It would therefore be necessary to establish a language that knows how to translate moral value into something computable by the machine. The perception of ethical value is a purely human capacity and the capacity to work with numerical values is instead the ability of the machine. But in the relationship between man and machine the true knower and bearer of value is the human part. The treatment on human dignity and human rights can only teach that it is man who must be protected in the relationship between man and machine.

This evidence provides us with the fundamental ethical imperative for the machine sapiens: doubt itself. This means having to put the machine in a position to nurture a certain sense of uncertainty. Every time the machine does not know whether it is safeguarding the human value with certainty it must request the action of man. This fundamental directive is obtained by introducing statistical paradigms within AIs. This capacity for uncertainty must be the heart of the machine's deciding, since by doing so the machine itself places the human at the center.

References

Ahl, B., L. Catá Backer, y Y. Chen. "Law and Social Credit in China." *China Review* 24, no. 3 (2024): 1–15

Angwin, J. et al. "Machine Bias." *ProPublica*, May 23, 2016.

Arendt, Hannah. *The Origins of Totalitarianism*. New York: Schocken Books, 1951.

Autoriteit Persoonsgegevens. "Belastingdienst beloofde ambtenaren niet te straffen om toeslagenaffaire." May 20, 2019.

Autoriteit Persoonsgegevens. "Werkwijze Belastingdienst in strijd met de wet en discriminerend." July 17, 2020.

Castro, C. "What's Wrong with Machine Bias." *Ergo* 5, no. 15 (2019–2020).

Chin, J., and L. Lin. *Surveillance State*. New York: St. Martin's Press, 2022.

"China's Chilling Social Credit Blacklist." *Human Rights Watch*, December 12, 2017.

"China Releases Investigative Journalist After Almost Year in Jail." *Newsweek*, August 3, 2014.

Creemers, R. *China's Social Credit System: An Evolving Practice of Control*. SSRN, 2018.

Dai, X. "Toward a Reputation State: A Comprehensive View of China's Social Credit System Project." In *Social Credit Rating*, edited by O. Everling, 139–165. Wiesbaden: Springer, 2020.

Decision of the Italian Data Protection Authority (Garante per la protezione dei dati personali) No. 488 of November 24, 2016 [web doc No. 5796783].

Di Carpegna Brivio, E. *Pari dignità sociale e reputation scoring: per una lettura costituzionale della società digitale*. Turin: Giappichelli, 2024.

European Data Protection Board (EDPB). *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*. August 22, 2018.

Elliott, A. Fenwick. "China Is Banning People with Bad 'Social Credit' from Using Planes and Trains." *The Telegraph*, March 19, 2018.

Fenger, M., and R. Simonse. "The Implosion of the Dutch Surveillance Welfare State." *Social Policy & Administration* 58, no. 2 (2024): 264–276.

Galli, F. "Rating reputazionale e diritto alla spiegazione." *Labour & Law Issues* 9, no. 2 (2023): 62–97.

Garante per la protezione dei dati personali. "Scuola: rating reputazionale sotto la lente del Garante privacy. Avviata istruttoria su una piattaforma rivolta agli studenti." Press release, May 3, 2022.

Gatti, A. "L'algoritmo tra volontà e rappresentazione." *DPCE Online*, no. 3 (2020): 3457–3461.

Gatti, A. *Profilazione e diritti fondamentali*. Napoli: ES, 2025.

Hofmann, S. *Social Credit: Technology-Enhanced Authoritarian Control with Global Consequences*. International Cyber Policy Center, Policy Brief Report no. 6, 2018.

Johnson, S. "How China's 'Social Credit Score' Will Punish and Reward Citizens." *Big Think*, April 26, 2018.

Lin, L. Yu-Hsin, and C. Milhaupt. "China's Corporate Social Credit System: The Dawn of Surveillance State Capitalism?" *The China Quarterly* 256 (2023): 835–853.

Liu, C., and A. Rona Tas. "Trusting by Numbers: An Analysis of a Chinese Social Credit System Governance Infrastructure." *Critical Sociology* 51, no. 6 (September 2025): 1247–1265.

Loefflad, C., M. Chen, and H. Grossklags. "Reputational Discrimination and Fairness in China's Social Credit System." *Digital Government* 5, no. 4 (December 2024): 1–27.

Mac Sithing, D., and M. Siems. "The Chinese Social Credit System: A Model for Other Countries." *Modern Law Review* 82, no. 6 (2019): 1034–1071.

National Development and Reform Commission. *Social Credit Action Plan 2024–2025*. June 2024. https://www.chinalawtranslate.com/en/2024-2025social-credit-plan/

Parlementaire ondervragingscommissie Kinderopvangtoeslag. *Tweede Kamer der Staten-Generaal*. July 11, 2020.

Soro, Antonello. "La reputazione online tra principi di dignità e diritto all'oblio." *Huffington Post*, October 6, 2015.

Stradella, E. "Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale." *Consulta Online,* March 20, 2020. www.giurcost.org.

"Uitspraak Bulgarenfraude: hoe zat het ook alweer?" *RTL Nieuws*, May 19, 2015.

van Blomberg, M. "The Social Credit System in China's Rule of Law." *Mapping China Journal*, no. 2 (2018): 78–162.

Venice Commission, Council of Europe. *Opinion No. 1030/2021*.

Zuo, Z. *Governance by Algorithm: China's Social Credit System*. University of Cambridge Working Paper, June 16, 2020.

La gobernanza de datos y de IA como herramientas para gestionar los riesgos derivados de los sesgos

Data Governance and AI Governance as Tools to Manage the Risks Arising from Biases

Eduard Chaveli Donet
Especialista en Gobernanza, riesgos y cumplimiento en protección de datos e inteligencia artificial

Resumen:
En este artículo se aborda el concepto de sesgos, se diferencia de otros afines con los que tiene relación y a veces se confunden y se describen las etapas del ciclo de vida de los sistemas de IA identificando cómo pueden penetrar los sesgos en cada una de ellas para, a partir de ahí, poder gestionar los riesgos derivados de los mismos.

Tras una visión general de las obligaciones que para los diferentes sujetos dispone el RIA y que pueden permitir la gestión de los riesgos para los derechos fundamentales, nos hemos centrado en dos concretas. Primero, la gobernanza de los datos, que se refiere a cómo se gestionan los datos utilizados por los sistemas de IA de alto riesgo, y que tiene un impacto directo en los sesgos. Segundo, la gobernanza de la IA que permite ordenar las diversas obligaciones en torno a un sistema de gestión de tal forma que se planifiquen, operen, monitoricen y se proceda a una mejora continua, para lo cual se apuesta por utilizar un estándar como la BS ISO/IEC 42001:2023 (Information technology — Artificial intelligence — Management system, ahora ya norma española mediante la UNE-ISO/IEC: 2025), por los beneficios que se mencionan en el artículo.

Palabras clave:
Gobernanza de datos; Gobernanza de la IA; Sesgos en IA; Gestión de riesgos e IA; ISO 42001.

Abstract:
This article addresses the concept of bias, distinguishing it from other related concepts with which it is sometimes confused. It describes the stages of the AI systems lifecycle, identifying how bias can penetrate each stage in order to manage the resulting risks.

After an overview of the obligations imposed on different stakeholders by the RIA and which can enable the management of risks to fundamental rights, we have focused on two specific aspects. First, data governance, which refers to how the data used by high-risk AI systems is managed, and which has a direct impact on bias. Second, AI governance allows the various obligations to be organized around a management system in such a way that they are planned, operated, monitored and continuously improved, for which it is recommended to use a standard such as BS ISO/IEC 42001:2023 (Information technology — Artificial intelligence — Management system, now a Spanish standard through UNE-ISO/IEC: 2025), due to the benefits mentioned in the article.

Keywords:
Data Governance; AI Governance; Bias in AI; Risk Management and AI; ISO 42001.

1. Introducción

Existe actualmente una conciencia ampliamente extendida de las oportunidades que la IA está suponiendo y va a suponer para la sociedad y la economía. Su impacto posee una amplitud y profundidad superior al del software tradicional, atribuible a su capacidad de aprendizaje autónomo, su escala de despliegue y, en ocasiones, su inherente opacidad. Su progresiva penetración en todos los sectores y contextos (bien sea de forma directa o a través de procesos instrumentales) va a suponer un efecto multiplicador de las oportunidades, pero también un aumento de los sesgos y los posibles riesgos asociados, especialmente aquellos que inciden sobre los derechos fundamentales. Para ello realizaremos una aproximación mostrando no solo los riesgos que pueden suponer y cómo penetran en el ciclo de vida de la IA, sino aportando una visión de la posible oportunidad que ello también puede suponer para gestionarlos. Las tecnologías de IA no sólo reflejan los valores y decisiones de quienes las crean y utilizan, sino que pueden ser un propagador de los sesgos que conviven con nosotros "desde siempre". Sin embargo, el propio proceso de su diseño y el hecho de tenerlos identificados constituye el primer paso para su correcta gestión.

Asimismo, también se profundiza en algunos de los "controles" que el RIA contempla para su gestión, focalizando el análisis tanto en la gobernanza de los datos como de la IA, y planteando la regulación no como un elemento restrictivo, sino como un vector de oportunidad para la generación de confianza.

Pero antes de adentrarnos en la identificación de los riesgos y en su gestión, resulta imprescindible partir de una aproximación conceptual precisa del término sesgo y de su diferenciación respecto de otros afines con los que frecuentemente se confunde, tales como el error, la discriminación y la exclusión.

La RAE define sesgado/da como relacionado con información "tendenciosa" y ésta a su vez como *"que manifiesta parcialidad, obedeciendo a una tendencia o idea determinadas"*.

Por su parte, la Organización Internacional de Normalización (ISO) define sesgos como "diferencia sistemática de trato de determinados objetos, personas o grupos en comparación con otros"[1].

Por tanto, un resultado inexacto puede derivarse bien de un sesgo o bien de un error. La distinción fundamental reside en que los errores suelen ser aleatorios, no predecibles y mitigables mediante un incremento en el volumen de datos o mejoras en el hardware. En cambio, los sesgos en Inteligencia Artificial no son simples errores aleatorios, sino que obedecen a patrones sistemáticos.

A veces también se confunden los sesgos con las posibles consecuencias negativas que – en ocasiones, pero no siempre[2] – puedan tener los mismos.

[1] ISO/IEC 22989:2022(E) cláusula 3.5.4.

[2] La desviación de la verdad que se produce en los sesgos puede contribuir a resultados diversos: perjudiciales o discriminatorios, ser neutra, o incluso puede ser beneficiosa. Por ejemplo, en el ámbito de la selección de personas, por ejemplo, los sistemas de IA pueden tener muchos beneficios, pero los sesgos, pueden suponer consecuencias que pueden ser tanto negativas, positivas como neutras. Un caso de sesgo negativo que podría llegar a discriminación por razón del sexo podría ser el siguiente: un sistema de IA en el que se hayan utilizado datos de entrenamiento sesgados (por ejemplo, la búsqueda de un perfil que ha desempeñado históricamente mayormente un sexo) utilizará ese sesgo en la fase de inferencia y – por tanto – producirá un resultado discriminatorio hacia ese sexo puesto que el hecho de que históricamente hayan desempeñado ese rol en un sexo no significa que lo vayan o deban desempeñar mejor en el futuro las personas de ese sexo. Pero es posible también que dicho sesgo genere un resultado positivo. Por ejemplo, si el sistema de IA se ha entrenado con perfiles muy cualificados, el resultado (desde esa óptica) puede ser positivo pues ofrece a los candidatos más capacitados. Ahora bien, es posible que dicho "sesgo positivo de capacitación" al beber de datos históricos de una profesión basculada históricamente hacia un sexo o una clase social pueda producir un sesgo que sea negativo y discriminatorio.

Al examinar la posible incidencia negativa sobre derechos fundamentales, pueden identificarse ejemplos en distintos ámbitos. No obstante, en la práctica, los casos que han adquirido mayor relevancia pública suelen vincularse a la igualdad ante la ley y, por tanto, la discriminación y la generación de posibles injusticias. Con todo, que no todos los sesgos implican un trato discriminatorio, ni desemboca de forma automática en una vulneración de derechos[3]:

> "El sesgo se refiere a una diferencia sistemática en el tratamiento de ciertas personas o grupos, sin implicar necesariamente si esta diferencia es "correcta" o "incorrecta". Por el contrario, la discriminación y la equidad introducen un juicio de valor sobre los resultados de un tratamiento sesgado. Un sistema de IA sesgado puede producir resultados que pueden considerarse "discriminatorios" o "injustos", según el contexto y los valores aplicados".

Otra posible consecuencia de los sesgos es la exclusión. A diferencia de la discriminación, que supone una situación de desventaja[4] comparativa, la exclusión constituye una forma de desigualdad que se materializa al impedir a la persona o grupo acceder a determinados servicios o recursos[5].

En este sentido, no todo error puede considerarse sesgo, ni todo sesgo es necesariamente negativo, del mismo modo que no todos los sesgos negativos resultan discriminatorios, ni toda discriminación deriva en una situación de exclusión.

Por último y como un fin, pero también como un mecanismo para evitar las consecuencias negativas (injusticias) que puedan generar los sesgos, hay que hacer referencia a otro concepto diferente y relacionado que es la equidad. En el contexto de la IA, la injusticia puede entenderse como el "trato diferencial injustificado que beneficia preferentemente a ciertos grupos sobre otros"[6] y "la equidad, por lo tanto, es la ausencia de tal trato diferencial injustificado o prejuicio hacia cualquier individuo o grupo"[7].

La equidad no significa que deba de tratarse de forma distinta a diferentes personas o grupos, pero que sí que es posible que deba de hacerse ese trato diferente para precisamente conseguir corregir desequilibrios o una representación incorrecta que suponen una injusticia.

Una vez realizadas estas precisiones conceptuales básicas, podemos abordar una aproximación general a la clasificación de los sesgos sobre la base de los tres tipos que, según el NIST[8], afectan a los sistemas basados en Inteligencia Artificial, complementando dicha clasificación con aportaciones propias:

1. Los sesgos computacionales, entendidos como aquellos que pueden identificarse y cuantificarse a partir de un modelo de Inteligencia Artificial ya entrenado, son la punta del iceberg. Aunque se pueden tomar medidas para subsanarlos, no constituyen todas las fuentes de posibles sesgos que intervienen y surgen a partir de errores que resultan cuando la muestra no es representativa de la población. Estos sesgos surgen de factores sistemáticos y no aleatorios, y pueden producirse sin que exista prejuicio, parcialidad o intención discriminatoria alguna.

En los sistemas de IA, estos sesgos están presentes en los conjuntos de datos y procesos algorítmicos utilizados en el desarrollo de aplicaciones de IA y, a menudo, surgen cuando los algoritmos se entrenan con un tipo de datos y no puede extrapolar más allá de esos datos.

Aquí podemos encontrar varios tipos a su vez como el sesgo de selección de datos, que se produce

[3] Como se cita en esta guía práctica de Rhite: Suzanne Snoek e Isabel Barberá, *From Inception to Retirement: Addressing Bias Throughout the Lifecycle of AI Systems: A Practical Guide* (Rhite, 5 de septiembre de 2024). Vid. https://rhite.tech/files/From-Inception-to-Retirement-Addressing-Bias-Throughout-the-Lifecycle-of-AI-Systems.pdf

[4] Por ejemplo, considerar que alguien con una discapacidad no relacionada con el trabajo pueda ser peor candidato que otro que no tiene dicha discapacidad.

[5] Por ejemplo, pensemos en un sistema de IA que no contempla opciones de vehículos adaptados para personas con determinadas discapacidades y los excluye.

[6] ISO/IEC 22989:2022.

[7] Snoek y Barberá, *From Inception to Retirement*.

[8] El National Institute of Standards and Technology (NIST) - agencia del gobierno de los EE. UU. encargada de promover la innovación y la competencia industrial - publicó en marzo de 2022 la NIST Special Publication 1270 con el título "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" cuyo objetivo es ofrecer una guía para futuro estándares en identificación y gestión de sesgos. Reva Schwartz et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, NIST Special Publication 1270 (Gaithersburg, MD: National Institute of Standards and Technology, 2022), https://doi.org/10.6028/NIST.SP.1270

cuando los datos utilizados para entrenar un modelo no representan de manera equitativa la realidad[9]; o el sesgo algorítmico que se produce cuando los algoritmos favorecen ciertos resultados o a ciertos grupos[10].

2. Por su parte, el sesgo humano es el que tenemos cada uno de nosotros de manera implícita y que afecta a cómo percibimos e interpretamos la información que recibimos.

Los prejuicios humanos reflejan errores sistemáticos en el pensamiento humano basados en un número limitado de principios heurísticos y predicción de valores hasta operaciones de juicio más simples. Se conocen también como sesgos cognitivos al estar relacionados con la forma en que los seres humanos procesamos información y en cómo tomamos decisiones a partir de ella.

Dentro del sesgo humano podemos encontrar diversas categorías: como los sesgos de confirmación, que se producen cuando el sistema de IA se basa en creencias o asunciones preexistentes en los datos al existir una tendencia a buscar, interpretar y recordar aquellos datos que refuerzan dichas creencias u opiniones preexistentes[11]; los sesgos de anclaje, que se producen cuando hay una dependencia excesiva de la primera información o ancla[12]; el efecto halo, que valora a una persona o cosa en función de una característica sobresaliente[13]; y el sesgo de negatividad, que se produce cuando se da un peso mayor a la información negativa que a la positiva[14].

3. Y, por último, hay que considerar el conocido como sesgo sistémico, que es el que se encuentra integrado en la sociedad y en las instituciones por razones históricas. Este no tiene por qué ser el resultado de ningún prejuicio o prejuicio consciente, sino más bien de que la mayoría siga reglas o normas existentes (el racismo y el sexismo son los ejemplos más comunes).

Las distintas clases de discriminación y exclusión constituyen algunas de las posibles consecuencias negativas derivadas de los sesgos y, por tanto, riesgos asociados a su manifestación en sistemas de IA[15].

La tipología de posibles sesgos y riesgos son numerosos y diversos, aunque de ciertos sesgos no se habla porque no producen discriminación a pesar de poder tener efectos en decisiones que supongan consecuencias negativas, como por ejemplo para la salud.

Dado que vamos a poner especial foco en la posible afectación a los derechos fundamentales conviene recordar, aunque sea obvio, algo que ya hemos indicado en otra obra[16]: "*los derechos humanos no son "algo nuevo" y con los siglos se ha producido una evolución en cuanto a su número, los sujetos beneficiarios, así como en cuanto a su alcance territorial; aunque su aplicación real diste de ser verdaderamente global. Pero, incluso en donde existe una "cultura" y sistemas tuitivos de los derechos humanos (como es el caso de Europa), la evolución industrial primero y la tecnológica después aportaron oportunidades y riesgos para los mismos que tuvieron que ser gestionadas*".

[9] Por ejemplo, si un algoritmo de contratación se entrena principalmente con currículos de hombres, podría sesgar las decisiones hacia candidatos masculinos.

[10] Por ejemplo, un sistema de crédito que penaliza automáticamente a personas de bajos ingresos debido a su historial financiero podría tener un sesgo algorítmico.

[11] Por ejemplo, si un algoritmo de recomendación de películas solo sugiere géneros específicos a un usuario, podría reforzar sus preferencias anteriores.

[12] Por ejemplo, si un sistema de precios en línea muestra un precio inicial alto, los usuarios pueden percibir como "oferta" o precio barato cualquier precio inferior a ese anclaje inicial.

[13] Por ejemplo, asumir que un candidato con una universidad prestigiosa en su currículo es automáticamente más competente.

[14] Por ejemplo, imaginemos un sistema de detección de fraudes podría ser más propenso a identificar falsos positivos debido a este sesgo derivado de que está entrenado con información "negativa".

[15] Eduard Chaveli Donet, "Sesgos en la IA (I): La necesaria distinción entre sesgos y conceptos afines," *Telefónica Tech Blog*, 11 de febrero de 2025, https://telefonicatech.com/blog/sesgos-en-ia-parte-i-distincion-entre-sesgos-y-conceptos-afines

[16] Lorenzo Cotino Hueso y Pere Simón Castellano, dirs., *Tratado sobre el reglamento de inteligencia artificial de la Unión Europea* (Madrid: Aranzadi, 2024).

2. La introducción de los sesgos en las diferentes etapas del ciclo de vida de los sistemas de IA

Antes de adentrarnos en las medidas para la gestión de los riesgos que los sesgos pueden generar para los derechos fundamentales, entendemos apropiado realizar una aproximación previa a la forma en que dichos sesgos se identifican y se incorporan en los sistemas de IA. El NIST considera que las organizaciones que diseñan y desarrollan la tecnología de IA utilizan el ciclo de vida de la IA para realizar un seguimiento de sus procesos y asegurar que la tecnología sea funcional, pero no necesariamente identifican posibles riesgos y daños y los gestionan.

Por otro lado, los enfoques actuales sobre los sesgos tienden a clasificarlos en función de su tipología (es decir, estadístico, cognitivo) o caso de uso y sector industrial (es decir, contratación, atención sanitaria, etc.). No obstante, estas aproximaciones pueden limitar la perspectiva necesaria para gestionar eficazmente el sesgo, entendido como un fenómeno intrínsecamente contextual. Por ello, el documento del NIST propone un enfoque para gestionar y reducir los efectos de los sesgos perjudiciales en todos los contextos teniendo en cuenta los "lugares o momentos" clave dentro de las etapas del ciclo de vida de un sistema de IA.

Identificar las fuentes de sesgo es el primer paso en cualquier estrategia de mitigación. Dichas fuentes pueden aparecer en múltiples fases del sistema de IA, lo que exige comprender con claridad cuáles son las etapas del ciclo de vida de los sistemas de IA y, a su vez, ponerlo en relación con el rol y tareas concretas que tienen los diferentes operadores. No es lo mismo cuando hablamos de un proveedor[17] que de un responsable del despliegue"[18] , por citar dos ejemplos muy claros. Por ello a continuación, realizaremos una visión general que hay que aterrizar en cada caso.

El desarrollo de un sistema de IA tiene unas fases básicas y, aunque hay diferentes aproximaciones o formas de englobarlas[19], consideramos que esta es la más clara de entender y la que sirve mejor para aterrizar los sesgos en cada una de ellas:

FASE 1. FASE DE INICIO: PRE-DISEÑO O DEFINICIÓN DEL ALCANCE

Los sistemas de IA comienzan en la fase de pre-diseño en la que:

a. Se identifica el problema que se quiere resolver con el sistema de IA (es decir: los objetivos). Si los objetivos del sistema están influenciados por prejuicios, el sistema reflejará esos sesgos. Por ejemplo, si se decide que un sistema de contratación debe priorizar candidatos de ciertas universidades que podemos calificar como prestigiosas, puede excluir a candidatos igualmente cualificados de otras universidades.

Por tanto, nos encontramos aquí con sesgos institucionales o sistémicos, por lo que hay que revisar esos posibles perjuicios introduciendo diversas medidas que van desde la revisión de los datos utilizados hasta la evaluación de los posibles impactos. Para ello hay que involucrar a expertos en áreas como, por ejemplo, la ética y los derechos fundamentales.

b. Se recopilan los requisitos funcionales (qué debe hacer el sistema) y no funcionales del sistema (cómo debe comportarse el sistema). Aquí podemos incluir la recopilación de datos en cuanto al análisis de requisitos (determinar qué datos son necesarios para resolver el problema) y la recopilación preliminar (obtener datos iniciales para entender mejor el ámbito y los desafíos a los que se enfrenta).

Si los datos recopilados no son representativos de la población objetivo, el modelo aprenderá patrones incorrectos. Por ejemplo, si se recopilan datos de salud solo de una región específica, el sistema puede no funcionar bien en otras regiones con diferentes características demográficas.

[17] En los términos del RIA "*aquellas personas que desarrollen un sistema de IA o modelo de IA de uso general para introducirlo en el mercado de la Unión Europea bajo su propio nombre o marca comercial*". Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial), art. 3.

[18] Según el RIA "*toda persona que utiliza un sistema de IA bajo su propia autoridad en el ámbito profesional*". Así, toda organización que incorpore un sistema de IA en sus procesos actuaría como responsable del despliegue de dicho sistema. Reglamento (UE) 2024/1689, art. 3.

[19] Sobre las fases de un sistema de IA pueden consultarse, por ejemplo: Keyrus, "El ciclo de vida de la inteligencia artificial," *Keyrus Insights*, https://keyrus.com/sp/es/insights/el-ciclo-de-vida-de-la-inteligencia-artificial-alcance-diseno-de-modelos-y/; Enzyme Advising Group, "Las tres fases para poner en marcha tu modelo de IA," *Enzyme Blog*, https://enzyme.biz/blog/el-ciclo-de-vida-de-ia-pasos-para-poner-en-marcha-tu-modelo; o una aproximación más detallada en la figura 3 de *ISO/IEC 22989:2023, Artificial Intelligence — Artificial Intelligence Concepts and Terminology*.

Se trata de trampas de abstracción derivadas de traducir del "mundo real al sistema de IA" y "sus entradas y salidas se simplifican en exceso o se ignoran".

Las principales trampas de abstracción incluyen la de formalismo, la del efecto dominó y la del solucionismo.

c. Se analiza la viabilidad tanto técnica como económica y operativa del proyecto.

d. Se seleccionan las tecnologías para desarrollar el sistema de IA.

f. Y se planifica el proyecto

FASE 2. FASE DE DISEÑO Y DESARROLLO

En esta etapa del ciclo de vida de la IA se llevan a cabo decisiones de calado como si se van a realizar determinados desarrollos o se van a adquirir, si se utilizarán soluciones de código abierto o propietario etc.

Los principales hitos son los siguientes:

1. Se analizan los requisitos y datos disponibles, lo que supone incluir la recopilación, limpieza y preparación de los datos concretos que son necesarios para entrenar el modelo de IA.

2. Se diseña y desarrolla el modelo de IA, lo que incluye la selección de algoritmos y técnicas de aprendizaje.

3. Se lleva a cabo el entrenamiento del modelo con los datos y se ajustan los parámetros para optimizar su rendimiento.

4. Se prueba y valida el modelo para asegurar que cumple con los requisitos y funciona correctamente.

5. Se realizan iteraciones para mejorar el modelo basado en los resultados de las pruebas.

Dada la importancia del diseño en el resultado, los sesgos de validez del constructo (de la construcción teórica para entender el problema) son especialmente importantes en esta fase. Este se produce cuando una variable no mide con previsión el constructo que se quiere representar para la construcción del sistema de IA, cuando hablamos de problemas complejos. Por ejemplo: imaginemos que confundimos el estatus socioeconómico con un elemento único como los ingresos, cuando realmente hay otros que pueden afectarlo como la educación, la riqueza, ocupación o prestigio, por ejemplo. Por tanto, hay que tener en cuenta diversas medidas de dichos elementos complejos y considerar diversas "formas" de interpretarlos como, por ejemplo, las derivadas de diferentes visiones culturales.

También es posible que los algoritmos hayan sido diseñados de manera que favorezcan ciertos resultados[20].

Por otro lado, aquí también se procede a la comprensión y preparación de los datos, siendo el sesgo de representación el más común y el más tratado en esta fase. Si el modelo se entrena con datos sesgados, aprenderá esos sesgos[21]. Para ello hay que garantizar la representación correcta, pudiendo utilizar, por ejemplo, técnicas de muestreo.

Otros sesgos importantes en esta fase son el sesgo de medición, el sesgo histórico, el sesgo de selección (de los que hay muchos ejemplos[22]) y el sesgo de etiquetado.

[20] Por ejemplo, un algoritmo de recomendación de empleo que prioriza ciertos términos en los currículums (como "agresivo", pensemos en "comercial agresivo" o "ejecutivo agresivo") puede sesgarse hacia candidatos masculinos, ya que estos términos son más comúnmente utilizados por hombres o por ejemplo términos como persona empática y colaborativa, más asociados a mujeres.

[21] Por ejemplo: 1.Si la población objetivo-definida no representa correctamente la población de uso posterior. Por ejemplo, un sistema de reconocimiento facial entrenado con imágenes mayoritariamente de personas de piel clara puede tener dificultades para reconocer a personas de piel más oscura. 2.O si hay grupos subrepresentados en dicha población puede no funcionar bien parta dicho subgrupo.

[22] Algunos ejemplos de sesgos de selección son el de muestreo, el de autoselección, el sesgo de cobertura y el sesgo de observación.

Durante la fase de desarrollo se construyen los modelos seleccionados y se entrenan los datos.

Los principales sesgos de esta fase son los sesgos de algoritmo. La principal característica de este sesgo es que no está en los datos, sino en el propio algoritmo. Un ejemplo de ello es que un algoritmo de selección de personas utilice datos de entrenamiento equilibrados asignando más peso a algún criterio que no tenga que ver con el posible rendimiento.

Hay diversos tipos de sesgos de algoritmo, pero en esta fase algunos que inciden en el desarrollo son los siguientes: el sesgo de agregación, el sesgo de variable omitida y el sesgo de aprendizaje.

Como dice el NIST, al final de la fase de diseño y antes del despliegue es necesaria una evaluación exhaustiva de la mitigación de sesgo para garantizar que el sistema se mantenga dentro de los límites preespecificados, lo que debe incluir:

- Las fuentes de sesgo identificadas
- Las técnicas de mitigación implementadas
- Evaluaciones de desempeño relacionadas antes de que el modelo pueda lanzarse para su implementación.

Como medidas para solucionar dichos riesgos, el NIST (en su informe citado) menciona por ejemplo el "desafío cultural efectivo" (*cultural effective challenge*), una práctica que busca crear un entorno en el que los desarrolladores de tecnología puedan desafiar y cuestionar activamente los pasos en el modelado y la ingeniería para ayudar a erradicar los sesgos estadísticos y los sesgos inherentes a la toma de decisiones humanas. Aunque lo hemos incardinado en esta fase debería ser iterativo. Como ya hemos sostenido anteriormente[23], pensamos que la existencia de una "parada formal", quizá incluso su constancia en un informe como sucede por ejemplo en las evaluaciones de impacto en protección de datos conforme al RGPD o las EIDF del RIA, sería una buena forma de "obligar a realizar dicha parada" y a que tuviese el calado oportuno. Si derivado de ello se constatase el sesgo en los algoritmos y el impacto que ello podría tener podría/debería incluso evitarse avanzar en las siguientes fases.

FASE 3. FASE DE PRUEBA Y EVALUACIÓN (VERIFICACIÓN Y VALIDACIÓN ANTES DEL DESPLIEGUE)

En esta fase:

1. Se monitoriza el rendimiento del modelo previo al despliegue. Si las métricas de evaluación no consideran la equidad, el modelo puede parecer preciso, pero ser injusto y una vez desplegado posteriormente, el sistema puede perpetuar y amplificar los sesgos existentes. Por ejemplo, un sistema de recomendación de préstamos, que ha sido entrenado con datos históricos, puede continuar discriminando a ciertos grupos si esos datos reflejan prácticas discriminatorias pasadas.

2. Se actualiza el modelo, se ajusta con datos de entrenamiento y validación, y se procede a las mejoras necesarias.

En esta fase se puede producir un sesgo denominado sesgo de evaluación, que es cuando los procedimientos o las métricas a utilizar para evaluar el modelo no están alineadas con el modelo a implementar.

Por tanto, hay que adoptar medidas como, por ejemplo, evaluar las métricas, los ajustes de datos e idealmente esta evaluación debería realizarse junto con otras partes interesadas para garantizar que todos los problemas previamente identificados se resuelvan a satisfacción de todos.

FASE 4. FASE DE DESPLIEGUE O IMPLEMENTACIÓN

Es en esta fase en la que los responsables del despliegue ya trabajan con esta tecnología, pues pasan del desarrollo a su implantación en el entorno de producción. Hay que tener en cuenta

[23] Eduard Chaveli Donet, "Sesgos en la IA (V): Introducción de los riesgos en el ciclo de vida de los sistemas de IA (parte 1)," *Telefónica Tech Blog*, 22 de abril de 2025, https://telefonicatech.com/blog/sesgos-ia-v-introduccion-riesgos-en-ciclo-de-vida-sistemas-de-ia-parte-1

especialmente el sesgo de implementación, que se produce si el sistema se implementa en un entorno que no refleja las condiciones del entrenamiento, lo que supone que puede comportarse de manera sesgada. Por ejemplo, un sistema de traducción automática entrenado principalmente con textos formales puede no funcionar bien con lenguaje coloquial.

Las trampas de abstracción (de las que ya hemos hablado) también son propias de esta fase.

FASE 5. FASE DE OPERACIÓN Y MONITORIZACIÓN

Cuando los sistemas están en producción (operando) se deben de monitorizar y realizar los cambios necesarios tanto a nivel de hardware y software, como efectuar ajustes en el propio algoritmo, en los datos utilizados (añadiendo, limpiando etc.), etc.

En los sistemas que utilizan aprendizaje continuo[24] se producen más estos sesgos a diferencia de los sistemas de IA en los que no[25].

En esta fase se puede producir el conocido como "bucle de retroalimentación de refuerzo", lo que conlleva que al volver a entrenar el modelo con los nuevos datos y partiendo de un sesgo inicial no solucionado, se pueda propagar el sesgo, a lo que se puede añadir, por ejemplo, el sesgo de automatización que puede tener un efecto multiplicador. Para ello hay que establecer mecanismos de retroalimentación continua para identificar posibles sesgos y corregirlos en tiempo real.

FASE 6. VALIDACIÓN CONTINUA

La "validación continua" está interrelacionada con el monitoreo y a veces se superponen, pero en teoría son cosas diferentes:

1. En el monitoreo se supervisa el rendimiento del modelo de IA en producción en tiempo real revisando métricas para ver si hay alguna que detecta posibles anomalías. Se trata de asegurar que el modelo continúa funcionando bien y de detectar posibles problemas cuanto antes.

2. En la validación continua se trata de evaluar regularmente el modelo con nuevos datos para ver si continúa siendo preciso.

3. La reevaluación es una revisión completa y más profunda en la que se analiza periódicamente el modelo completo y su rendimiento general para ver si se requieren cambios más importantes.

En consecuencia, la validación continua se puede realizar en sistemas de IA donde no aplica el aprendizaje continuo para, por ejemplo "detectar desviaciones de datos, de conceptos o para detectar cualquier mal funcionamiento técnico" (ISO/IEC 5338), pero es especialmente relevante con nuevos datos, por lo que es fundamental en los supuestos de aprendizaje continuo donde el reentrenamiento existe aunque no sea explícito. En los sistemas con aprendizaje continuo, los modelos integran nuevos datos de forma continua sin un reentrenamiento explícito, por lo que es fundamental: comprobar la coherencia de los datos en producción con los iniciales del entrenamiento y tener que actualizar los propios datos de prueba.

Por tanto, los principales sesgos en esta fase son los sesgos de los datos, de entre los que cabe destacar los de representación, selección, medición, el de etiquetado y el de *proxies*, por lo que habrá que poner especial foco en las medidas para gestionarlos en esta fase.

FASE 7. RE-EVALUACIÓN

A diferencia del monitoreo y validación continua que se refieren a ajustes constantes cada uno con la finalidad que hemos visto, la de reevaluación es más profunda.

Aparte de los sesgos de evaluación y las trampas de abstracción que ya conocemos, y que en

[24] Pensemos por ejemplo en los asistentes virtuales como Alexa de Amazon o el asistente de Google, por citar varios que aprenden y se actualizan con las interacciones de los usuarios.

[25] Por ejemplo, los "sistemas de calculadoras matemáticas" para operaciones complejas que se basan en reglas predefinidas u otros que se basan en reglas predefinidas y no aprenden de forma continua.

estas fases pueden servir para refinar el sistema con decisiones, hay varios propios de esta fase:

1. Falacia del costo volcado a la suma: es la tendencia de las personas u organizaciones a seguir invirtiendo en un proyecto o esfuerzo en el que ya han invertido recursos significativos (dinero, tiempo, esfuerzo, reputación etc.), incluso cuando los costos actuales superan los beneficios potenciales y de hecho cuanto mayor es el costo volcado mayor es el sesgo.

2. Y aunque tienen relación hay otro sesgo diferente que es el sesgo de *estatus quo*, que se refiere a la preferencia por mantener el estado actual de las cosas.

En ambos casos es fundamental que las partes interesadas lo conozcan, reconozcan y tomen decisiones al respeto.

FASE DE RETIRO

Incluso si se toma la decisión de retirar el sistema, lo que se puede deber a diferentes motivos (no sirve a los propósitos, se ha buscado otra solución, se entiende que no es justo etc.), ello puede producir un sesgo conocido como sesgo histórico (o sesgo de legado) dado que el sistema se ha entrenado con datos históricos sesgados que se replican. Un ejemplo son los algoritmos de recomendación de noticias que se puedan basar en aquellas más relevantes, aunque quizá no sean más verídicas o contrastadas. Obviamente al que era usuario de ese sistema ya no le afectará, pero que les afectará a otros usuarios del sistema de IA que lo adquieran o usen.

3. Gestión de los riesgos derivados de los sesgos. Especial referencia a la gobernanza de los datos y gobernanza de la IA

3.1. Introducción

Antes de adentrarnos en las medidas previstas en el RIA para gestionar los riesgos que los sesgos pueden generar para los derechos fundamentales, resulta necesario formular algunas consideraciones generales previas sobre los mismos.

En primer lugar, que cuando hablamos de riesgos es necesario distinguir tres categorías diferenciadas: los riesgos del propio sistema de IA, los riesgos del sistema de gestión de IA (especialmente relevante si hablamos de un sistema certificado bajo un estándar, en concreto la BS ISO/IEC 42001:2023 (Information technology — Artificial intelligence — Management system, ahora ya norma española mediante la UNE-ISO/IEC: 2025), en adelante ISO 42001; y, por último, los riesgos para los derechos fundamentales.

Mientras que en los dos primeros casos el análisis se orienta principalmente al "dolor" organizativo (aunque obviamente pueda irradiar a terceros), en el caso de las evaluaciones de impacto en derechos fundamentales, el enfoque se desplaza a la posible afectación a terceros (individuos o colectivos).

En segundo lugar, el enfoque respecto de los riesgos del RIA, según se desprende de su Considerando 14 – consiste en "*introducir un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA" adaptando "el tipo y el contenido de tales normas a la intensidad y el alcance de los riesgos que pueden generar los sistemas de IA*". Esto lo hace el RIA mediante la siguiente fórmula que algunos autores han criticado[26]: "*prohibir determinadas prácticas inaceptables de inteligencia artificial, establecer requisitos para los sistemas de IA de alto riesgo y obligaciones para los operadores correspondientes, y establecer obligaciones de transparencia para determinados sistemas de IA*".

Este enfoque basado en riesgos implica que el propio RIA incorpora ya una medida normativa de gestión del riesgo: la prohibición de ciertos sistemas y la diferenciación entre niveles de riesgo.

[26] Mantelero indicaba que *"Aunque esto es eficaz en términos de impacto político y aceptabilidad, es una forma débil de prevención de riesgos. La Propuesta hace una distinción bastante rígida entre el riesgo de alto nivel y el resto, no proporcionando ninguna metodología para evaluar el primero, y eximiendo en gran medida al segundo de cualquier mitigación (con la limitada excepción de las obligaciones de transparencia en ciertos casos)"*. Alessandro Mantelero, *Artificial Intelligence and Data Protection: Challenges and Possible Remedies* (Estrasburgo: Consejo de Europa, 2019), 173.

Asimismo, existen ejemplos históricos de discriminación vinculada a sesgos (tanto reales como otros recreados en obras de ficción[27]) que ya no deberían de producirse en el marco del RIA, pues los prohíbe "de plano".

La tercera reflexión necesaria es que el riesgo cero, en términos generales, no existe. Tampoco en el ámbito de la IA. Por ello, debemos de abordar el tema del nivel de riesgo aceptable. Como hemos mantenido ya anteriormente[28] el artículo 9.4. del RIA señala que las medidas de gestión de riesgos "*serán tales que el riesgo residual pertinente asociado a cada peligro, así como el riesgo residual global de los sistemas de IA de alto riesgo, se consideren aceptables*". Por tanto *"asumiendo lege data la utilización de la metodología de gestión de riesgos que ha establecido el RIA y asumiendo también que no pueden tolerarse impactos altos en derechos fundamentales, el riesgo aceptable será por definición bajo*". En consecuencia, deberán de adoptarse salvaguardas y controles que permitan reducir el riesgo inicial hasta situarlo por debajo del umbral de tolerancia definido con carácter previo.

La protección de los derechos fundamentales constituye, junto con la salud y la seguridad, uno de los objetivos esenciales del RIA. En consecuencia, esta protección debe de integrarse en todas las fases del ciclo de vida de los sistemas de IA, como parte inherente del enfoque basado en el riesgo. de manera distinta según el tipo de sistemas, y el operador involucrado, lo que podemos ver de forma general en esta tabla (referida a los sistemas de alto riesgo):

Obligaciones sistemas de alto riesgo	Proveedor	Responsable despliegue	Representante autorizado	Importador	Distribuidor
Art 8 Cumplimiento de los requisitos sistemas Alto riesgo (Art 8 a 15)	X				
Art. 9 gestión de riesgos	X				
Art. 10 Gobernanza de datos	X				
Art. 16 Obligaciones de los proveedores de sistemas de IA de alto riesgo	X				
Art. 17 Sistema de gestión de la calidad	X				
Art. 18 Conservación de la documentación	X				
Art 19. Archivos de registro generados automáticamente	X				
Art 20. Medidas correctoras y obligación de información	X				
Art .21 Cooperación con las autoridades competentes	X				
Art. 22 Representantes autorizados de los proveedores	X		X		
Art. 23 Obligaciones de los importadores				X	X
Art. 24 Obligaciones de los distribuidores					X
Art. 25 Responsabilidades cadena de valor de la IA		X	X	X	X
Art. 26 Obligaciones responsables del despliegue		X			
Art. 27 EIDF		X			
Art. 43 Evaluación de la conformidad	X				

[27] Por ejemplo, sistemas que permitan evaluar o predecir la probabilidad de que una persona física cometa una infracción penal; o los más recientes intentos de poner de moda en algunos entornos los Sistemas de puntuación social.

[28] Eduard Chaveli Donet, "La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento", en *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, dirs. Lorenzo Cotino Hueso y Pere Simón Castellano (Las Rozas, Madrid: Editorial Aranzadi, 2024), 495–533.

Obligaciones sistemas de alto riesgo	Proveedor	Responsable despliegue	Representante autorizado	Importador	Distribuidor
Art. 47 Declaración UE de conformidad	X				
Art. 48 Marcado CE	X				
Art. 49 Registro	X	X	X		
Art. 52 transparencia hacia los usuarios		X			
Art. 72 Vigilancia poscomercialización (Alto riesgo)	X	X			
Art. 73 Notificación de incidentes graves	X	X			
Art. 86 Derecho a explicación de decisiones tomadas individualmente		X			

Fuente. Curso RIA de la AEC. Equipo Govertis, part of Telefónica Tech.

Tras la lectura sistemática del RIA, se aprecia que existen obligaciones estrechamente vinculadas con la gestión de los riesgos para los derechos fundamentales: los análisis de riesgos como parte de los sistemas de gestión y, en su caso, de los procesos de evaluación de conformidad; las exigencias adicionales impuestas a los modelos de propósito general con riesgo sistémico (como disponer de documentación técnica, políticas de gobernanza y supervisión), en las que aunque no exista una referencia explícita es normal tenerlos en cuenta; por supuesto en las EIDF, cuando son requeridas y dónde sí que se pone absoluto foco en ellos; y también en el caso de la supervisión humana puesto que el elemento humano, que está en la base de los sesgos, también puede desempeñar un papel decisivo mediante vigilancia humana, la participación de las partes interesadas o la intervención de los profesionales cualificados. Finalmente, la calidad de los datos y - por tanto - su gobernanza, constituye la herramienta fundamental a la que dedicamos el apartado siguiente.

3.2. La gobernanza de datos como herramienta para gestionar los riesgos derivados de los sesgos

La importancia y el valor de los datos —tanto personales como no personales— para la economía y la sociedad no constituye un fenómeno reciente. Debe recordarse que el propio Reglamento General de Protección de Datos (RGPD) no se limita a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales, sino que también promueve la libre circulación de estos datos dentro de la Unión Europea. Tal y como enfatiza el considerando 6 del RGPD, "*la tecnología ha transformado tanto la economía como la vida social, y ha de facilitar aún más la libre circulación de datos personales dentro de la Unión y la transferencia a terceros países y organizaciones internacionales*", aunque "*garantizando al mismo tiempo un elevado nivel de protección de los datos personales*".

Este equilibrio entre protección y circulación ha guiado múltiples iniciativas europeas en la última década para construir un verdadero mercado único de datos, entre ellas la Estrategia Europea de Datos que en 2020 marcó el rumbo hacia una economía basada en datos interoperables, accesibles y reutilizables, a la que se suman otros instrumentos y normas clave aprobados con posterioridad[29]

[29] La Brújula Digital 2030 y el Decenio Digital Europeo, que fijan metas concretas para la transformación digital de Europa en ámbitos como conectividad, competencias digitales, digitalización de servicios públicos y empresas. El Reglamento (UE) 2022/868 de Gobernanza de Datos (Data Governance Act), que establece mecanismos para compartir datos entre sectores públicos y privados de forma segura, voluntaria y bajo condiciones claras. El Reglamento (UE) 2023/2854 sobre normas armonizadas para un acceso justo a los datos y su utilización (Reglamento de Datos), que garantiza que los usuarios de productos conectados puedan acceder a los datos que generan, fomenta la equidad contractual y refuerza la seguridad jurídica. El Reglamento (UE) 2018/1807 sobre la libre circulación de datos no personales, que complementa al RGPD al eliminar obstáculos al movimiento de datos industriales, de IoT o empresariales dentro del mercado único. El Reglamento (UE) 2025/327 sobre el Espacio Europeo de Datos de Salud (EEDS), que crea un marco para el acceso transfronterizo a datos sanitarios electrónicos y su reutilización ética para investigación, innovación y formulación de políticas, con especial atención a la protección de datos sensibles; Reglamento (UE) 2024/2847 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, relativo a los requisitos horizontales de ciberseguridad para los productos con elementos digitales y por el que se modifica el Reglamento (UE) nº 168/2013 y el Reglamento (UE) 2019/1020 y la Directiva (UE) 2020/1828 (Reglamento de Ciberresiliencia, CRA) que pretende garantizar que los productos que generan, procesan o transmiten datos lo hagan de forma segura lo que refuerza la confianza en la infraestructura digital, lo cual es esencial para que los datos puedan circular, compartirse y reutilizarse sin poner en riesgo a los usuarios ni a las empresas.

así como otros que están fase de elaboración actualmente.

Este ecosistema normativo refleja una visión europea que reconoce el valor estratégico de los datos, sin renunciar a la protección de los derechos fundamentales. La gobernanza de los datos -y por extensión, de las tecnologías que los procesan, como la inteligencia artificial- se configura como un elemento clave para garantizar una transformación digital competitiva, pero a la vez respetuosa con los derechos fundamentales.

Para garantizar ese mercado único de datos (gobernanza a nivel "macro") es imprescindible, como sostiene Loza[30], que las organizaciones cuenten con una sólida gobernanza de datos a nivel interno (nivel "micro"), la cual además permitirá avanzar hacia la gobernanza de la inteligencia artificial".

No vamos a entrar en analizar la aproximación detallada al concepto de gobernanza que tiene múltiples aproximaciones - algunas procedentes del ámbito- como el de las TI, pero sí dar unas pinceladas sobre lo que se entiende por gobernanza de los datos y aterrizarla a los sistemas de IA.

Sostiene también esta autora que "*no existe una definición unívoca o normativa para el concepto de Gobernanza de datos*" y que este concepto ha ido evolucionando de referirse al contexto interno de las organizaciones sólo en lo relativo al control y gestión de sus datos a evolucionar hacia un concepto más amplio y elaborado como el que propone el Instituto de Gobernanza de Datos (*Data Governance Institute*), que lo define[31] como "*el ejercicio de la toma de decisiones y la autoridad en asuntos relacionados con los datos", y de forma más amplia, como "un sistema de derechos de decisión y responsabilidades para los procesos relacionados con la información, ejecutados según modelos acordados que describen quién puede tomar qué acciones con qué información, y cuándo, bajo qué circunstancias, utilizando qué métodos*".

Por su parte la Agencia Española de Protección de Datos (AEPD) afirmó ya en 2020 que "*la gobernanza de datos, o data governance, es la estrategia para la correcta administración y gestión de la política de datos en la organización*".

Y que es "*una parte importante de esa política ha de ser la política de protección de datos establecida en el artículo 24 del RGPD y la necesidad de adoptar dichas políticas expresada en el considerando 78*", y concreta los objetivos de la gobernanza de datos, los específicos a añadir cuando se traten datos personales, así como determinados factores de éxito de esta[32].

A partir de las aportaciones doctrinales, pueden sintetizarse las siguientes características fundamentales de la gobernanza de datos[33]:

1. Poner en valor los datos como un activo de la organización que debe gestionarse.

2. Establecer responsabilidades, tanto a nivel estratégico, táctico como operativo.

3. Definir pautas y normas para velar por la calidad de los datos y su uso adecuado.

4. Existencia de un liderazgo estratégico de la dirección que no dependa de un exclusivo departamento o área de la compañía, de forma que, como sistema transversal, sea coherente con los objetivos y cultura de la organización y, por supuesto, con la normativa vigente.

5. Y añadiría que la propia gobernanza de datos exige que se integre con políticas de transformación digital, seguridad y cumplimiento normativo para que dicho flujo de datos sea seguro.

Vamos a centrarnos ahora en la Gobernanza de los datos en sistemas de IA en el marco del RIA

[30] María Loza Corera, "Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento", en *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, dirs. Lorenzo Cotino Hueso y Pere Simón Castellano (Las Rozas, Madrid: Editorial Aranzadi, 2024), 567–591

[31] Data Governance Institute, "Definitions of Data Governance," *datagovernance.com*, consultado el 8 de marzo de 2026, https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/

[32] Agencia Española de Protección de Datos, "Gobernanza y política de protección de datos," *Blog de la AEPD*, 2 de septiembre de 2020, https://www.aepd.es/prensa-y-comunicacion/blog/gobernanza-y-politica-de-proteccion-de-datos

[33] Las tres primeras definiciones proceden de Manuel Salvador Serna, "Inteligencia artificial y gobernanza de datos en la Administración Pública: sentando las bases para su integración a nivel corporativo", en *Repensando la administración pública: administración digital e innovación pública* (Madrid: INAP, 2021), 126–148, a las que hemos añadido el matiz de que entendemos que dichas responsabilidades deben definirse en los diferentes niveles organizativos; la cuarta definición procede de María Loza Corera, Loza Corera, "Datos y gobernanza de datos…".

sin desconocer que existen otros en el ámbito del conocido como *soft law*[34]. En conjunto, estas iniciativas configuran un ecosistema internacional convergente hacia una gobernanza responsable de los datos y de la IA.

Como se ha indicado, el RIA (en su aproximación a riesgos) clasifica los sistemas de IA en diferentes niveles y para cada uno dispone una serie de obligaciones. En concreto, la Gobernanza de datos se regula en el artículo 10, enmarcado en el capítulo II del Título III dedicado a los Sistemas de IA alto riesgo. Sus obligaciones se aplican con independencia de si se tratan datos personales o no, lo que refleja evidencia su carácter transversal. En un sistema basado en datos, no disponer de datos adecuados (de entrenamiento, de validación y prueba etc.) puede generar sesgos que a su vez pueden amplificarse por mecanismos de retroalimentación del propio sistema. Obviamente, muchos sistemas de IA tratan datos de carácter personal, por lo que el RGPD directamente aplicable, aunque el articulado del RIA no dispone expresamente el cumplimiento de ningún principio de protección datos, cosa que sí que hacen los Considerandos. Por ejemplo, el Considerando 67, que hace referencia a la transparencia sobre el fin original de la recopilación de datos; o el Considerando 69, que se refiere a la necesidad de garantizar el derecho a la protección de los datos personales a lo largo de todo el ciclo de vida del sistema de IA; mencionando de forma expresa los principios de minimización de datos y de protección de datos desde el diseño y por defecto, cuando el sistema trate datos personales.

El Considerando 67 también está relacionado con gobernanza de los datos cuando al disponer que "*los conjuntos de datos para el entrenamiento, la validación y la prueba, incluidas las etiquetas, deben ser pertinentes, lo suficientemente representativos y, en la mayor medida posible, estar libres de errores y ser completos en vista de la finalidad prevista del sistema",* añadiendo de forma complementaria el Considerando 69 que *"las medidas adoptadas por los proveedores para garantizar el cumplimiento de estos principios podrán incluir no solo la anonimización y el cifrado, sino también el uso de una tecnología que permita llevar los algoritmos a los datos y el entrenamiento de los sistemas de IA sin que sea necesaria la transmisión entre las partes ni la copia de los datos brutos o estructurados, sin perjuicio de los requisitos en materia de gobernanza de datos establecidos en el presente Reglamento*".

Por otra parte, cuando el artículo 10 exige que los datos estén exentos de errores y sean completos con vistas a la finalidad prevista, entendemos que se está haciendo referencia directa al principio de exactitud, en consonancia con la que afirma la AEPD[35].

Como vemos, la conexión de ambas normas en materia de gobernanza se puede dar en diferentes puntos, pero merece especial referencia el considerando 70 del RIA, que desarrolla el artículo 10.5, al disponer la posibilidad excepcional bajo ciertos requisitos adicionales a las garantías del RGPD, de que los proveedores de dichos sistemas puedan tratar categorías especiales de datos personales siempre que ofrezcan las garantías adecuadas en relación con los derechos y las libertades fundamentales de las personas físicas siempre que además de las disposiciones establecidas en el RGPD, la Directiva (UE) 2016/680 y el Reglamento (UE) 2018/1725, cumplan todas las condiciones: (i) que no pueda realizarse mediante datos sintéticos, anónimos u otros datos; (ii) que las categorías especiales de datos personales tratados se sujeten a limitaciones técnicas en cuanto a la reutilización y a las medidas más avanzadas de seguridad (iii) se sujeten a medidas que garanticen la seguridad y protección de los datos personales tratados (iv) no se transmitirán, transferirán ni serán accesibles de otro modo a terceros; (v) se suprimirán una vez que se haya corregido el sesgo o los datos personales hayan llegado al final de su período de conservación.

Por último, nos parece especialmente relevante la contundencia con la que el Considerando 63

[34] Podemos citar, entre otras: Las OECD AI Principles (2019), adoptadas por más de 40 países, que promueven una IA centrada en el ser humano, responsable y transparente. La OCDE ha desarrollado además el OECD Framework for the Classification of AI Systems, que incluye criterios sobre gobernanza de datos y riesgos. En el ámbito europeo, además del RIA, se han publicado documentos como las Ethics Guidelines for Trustworthy AI del Grupo de Expertos de Alto Nivel en IA de la Comisión Europea (2019), que identifican la gobernanza de datos como uno de los requisitos clave para una IA fiable. La UNESCO Recommendation on the Ethics of Artificial Intelligence (2021), que establece principios globales sobre gobernanza, protección de datos, equidad y supervisión humana, con especial atención a los contextos culturales y sociales. El Artificial Intelligence Risk Management Framework (AI RMF) del NIST (EE.UU.), publicado en 2023, que ofrece un enfoque voluntario para identificar, evaluar y gestionar riesgos asociados al uso de IA, incluyendo la gobernanza de datos como uno de sus pilares fundamentales. Las Governance Guidelines for the Implementation of AI Principles del Gobierno de Japón, que proporcionan recomendaciones prácticas para aplicar principios éticos en el desarrollo y uso de sistemas de IA, con énfasis en la trazabilidad, la transparencia y la calidad de los datos. Así como otras iniciativas como la Global Partnership on AI (GPAI), que impulsa investigaciones y recomendaciones sobre gobernanza responsable de la IA, incluyendo la gestión de datos, la equidad algorítmica y la interoperabilidad.

[35] Agencia Española de Protección de Datos (AEPD), "Inteligencia artificial: principio de exactitud en los tratamientos," *Blog de la AEPD*, 31 de mayo de 2023, https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-principio-de-exactitud-en-los-tratamientos

indica algo que pudiera parecer obvio, pero que conviene "subrayar" mediante dicho considerando: "*No debe entenderse que el presente Reglamento constituye un fundamento jurídico para el tratamiento de datos personales, incluidas las categorías especiales de datos personales, en su caso, salvo que el presente Reglamento disponga específicamente otra cosa*", que es el supuesto analizado en el párrafo anterior.

3.3. La gobernanza de la IA como herramienta para gestionar los riesgos derivados de los sesgos

La gobernanza de los datos constituye una dimensión esencial de un marco más amplio: la gobernanza de la IA. Mientras que la gobernanza de datos en el RIA se centra en cómo se gestionan los datos utilizados por los sistemas de IA, especialmente los de alto riesgo, la gobernanza de la IA abarca todo el ciclo de vida (desde el diseño hasta el despliegue y supervisión) y diversas tareas que deben de estar "correctamente ordenadas".

El RIA no ofrece una definición de gobernanza de la IA ni tampoco encontramos una definición en la ISO *Information technology-Artificial intelligence -Artificial intelligence concepts and terminology ISO/IEC 22989*, pero puede entenderse como el conjunto de políticas, procesos, roles y controles que una organización establece para garantizar que sus sistemas de IA se diseñan, desarrollan, despliegan y supervisan de forma responsable.

Hablar de Gobernanza de la IA supone,por tanto,definir un Sistema de Gestión de Inteligencia Artificial (SGIA) e implementarlo, siguiendo el enfoque PDCA (plan, do, check, act), permitiendo una gestión proactiva y demostrable del riesgo, incluyendo los sesgos.

La adopción de un estándar, como por ejemplo la citada ISO 42001, se perfila como la opción óptima por varias razones: su carácter de estándar, lo que supone que haya sido analizada y contrastada; por su validez "global", lo que permite tener en cuenta buenas prácticas aplicables a diferentes países y a contextos multinacionales; y también porque el RIA reconoce y fomenta el uso de herramientas válidas para demostrar cumplimiento, especialmente en sistemas de alto riesgo, que pueden ser adoptadas como base para normas armonizadas europeas si lo aprueba la comisión, lo que pudiera suceder en un futuro este caso, y tal y como ha sucedido en otros precedentes[36]. En esos supuestos el reglamento establece los requisitos legales generales y la norma ISO proporciona un modelo que aterriza para ayudar a cumplir con esos requisitos de forma complementaria. Otro motivo por el que apostar por esta norma ISO es porque no se limita a sistemas de alto riesgo y se puede aplicar a cualquier tipo de operador (principalmente proveedor y responsable del despliegue), sea grande o pequeño y sea cual sea el sector de su actividad. Adicionalmente, al tratarse de una norma certificable permite además demostrar confianza ante posibles clientes y mercado, ciudadanos y sociedad.

Tomando como referencia la ISO/IEC 42001, el Sistema de Gestión de la Inteligencia Artificial (SGIA) puede estructurarse de manera coherente con la lógica de mejora continua propia de los sistemas de gestión conforme al tradicional ciclo PDCA (Plan–Do–Check–Act) y que cada actor tenga su propio SGIA, pero alineado y complementario. Por ejemplo, el proveedor se enfoca en el diseño ético y técnico, mientras que el usuario o responsable del despliegue se centra en el uso responsable y el impacto real en su caso concreto. Por poner un par de ejemplos centrados en la fase de plan: En relación con la obligación de disponer de una política de IA, el proveedor define principios éticos, transparencia, mitigación de sesgos desde el diseño y el responsable del despliegue adapta la política a su contexto de uso. Si hablamos de roles y responsabilidades, el proveedor asigna responsables de diseño ético, gestión de datos, validación de modelos y el responsable del despliegue asigna responsables de supervisión, uso seguro, atención a impactos en personas.

Asimismo, cada organización debe de adaptarlo a su tamaño, actividad etc. y en función de todo lo anterior aplicarán una serie de controles u otros (vid anexo 1 de la norma) que serán reflejados en la declaración de aplicabilidad, e incluso pueden tener que ampliarse. Si hablamos de sesgos vemos algunos ejemplos de controles específicos como el control A.6.3.1 (evaluación de impacto de IA), que promueve identificar impactos en derechos fundamentales desde el diseñ;, o el control A.6.4.1 (gestión de riesgos de IA), que obliga a establecer procesos para detectar y mitigar sesgos. Por último, el hecho de que disponga de una estructura armonizada de alto nivel con otras normas ISO (sin perjuicio de peculiaridades propias de la norma), que además es habitual y posible que

[36] Por ejemplo, en el ámbito sanitario tenemos la ISO 13485: 2016 (adaptada por la UNE-EN ISO 13485:2016) de "Gestión de calidad en productos sanitarios" en el Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios; que fue reconocida en la decisión de Ejecución (UE) 2021/1182 de la Comisión, de 16 de julio de 2021.

se hayan desplegado en las organizaciones que acometan un proyecto de ISO 42001, permite la integración coherente y eficiente de los diversos sistemas de gestión. Ejemplos claros y "cercanos" son desde la más general ISO 9001:2015 (sobre Sistemas de Gestión de Calidad) a otras más específicas del ámbito TI como por ejemplo la ISO/IEC 27001:2022 (Gestión de la Seguridad de la Información -SGSI- o la ISO/IEC 27701:2025, que es la versión actualizada del Sistema de Gestión de Información de Privacidad (PIMS).

4. Consideraciones finales

A modo de conclusiones realizamos las siguientes consideraciones finales:

1. Es fundamental conocer el concepto de sesgos, diferenciarlo de otros afines y comprender cómo puede introducirse en los sistemas de IA a través de las diferentes etapas de su ciclo de vida. Solo desde ese conocimiento es posible identificar sesgos y riesgos asociados, puesto que este es el primer paso para poder gestionarlos.

2. La legislación y particularmente el RIA proporciona un conjunto de "herramientas" (materializadas como obligaciones para los distintos operadores), que guardan una relación clara con la gestión de los riesgos para los derechos fundamentales. Entre ellas destacan, por especial relevancia, las Evaluaciones de Impacto en Derechos Fundamentales (EIDF), sin perjuicio de otras obligaciones que también contribuyen a mitigar dichos riesgos.

3. De entre dichas obligaciones, adquiere especial relevancia la gobernanza de los datos en el RIA, que aborda cómo se gestionan los datos utilizados por los sistemas de IA de alto riesgo. La calidad de los datos se encuentra estrechamente vinculada a la aparición o amplificación de sesgos, por lo que su adecuada gestión constituye un elemento nuclear para la fiabilidad del sistema.

4. Finalmente, todas estas obligaciones deben de integrarse en un sistema de gestión estructurado, de tal forma que permita planificar, operar, supervisar y mejorar de forma continua los procesos asociados. En este sentido, se ha puesto el acento en el valor de apoyarse en un estándar como la ISO 42001, por los beneficios y las ventajas que ofrece para la ordenación coherente y verificable de la gobernanza de la IA como se ha abordado a lo largo del artículo.

Bibliografía

Agencia Española de Protección de Datos. "Gobernanza y política de protección de datos." *Blog de la AEPD*. 2 de septiembre de 2020. https://www.aepd.es/prensa-y-comunicacion/blog/gobernanza-y-politica-de-proteccion-de-datos

"Inteligencia artificial: principio de exactitud en los tratamientos." *Blog de la AEPD*. 31 de mayo de 2023. https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-principio-de-exactitud-en-los-tratamientos

Chaveli Donet, Eduard. "La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento". En *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, dirigido por Lorenzo Cotino Hueso y Pere Simón Castellano, 495–531. Las Rozas (Madrid): Editorial Aranzadi, 2024.

"Sesgos en la IA (I): La necesaria distinción entre sesgos y conceptos afines." *Telefónica Tech Blog*, 11 de febrero de 2025. https://telefonicatech.com/blog/sesgos-en-ia-parte-i-distincion-entre-sesgos-y-conceptos-afines

"Sesgos en la IA (V): Introducción de los riesgos en el ciclo de vida de los sistemas de IA (parte 1)." *Telefónica Tech Blog*. 22 de abril de 2025. https://telefonicatech.com/blog/sesgos-ia-v-introduccion-riesgos-en-ciclo-de-vida-sistemas-de-ia-parte-1

Cotino Hueso, Lorenzo, y Pere Simón Castellano, dirs. *Tratado sobre el reglamento de inteligencia artificial de la Unión Europea*. Madrid: Aranzadi, 2024.

Data Governance Institute. "Definitions of Data Governance." *datagovernance.com*. Consultado el 8 de marzo de 2026. https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/

ISO/IEC. ISO/IEC 22989:2023. Artificial Intelligence — Artificial Intelligence Concepts and Terminology. Geneva: International Organization for Standardization, 2023.

Loza Corera, María. "Datos y gobernanza de datos y conexiones con principios protección de datos en el artículo 10 del Reglamento". En *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, dirigido por Lorenzo Cotino Hueso y Pere Simón Castellano, 567–591. Las Rozas (Madrid): Editorial Aranzadi, 2024.

Mantelero, Alessandro. *Artificial Intelligence and Data Protection: Challenges and Possible Remedies*. Estrasburgo: Consejo de Europa, 2019.

Salvador Serna, Manuel. "Inteligencia artificial y gobernanza de datos en la Administración Pública: sentando las bases para su integración a nivel corporativo". En *Repensando la administración pública: administración digital e innovación pública*, 126–148. Madrid: Instituto Nacional de Administración Pública, 2021

Schwartz, Reva, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, y Patrick Hall. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. NIST Special Publication 1270. Gaithersburg, MD: National Institute of Standards and Technology, 2022. https://doi.org/10.6028/NIST.SP.1270

Snoek, Suzanne, e Isabel Barberá. *From Inception to Retirement: Addressing Bias Throughout the Lifecycle of AI Systems: A Practical Guide*. Rhite, 2024.

Estudios de privacidad

Capacidades para la era de la IA agéntica: gobernanza epistémica y co-inteligencia humano-máquina

Capabilities for the age of agentic AI: epistemic governance and human–machine Co-Intelligence

Áurea Rodríguez López
Especialista en estrategias de innovación, escritora, estudiante de doctorado, UIC, directora corporativa Eurecat

Resumen:
La transición desde la automatización clásica hacia sistemas de inteligencia híbrida se acelera con la irrupción de la IA agéntica: arquitecturas capaces de planificar, decidir y ejecutar flujos de trabajo mediante el uso de herramientas y datos en bucle. Este cambio desplaza el foco desde la pregunta por la sustitución o la mera "ayuda" al trabajo hacia un problema más fino: qué capacidades deben adquirir personas, equipos y organizaciones para delegar, supervisar y aprender con agentes cada vez más competentes sin perder criterio, trazabilidad ni control. El artículo propone un marco de "capacidades para la co-inteligencia" que integra: (i) capacidades cognitivas (formulación de pro-blemas, vigilancia epistémica, verificación y calibración), (ii) capacidades socio-técnicas (diseño de instrucciones, orquestación de herra-mientas, evaluación continua y recuperación ante fallos) y (iii) capacidades de gobernanza (gestión de datos, registro de decisiones, auditoría y rendición de cuentas). A partir de evidencia empírica sobre productividad y nivelación de habilidades, y de la literatura sobre offloading cognitivo, se argumenta que la adopción madura de la IA agéntica depende menos de "modelos más inteligentes" que de prácticas repetibles que conviertan la delegación en aprendizaje y reduzcan la exposición a errores encadenados, inyección de instrucciones y sobre-autonomía. El texto concluye con recomendaciones para el diseño organizativo y la formación: alfabetización agéntica, evaluación como hábito, arquitectura de permisos, y cultura de trazabilidad. La regulación, en particular, las exigencias de supervisión humana del AI Act, se interpreta como un marco habilitador que obliga a operacionalizar capacidades, no como un sustituto de ellas.

Palabras clave:
IA agéntica; Inteligencia híbrida; Capacidades; Gobernanza de IA; Co-inteligencia humano-máquina.

Abstract:
The shift from traditional automation to hybrid intelligence is accelerating with the rise of agentic AI: architectures that can plan, decide, and execute workflows by looping over tools and data. This transformation reframes the debate from substitution versus augmentation to a more precise question: which capabilities do individuals, teams, and organizations need in order to delegate, supervise, and learn with increasingly competent agents without losing judgment, traceability, or control? This article proposes a "co-intelligence capability" framework integrating (i) cognitive capabilities (problem framing, epistemic vigilance, verification and calibration), (ii) socio-technical capabilities (instruction design, tool orchestration, continuous evaluation, and failure recovery), and (iii) governance capabilities (data management, decision logging, auditing, and accountability). Drawing on empirical evidence on productivity and skill leveling, and on the cognitive offloading literature, it argues that mature adoption of agentic AI depends less on "smarter models" than on repeatable practices that turn delegation into learning while reducing exposure to cascading errors, instruction injection, and excessive autonomy. The article concludes with recommendations for organizational design and training: agentic literacy, evaluation as a routine, permission architecture, and a culture of traceability. Regulation, particularly the human oversight requirements in the EU AI Act, is interpreted as an enabling scaffold that forces capability operationalization rather than replacing it.

Keywords:
Agentic AI; Hybrid intelligence; Capabilities; AI governance; Human–machine co-intelligence.

1. Introducción

En apenas dos años, la conversación pública sobre la inteligencia artificial (IA) ha pasado de la "automatización" a la "aumentación" y ahora, a la "colaboración". La IA generativa, inicialmente utilizada como asistente conversacional o herramienta de redacción, está dando paso a sistemas más autónomos que planifican y ejecutan acciones: consultan fuentes, llaman a herramientas, actualizan registros, crean artefactos y devuelven resultados con resultados superiores a los humanos en algunos casos. Este giro se suele describir como IA agéntica, un término que alude a la capacidad de un sistema para orientar su conducta hacia un objetivo a lo largo de varios pasos, interactuando con un entorno mediante herramientas y feedback.

Este cambio se entiende mejor si se adopta el paradigma de la inteligencia híbrida: configuraciones socio-técnicas en las que agentes humanos y algorítmicos co-producen resultados y donde el rendimiento emerge de su interacción, no de cada parte por separado[1]. En lugar de preguntar "qué tareas hace la IA", conviene preguntar "qué tareas se vuelven delegables" y "qué tareas se vuelven críticas" cuando la IA asume partes del flujo. La IA agéntica empuja esta lógica porque deja de limitarse a sugerir contenido, sino que puede explorar el espacio de soluciones, seleccionar herramientas y ejecutar pasos intermedios, de modo que la interacción humana se desplaza hacia el diseño del encargo, la autorización de acciones y la verificación.

Para evitar confusiones conceptuales, hablaremos de co-inteligencia humano-máquina para referirnos a procesos en los que la IA participa en la generación de hipótesis, en la búsqueda de evidencias y en la redacción o cálculo, mientras que el humano conserva responsabilidad sobre los objetivos, la selección de evidencias decisivas y la decisión final. Esta definición se alinea con la idea de la IA como "copiloto" y con recomendaciones industriales que insisten en que el valor depende del diseño de la interacción, de la calidad de los datos y de la capacidad de evaluación, no solo del tamaño del modelo.

[1] Dominik Dellermann, Philipp Ebel, Matthias Söllner y Jan Marco Leimeister, "Hybrid Intelligence," *Business & Information Systems Engineering* 61, no. 5 (2019): 637–643, https://doi.org/10.1007/s12599-019-00595-2

Esta evolución no convierte automáticamente a la máquina en un sujeto moral o jurídico; "agencia" aquí es un rasgo funcional, no una atribución de personalidad. Sin embargo, sí altera la economía cognitiva del trabajo del conocimiento. Cuando el sistema deja de ser una calculadora avanzada y empieza a ejecutar tareas compuestas, con acceso a herramientas y a datos, el papel humano se desplaza: menos teclear y más definir objetivos, fijar restricciones, verificar resultados, decidir excepciones y aprender del desempeño. La diferencia no es solo de grado (más automatización) sino de tipo: cambia la unidad de trabajo delegable, que ya no es una tarea discreta, sino un proceso.

En este contexto, el debate "sustitución versus aumento" se queda corto por dos motivos. Primero, porque la sustitución no depende solo de la capacidad del modelo, sino del ensamblaje socio-técnico (datos, herramientas, procesos, incentivos) que hace que una tarea sea delegable de forma fiable. Segundo, porque el aumento puede degradar capacidades humanas si se convierte en externalización sistemática del pensamiento (cognitive offloading) sin prácticas de verificación y aprendizaje. El problema ya no es si la IA "ayuda", sino si ayuda de un modo que preserve o incluso fortalezca el criterio.

La hipótesis central de este artículo es que la adopción eficaz de IA agéntica requiere un conjunto de capacidades específicas, distribuibles entre individuos, equipos y organización, que hacen posible la co-inteligencia: la construcción conjunta de conocimiento y decisiones por agentes humanos y algorítmicos. La propuesta implica un desplazamiento analítico: desde derechos y obligaciones como punto de partida hacia capacidades como unidad de diseño. El derecho sigue siendo relevante, pero en un sentido instrumental: establece un suelo de garantías y un lenguaje de responsabilidad, pero no resuelve por sí mismo qué prácticas, competencias y artefactos deben existir para que la supervisión sea efectiva. En un sistema agéntico, "tener un humano en el bucle" no es una propiedad; es el resultado de un diseño.

2. De la IA generativa a la IA agéntica: qué cambia

Para entender el giro agéntico conviene distinguir entre tres niveles de integración. En el nivel 1, la IA actúa como motor de producción: genera texto, código o resúmenes a partir de una instrucción única. En el nivel 2, la IA se integra en un flujo guiado: se encadena con recuperación de información, plantillas, validaciones y reglas; el recorrido está determinado por el software. En el nivel 3, el sistema decide el recorrido: el modelo gestiona la ejecución del flujo de trabajo, elige herramientas y controla la iteración (plan-acción-observación-replanificación) hasta cumplir un criterio de parada. Este tercer nivel se aproxima a lo que la literatura industrial llama "agentes" o "sistemas agénticos"[2].

Dos rasgos distinguen este nivel 3. El primero es la autonomía operacional: el agente decide qué pasos dar, no solo cómo redactar un paso ya definido. El segundo es la ampliación del "espacio de acción" mediante herramientas: el agente puede consultar una base de datos, ejecutar código, escribir en un CRM o lanzar una búsqueda. La combinación de autonomía y herramientas crea un sistema con capacidad de actuación, por lo que el error ya no es únicamente semántico (una respuesta incorrecta) sino operativo (una acción incorrecta).

2.1. Arquitectura agéntica: planificar, actuar, observar y corregir

Los agentes operan, típicamente, en bucles. Una formulación simple es plan-act-observe: el sistema propone un plan, ejecuta una acción (por ejemplo, llamar a una herramienta), observa el resultado y decide el siguiente paso. En versiones más robustas, el bucle incorpora una fase de crítica o reflexión: el agente evalúa si el resultado satisface el criterio y, si no, re-intenta o replantea el plan. Este patrón, documentado en guías de implementación, explica por qué los agentes pueden resolver tareas abiertas sin que el programador especifique todos los pasos: la búsqueda de una trayectoria adecuada se hace en tiempo de ejecución.

Este diseño tiene implicaciones directas para la supervisión. En un sistema de una sola respuesta, la supervisión se concentra en el output. En un sistema agéntico, la supervisión puede situarse en el plan (validar antes de actuar), en las herramientas (limitar permisos), en los checkpoints (revisiones parciales) o en el criterio de parada (cuándo el agente debe considerar la tarea "terminada"). La pregunta "¿dónde pongo al humano?" se vuelve una decisión de ingeniería: depende del coste de verificación y del daño potencial de una acción incorrecta.

[2] Véanse, entre otras, OpenAI, "A Practical Guide to Building Agents" (documento PDF), consultado el 31 de diciembre de 2025, https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf; y Anthropic, "Building Effective Agents," 19 de diciembre de 2024, https://www.anthropic.com/research/building-effective-agents.

2.2. Memoria, contexto y límites: por qué "más inteligencia" no elimina el riesgo

Los modelos actuales trabajan con ventanas de contexto finitas y con memorias externas diseñadas por el sistema (por ejemplo, un registro de trabajo o una base documental). En IA agéntica, la memoria es una pieza crítica: determina qué información se conserva, cómo se resume y qué se olvida. Las decisiones de memoria afectan tanto a la calidad como a la seguridad: un agente que "recuerda" demasiado puede exponer datos; uno que "olvida" demasiado puede repetir errores o perder restricciones. Además, la capacidad de un agente para operar en entornos reales depende de herramientas que no son perfectas: buscadores con sesgos, bases de datos incompletas, APIs con errores, documentos ambiguos. El punto de vista de capacidades ayuda a evitar el espejismo de la "superinteligencia práctica": lo que importa es el sistema completo, incluido el diseño de herramientas, permisos, registros y métricas.

Desde una perspectiva de diseño, el salto al nivel 3 reorganiza la arquitectura del trabajo en torno a tres componentes: modelo, herramientas e instrucciones.[3] El modelo aporta capacidad de razonamiento y planificación; las herramientas permiten actuar; y las instrucciones definen objetivos, límites, criterios de calidad y condiciones de parada. El desempeño no depende solo del modelo "más inteligente", sino del acoplamiento entre estos componentes: herramientas mal diseñadas o mal documentadas aumentan errores; instrucciones ambiguas amplifican desviaciones; y un modelo muy capaz sin permisos adecuados puede ser más riesgoso que uno menos capaz con guardarraíles.

La IA agéntica introduce además una nueva unidad de evaluación. En IA generativa, se evalúa una respuesta (un texto) respecto a un criterio. En IA agéntica, se evalúa una trayectoria: una secuencia de decisiones y acciones. Esto obliga a pensar en métricas de proceso: número de herramientas llamadas, tasa de errores, tiempo por paso, cobertura de evidencias, y calidad del registro. En la práctica, este enfoque desplaza el control desde el "resultado final" hacia la trazabilidad de la ejecución.

Un riesgo específico de los sistemas agénticos es la "agencia excesiva". Si el agente puede ejecutar acciones con pocos frenos, un error de interpretación o una instrucción maliciosa (por ejemplo, *prompt injection* indirecta en un documento que el agente lee) puede traducirse en acciones no deseadas. Marcos de seguridad recientes para aplicaciones con modelos de lenguaje identifican la agencia excesiva y la inyección de instrucciones como riesgos centrales.[4] Por eso, la seguridad deja de ser un asunto exclusivamente técnico y se convierte en una práctica operativa: delimitar permisos, aislar entornos, diseñar herramientas ergonómicas para agentes y crear registros auditables. Un ejemplo ayuda. Supongamos un agente encargado de preparar un informe de diligencia debida sobre un proveedor. En un enfoque no agéntico, el profesional pide un resumen, revisa y corrige. En un enfoque agéntico, el agente: (i) consulta bases internas, (ii) busca información pública, (iii) extrae indicadores, (iv) redacta un borrador, (v) genera una matriz de riesgos, y (vi) propone recomendaciones. En cada paso puede equivocarse: confundir identidades, interpretar erróneamente una norma, o incorporar una instrucción maliciosa encontrada en una web. El resultado final puede parecer impecable; el problema puede estar en la trayectoria. La pregunta clave es: ¿dispone el usuario de medios para inspeccionar esa trayectoria sin rehacer todo el trabajo?

Esa pregunta conduce a la tesis del artículo: operar con IA agéntica exige capacidades específicas que hacen viable la inspección, la verificación y el aprendizaje. Sin ellas, la organización puede obtener velocidad a corto plazo a costa de fragilidad a largo plazo.

3. Impactos en el trabajo del conocimiento: productividad, nivelación y *offloading*

La evidencia empírica sobre IA generativa en entornos de trabajo sugiere incrementos de productividad y cambios en la distribución de habilidades. Un estudio de referencia en un entorno de atención al cliente reporta mejoras medias relevantes y heterogéneas, con mayores beneficios para perfiles menos experimentados, y ganancias más modestas, o incluso pequeñas caídas de calidad, para perfiles expertos.[5] Este patrón se interpreta como "nivelación": el sistema difunde

[3] OpenAI, "New Tools for Building Agents," 11 de marzo de 2025, https://openai.com/index/new-tools-for-building-agents/.

[4] OWASP Foundation, "OWASP Top 10 for Large Language Model Applications (Version 1.1)," consultado el 31 de diciembre de 2025, https://owasp.org/www-project-top-10-for-large-language-model-applications/.

[5] Erik Brynjolfsson, Danielle Li y Lindsey Raymond, "Generative AI at Work," *The Quarterly Journal of Economics* 140, n.º 2 (2025): 889–942, https://doi.org/10.1093/qje/qjae044.

buenas prácticas, reduce la brecha entre novatos y expertos y acelera el aprendizaje inicial.

Esta nivelación no es un detalle: reconfigura la idea de "competencia" en el trabajo del conocimiento. Si un agente puede redactar, sintetizar y argumentar con un nivel razonable, el valor diferencial humano se desplaza hacia tareas menos visibles: definir el problema, seleccionar evidencias, negociae, detectar supuestos erróneos, y decidir bajo incertidumbre. En suma, se desplaza desde la producción hacia el juicio.

Este desplazamiento tiene dos efectos prácticos. Primero, hace que la experiencia sea menos visible: el experto aporta valor no tanto por producir un texto impecable, sino por saber qué preguntar, qué no creer y qué evidencia exigir. Segundo, cambia el "punto de fallo" de los procesos. Si antes el fallo típico era una mala redacción o un cálculo erróneo, ahora el fallo típico es un objetivo mal especificado o un supuesto no detectado que se amplifica a lo largo del flujo agéntico.

Desde el punto de vista organizativo, esto implica que la adopción de IA no es solo una cuestión de licencias o de herramientas, sino de rediseño de prácticas de calidad y de las propias organizaciones. Las organizaciones con mayor madurez tienden a formalizar criterios: definen qué decisiones exigen evidencia primaria, qué tareas requieren revisión por pares, qué umbrales activan escalado y qué métricas se monitorean (tasa de correcciones, incidencias de datos, retrabajo). La IA agéntica acelera la necesidad de estas prácticas.

Además, la nivelación de habilidades puede coexistir con una "polarización" de capacidades. Mientras las tareas de producción se democratizan, aumentan los retornos de las capacidades de verificación, orquestación y gobernanza. Esto sugiere una reconfiguración de la formación: ya no basta con enseñar a "hacer" (redactar, resumir), sino a "asegurar" (verificar, auditar) y a "diseñar" (definir objetivos, restricciones y evaluaciones).

Sin embargo, la misma lógica que facilita el aprendizaje puede favorecer la dependencia. La literatura sobre la descarga (*offloading*) cognitivo describe cómo las personas delegan memoria, cálculo y razonamiento a artefactos externos (desde una libreta hasta internet) y cómo esa delegación puede ser adaptativa o perjudicial según el contexto.[6] Con IA generativa, la externalización es más profunda: el sistema no solo almacena o busca, sino que produce inferencias, sugiere decisiones y completa razonamientos. Además, la interfaz conversacional hace que la delegación parezca diálogo: el usuario siente que "piensa con" el sistema incluso cuando está aceptando resultados sin reconstruirlos.

Tres mecanismos explican por qué la descarga cognitiva puede erosionar capacidades. Primero, reduce la práctica deliberada: si el sistema escribe o calcula por nosotros, practicamos menos la habilidad subyacente. Segundo, altera la memoria: cuando sabemos que un recurso externo está disponible, codificamos menos información interna (el "efecto Google"). Tercero, cambia la metacognición: la facilidad de obtener respuestas reduce la calibración del propio conocimiento y puede aumentar la confianza injustificada. Estos efectos no son inevitables, pero emergen cuando la organización incentiva la velocidad sin crear contrapesos.

La IA agéntica intensifica el dilema. Cuando el agente encadena pasos y devuelve un resultado final, el usuario ve menos del proceso y, por tanto, tiene menos oportunidades de detectar errores. La "fluidez" del resultado induce una ilusión de corrección: suena competente, luego debe estarlo. Este fenómeno se agrava en tareas con verificación costosa o ambigua (estrategia, análisis normativo, diagnóstico organizativo), donde no existen pruebas automáticas equivalentes a un test unitario.

En contraste, hay dominios donde el enfoque agéntico puede fortalecer capacidades: programación con tests, análisis con datos reproducibles, o investigación con trazabilidad de fuentes. La diferencia es la verificabilidad. Cuando se puede verificar sin rehacer el trabajo, se puede confiar sin fe. Por eso, la co-inteligencia madura se parece a la ingeniería: produce resultados y, al mismo tiempo, genera evidencia de que el resultado es bueno.

La pregunta relevante, por tanto, no es si la IA aumenta la productividad, en muchos casos lo hace, sino bajo qué condiciones esa productividad se convierte en aprendizaje y no en desentrenamiento. Ahí es donde el enfoque de capacidades aporta ventaja: permite describir prácticas concretas para obtener valor sin perder criterio. En lugar de pedir a las personas que "confíen menos", se trata de diseñar rutinas, herramientas y métricas que hagan la verificación viable y, por tanto, incentivada.

[6] Evan F. Risko y Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences* 20, n.º 9 (2016): 676–688; Betsy Sparrow, Jenny Liu y Daniel M. Wegner, "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips," *Science* 333, n.º 6043 (2011): 776–778.

4. Un marco de capacidades para la co-inteligencia

Proponemos siete familias de capacidades que, en conjunto, permiten operar con IA agéntica de forma fiable. No son rasgos individuales "de talento": son combinaciones de conocimientos, prácticas y artefactos que pueden distribuirse entre personas, equipos, procesos y tecnología. En sistemas híbridos, las capacidades se co-construyen: lo que hace posible el juicio humano no es solo el "cerebro", sino el acceso a evidencias, registros, contra-ejemplos y mecanismos de control. La figura conceptual es una "superficie de control": cuanto más autonomía se delega, más importante es disponer de superficies donde intervenir.

Antes de detallar las familias de capacidades, conviene formular cuatro principios de diseño que las atraviesan:

(1) Verificabilidad antes que elocuencia. Un agente útil no es el que "suena" convincente, sino el que produce evidencias y se deja comprobar. En tareas donde no existe verificación razonable, la autonomía debe ser baja.

(2) Trazabilidad como condición del aprendizaje. La organización solo mejora el sistema si puede reconstruir qué ocurrió. Por eso, el registro (prompts, herramientas, datos, decisiones) no es burocracia: es memoria institucional.

(3) Autonomía graduada. No hay un único modo de uso. La autonomía debe ajustarse al riesgo y a la reversibilidad de la acción. Un agente puede tener autonomía alta en borradores y baja en aprobaciones o comunicaciones externas.

(4) Separación de roles. La co-inteligencia se fortalece cuando se separa generación y crítica. Un agente genera hipótesis; otro evalúa. Un humano decide. Esta separación reduce sesgos y favorece la calibración.

Estos principios pueden traducirse en una lista operativa de preguntas, útil como checklist al desplegar un agente:

- ¿Qué decisión o resultado produce el agente y qué daños causaría un error
- ¿Qué evidencias debe aportar para que un humano pueda verificarlo sin rehacer todo el trabajo
- ¿Qué acciones puede ejecutar y qué permisos requiere? ¿Hay acciones irreversibles?
- ¿Qué registros se guardan para permitir auditoría y aprendizaje?

Con estos principios en mente, pasamos a las capacidades.

4.1. Capacidad de formulación de problemas y objetivos

En contextos agénticos, lo difícil no es "pedir" una respuesta, sino especificar qué cuenta como éxito. Esta capacidad incluye: delimitar el problema, identificar restricciones (legales, éticas, presupuestarias), descomponer objetivos en subtareas y traducir criterios de calidad en instrucciones verificables. Se parece a la ingeniería de requisitos, pero aplicada a tareas cognitivas y socio-organizativas.

Una práctica básica es la especificación de salida: antes de ejecutar, describir el formato, el nivel de evidencia y los supuestos permitidos. Por ejemplo: "Redacta un informe de 1.500 palabras con tres opciones estratégicas. Para cada opción, incluye (a) supuestos explícitos, (b) evidencia citada, (c) riesgos y mitigaciones, y (d) condiciones de reversibilidad". Esta especificación reduce alucinaciones porque obliga a anclar el texto en estructura y evidencia.

Otra práctica es el "contrato de incertidumbre": pedir al agente que señale qué partes dependen de supuestos no verificados y qué datos faltan. En IA agéntica, esto se extiende a la planificación: el agente debe proponer un plan y pedir confirmación cuando el plan implique acciones irreversibles o acceso a datos sensibles. La formulación, así, no es solo una instrucción; es un diálogo de diseño.

4.2. Capacidad de delegación y orquestación

Delegar a un agente no es "ceder" la tarea: es asignar una sub-tarea con límites, permisos y checkpoints. Esta capacidad comprende: seleccionar herramientas adecuadas, definir ám-

bitos de acceso (principio de mínimo privilegio), usar patrones de orquestación (por ejemplo, "orquestador-trabajadores" o "evaluador-optimizador") y coordinar múltiples agentes especializados.[7]

En términos prácticos, implica construir un mapa de trabajo donde cada agente tiene: rol, objetivo, conjunto de herramientas autorizadas, criterio de parada y mecanismo de escalado. Un patrón útil es separar "agentes de generación" y "agentes de crítica". En escritura, por ejemplo, un agente produce un borrador y otro lo revisa con criterios explícitos. En investigación, un agente propone fuentes y otro valida que las fuentes son primarias y relevantes.

La delegación también exige comprender costes y latencias. Los agentes agénticos pueden ser caros: cada iteración consume recursos. Por eso, parte de la capacidad es saber cuándo no usar agentes. Un principio defendido en guías industriales es comenzar con soluciones simples y aumentar complejidad solo cuando demuestre mejorar resultados.[8] Delegar bien es también decidir qué delegar.

4.3. Capacidad de vigilancia epistémica y verificación

La vigilancia epistémica es la habilidad de evaluar la fiabilidad de una información o un razonamiento en función de su fuente, coherencia y plausibilidad. En humanos, opera a través de heurísticas; en co-inteligencia, debe complementarse con prácticas formales: trazabilidad de fuentes, contrastación con evidencia primaria y verificación de afirmaciones críticas. La literatura sobre comunicación y cognición destaca que la confianza no es global: es una evaluación situada y revisable[9].

En IA agéntica, esta capacidad se traduce en protocolos. Proponemos una regla operativa: toda afirmación "decisiva" (aquella que, si es falsa, cambia la decisión) debe estar respaldada por una fuente primaria o por un dato reproducible. Esto obliga a diseñar al agente para que cite fuentes y a diseñar herramientas para devolver contexto suficiente. Una cita sin contexto es una falsa trazabilidad.

La verificación puede ser humana o automática. En software, tests. En análisis de datos, reproducibilidad (scripts y datasets). En regulación, consulta de textos oficiales. En tareas blandas, triangulación (múltiples fuentes y perspectivas). El objetivo no es eliminar incertidumbre, sino reducirla y localizarla.

4.4. Capacidad de calibración y metacognición

Incluso con verificación, muchas tareas tienen incertidumbre residual. Calibrar consiste en ajustar la confianza subjetiva al desempeño real: saber cuándo el sistema suele fallar, en qué dominios es fiable y qué señales anticipan errores. Esta capacidad requiere métricas y feedback: sin medición, la confianza se convierte en opinión.

En agentes, la calibración se apoya en evals y en análisis de fallos similares a los de ingeniería de software. Guías recientes insisten en construir evaluaciones para herramientas y agentes y en analizar transcripciones para detectar patrones de confusión[10]. La organización que adopta IA agéntica debería mantener un repositorio de "casos de fallo", etiquetados por causa (datos faltantes, ambigüedad, herramienta mal definida, *prompt injection*, etc.) y por severidad.

La metacognición incluye también saber "qué no sé". Paradójicamente, un sistema muy elocuente puede ocultar ignorancia. Por eso, una práctica útil es pedir estimaciones probabilísticas o rangos, y comparar esas estimaciones con resultados reales. Con el tiempo, se construye una "tabla de calibración" por dominio: por ejemplo, en redacción de comunicaciones, el agente es fiable; en interpretación jurisprudencial reciente, requiere verificación intensa; en síntesis de políticas internas, depende de si tiene acceso al repositorio correcto.

[7] Anthropic, "Building Effective Agents.".

[8] OpenAI, "A Practical Guide to Building Agents."

[9] Dan Sperber et al., "Epistemic Vigilance," *Mind & Language* 25, n.º 4 (2010): 359–393.

[10] National Institute of Standards and Technology (NIST), *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1) (Gaithersburg, MD: NIST, 2024), https://doi.org/10.6028/NIST.AI.600-1.

4.5. Capacidad de resiliencia cognitiva frente a la descarga cognitiva

La descarga cognitiva no es mala en sí, lo es cuando sustituye la práctica necesaria para conservar pericia. Resiliencia cognitiva significa diseñar un uso de IA que preserve el entrenamiento: alternar modos (con y sin IA), practicar verificación activa (reconstruir el argumento con palabras propias), y mantener espacios de trabajo donde la persona produce sin ayuda para consolidar memoria y comprensión.

Proponemos tres técnicas simples. (i) Doble entrega: primero un resultado asistido y luego una versión "explicada" por la persona, donde se justifiquen supuestos y se expongan evidencias. (ii) Inversión de roles: pedir al agente que critique un razonamiento humano, pero no que lo sustituya; esto fomenta aprendizaje deliberado. (iii) "Ayuda tardía": en tareas de aprendizaje, retrasar el uso de IA hasta haber intentado resolver por cuenta propia. Estas prácticas mantienen la agencia cognitiva y reducen la tentación de aceptar sin comprender.

4.6. Capacidad de seguridad operativa en sistemas agénticos

Los agentes amplían la superficie de ataque porque consumen entradas externas y ejecutan acciones. La seguridad se convierte en competencia transversal: reconocer prompt injection, separar instrucciones del contenido, aislar herramientas, registrar acciones y limitar permisos. Marcos como el OWASP Top 10 para aplicaciones con modelos de lenguaje enumeran riesgos específicos (inyección de prompts, manejo inseguro de salidas, agencia excesiva) y ofrecen un vocabulario útil para traducir seguridad a checklist[11].

En entornos agénticos, el principio clave es la separación de dominios: el agente puede leer contenido no confiable, pero no debe tratarlo como instrucciones. Esto implica filtros, delimitadores, y políticas de "no ejecutar" sin confirmación humana para acciones irreversibles. También implica sandboxes: probar agentes en entornos aislados con datos sintéticos antes de permitir acceso a sistemas reales.

La seguridad incluye además el diseño de herramientas. Herramientas "ergonómicas" para agentes devuelven información de alto valor, no volúmenes masivos; usan nombres y parámetros claros; y evitan exponer identificadores opacos. La literatura de ingeniería agéntica destaca que herramientas bien diseñadas reducen alucinaciones y errores, porque el agente dispone de affordances más naturales[12]. En este sentido, la seguridad no es un parche: es un diseño de interfaz para sistemas no deterministas.

4.7. Capacidad de gobernanza: datos, trazabilidad y rendición de cuentas

La última familia de capacidades es organizativa. Incluye: gobierno de datos (calidad, licitud, minimización), trazabilidad de decisiones (logs, versionado de prompts, modelos y herramientas), auditoría y mecanismos de rendición de cuentas. En Europa, el AI Act incorpora obligaciones de supervisión humana para sistemas de alto riesgo que, en la práctica, exigen operacionalizar esta capacidad: no basta con "un humano en el bucle" si no tiene información, autoridad y tiempo para intervenir[13]. La gobernanza se concreta en artefactos. Por ejemplo: (i) registro de versiones de modelos y prompts, (ii) bitácora de acciones de herramientas, (iii) políticas de datos (qué se puede usar, cómo se anonimiza, cuánto se retiene), (iv) criterios para escalado a revisión humana, y (v) procedimientos de respuesta a incidentes. Estos artefactos permiten que el sistema sea auditable y, por tanto, mejorable. Sin ellos, los fallos se repiten porque no se aprende de forma institucional.

5. Gobernanza como infraestructura de capacidades: del "control" a la trazabilidad

En la adopción temprana de IA generativa, muchas organizaciones han confundido control con prohibición o con controles superficiales (por ejemplo, exigir que alguien "revise" al final). En IA

[11] OWASP Foundation, "OWASP Top 10 for Large Language Model Applications (Version 1.1)."

[12] Anthropic, "Engineering at Anthropic: Writing Tools for Agents," 18 de diciembre de 2024, https://www.anthropic.com/engineering/writing-tools-for-agents/.

[13] Unión Europea, *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Artificial Intelligence Act)*, DO L, 2024/1689, 12 de julio de 2024, art. 14.

agéntica, el control efectivo es trazabilidad. Si el sistema actúa, debe quedar constancia de qué herramientas usó, con qué datos, bajo qué instrucciones, qué criterios activaron cada acción y qué incertidumbres estaban presentes. Sin trazabilidad, no hay aprendizaje (porque no se puede diagnosticar) y no hay responsabilidad (porque no se puede atribuir).

Una forma de operacionalizar esta idea es distinguir entre trazabilidad de evidencia (fuentes y datos), de decisión (prompts, reglas y umbrales) y de ejecución (acciones y llamadas a herramientas). Estas trazas conforman el expediente de una decisión híbrida y permiten auditoría, revisión y mejora; además, habilitan una supervisión selectiva, basada en señales, en lugar de una revisión exhaustiva e inviable.

Proponemos entender la gobernanza como un bloque de capacidades distribuida en cuatro niveles.

a) Nivel individual: alfabetización agéntica. Comprender qué es un agente, qué puede y qué no puede hacer, y cómo se degrada bajo incertidumbre, ambigüedad o inputs adversarios. Saber formular objetivos y exigir evidencia.

b) Nivel de equipo: protocolos compartidos. Plantillas de instrucciones, criterios de calidad, listas de verificación de seguridad, repertorio de evals, roles de verificación y revisión cruzada.

c) Nivel organizativo: infraestructura y políticas. Registro central de prompts y modelos, control de accesos, auditoría, gestión de incidentes, y un catálogo de herramientas con documentación diseñada para agentes. Aquí se decide qué tareas se automatizan, con qué grados de autonomía y qué indicadores de riesgo activan supervisión reforzada.

d) Nivel ecosistémico: interoperabilidad y estándares. Modelos de documentación (por ejemplo, para describir herramientas), prácticas de evaluación comparables y marcos regulatorios que armonicen expectativas.

Esta pila permite reencuadrar el papel de la regulación. El AI Act no define "competencias", pero al exigir supervisión humana en sistemas de alto riesgo, obliga a crear capacidades de intervención real: asignación de responsabilidad, formación adecuada, interfaz que permita interpretar el funcionamiento y mecanismos de parada. En otras palabras, obliga a que la supervisión sea un diseño, no una intención. En paralelo, marcos como el NIST AI RMF y su perfil para IA generativa ofrecen un vocabulario operativo para gestionar riesgos (gobernar, mapear, medir, gestionar) y pueden integrarse como prácticas organizativas[14].

El punto crucial es que gobernar no es frenar; es hacer que el sistema sea mejorable. En IA agéntica, los fallos no se eliminan por decreto, sino por iteración: se detectan, se clasifican, se convierten en evals, se ajustan instrucciones o herramientas, y se re-prueba. La gobernanza es el ciclo de aprendizaje institucional.

6. Formación y diseño organizativo para equipos con agentes

Si la IA agéntica se integra como "un empleado digital", la formación no puede limitarse a "cómo escribir prompts". Requiere un currículum de capacidades distribuido entre técnica, cognición y gobernanza. Proponemos cinco líneas de formación aplicables a todo tipo de organizaciones públicas o privadas.

1) Pensamiento de sistemas y descomposición de tareas. Aprender a convertir problemas difusos en flujos verificables, identificar dependencias, datos críticos y puntos de fallo. Esto incluye modelar el proceso antes de automatizarlo.

2) Verificación y evaluación. Entrenar en lectura crítica, contraste de fuentes, construcción de pruebas y uso de evals. La habilidad central es convertir un criterio de calidad en un test repetible. En redacción, listas de verificación; en datos, reproducibilidad; en software, tests; en decisiones, auditoría de supuestos.

3) Orquestación y herramientas. Conocer patrones de agentes (chaining, routing, orquestador-trabajadores) y aprender a diseñar herramientas ergonómicas para agentes, con entradas claras y salidas de alto valor informativo . Esto requiere pensar como diseñador de interfaces para un "usuario" no determinista.

[14] National Institute of Standards and Technology (NIST), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1) (Gaithersburg, MD: NIST, 2023), https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

4) Seguridad y privacidad operativa. Reconocer riesgos de inyección, filtraciones, uso indebido de datos y sobre-autonomía; aprender a usar sandboxes y entornos de prueba antes de producción; y entender que la privacidad es parte de la calidad del sistema.

5) Metacognición y aprendizaje deliberado. Diseñar rutinas personales y de equipo que mantengan habilidades internas. Por ejemplo, alternar ciclos de trabajo asistido y no asistido, y mantener "cuadernos de decisiones" donde se registren supuestos, evidencias y razonamientos.

Estas líneas pueden traducirse a programas formativos orientados a práctica, con casos de fallo, conjuntos de prueba y revisión de decisiones.

En paralelo, el diseño organizativo debe adaptarse. Equipos con agentes funcionan mejor cuando se redefinen responsabilidades: quién autoriza acciones, quién mantiene las evals, quién gestiona incidentes, quién decide cuándo actualizar modelos. También necesitan interfaces de confianza: visualizaciones de fuentes, trazas de herramientas y explicaciones estructuradas que permitan inspección. La transparencia no es un valor abstracto; es una condición para que el juicio humano sea posible.

Proponemos tres recomendaciones de diseño:

- Diseñar para la inspección: todo flujo agéntico debe producir un "paquete de evidencia" que permita revisión rápida (fuentes, decisiones, ejecución). Si no se puede inspeccionar, no se debe delegar con autonomía alta.
- Diseñar para el aprendizaje: cada fallo relevante debe convertirse en un caso de prueba. Esto crea una memoria organizativa y evita que el sistema repita errores con aparente seguridad.
- Diseñar para la reversibilidad: las acciones irreversibles (enviar, borrar, aprobar) deben requerir confirmación humana o mecanismos de doble control. La autonomía debe crecer solo donde la reversibilidad y la verificabilidad lo permitan.

7. Aplanamiento organizativo y equipos híbridos en la era de la IA agéntica

La irrupción de sistemas de IA cada vez más capaces y, en particular, de la IA agéntica que no solo "asiste" sino que planifica, encadena subtareas y ejecuta acciones, reconfigura la morfología de la organización. El efecto más visible es un aplanamiento estructural: disminuye la necesidad de capas intermedias dedicadas a coordinar, consolidar información, traducir decisiones en tareas y verificar entregables.

En paralelo, emerge la forma de trabajo dominante de la co-inteligencia: el equipo híbrido, entendido no como un grupo humano que usa herramientas, sino como una unidad operativa compuesta por personas y agentes con funciones diferenciadas, integradas por reglas de delegación, control y trazabilidad.

7.1. Por qué las organizaciones se aplanan: de la escasez de ejecución a la escasez de criterio

En los modelos organizativos tradicionales, buena parte de la jerarquía se justificaba por la escasez de capacidad de producción y por los costes de coordinación. El "mando" era, a la vez, mecanismo de asignación de tareas, control de calidad y canal de información. Cuando la IA agéntica abarata drásticamente la elaboración de borradores, el análisis preliminar, la síntesis documental, la generación de alternativas y, en algunos casos, la ejecución operativa, la organización experimenta un desplazamiento del cuello de botella: hacer deja de ser lo crítico; lo crítico pasa a ser decidir bien.

Este cambio tiene dos consecuencias. Primero, las capas intermedias centradas en supervisión micro-operativa pierden centralidad, porque la organización puede instrumentar controles automáticos (reglas, pruebas, trazas, auditorías) y verificación por muestreo en puntos de control. Segundo, la dirección se transforma: de una lógica de vigilancia del trabajo humano a una lógica de diseño del sistema socio-técnico (estándares, métricas, umbrales de riesgo, protocolos de escalado). En términos de gobernanza, el aplanamiento no implica ausencia de control, sino control reubicado: menos control "por presencia" y más control "por arquitectura".

7.2. Del organigrama al flujo decisional: autoridad distribuida y puntos de escalado

El aplanamiento organizativo no opera como una simple reducción de mandos, sino como una reingeniería del trabajo en torno a flujos decisionales. La autoridad se desplaza hacia los bordes de la organización en aquellas tareas que cumplen dos condiciones: (i) son repetitivas o suficientemente estandarizables y (ii) su perfil de riesgo permite delegación bajo reglas. En ese terreno, los equipos cercanos al problema pueden operar con mayor autonomía porque disponen de capacidades algorítmicas que antes estaban fragmentadas en funciones de soporte (análisis, documentación, control de calidad, reporting).

Sin embargo, el aplanamiento solo es sostenible si se institucionalizan mecanismos de escalado: reglas que determinan cuándo una decisión debe elevarse a revisión humana de mayor nivel o a un comité de riesgo. No se trata de crear burocracia, sino de fijar condiciones objetivas: impacto económico, afectación reputacional, incertidumbre, datos sensibles, cambios de política, desviación respecto de patrones previos. La organización co-inteligente se reconoce por este diseño: autonomía amplia en lo delegable y deliberación reforzada en lo crítico.

7.3. Equipos híbridos: de "herramientas" a "colegas operativos" con límites

El equipo híbrido es la respuesta funcional a la IA agéntica. En lugar de un equipo humano que "consulta" una IA, se consolida una configuración en la que distintos agentes desempeñan roles operativos relativamente estables (búsqueda y síntesis, generación de opciones, documentación, pruebas, monitorización), mientras que los humanos concentran el criterio, la priorización y la responsabilidad.

La hibridación introduce roles que, aunque pueden recaer en personas existentes, se vuelven explícitos:

- Propietario de decisión: asume la responsabilidad final y define los criterios de aceptabilidad (calidad, riesgos, trade-offs).
- Orquestador: descompone objetivos en subtareas, asigna agentes y establece límites; gestiona dependencias y "handoffs".
- Verificador: contrasta fuentes, valida cifras, detecta incoherencias y alucinaciones, revisa supuestos y trazabilidad.
- Arquitecto de calidad: convierte estándares en pruebas (rúbricas, checklists, tests de regresión, "definition of done"), y monitoriza deriva.

Estos roles no son decorativos: permiten que el aplanamiento no derive en descontrol. La organización co-inteligente no elimina la supervisión; la convierte en supervisión distribuida y formalizada mediante reglas y métricas.

7.4. Métricas y disciplina operativa: fiabilidad, trazabilidad y tiempo de verificación

En equipos híbridos, medir solo "output" es insuficiente y, de hecho, peligroso: incentiva velocidad sin garantías. Por ello, las organizaciones co-inteligentes introducen métricas de calidad estructural: tasa de errores detectados en verificación, densidad y calidad de evidencia, trazabilidad de decisiones (qué información se usó, qué agente intervino, qué controles se aplicaron), y tiempo de verificación relativo al tiempo de generación. La eficiencia relevante no es producir más, sino producir mejor con menos fricción de verificación.

Este giro métrico opera como condición de posibilidad del aplanamiento. Si la calidad está instrumentada, no hace falta interponer capas jerárquicas para garantizarla; basta con diseñar el sistema para que la calidad sea un "resultado por defecto".

7.5. Una tesis organizativa: menos jerarquía de vigilancia, más jerarquía de sentido

En síntesis, la organización co-inteligente se aplana porque la IA agéntica reduce costes de coordinación y multiplica capacidad de ejecución, pero se mantiene, e incluso se refuerza, una jerarquía distinta: la jerarquía del sentido y el criterio. La dirección ya no se legitima por controlar tareas, sino por fijar prioridades, diseñar umbrales, construir estándares y garantizar que la delegación algorítmica no erosione la calidad epistémica del trabajo. Los equipos híbridos, por

su parte, no son un "añadido tecnológico", sino la nueva unidad de producción: una cooperación sistemática entre capacidades humanas (juicio, responsabilidad, valores) y capacidades algorítmicas (escala, velocidad, exploración de alternativas), gobernada por reglas de delegación, verificación y aprendizaje institucional.

8. Conclusiones

La IA agéntica representa un cambio de fase en la integración de IA en el trabajo del conocimiento. El foco deja de ser la automatización de tareas aisladas y pasa a ser la delegación de flujos completos con capacidad de acción. Esto amplifica el potencial de productividad y nivelación de habilidades, pero también los riesgos de errores encadenados, inyección de instrucciones y sobre-autonomía. Para navegar esta transición, hemos propuesto tratar la co-inteligencia como un problema de capacidades: qué debe saber hacer una persona, qué debe institucionalizar un equipo y qué debe soportar una organización para que la delegación sea segura y útil. El marco de formulación, orquestación, verificación, calibración, resiliencia cognitiva, seguridad y gobernanza, ofrece un lenguaje operativo para diseñar prácticas y políticas. En este enfoque, la regulación europea sobre supervisión humana funciona como un marco habilitador: obliga a construir capacidades de intervención real y a sostener la trazabilidad que hace posible la rendición de cuentas.

La conclusión es que, cuanto más inteligente y autónoma sea la IA, más valiosas se vuelven las capacidades humanas de criterio, verificación y diseño de objetivos. No porque la máquina "no pueda" producir razonamientos, sino porque el valor social y organizativo del razonamiento depende de la responsabilidad, del aprendizaje y de la capacidad de dar razones. En la era de la IA agéntica, la agencia humana se preserva menos por proclamación y más por diseño.

Declaración sobre el uso de IA

Este manuscrito ha sido revisado y reescrito con apoyo de un sistema de IA generativa utilizado como herramienta de edición, reorganización y mejora estilística. La responsabilidad final del contenido, la selección de fuentes y la coherencia argumental corresponde a la autora.

Bibliografía

Anthropic. "Building Effective Agents." 19 de diciembre de 2024. https://www.anthropic.com/research/building-effective-agents.

"Engineering at Anthropic: Writing Tools for Agents." 18 de diciembre de 2024. https://www.anthropic.com/engineering/writing-tools-for-agents.

Brynjolfsson, Erik, Danielle Li y Lindsey Raymond. "Generative AI at Work." *The Quarterly Journal of Economics* 140, n.º 2 (2025): 889–942. https://doi.org/10.1093/qje/qjae044.

Dellermann, Dominik, Philipp Ebel, Matthias Söllner y Jan Marco Leimeister. "Hybrid Intelligence." *Business & Information Systems Engineering* 61, n.º 5 (2019): 637–643. https://doi.org/10.1007/s12599-019-00595-2.

National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology, 2023. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. Gaithersburg, MD: National Institute of Standards and Technology, 2024. https://doi.org/10.6028/NIST.AI.600-1.

OpenAI. "A Practical Guide to Building Agents." Documento PDF. Consultado el 31 de diciembre de 2025. https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf.

"New Tools for Building Agents." 11 de marzo de 2025. https://openai.com/index/new-tools-for-building-agents/.

OWASP Foundation. "OWASP Top 10 for Large Language Model Applications (Version 1.1)." Consultado el 31 de diciembre de 2025. https://owasp.org/www-project-top-10-for-large-language-model-applications/.

Risko, Evan F., y Sam J. Gilbert. "Cognitive Offloading." *Trends in Cognitive Sciences* 20, n.º 9 (2016): 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Sparrow, Betsy, Jenny Liu y Daniel M. Wegner. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333, n.º 6043 (2011): 776–778. https://doi.org/10.1126/science.1207745.

Sperber, Dan, Hugo Mercier, Christophe Origgi y otros. "Epistemic Vigilance." *Mind & Language* 25, n.º 4 (2010): 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x.

Unión Europea. *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Artificial Intelligence Act)*. DO L, 2024/1689, 12 de julio de 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Neurotecnologías, neurodatos y el futuro de la humanidad

Neurotechnologies, Neurodata, and the Future of Humanity

Nelson Remolina Angarita
Profesor de Derecho
Universidad de los Andes (Colombia)
Director del Grupo de Estudios en Internet, Comercio Electrónico, Telecomunicaciones e Informática (GECTI)

Resumen:
El texto aporta insumos para llamar la atención sobre los eventuales riesgos negativos que genera el uso indebido de las neurotecnologías y los neurodatos. A partir de documentos internacionales, pero especialmente los informes de la relatoría de privacidad de la ONU de 2025, busca que el lector reflexione y actúe con el objetivo de definir el tipo de sociedad que queremos.

Palabras clave:
Neurotecnologías; neurodatos; dignidad humana; cerebro; derechos humanos.

Abstract:
The text provides insights to raise awareness about the potential negative risks arising from the misuse of neurotechnologies and neural data. Based on international documents, particularly the reports from the UN Special Rapporteur on Privacy in 2025, it encourages the reader to reflect and take action with the aim of defining the kind of society we want.

Keywords:
Neurotechnologies; neural data; human dignity; brain; human rights.

1. Reflexión inicial

Mediante la resolución 3384 de 1975[1] de la Organización de las Naciones Unidas (en adelante, ONU) se ha puesto de presente que "el progreso científico y tecnológico se ha convertido en uno de los factores más importantes del desarrollo de la sociedad humana" porque "crea posibilidades cada vez mayores de mejorar las condiciones de vida de los pueblos y las naciones". Pero, al mismo tiempo, "puede en ciertos casos dar lugar a problemas sociales, así como amenazar los derechos

[1] Proclamada por la Asamblea General en su resolución 3384 (XXX), de 10 de noviembre de 1975 sobre la utilización del progreso científico y tecnológico en interés de la paz y en beneficio de la humanidad. Organización de las Naciones Unidas, *Declaración sobre la Utilización del Progreso Científico y Tecnológico en Interés de la Paz y en Beneficio de la Humanidad*, Resolución 3384 (XXX), 10 de noviembre de 1975. https://www.ohchr.org/es/instruments-mechanisms/instruments/declaration-use-scientific-and-technological-progress-interests

humanos y las libertades fundamentales del individuo". Concretamente, señala dicha resolución que "los logros científicos y tecnológicos pueden entrañar peligro para los derechos civiles y políticos de la persona o del grupo y para la dignidad humana".

En un entorno donde la tecnología avanza a un ritmo vertiginoso, el uso de las tecnologías en pro de la humanidad y la protección de los derechos humanos son dos pilares esenciales para fijar las bases de una sociedad humana, confiable, incluyente y centrada en el ser humano. Las innovaciones como las neurotecnologías generan muchos beneficios, pero, al mismo tiempo, plantean desafíos sin precedentes que exigen una respuesta clara para preservar la dignidad humana y aspectos centrales para la humanidad.

Las neurotecnologías generan grandes expectativas para dar mejores soluciones a los problemas de salud mental. Son una esperanza para más de un tercio de la población mundial que, según los expertos, padecen o padecerán afectaciones a su salud mental. Pero, al mismo tiempo, su uso puede causar efectos no deseados (al menos para el autor de estas notas) como, por ejemplo, programar seres humanos para que piensen o actúen de determinada manera (seres humanos como títeres programados por otros seres humanos). Es decir, a la par que se programan robots, se programarán personas naturales.

En otras palabras, es posible que tengamos seres humanos con el cerebro artificialmente programado para fines diferentes a la salud mental. A ellos se suman los seres humanos con el cerebro natural no manipulado con neurotecnologías. Habrá seres humanos de primera y de segunda desde la perspectiva de la capacidad de su cerebro.

Creo que, en últimas, estamos frente a la creación de una nueva generación de seres humanos diferente a la que hemos conocido por muchos siglos. Se trata de los seres humanos cerebralmente modificados mediante neurotecnologías. No serán personas cuyos órganos son fruto del proceso natural (por obra y gracia de la naturaleza), sino sujetos con el cerebro modificado o potencializado cerebralmente de manera artificial para fines diferentes a los estrictamente necesarios para dar respuesta a desafíos de salud mental.

Y sobre esto tiene mucho que decir y hacer el regulador, el derecho, los jueces, la academia y la sociedad en general. De esto dependerá el presente y futuro de la humanidad. Por eso, la Global Privacy Assembly -GPA-, entre otros, recomienda a los legisladores y a los responsables de las políticas públicas que, entre otras, establezcan reglas claras que protejan la dignidad humana, la identidad de todos los seres humanos, así como el respeto de sus derechos y libertades fundamentales frente al tratamiento de los neurodatos y el uso de las neurotecnologías.[2]

Vale la pena reflexionar sobre lo siguiente que planteamos en 2024[3]: ¿Qué tipo de sociedad queremos? ¿Todo lo tecnológicamente posible es socialmente deseable? ¿Los creadores de tecnología seguirán siendo quienes definan el alcance de los derechos humanos y el destino de la humanidad? ¿Es correcto que se manipule artificialmente el cerebro para hacer que el ser humano se comporte como una marioneta? ¿Hasta qué punto es ético y humano cambiar la información mental de los seres humanos? ¿Es ético desarrollar seres humanos aumentados cognitivamente? En caso positivo, ¿a quiénes sí y a quiénes no? ¿Es ético implantar en el cerebro sesgos mediante herramientas tecnológicas como algoritmos de inteligencia artificial que utilizan en neurotecnologías?[4]

Queda mucho por responder, pero, sobre todo, mucho por hacer para evitar que las neurotecnologías y la neurodata se utilice en detrimento del ser humano, sus derechos, la dignidad humana, la sociedad y la humanidad.

Reitero lo manifestado en 2015 en el sentido de que "no es sensato plantear las cosas en términos de «tecnofobia» o «tecnofascinación». Siempre será bienvenida la innovación y el desarollo fundado en el respeto de los derechos humanos y la dignidad humana. La «tecnofascinación» ciega y acrítica conlleva muchos riesgos. No debemos ser sujetos sumisos y conformistas con lo que algunos quieren hacer con nuestros derechos y nuestras vidas. Si dejamos que nos lleve la corriente es

[2] Global Privacy Assembly. *Resolution on principles regarding the processing of personal information in neuroscience and neurotechnology*. 46th Closed Session of the Global Privacy Assembly. November 2024: 10. En: https://globalprivacyassembly.org/wp-content/uploads/2024/11/Resolution-on-Neurotechnologies.pdf

[3] Nelson Remolina Angarita, "Neuro reflexión: hacia una declaración universal sobre las neuro tecnologías y los derechos humanos", en *En defensa de los neuroderechos*, ed. por Moisés Sánchez, Ciro Colombara y Natalia Monti, (Chile: Fundación Kamanau, 2024), 227.

[4] Todas las preguntas sobre neurotecnologías y neuroderechos fueron tomadas o adaptadas a partir de la conferencia del 20 de octubre de 2021 impartida por Rafael Yuste: "Las neurotecnologías y sus consecuencias éticas y sociales". Disponible en: https://www.youtube.com/live/mqfghQJAB2w?si=nWVHuzDfo37DM5EB

factible que nuestros derechos cada vez sean menos —o desaparezcan— y que nos toque pagar por ellos. Es necesario despertar y reaccionar para reivindicar el respeto de nuestros derechos"[5].

Con lo anterior en mente, este texto deja en manos del lector información y reflexiones para que construya su criterio sobre lo que planteamos anteriormente: ¿qué tipo de sociedad queremos? Y ¿todo lo tecnológicamente posible con las neurotecnologías es socialmente deseable?

No se suministran las respuestas a los problemas, sino con algunos insumos de reflexión para responderlas.

2. De la importancia del cerebro, las neurotecnologías y los neurodatos

El cerebro es maravilloso y determinante de, entre otras, la conducta de los seres humanos, su personalidad, esencia e impronta que los hace únicos y diferentes de los demás. En este sentido, señalan Manes y Niro que "si nos hicieran un trasplante de riñón o de pulmón, seguiríamos siendo nosotros mismos. Pero si nos cambiaran el cerebro, nos convertiríamos en personas distintas"[6]. Dichos autores destacan la grandeza de nuestro cerebro en los siguientes términos: "el cerebro humano es la estructura más compleja en el universo. (…) El cerebro dicta toda nuestra actividad mental -desde procesos inconscientes, como respirar, hasta los pensamientos filosóficos más elaborados- y contiene más neuronas que las estrellas existentes en la galaxia"[7].

Según la *Neurorights foundation* el "cerebro humano es diferente a cualquier otro órgano, ya que genera todas nuestras actividades mentales y cognitivas. Los datos que produce son diferentes a los de cualquier otro tipo, ya que reflejan el procesamiento mental. Los datos neuronales, que se refieren a la información que refleja directamente la actividad del sistema nervioso central o periférico de un individuo, son, por lo tanto, capaces de revelar información sumamente sensible sobre las personas de quienes se recopilaron, incluyendo información identificable sobre su salud mental, salud física y procesamiento cognitivo"[8].

Las neurotecnologías, por su parte, son "métodos, herramientas o dispositivos para registrar la actividad cerebral o para cambiarla"[9]. Según Rafael Yuste, miembro del *Neuro Technology Center* (NTC) de la Universidad de Columbia, "la neurotecnología es importante porque el cerebro no es un órgano más del cuerpo, sino el órgano que genera toda la actividad mental y cognitiva de los seres humanos. Nuestros pensamientos, nuestras percepciones, nuestras emociones, nuestras memorias, incluso el subconsciente… todo surge de la actividad coordinada de circuitos neuronales dentro de nuestro cerebro. Y con la neurotecnología, por primera vez podemos adentrarnos en estos circuitos neuronales, registrar su actividad y cambiarla"[10]. Estas presentan importantes beneficios, como mejorar la comprensión del cerebro y desarrollar tratamientos para enfermedades neurológicas. El citado científico pone de presente los beneficios del uso de neurotecnologías como, entre otros, los siguientes: (i) realizar "investigaciones para descubrir cómo funciona el cerebro y cuál es la base científica de la mente humana"[11]; (2) "diagnosticar, entender, y diseñar nuevas terapias para las enfermedades cerebrales tanto neurológicas, neurodegenerativas o psiquiátricas. Enfermedades como el Alzheimer, la esquizofrenia, el Parkinson, la epilepsia, la discapacidad mental, el ictus, la esclerosis lateral, la depresión, la ansiedad, etc… Estas enfermedades cerebrales afectan de una

[5] Nelson Remolina Angarita, *Recolección Internacional de Datos Personales. Un reto del mundo post-internet* (BOE: Madrid, 2015), 375. Texto citado en las notas de pie de página 5 y 6 del siguiente informe de la ONU: A/80/283 del 31 de julio de 2025: "Elementos para crear una ley modelo sobre neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad". Puede consultarse en: https://www.ohchr.org/es/documents/thematic-reports/a80283-report-special-rapporteur-right-privacy-ana-brian-nougreres

[6] Facundo Manes y Mateo Niro, *Usar el cerebro. Conocer nuestra mente para vivir mejor*, 4ª ed., (Planeta: Buenos Aires, 2014),8.

[7] Ibídem.

[8] Jared Genser, Stephen Damianos, y Rafael Yuste, *Safeguarding Brain Data: Assessing the Privacy Practices of Consumer Neurotechnology Companies* (Neurorights foundation, 2024) 3. Disponible en: https://perseus-strategies.com/wp-content/uploads/2024/04/FINAL_Consumer_Neurotechnology_Report_Neurorights_Foundation_April-1.pdf

[9] Rafael Yuste, "Un paso histórico", en *En defensa de los neuroderechos*, ed. por Moisés Sánchez, Ciro Colombara y Natalia Monti, (Chile: Fundación Kamanau, 2024), 7.

[10] Ibídem.

[11] Ibídem.

manera cada vez mayor a un gran porcentaje de los ciudadanos y son la lacra de la humanidad"[12]; y (3) Fomentar "la creación de dispositivos de interfaz cerebro computadora, que permitan la conexión directa con el internet, y forme la base de una industria nueva, con grandes beneficios económicos y también a los consumidores"[13].

Los neurodatos, por su parte, hacen referencia a la información que se obtiene del sistema nervioso central y periférico de una persona mediante el uso de neurotecnologías. Según la Red Iberoamericana de Protección de Datos (en adelante, RIPD) "Los datos cerebrales o neurodatos" contienen del sistema nervioso y del cerebro que es única y personal. Los neurodatos pueden permitir "conocer los procesos cerebrales en "tiempo real", lo que permite el registro directo de procesos asociados con la personalidad, el estado de ánimo, los comportamientos, los pensamientos o los sentimientos"[14].

En 2024, la RIPD recalcó que el cerebro será un identificador tan único como la huella dactilar o el genoma" y que "los avances técnicos y científicos no se encuentran libres de errores, tendencias, sesgos, interpretaciones políticas o religiosas o prejuicios, por lo que puede llevar a situaciones de neurodiscriminación." Por ende, "todo tratamiento que incluya neurodatos se considerará un tratamiento de alto riesgo de datos personales"[15].

Estima la *Neurorights foundation* que "en los próximos años, la sensibilidad de los datos neuronales se intensificará a medida que aumenten las inversiones del sector privado, los gobiernos e iniciativas similares. Esto se traducirá en mejoras en las capacidades técnicas de la neurotecnología, lo que permitirá una mayor resolución de los escáneres cerebrales y la recopilación de conjuntos de datos cerebrales más amplios.

Por otro lado, la inteligencia artificial generativa acelerará la capacidad de decodificar con precisión estos escáneres. Mientras tanto, las neurotecnologías implantables ya pueden decodificar con precisión el lenguaje y las emociones, y los dispositivos portátiles también están comenzando a incorporar algunas de estas capacidades. Estos avances tienen implicaciones significativas para la privacidad mental, lo que pone de relieve la importancia crucial de comprender las prácticas de privacidad y la protección del usuario que ofrecen las empresas de neurotecnología de consumo"[16].

3. Desafíos derivados de las neurotecnologías respecto de los derechos humanos

El uso indebido de las neurotecnologías también generan riesgos significativos como, entre otros, los siguientes que la ONU, citando a Yuste, menciona en el informe A/HRC/58/58 del 16 de enero de 2025[17]:

- Usar las neurotecnologías para fines contrarios a la dignidad humana. Con éstas se puede descodificar y alterar la actividad cerebral, lo cual genera problemas/retos éticos, jurídicos y sociales muy profundos ya que se podría cambiar la esencia del ser humano y manipularlo / alterarlo.

- Modificar artificialmente los seres humanos. Los hallazgos científicos en neurociencias y su aplicación a través de diversas neurotecnologías tienen el potencial de alterar algunas características humanas fundamentales, como la autonomía, la responsabilidad moral, el libre albedrío, la dignidad, la identidad, la vida mental privada, la comprensión de los individuos como

[12] Ibídem 7-8.

[13] Ibídem 8.

[14] RIPD, "Declaración sobre neurodatos", aprobada en sesión cerrada del encuentro de la Red Iberoamericana de Protección de Datos, en la Antigua, Guatemala, el 25 de septiembre de 2023.

[15] RIPD, "Declaración sobre neurotecnologías y neurodatos en el marco de la normativa de protección de datos", aprobada en sesión cerrada del encuentro de la Red Iberoamericana de Protección de Datos, en Cartagena, Colombia, el 29 de mayo de 2024: 2-3.

[16] Genser, Damianos y Yuste, *Safeguarding Brain Data*.

[17] Cfr. ONU. Documento A/HRC/58/58: Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad - Informe de la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères. 16 de enero de 2025. Disponible en: https://www.ohchr.org/es/documents/thematic-reports/ahrc5858-foundations-and-principles-regulation-neurotechnologies-and

entidades atadas por sus cuerpos, la integridad y la seguridad corporal.[18]

- Generar daños físicos al ser humano y la manipulación mental. También pueden producir daños físicos asociados con los procedimientos invasivos de colocación de los dispositivos para mejoramiento o para la interfaz cerebro-máquina. daño tisular y deterioro de la función motora (vulneración al derecho a la integridad mental).[19]

- Indebido tratamiento de los neurodatos y uso de los mismos para fines contrarios a la dignidad humana o no autorizados por la ley. El "secuestro cerebral" puede implicar el robo de información (violación del derecho a la privacidad mental). También existe la posibilidad de ingreso de virus, o que los dispositivos neuronales conectados a internet posibiliten que individuos u organizaciones (hackers, corporaciones o agencias gubernamentales) rastreen o, incluso, manipulan la experiencia mental de un individuo[20].

Resalta la ONU que "pese a los beneficios en la salud mental que traerán las neurotecnologías, existe el temor que con esa tecnología se pueda, no sólo conocer lo que piensan las personas (que por ahora es un secreto), sino manipular cerebralmente seres humanos. Por eso, desde hace poco se vienen gestando los neuroderechos que tienen como finalidad lo siguiente:

- No perder la privacidad que tenemos respecto de nuestro cerebro (lo que pensamos)

- Derecho a ser como soy: derecho al yo, a mi identidad cerebral natural.

- Derecho a decidir por mí mismo, sin ser artificialmente manipulado o programado

- Neurotecnologías neutrales. No sesgadas. Que no se implanten sesgos en nuestro cerebro.

- Acceso equitativo a las neurotecnologías"[21]

4. Del caso chileno

Por ser referentes en Latinoamérica y en el mundo, nos referiremos a la reforma constitucional chilena sobre desarrollo científico y tecnológico y al caso Girardi vs Emotiv Inc. De 2023

4.1. Reforma constitucional que otorga relevancia a la información de la actividad cerebral

El día 14 de octubre del año 2021 se promulgó en Chile la Ley N° 21.383[22] que modifica Constitución Política en los siguientes términos:

> "Artículo único.- Modifícase el número 1° del artículo 19 de la Constitución Política de la República, de la siguiente forma:
> (...)
>
> 2) Agrégase el siguiente párrafo final, nuevo:
> "El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella;".

[18] Rafael Yuste et al., "Four Ethical Priorities for Neurotechnologies and AI," *Nature* 551, no. 7679 (2017): 159–163.

[19] Ibídem.

[20] Ibídem.

[21] Cfr. ONU. Documento A/HRC/58/58: Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad - Informe de la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères. 16 de enero de 2025. Disponible en: https://www.ohchr.org/es/documents/thematic-reports/ahrc5858-foundations-and-principles-regulation-neurotechnologies-and

[22] Biblioteca del Congreso Nacional de Chile, *Ley N.º 21.383*, https://www.bcn.cl/leychile/navegar?idNorma=1166983

La primera parte del artículo sigue parte de los mensajes del texto de la resolución 3384 de 1975 de la ONU sobre la utilización del progreso científico y tecnológico en interés de la paz y en beneficio de la humanidad, en donde, entre otras, se acordó que:

> 1. "Todos los Estados promoverán la cooperación internacional con objeto de garantizar que los resultados del progreso científico y tecnológico se usen en pro del fortalecimiento de la paz y la seguridad internacionales, la libertad y la independencia, así como para lograr el desarrollo económico y social de los pueblos y hacer efectivos los derechos y libertades humanos de conformidad con la Carta de las Naciones Unidas.
>
> 2. Todos los Estados tomarán medidas apropiadas a fin de impedir que los progresos científicos y tecnológicos sean utilizados, particularmente por órganos estatales, para limitar o dificultar el goce de los derechos humanos y las libertades fundamentales de la persona consagrados en la Declaración Universal de Derechos Humanos, en los Pactos Internacionales de derechos humanos y en otros instrumentos internacionales pertinentes.
>
> 3. Todos los Estados adoptarán medidas con objeto de garantizar que los logros de la ciencia y la tecnología sirvan para satisfacer las necesidades materiales y espirituales de todos los sectores de la población."[23] (Destaco)

La reforma constitucional propende porque las tecnologías y la ciencia estén al servicio del ser humano con especial referencia a la protección de la vida y la integridad física o psíquica. Esto es supremamente relevante porque la Constitución señala expresamente para que queremos el desarrollo tecnológico y, por ende, fija un rumbo a seguir por parte de los científicos y los creadores de tecnologías.

La segunda parte hace referencia a dos cosas:

- La necesidad de que por ley se establezcan pautas para el uso de la ciencia y la tecnología en los seres humanos. Es necesario que en la ley se indiquen los requisitos, condiciones y restricciones para dicho efecto. El uso de la tecnología en los seres humanos no es libre ni arbitrario; no puede depender únicamente de la voluntad de los científicos o los creadores de dichas tecnologías. Por eso, es fundamental que la ley establezca de manera clara y precisa las pautas, requisitos, condiciones y restricciones para su aplicación, garantizando un uso responsable, ético y en beneficio de las personas, evitando decisiones unilaterales que puedan poner en riesgo la integridad y derechos de los individuos.

- La importancia de proteger la actividad cerebral y la información actividad proveniente de ella, dentro de la cual se encuentran los neurodatos. Proteger lo anterior (la actividad cerebral y la información proveniente de ella) es fundamental porque estos datos representan aspectos íntimos y esenciales de la identidad, la autonomía y los derechos fundamentales de las personas. Su protección garantiza que no se vulneren la aspectos esenciales del ser humano cuyo uso abusivo o manipulación indebida pueden afectar el corazón de la dignidad humana y el futuro de la humanidad.

En síntesis, la reforma constitucional de 2021, marca un hito al crear pautas sobre el desarrollo científico y tecnológico en pro de la humanidad, alineándose con recomendaciones de la ONU de mediados de los años setenta. Esta reforma no solo traza un rumbo claro para científicos y tecnólogos, sino que también establece un precedente significativo para otros países, subrayando la responsabilidad ética y social del avance tecnológico. En otras palabras, dicha reforma constitucional crea un precedente global al regular el uso de la tecnología en la actividad cerebral y la información proveniente del cerebro, resaltando el alto nivel de diligencia y responsabilidad en el uso de las tecnologías. Su enfoque marca un camino para que otros países alineen el avance tecnológico con el bienestar humano.

4.2. La decisión judicial del caso Girardi vs Emotiv Inc.

El 9 de agosto de 2023 la Corte Suprema de Chile[24] emitió el fallo conocido Girardi vs Emotiv Inc. Aunque en éste no se hace un desarrollo profundo sobre las neurotecnologías y los neuroderechos,

[23] Proclamada por la Asamblea General en su resolución 3384 (XXX), de 10 de noviembre de 1975 sobre la utilización del progreso científico y tecnológico en interés de la paz y en beneficio de la humanidad. El texto oficial puede consultarse: https://www.ohchr.org/es/instruments-mechanisms/instruments/declaration-use-scientific-and-technological-progress-interests

[24] República de Chile, Corte Suprema, Tercera Sala, *Girardi Lavín v. Emotiv Inc.*, sentencia de 9 de agosto de 2023.

la decisión es relevante porque es pionera en dar comienzo al desarrollo de los neuroderechos ante los jueces. Con esa decisión se ha el primer paso hacia el derecho de los jueces, las neurotecnologías, los derechos humanos y la dignidad humana.

Su importancia es de gran calado no solo por ser la primera decisión sobre neurotecnologías sino porque pone de presente los retos a que se expone la humanidad frente al desarrollo tecnológico. La sentencia decide sobre la acción constitucional de protección de un ciudadano chileno frente a Emotiv Inc.[25], con ocasión de la venta y uso en Chile del dispositivo "Insight". Según el denunciante, no se protege adecuadamente la privacidad de la información cerebral de los usuarios del citado aparato, lo cual vulnera los derechos previstos en los numerales 1, 4, 6 y 24 del artículo 19 de la Constitución Política de la República de Chile. De conformidad con los hechos descritos en la sentencia, "El producto por el que se recurre, Insight, consiste en un dispositivo inalámbrico funciona como una vincha con sensores que recaban información sobre la actividad eléctrica del cerebro, obteniendo datos sobre gestos, movimientos, preferencias, tiempos de reacción y actividad cognitiva de quien lo usa."[26]

Manifiesta el demandante que "por el uso del dispositivo y el almacenamiento de su información cerebral por la empresa, se ha expuesto a riesgos que comprenden: (i) La reidentificación; (ii) La piratería o hackeo de datos cerebrales; (iii) Reutilización no autorizada de los datos cerebrales; (iv) Mercantilización de los datos cerebrales; (v) Vigilancia digital; (vi) Captación de datos cerebrales para fines no consentidos por el individuo, entre otros; además de vulnerarse lo dispuesto en el artículo 11 de la Ley N° 19.628, sobre la debida diligencia en el cuidado de datos personales a la que se encuentran obligados los responsables de registros o bases de datos personales, y lo señalado en el artículo 13 de la misma ley, sobre el derecho de las personas a la cancelación o bloqueo de sus datos personales, ya que, aun cuando la cuenta de usuario de Emotiv se encuentre cerrada, la empresa recurrida retiene información cerebral para propósitos de investigación científica e histórica."[27]

La empresa demandada, por su parte, señaló que "su producto Insight consiste en un dispositivo de neurotecnología no invasiva, sin fines terapéuticos de tipo electroencefalograma móvil, diseñado para la autocuantificación, investigación de campo, no vendiéndose como dispositivo médico"[28]. También alega que el demandante "omite señalar que el producto y su instalación contienen una detallada explicación de los términos y condiciones tanto del producto como del servicio contratado, donde se le solicita su consentimiento expreso para el tratamiento de sus datos personales y cerebrales, que fue otorgado por el actor."[29] Culmina manifestando que en el recurso no se "menciona la manera concreta en la que se estarían vulnerando las garantías constitucionales enunciadas por el recurrente, sino que se enumeran una serie de riesgos hipotéticos, en abstracto, a los que está sujeta cualquier plataforma tecnológica o de servicios que realice tratamientos de datos personales"[30]

Con ocasión de la decisión, Corte puso de presente, entre otras, lo siguiente:

Primero: Con la reforma constitucional introducida mediante la citada ley 21.383 de 2021 "se materializó la especial preocupación del constituyente en el tema la neurotecnología y los Derechos Humanos".[31]

Segundo: El Estado debe poner especial atención y cuidado frente al "desarrollo de nuevas tecnologías que involucran cada vez más aspectos de la persona humana, aspectos que era impensable hace algunos años que pudieran conocerse"[32]. Para la Corte, el Estado debe "prevenir y anticiparse a sus posibles efectos, además de proteger directamente la integridad humana en

[25] Según la sentencia y lo expuesto por el demandante, Emotiv Inc es una "empresa de bioinformática y tecnología que desarrolla y fabrica productos de electroencefalografía portátil, junto con neuroauriculares, kits de desarrollo de software, software, aplicaciones móviles y productos de datos, con sede en San Francisco, Estados Unidos."

[26] República de Chile. Corte Suprema. Tercera Sala. Acción constitucional de protección de Guido Girardi. Lavín, en contra de la empresa Emotiv Inc. Agosto 9 de 2023, pág 2.

[27] Ibid. Págs. 2-3.

[28] Ibid. Pág. 4.

[29] Ibid. Pág. 4.

[30] Ibid. Pág. 4.

[31] Ibid. Pág. 7.

[32] Ibid. Pág. 12.

su totalidad, cuestión que incluye su privacidad y confidencialidad y los derechos propios de la integridad psíquica y del sujeto de experimentación científica."[33]

Tercero: Es necesario que las autoridades competentes, previo a la comercialización de productos como el del presente caso, analicen los riesgos que pueden generar estas tecnologías para el ser humano y sus derechos. En el caso, el aspecto determinante de la decisión no fue un debate en centrado en las neurotecnologías, sino en la forma como las mismas se introdujeron a territorio chileno porque la Corte encontró que el dispositivo Insight no cumplía un requisito aduanero para comercializarse en Chile. Por esta razón, se acogió la acción constitucional y se ordenó al Instituto de Salud Pública y la autoridad aduanera chilena que evalúen los antecedentes y dispongan lo que en derecho corresponda "a efectos que la comercialización y uso del dispositivo Insight y el manejo de datos que de él se obtengan se ajuste estrictamente a la normativa aplicable en la especie y reseñada en esta sentencia"[34].

Cuarto: Concluye al Corte que se vulneraron las "garantías constitucionales contenidas en los numerales 1 y 4 del artículo 19 de la Constitución Política de la República, que se refieren a la integridad física y psíquica y de derecho a la privacidad, en los términos expuestos en el presente fallo en los considerandos precedentes, al comercializarse el producto Insight sin contar con todas las autorizaciones pertinentes, y no habiendo sido evaluado y estudiado por la autoridad sanitaria a la luz de lo expresado."[35] Por ende, la Corte ordenó a Emotiv Inc. eliminar "toda la información que se hubiera almacenado en su nube o portales, en relación con el uso del dispositivo por parte del recurrente".

El fallo judicial resucita la discusión sobre los desarrollos tecnológicos y su uso. En últimas, invita a repensar sobre el tipo de sociedad que queremos y a decidir si todo lo tecnológicamente posible es socialmente deseable.

Para dar respuesta a lo anterior, existen directrices de antaño como, entre otras, la precitada resolución de 1975[36] de la ONU sobre la utilización del progreso científico y tecnológico en interés de la paz y en beneficio de la humanidad. Mediante la misma se reconoce que "el progreso científico y tecnológico se ha convertido en uno de los factores más importantes del desarrollo de la sociedad humana" porque "crea posibilidades cada vez mayores de mejorar las condiciones de vida de los pueblos y las naciones". Pero, al mismo tiempo, "puede en ciertos casos dar lugar a problemas sociales, así como amenazar los derechos humanos y las libertades fundamentales del individuo". Concretamente, señala dicha resolución que "los logros científicos y tecnológicos pueden entrañar peligro para los derechos civiles y políticos de la persona o del grupo y para la dignidad humana". Por eso es inaplazable adoptar medidas para evitar las consecuencias negativas de algunos desarrollos tecnológicos frente a la sociedad en general, los derechos humanos y la dignidad humana. En línea con lo anterior, en la precitada resolución se acuerda, entre otros, lo que sigue a continuación:

"7. Todos los Estados adoptarán las medidas necesarias, incluso de orden legislativo a fin de asegurarse de que la utilización de los logros de la ciencia y la tecnología contribuya a la realización más plena posible de los derechos humanos y las libertades fundamentales sin discriminación alguna por motivos de raza, sexo, idioma o creencias religiosas. (Destaco)

"8. Todos los Estados adoptarán medidas eficaces, incluso de orden legislativo, para impedir y evitar que los logros científicos se utilicen en detrimento de los derechos humanos y las libertades fundamentales y la dignidad de la persona humana." (Destacado).

5. Neurotecnologías y genoma humano como insumo para actualizar la Declaración Universal de los Derechos Humanos

Vivimos en constante riesgo de que las tecnologías no se usen en pro del ser humano sino en contra del mismo. Esto no es nuevo, pero quizá lo que está sucediendo con reflexiones sobre las neurotecnologías y los neuroderechos va a ser similar con lo que aconteció con los desarrollos genéticos.

[33] Ibid. Pág. 12.

[34] Ibid. Pág. 14.

[35] Ibid. Pág 13.

[36] Proclamada por la Asamblea General en su resolución 3384 (XXX), de 10 de noviembre de 1975. El texto oficial puede consultarse en: https://www.ohchr.org/es/instruments-mechanisms/instruments/declaration-use-scientific-and-technological-progress-interests

Como es sabido, la ONU emitió la Declaración Universal sobre el genoma humano y los derechos humanos[37]. En ella, por ejemplo, se establece lo siguiente que, por su importancia, se trascribe:

- "Una investigación, un tratamiento o un diagnóstico en relación con el genoma de un individuo, sólo podrá efectuarse previa evaluación rigurosa de los riesgos y las ventajas que entraña y de conformidad con cualquier otra exigencia de la legislación nacional" (Literal a) del artículo 5)

- "Ninguna investigación relativa al genoma humano ni sus aplicaciones, en particular en las esferas de la biología, la genética y la medicina, podrán prevalecer sobre el respeto de los derechos humanos, de la libertades fundamentales y de la dignidad humana de los individuos o, si procede, de los grupos humanos" (Artículo 10)

- "No deben permitirse las prácticas que sean contrarias a la dignidad humana, como la clonación con fines de reproducción de seres humanos." (Artículo 11)

- "Toda persona debe tener acceso a los progresos de la biología, la genética y la medicina en materia de genoma humano, respetándose su dignidad y derechos" (Literal a) del artículo 12)

- "Las aplicaciones de la investigación sobre el genoma humano, en particular en el campo de la biología, la genética y la medicina, deben orientarse a aliviar el sufrimiento y mejorar la salud del individuo y de toda la humanidad" (Literal b) del artículo 12)

- "Las consecuencias éticas y sociales de las investigaciones sobre el genoma humano imponen a los investigadores responsabilidades especiales de rigor, prudencia, probidad intelectual e integridad, tanto en la realización de sus investigaciones como en la presentación y explotación de los resultados de éstas. Los responsables de la formulación de políticas científicas públicas y privadas tienen también responsabilidades especiales al respecto." (Artículo 13)

- "Los Estados tomarán las medidas apropiadas para fijar el marco del libre ejercicio de las actividades de investigación sobre el genoma humano respetando los principios establecidos en la presente Declaración, a fin de garantizar el respeto de los derechos humanos, las libertades fundamentales y la dignidad humana y proteger la salud pública. Velarán por los resultados de esas investigaciones no puedan utilizarse con fines no pacíficos." (Artículo 15)

- "Los Estados reconocerán el interés de promover, en los distintos niveles apropiados, la creación de comités de ética independientes, pluridisciplinarios y pluralistas, encargados de apreciar las cuestiones éticas, jurídicas y sociales planteadas por las investigaciones sobre el genoma humano y sus aplicaciones" (Artículo 16).

Existen otros temas muy importantes en la citada resolución, pero los mencionados son elementos relevantes para que no solo se actualice la Declaración Universal de los Derechos Humanos incluyendo los neuroderechos, sino para que la ONU expida urgentemente una Declaración Universal sobre las neurotecnologías y los derechos humanos.

6. Algunos insumos internacionales para la regulación y la toma de decisiones judiciales

La regulación sobre cualquier tecnología debe tener presente la labor de armonización internacional porque las tecnologías impactan globalmente y no tienen fronteras. Usualmente se diseñan y fabrican en unos países y se usan en otros. Un enfoque exclusivamente local o territorial sería insuficiente para abordar correctamente los desafíos de las neurotecnologías frente a la humanidad.

En línea con lo anterior, varias organizaciones internacionales han emitidio documentos relevantes sobre la materia, la "Declaración de principios interamericanos en materia de neurociencias, neurotecnologías y derechos humanos", aprobada por el Comité Jurídico Interamericano de la Organización de Estados Americanos (OEA)[38].

[37] El texto oficial puede consultarse en: https://www.ohchr.org/es/instruments-mechanisms/instruments/universal-declaration-human-genome-and-human-rights#:~:text=a)%20Cada%20individuo%20tiene%20derecho,car%C3%A1cter%20%C3%BAnico%20y%20su%20diversidad

[38] Cfr. Organización de Estados Americano (OEA). Comité Jurídico Interamericano. Declaración de principios interamericanos en materia de neurociencias, neurotecnologías y derechos humanos. 102 PERÍODO ORDINARIO DE SESIONES OEA/Ser. Q. 6 – 10 de marzo, 2023 CJI/RES. 281 (CII-O/23) corr.1. Rio de Janeiro, Brasil 9 marzo 2023.

Se suma a lo anterior, los lineamientos y directrices de la Organización de las Naciones Unidas (ONU)[39]; la Red Iberoamericana de protección de datos (RIPD)[40]; la Global Privacy Assembly[41]; la Organisation for Economic Cooperation and Development (OCDE)[42]; el Consejo de Europa[43]; la Organización de Estados Americanos (OEA)[44]; la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO)[45] y el Parlamento Latinoamericano y Caribeño (PLC)[46].

Otros documento recientes son los dos informes de la relatora de privacidad de la ONU emitidos en 2025, cuyos principales aspectos destacamos a continuación.

Informe A/HRC/58/58 del 16 de enero de 2025: Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad

En el informe A/HRC/58/58[47], se invitó a los Estados para que:

- Promuevan prácticas éticas en el uso de neurotecnologías que aseguren el debido tratamiento de los neurodatos y eviten el indebido uso de las neurotecnologías
- Fomenten la educación ciudadana sobre neurotecnologías y neurodatos para que le permita a las personas comprender el impacto de las neurotecnologías, adoptar decisiones informadas, conocer sus derechos y exigir el respeto de los mismos.
- Regulen las neurotecnologías y el tratamiento de los neurodatos. Allí se enfatiza la importancia de expedir normas locales que se anticipen o mitiguen los riesgos asociados al uso indebido de las neurotecnologías.
- Incorporar los fundamentos y principios desde el derecho a la privacidad sugeridos en el informe en comento. Es decir, los siguientes: la protección de la dignidad humana, el consentimiento informado, la ética en el diseño, el principio de precaución y la no discriminación. Estos principios asegurarán un equilibrio entre la innovación tecnológica en neurotecnologías y la protección de

[39] ONU, documento A/HRC/RES/51/3 del 13 de octubre de 2022. La neurotecnología y los derechos humanos.; ONU, Informe A/79/173 del 17 de julio de 2024 sobre actualización de la resolución 45/95 de la Asamblea General, de 14 de diciembre de 1990, titulada "Principios rectores sobre la reglamentación de los ficheros computarizados de datos personales"; ONU, documento A/HRC/57/61 del 8 de agosto de 2024. Efectos, oportunidades y retos de la neurotecnología en relación con la promoción y la protección de todos los derechos humanos, y ONU, Informe A/HRC/58/58 del 16 de enero de 2025: Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad.

[40] RIPD, 2024. Declaración sobre neurotecnologías y neurodatos en el marco de la normativa de protección de datos, aprobada en sesión cerrada del encuentro de la Red Iberoamericana de Protección de Datos, en Cartagena, Colombia el 29 de mayo de 2024); RIPD, 2023. Declaración sobre neurodatos, aprobada en sesión cerrada del encuentro de la Red Iberoamericana de Protección de Datos, en la Antigua, Guatemala el 25 de septiembre de 2023.

[41] Cfr. Global Privacy Assembly. Resolution on principles regarding the processing of personal information in neuroscience and neurotechnology. 46th Closed Session of the Global Privacy Assembly.

[42] Cfr. Organisation for Economic Cooperation and Development (2019) Recommendation of the Council on Responsible Innovation in Neurotechnology. En: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457#mainText

[43] Council of Europe (2021) Report 'Common Human Rights challenges raised by different applications of neurotechnologies in the biomedical field', October.

[44] OEA, 2023. Declaración de principios interamericanos en materia de neurociencias, neurotecnologías y derechos humanos. Aprobada en marzo de 2023 por el Comité Jurídico Interamericano de la Organización de Estados Americanos; OEA, 2021. Principios actualizados sobre la privacidad y la protección de datos personales, con anotaciones expedidos el 9 de abril de 2021 por el Comité Jurídico Interamericano (CJI), órgano consultivo de la Organización de Estados Americanos (OEA). Estos principios fueron aprobados por la Asamblea General de la OEA en noviembre de 2021; OEA, 2021. Declaración sobre neurociencia, neurotecnologías y derechos humanos: nuevos desafíos jurídicos para las américas. Aprobada en agosto de 2021 por el Comité Jurídico Interamericano de la Organización de Estados Americanos.

[45] UNESCO, 2023. Unveiling the neurotechnology landscape. Scientific advancements, innovations and major trends. En: https://unesdoc.unesco.org/ark:/48223/pf0000386137; UNESCO, 2021. Recomendación sobre la ética de la inteligencia artificial; UNESCO, 2021. Ethical issues of neurotechnology: report, adopted in December 2021.

[46] PLC, 2023. Ley Modelo de Neuroderechos para América Latina y el Caribe (Panamá 19 y 20 de mayo 2023).

[47] ONU. A/HRC/58/58: Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad - Informe de la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères. Texto oficial publicado en: https://www.ohchr.org/es/documents/thematic-reports/ahrc5858-foundations-and-principles-regulation-neurotechnologies-and

los derechos humanos con especial referencia a la privacidad y el debido tratamiento de los neurodatos.

Previo a lo anterior, en el informe se analizaron documentos relevantes sobre las neurotecnologias y tratamiento de datos en la ONU[48] en donde se destaca la importancia de dichas tecnologías, los riesgos por su uso indebido y la importancia de establecer marcos regulatorios que protejan los derechos humanos frente a las neurotecnologías. Posteriormente se definió las neurotecnologías y destaco sus beneficios[49] junto con algunos riesgos[50] frente a la privacidad, la dignidad humana y la sociedad. Adicionalmente se refirió a los datos cerebrales o neurodatos como una categoría especial de datos personales que requiere un tratamiento ético, profesional y diligente para garantizar la protección de las personas y salvaguardar su dignidad humana.

Se destacó que pese a los beneficios en la salud mental que traerán las neurotecnologías, existe el temor de que se pueda manipular cerebralmente seres humanos y se hizo referencia al derecho a "ser como soy: derecho al yo, a mi identidad cerebral natural" y a "decidir por mí mismo, sin ser artificialmente manipulado o programado"

También se refirió a los siguientes principios relevantes que deben tenerse presente en la regulación del uso de neurotecnologías y tratamiento de los neurodatos:

1. Dignidad humana.
2. Datos neuronales como datos personales altamente sensibles
3. Privacidad mental y consentimiento para tratar neurodatos
4. Ética y protección de los derechos humanos desde el diseño y por defecto en el desarrollo y uso de neurotecnologías
5. Precaución
6. Responsabilidad demostrada y seguridad en el tratamiento de neurodatos
7. No discriminación
8. Protección efectiva de los derechos de las personas frente al tratamiento de los neurodatos.

Dentro de los pilares principales o principios fundamentales que sugiere la ONU sobresalen los siguientes:

La dignidad humana como eje central del diseño y uso de las neurotecnologías.

La dignidad humana es el principal fundamento que irradia el diseño y uso de las neurotecnologías. Para la ONU, "No deben permitirse las prácticas que sean contrarias a la dignidad humana. Toda persona debe tener acceso a los progresos de las neurotecnologías, respetándose su dignidad y derechos. El Estado promoverá un enfoque basado en el respeto a la dignidad humana y los derechos humanos en el diseño, desarrollo, implementación, comercialización, evaluación y uso de las neurotecnologías."

Ética y protección de los derechos humanos desde el diseño y por defecto en la creación y uso de neurotecnologías.

La ética es otro componente esencial de las recomendaciones de la ONU, la cual sugiere que "la ética desde el diseño y por defecto debe irradiar el esquema, desarrollo y uso de

[48] Particularmente se refirió a los siguientes documentos: (1) Documento A/HRC/RES/51/3 del 13 de octubre de 2022. La neurotecnología y los derechos humanos.; (2) Informe A/79/173 del 17 de julio de 2024. Propuesta de actualización de la Resolución 45/95 de la Asamblea General, de 14 de diciembre de 1990, titulada" principios rectores para la reglamentación de los ficheros computadorizados de datos personales", y (3) Documento A/HRC/57/61 del 8 de agosto de 2024. Efectos, oportunidades y retos de la neurotecnología en relación con la promoción y la protección de todos los derechos humanos.

[49] Dentro de los beneficios se destacaron los siguientes: Realizar investigaciones para descubrir cómo funciona el cerebro y cuál es la base científica de la mente humana; Diagnosticar, entender, y diseñar nuevas terapias para las enfermedades cerebrales tanto neurológicas, neurodegenerativas o psiquiátricas. Enfermedades como el Alzheimer, la esquizofrenia, el Parkinson, la epilepsia, la discapacidad mental, el ictus, la esclerosis lateral, la depresión, la ansiedad, etc… Estas enfermedades cerebrales afectan de una manera cada vez mayor a un gran porcentaje de los ciudadanos y son la lacra de la humanidad, y Fomentar la creación de dispositivos de interfaz cerebro computadora, que permitan la conexión directa con el internet, y forme la base de una industria nueva, con grandes beneficios económicos y también a los consumidores.

[50] Se destacaron los siguientes riesgos: (1) Usar las neurotecnologías para fines contrarios a la dignidad humana; (2) . Modificar artificialmente los seres humanos; (4) Generar daños físicos al ser humano y la manipulación mental; Tratar Indebidamente de los neurodatos y usarlos para fines contrarios a la dignidad humana o no autorizados por la ley.

los productos o procesos de investigación sobre el cerebro, las neurotecnologías y los neurodatos." Adicionalmente, promueve "un enfoque basado en derechos humanos en el desarrollo de las neurotecnologías, buscando garantizar la protección integral y el respeto a los mismos a partir del diseño de las neurotecnologías, sus modos de investigación, como en su implementación, comercialización, evaluación y uso."

Principio de precaución en el uso, desarrollo e implementación de neurotecnologías.

El principio de precaución refuerza el enfoque preventivo en materia de derechos humanos de manera que busca evitar o controlar riesgos graves o irreversibles que puedan comprometer la dignidad humana, la integridad personal, la privacidad mental y otros derechos fundamentales, incluso cuando no exista certeza científica sobre la magnitud de dichos riesgos.

Informe A/80/283 del 31 de julio de 2025: Elementos para crear una ley modelo sobre neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad.

El informe A/80/283[51] es complementario del mencionado anteriormente (A/HRC/58/58) y, como su nombre lo indica, tiene como objetivo sugerir los fundamentos para la creación de una ley modelo sobre neurotecnologías desde la perspectiva del derecho a la privacidad.

Allí se reiteran los principios señalados anteriormente respecto del informe A/HRC/58/58 y se desarrollan los siguientes:

Identidad cerebral, autonomía, privacidad de la actividad neuronal y no manipulación cerebral.

Para la ONU, es imprescindible preservar la identidad cerebral natural de manera que excepcionalmente se puede alterar la misma. Por eso, se debería prohibir cualquier manipulación artificial del cerebro o de la información neuronal, excepto cuando se realice con los siguientes fines: (a) Protección de la salud; b) Diagnóstico, tratamiento, rehabilitación o paliación de enfermedades, en el marco del derecho fundamental a la salud, y c)Investigación científica en los campos de la biología, psicología y medicina, orientada a aliviar el sufrimiento o mejorar la salud, siempre que se realice conforme a las normas éticas y legales aplicables. En suma, señala la ONU que "cada persona debe tener el control exclusivo sobre su identidad neuronal, asegurando la autodeterminación y la libertad de pensamiento. En otras palabras, toda persona tiene derecho a decidir sobre su identidad cerebral natural, y a que su cerebro no sea manipulado artificialmente de forma que se alteren sus decisiones o personalidad, salvo en los casos expresamente autorizados por la ley y en cumplimiento de estrictas normas éticas."

Aplicación terapéutica exclusiva respecto al aumento de las capacidades cognitivas para evitar seres humanos de primera y de segunda categoría.

Se enfatiza, de una parte, la necesidad de utilizar las neurotecnologías únicamente "para fines médicos, incluyendo la promoción de la salud, la prevención, el diagnóstico, el tratamiento, la rehabilitación y los cuidados paliativos de las enfermedades" . Y, de otra parte, se hace un llamado a evitar "que se creen seres humanos de primera y de segunda. Los de primera serían que poseen cerebro artificialmente mejorado y los de segunda lo que tienen cerebro natural."

Integridad e intimidad neurocognitiva.

Se recomienda "prohibir alterar la libertad de pensamiento y conciencia o convertir al individuo en dependiente de un tercero. En otras palabras, no se debe permitir la manipulación cerebral para convertir a los seres humanos en títeres o sujetos manipulados cerebralmente".

7. Conclusión

Como mencioné al inicio, no hay respuestas perfectas ni conclusiones definitivas sobre estos temas; solo contamos con insumos que deberíamos tener en cuenta para abordar correctamente

[51] ONU. A/80/283: Informe de la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères - Elementos para crear una ley modelo sobre neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad. El texto oficial del informe puede consultarse en: https://www.ohchr.org/es/documents/thematic-reports/a80283-report-special-rapporteur-right-privacy-ana-brian-nougreres

estas cuestiones.

Las decisiones que se tomen en ese momento serán fundamentales, ya que marcarán el rumbo de la sociedad que queremos construir y definirán los límites necesarios para que los avances tecnológicos y el uso de la tecnología no perjudiquen la dignidad humana ni la esencia de la humanidad.

Queda mucho por hacer. Aún es tiempo de tomar decisiones sensatas, inteligentes y éticas que sean determinantes para definir el tipo de sociedad que queremos y que no deseamos.

Bibliografía

Brian Nougrères, Ana. *Elements for the Development of a Model Law on Neurotechnologies and the Processing of Neurodata from the Perspective of the Right to Privacy*. A/80/283. New York: United Nations General Assembly, 2025.

Foundations and Principles for the Regulation of Neurotechnologies and the Processing of Neurodata from the Perspective of the Right to Privacy. A/HRC/58/58. Geneva: United Nations Human Rights Council, 2025.

Genser, Jared, Stephen Damianos y Rafael Yuste. *Safeguarding Brain Data: Assessing the Privacy Practices of Consumer Neurotechnology Companies*. New York: Neurorights Foundation, 2024.

Global Privacy Assembly. "Resolution on Neurotechnologies, Human Rights, and Data Protection." 46th Global Privacy Assembly, 2024.

Manes, Facundo y Mateo Niro. *Usar el cerebro: Conocer nuestra mente para vivir mejor*. 4th ed. Buenos Aires: Planeta, 2014.

Organización de las Naciones Unidas. *Declaración sobre la Utilización del Progreso Científico y Tecnológico en Interés de la Paz y en Beneficio de la Humanidad*. Resolución 3384 (XXX), 10 de noviembre de 1975.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). *Declaración Universal sobre el Genoma Humano y los Derechos Humanos*. París: UNESCO, 1997.

Organización de los Estados Americanos, Comité Jurídico Interamericano. *Declaración de Principios Interamericanos en Materia de Neurociencias, Neurotecnologías y Derechos Humanos*. CJI/RES. 281 (CII-O/23) corr.1. Río de Janeiro: OEA, 2023.

Red Iberoamericana de Protección de Datos. *Declaración de la RIPD sobre Neurotecnologías y Neurodatos*. 2024.

Remolina Angarita, Nelson. ***Recolección Internacional de Datos Personales: Un reto del mundo post-internet***. Madrid: BOE, 2015.

"Neuro reflexión: hacia una declaración universal sobre las neurotecnologías y los derechos humanos." En *En defensa de los neuroderechos*, edited by Moisés Sánchez, Ciro Colombara, and Natalia Monti. Chile: Fundación Kamanau, 2024.

Yuste, Rafael. "Un paso histórico." En *En defensa de los neuroderechos*, editado por Moisés Sánchez, Ciro Colombara y Natalia Monti. Chile: Fundación Kamanau, 2024.

Las tres sendas hacia la privacidad

The Three Paths to Privacy

Pilar Vargas Martínez
Profesora sustituta de Derecho Internacional
Universidad Complutense de Madrid

Resumen:
En materia de gobernanza del dato, el análisis comparado de las políticas seguidas por la Unión Europea, los Estados Unidos y la República Popular China puede representarse a través de tres metáforas históricas con lógicas regulatorias bien diferenciadas: el Camino de Santiago, la Ruta 66 y la Ruta de la Seda. El Camino de Santiago, caracterizado por ser una red supranacional de protección, hospitalidad y reglas comunes refleja el enfoque de la Unión Europea, donde la privacidad se configura como un derecho fundamental. Así, el peregrino medieval y el ciudadano digital comparten un marco que busca protegerlos durante su tránsito, lo que asegura que la movilidad (física o digital) se mantenga como un elemento central de la cohesión y la identidad europeas. Por su parte, la Ruta 66 estadounidense simbolizó un modo de entender la movilidad asociado a la libertad individual y al espíritu empresarial. En cambio, en la Ruta 66 las normas eran mínimas y el propio conductor asumía la responsabilidad de gestionar riesgos y oportunidades. Esta lógica es coherente con la tradición regulatoria estadounidense, caracterizada por una intervención estatal limitada y una marcada preferencia por la autorregulación y el libre mercado. La Ruta de la Seda, al ser una vía histórica de intercambio controlado y herramienta de proyección estratégica del poder imperial, permite entender el modelo chino, basado en la soberanía digital, la centralidad del Estado y la primacía de la seguridad nacional. Su legislación articula un régimen que combina protección formal del individuo con un marcado autoritarismo. En conjunto, estos tres modelos revelan un panorama global fragmentado, donde conviven enfoques garantistas, liberal-mercantilistas y soberanistas, lo que produce tensiones estructurales en la gobernanza del dato y en su viabilidad regulatoria.

Palabras clave:
Protección de datos, privacidad, Estados Unidos, Unión Europea, China.

Abstract:
In the field of data governance, a comparative analysis of the policies pursued by the European Union, the United States, and the People's Republic of China can be represented through three historical metaphors with clearly differentiated regulatory logics: the Camino de Santiago, Route 66, and the Silk Road. The Camino de Santiago, characterized as a supranational network of protection, hospitality, and shared rules, reflects the approach of the European Union, where privacy is conceived as a fundamental right. In this sense, the medieval pilgrim and the digital citizen share a framework designed to protect them during their journey, ensuring that mobility (whether physical or digital) remains a central element of European cohesion and identity. By contrast, the American Route 66 symbolized a conception of mobility associated with individual freedom and entrepreneurial spirit. Along Route 66, rules were minimal and the driver assumed responsibility for managing risks and opportunities. This logic is consistent with the U.S. regulatory tradition, characterized by limited state intervention and a strong preference for self-regulation and the free market. The Silk Road, as a historical route of controlled exchange and a tool for projecting imperial strategic power, helps explain the Chinese model, which is based on digital sovereignty, the centrality of the state, and the primacy of national security. Chinese legislation establishes a regime that combines formal protection of the individual with a pronounced authoritarian dimension. Taken together, these three models reveal a fragmented global landscape in which rights-based, liberal–market-oriented, and sovereignist approaches coexist, generating structural tensions in data governance and in its regulatory viability.

Keywords:
Data protection, privacy, United States, European Union, China.

1. Las tres sendas hacia la privacidad

La Unión Europea (UE), los Estados Unidos (EE. UU.) y la República Popular China (China) son los tres principales actores en el comercio mundial de servicios, incluidos los servicios digitales. De hecho, la UE y los EE.UU. son los principales socios comerciales en la exportación e importación de servicios digitales. En 2019, los servicios digitales representaron una parte muy significativa del comercio de servicios de EE.UU,[1] donde aproximadamente el 59 % de todas las exportaciones de servicios y el 50 % de sus importaciones de servicios estaban relacionados con servicios habilitados digitalmente, lo que representó también una gran mayoría del superávit mundial estadounidense en comercio de servicios[2].

En la UE, los servicios digitales constituyen una parte importante del comercio de servicios del bloque con terceros países. La UE es uno de los mayores exportadores e importadores de servicios del mundo, con alrededor del 25 % del comercio mundial de servicios, y los servicios digitales se encuentran entre los segmentos más dinámicos de este sector[3]. En cuanto a las exportaciones de servicios de la UE, constituyen uno de los principales destinos, al representar cerca de una quinta parte de las exportaciones totales de servicios de la UE en 2023. En sentido inverso, se estima que las exportaciones de servicios digitales de los EE.UU. con destino la UE representan entre un 25 % y un 30 % del total de sus exportaciones en servicios digitales.

Por su parte, China ha crecido rápidamente como actor global en comercio, especialmente en bienes manufacturados; sin embargo, a diferencia del sector de productos digitales, en servicios digitales persiste una participación menor en comparación con la UE y EE. UU. En consecuencia, los datos disponibles muestran que el comercio de servicios de China es aún más limitado y, del mismo modo, sus exportaciones de servicios digitales exponen una fracción menor del total mundial comparado con los líderes transatlánticos[4].

En ese orden de ideas, resulta evidente que el comercio digital se ha convertido en un elemento central de las relaciones económicas y comerciales globales. A pesar de ello, han existido presiones políticas (especialmente desde la UE) para restringir el flujo transfronterizo de datos entre los principales bloques y, con este, de numerosas operaciones comerciales, lo que argumenta el mantenimiento de estándares de privacidad y explica, por defecto, su creciente intervención restrictiva en el pasado. A su vez, los EE. UU. han liderado recientemente una ofensiva en el sentido inverso, que podría haber calado en el regulador europeo. Aún así, cabe esgrimir que la contraposición absoluta entre privacidad y comercio resulta engañosa.

En ese sentido, se debe partir de la base de que, si bien el derecho humano a la privacidad está universalmente reconocido, su alcance sobre la protección de datos personales no cuenta con una lectura uniforme, dado que mientras en algunos lugares constituye un derecho fundamental, en otros carece de esta consideración. A la luz de este contexto, se arguye que regulación nacional de la protección de la privacidad y de los datos personales, como la de cualquier objetivo legítimo de interés público nacional, está ligada a los valores constitucionales y éticos que persiguen los

[1] Según diferentes informes del BEA y estudios especializados sobre comercio digital los Servicios habilitados digitalmente (*digitally deliverable services*): Representan aprox. el 60 % de todas las exportaciones de servicios de EE. UU; Representan aprox. el 52 % de todas las importaciones de servicios; y Generan más del 75 % del superávit total estadounidense en comercio de servicios. Esto significa que el superávit de EE. UU. en servicios proviene abrumadoramente del sector digital, no del sector industrial.

[2] *The Transatlantic Economy 2021: Digital Acceleration*, The Transatlantic Economy, 2021, https://transatlanticrelations.org/wp-content/uploads/2021/03/TA-economy-2021_CH4.pdf?utm_source=chatgpt.com

[3] Comisión Europea, "Servicios – estadísticas," acceso el 8 de marzo de 2026, https://trade.ec.europa.eu/access-to-markets/en/content/services-statistics?utm_source=chatgpt.com

[4] Jacques Delors Institute, *Europe in the World: Mapping the EU's Digital Trade—A Global Leader Hidden in Plain Sight?* (París: Jacques Delors Institute, 2023).

diferentes Estados[5].

Ahora bien, en EE. UU. no existe una normativa transversal específica para la protección de datos, sino normativas sectoriales al respecto. Sobre esto, la Constitución no menciona este derecho concreto en torno a la protección de los datos frente al uso que los responsables de su tratamiento hacen de ellos[6].

En efecto, más allá del reconocimiento por el Tribunal Supremo del derecho a la intimidad en el ámbito del derecho penal y en el de la salud reproductiva, se evidencia que la salvaguarda de la información personal de los estadounidenses reside, en gran medida, en la ley de protección del consumidor. En este sentido, no puede perderse de vista que, en sistemas jurídicos del entorno europeo, los datos personales y su protección se conciben como un derecho fundamental en sí mismo[7] mientras que, en otros, como los EE. UU., aunque se sitúe en el marco del respeto a la privacidad del individuo como derecho fundamental, se asocia esencialmente a la protección de los consumidores[8]. De tal modo, esto puede traducirse en distintas consideraciones de los datos personales; es decir, como un derecho en la UE frente a su catalogación como una mercancía o el propio objeto de una transacción económica en los EE.UU.

Sin embargo, aunque se discuta esta caracterización, quien recogió los datos puede, salvo en algunos ámbitos, disponer de ellos prácticamente como sí de un bien más se tratara[9]. De este modo, resulta evidente que reúnen cualidades distintas de las mercancías tradicionales al estar directamente vinculados con aspectos personales de los individuos. En consecuencia, su mercantilización, que busca obtener rendimientos económicos, ha dado lugar a un "capitalismo de vigilancia" (*surveillance capitalism*). Así, desde el punto de vista de las transferencias internacionales de datos, se advierte que la legislación sectorial estadounidense no contiene restricciones generales. Asimismo, la amplia legislación en materia de privacidad considerada recientemente tampoco preveía un cambio fundamental de este enfoque de "*laissez-faire*", y parece poco probable que se efectúe en el futuro.

Por el contrario, el Reglamento general de protección de datos (RGPD)[10] de la UE condiciona los flujos de datos de carácter personal desde el territorio de la UE a la existencia de garantías de privacidad que se desplacen con los datos hacia su lugar destino. Por lo tanto, si la Comisión Europea ha decidido que un tercer país concreto ofrece un nivel de protección "adecuado", los datos pueden circular libremente hacia este desde el territorio de la UE sin formalidades adicionales. Para todos los demás destinos, las salvaguardas deben incluirse en los contratos comerciales individuales de transferencia de datos (como cláusulas contractuales tipo).

En ese sentido, la UE ha aumentado gradualmente el número de decisiones de adecuación que ha emitido, lo que permite flujos de datos sin restricciones a los Estados favorecidos por estas; no obstante, los efectos en el comercio digital mundial siguen siendo relativamente modestos, dado que se han emitido 15 decisiones unilaterales de adecuación en 25 años de esfuerzos[11]. Asimismo,

[5] Carmen Otero García-Castrillón, *Protección de datos en la economía digital una aproximación desde la regulación del comercio internacional* (Pamplona: Thomson Reuters Aranzadi, 2021), 74.

[6] Solamente tendría relevancia la Cuarta Enmienda de la Constitución en lo que concierne a la recopilación y el uso de información de datos personales por parte de organismos gubernamentales.

[7] Paul Schwartz y Karl-Nikolaus Peifer, en "Structuring International Data Privacy Law", *International Data Privacy Law 9*, no. 1 (2019): 7-21, señalan "European data protection law is strongly anchored at the constitutional level. Its goal is to protect individuals from risks to personhood caused by the processing of personal data". En este sentido, se hacen eco de cómo Tribunal Constitucional Federal alemán -casos *Census* (1983) e IT *Privacy* (2008)- han cumplido un papel destacado en la conceptualización europea de la privacidad de los datos, al destacar cómo el tratamiento de datos de carácter personal puede amenazar la autonomía de la toma de decisiones individual y socavar "una comunidad democrática libre basada en la capacidad de sus ciudadanos para actuar y participar". El resultado es el concepto de "derecho a la autodeterminación informativa", una idea que la legislación europea sobre privacidad de datos ha adoptado buscando prevenir los riesgos causados por el tratamiento de datos personales.

[8] Otero García-Castrillón. "Protección…", 35

[9] *Ibid*, 48

[10] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos), DOUE L 119, 4 de mayo de 2016, 1–88.

[11] Andorra, Argentina, Canadá (solo para organizaciones comerciales), Islas Feroe, Guernsey, Israel, Isla de Man, Japón, Jersey, Nueva Zelanda, República de Corea del Sur, Suiza, Reino Unido, EE. UU. (bajo el marco EU-US Data Privacy Framework) y Uruguay. Serían 16 si se incluye la de la Organización Europea de Patentes, de menor relevancia en el contexto específico que se expone.

se han celebrado escasos acuerdos de comercio digital[12]. En efecto, la UE -por número de acuerdos de comercio digital firmados- se sitúa a la cola de sus grandes competidores en el comercio digital. Como resultado, se observa que una gran proporción de las transferencias de datos de carácter personal entre la UE y el resto del mundo requieren la meticulosa y económicamente ineficiente inclusión de cláusulas contractuales tipo en las contrataciones de servicios que implican de transferencia de datos.

Aunado a esto, se expone que, en la relación entre la UE y EE. UU. se han alcanzado notables progresos para la estabilización de las transferencias de datos. En primer lugar, se encuentra el acuerdo *Data Privacy Framework*[13] -que sustituye al antiguo *Privacy Shield*-; en segundo lugar, está la firma, tanto de la UE como de los EE.UU., de la Declaración de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) sobre el acceso de los gobiernos a los datos de carácter personal en poder de responsables o encargados del tratamiento del sector privado[14]. Sin embargo, para establecer un marco jurídico más estable entre ambas potencias aún queda margen para profundizar en la negociación bilateral y en la búsqueda de un acuerdo de comercio digital específico para cada sector, utilizando para ello el Consejo de Comercio y Tecnología (TTC), que podría trabajar para un entendimiento común sobre la evaluación de riesgos de transferencias internacionales de datos de carácter personal (TIAS por sus siglas en inglés *Transfer impact assessment*).

A la luz de lo expuesto, un pertinente punto de inicio para encontrar un entendimiento en el largo plazo entre la UE y los EE. UU. puede ser el planteamiento común -ya hoy compartido en cierta medida por europeos y estadounidenses y presente en una serie de instrumentos jurídicos internacionales sobre flujos de datos- sobre la responsabilidad proactiva de las entidades que tratan los datos personales de los individuos. En efecto, se plantea que estas entidades deben evaluar los riesgos asociados al tratamiento de datos que realicen, establecer políticas, procedimientos y mecanismos internos con el objeto de gestionarlos de forma segura y, en su caso, aportar pruebas de su cumplimiento a las autoridades de supervisión en la materia, así como a los interesados que ejerzan determinados derechos de protección de datos de carácter personal.

Al respecto, varias autoridades europeas de protección de datos han elaborado guías al efecto para los responsables y encargados del tratamiento. De tal modo, el objetivo de esta estrategia es doble: por un lado, promueve que estos evalúen los riesgos derivados de las transferencias internacionales de datos y de la potencial supervisión por parte de gobiernos extranjeros. Lo anterior se denomina TIA, tal y como se ha indicado. Por otro lado, fomentan el establecimiento de salvaguardias correspondientes mediante cláusulas contractuales tipo[15]. Por su parte, los EE. UU. no han adoptado un enfoque de diligencia equivalente al europeo, a pesar de los intentos infructuosos en este sentido.[16] En cambio, China ha establecido un sistema enfocado en la seguridad nacional antagónico a la filosofía occidental de un internet libre y, en consecuencia, de la liberalización de los flujos de datos.

En todo caso, esta situación revela cómo las distintas visiones de la privacidad -como un derecho del consumidor en los EE. UU. y como un derecho fundamental en la UE- inciden en el ámbito del comercio internacional. Tal vez por esta razón los EE.UU.[17] solían encabezar, hasta muy recientemente[18] tanto a escala bilateral como regional, la firma de acuerdos internacionales para la liberalización de intercambios comerciales en los que ha empezado a incorporarse la protección

[12] La UE ha concluido acuerdos con capítulos de comercio digital con Canadá, Singapur, Vietnam, Japón, Reino Unido, México, Chile, Mercosur y Nueva Zelanda (algunos de ellos con capítulos específicos sobre comercio digital o digital trade provisions). Además, en 2024-2025 la UE ha firmado acuerdos de comercio digital "self-standing" con Singapur y con Corea del Sur-lo que se considera el primer y segundo acuerdo de este tipo-.

[13] Comisión Europea, "EU–US Data Privacy Framework: Questions and Answers on the Adequacy Decision," 13 de diciembre de 2022, https://ec.europa.eu/commission/presscorner/detail/de/qanda_22_7632

[14] Organización para la Cooperación y el Desarrollo Económicos (OCDE), "Declaration on Government Access to Personal Data Held by Private Sector Entities," OECD Legal Instruments, 13 de mayo de 2023, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0487

[15] Comité Europeo de Protección de Datos (EDPB), *Recomendaciones 02/2020 sobre las garantías esenciales europeas para medidas de vigilancia*, adoptadas el 10 de noviembre de 2020, https://www.edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-022020-european-essential-guarantees_es

[16] A medida que consiguieran avanzar hacia una ley federal integral de privacidad y hacia un mayor compromiso con las transferencias de datos en entornos multilaterales, se verían inevitablemente empujados en esta dirección.

[17] El Acuerdo entre los EE. UU, México y Canadá (art. 19.11); y el Acuerdo de comercio digital entre los EE. UU. y Japón (art. 11) contienen amplias garantías sobre la capacidad de transferir datos a través de redes digitales transnacionales.

[18] EE. UU. recientemente ha protagonizado un cambio sustancial de postura en este ámbito lo que ha permitido que algunos estados asiáticos le tomen la delantera en materia de firma de acuerdos internacionales en este sentido.

de datos de carácter personal y de la privacidad. De forma reciente, esta cuestión se ha intentado incorporar sin éxito en las negociaciones comerciales multilaterales.

Asimismo los acuerdos de libre comercio de los EE. UU. y los de la UE[19] reconocen el derecho de los gobiernos a limitar las transferencias de datos si las consideran incompatibles con sus respectivas normativas internas, incluidas las de protección de la privacidad o las de seguridad nacional. Sin embargo, se observa que los acuerdos promovidos por EE. UU. poseen ventajas e insisten en que esas medidas reguladoras restrictivas sean las estrictamente necesarias[20]. Del mismo modo, los acuerdos estadounidenses hacen hincapié en que dichas medidas no constituyan una discriminación arbitraria o injustificable ni restricciones encubiertas al comercio de servicios entre Estados[21]. Para ello, este país incorpora referencias a las excepciones permitidas en virtud del Acuerdo General sobre el Comercio de Servicios (GATS)[22] de la Organización Mundial del Comercio (OMC). Por lo tanto, aunque la UE es parte del GATS y, sin perjuicio del compromiso de su cumplimiento, sus acuerdos comerciales bilaterales permiten expresamente a cada parte imponer las limitaciones que considere oportunas por razones de privacidad o protección de datos[23].

1.1. La Ruta 66 digital: la tecnología al servicio del mercado

La Ruta 66 es una conocida carretera federal creada en 1926 para cruzar los EE.UU. de costa a costa -con inicio en Chicago y final en California-, acercando la costa oeste a los emigrantes estadounidenses que soñaban con un futuro mejor. Aunque se concibió para ampliar las vías comerciales, terminó convirtiéndose en un símbolo de la libertad estadounidense. De tal modo, la Ruta 66 encarna el sueño americano y, por ello, adquiere una relevancia cultural significativa, al representar, entre otros aspectos, la nostalgia por una época de exploración y progreso.

De manera análoga a como la Ruta 66 simboliza esa sed de progreso y libertad, el enfoque estadounidense en materia de regulación de la economía digital se articula históricamente en torno a una fe inquebrantable en los mecanismos de mercado y a un escepticismo arraigado frente a la intervención gubernamental. Esta aproximación tecno-optimista -que hunde sus raíces en una cultura jurídica y política marcadamente liberal- se fundamenta en la convicción de que la innovación tecnológica se despliega con mayor eficacia allí donde el Estado limita su actividad regulatoria al mínimo indispensable. Desde esta perspectiva, la economía digital prospera principalmente gracias al dinamismo empresarial y a la autorregulación de las empresas tecnológicas, mientras que la intervención gubernamental se percibe como una carga que incrementa los costos y restringe el comportamiento innovador del sector privado.

Bajo esta visión tecno-libertaria, la regulación estatal no solo afecta negativamente la eficiencia de los mercados, sino que también podría socavar la libertad individual y la autonomía de empresas y ciudadanos. De este modo, la defensa de la innovación y del crecimiento económico opera como justificación económica para la no intervención, mientras que la protección de la autonomía personal y de la libertad individual funciona como argumento político orientado a minimizar el papel del Estado en la economía digital.

Estas convicciones se encuentran profundamente arraigadas en el régimen jurídico estadounidense. Ahora bien, el elemento normativo que mejor encarna este *ethos* regulatorio es la Sección 230 de

[19] Por el momento China presenta una posición expectante en materia de firma de acuerdos bilaterales y multilaterales en este ámbito, priorizando compromisos genéricos de *soft law* alineados con los principios que este Estado promueve.

[20] Aunque sus acuerdos no definen el criterio de necesidad, sí se hacen menciones concretas a "evitar carga regulatoria innecesaria en las transacciones electrónicas", – acorde con las restricciones permitidas en el art. XIV c) ii) y el art. XIV bis 1.b) del GATS – "asegurar que las restricciones a los flujos transfronterizos de información personal son necesarias y proporcionales a los riesgos presentados" o a que – acorde con las restricciones permitidas en el art. XIV a) y el art. XIV bis 1.b) del GATS – "ninguna parte prohibirá o restringirá la transferencia transfronteriza de información [...] cuando esta actividad sea para la realización de un negocio de una persona cubierta. Esto no impide que una parte adopte o mantenga una medida incompatible que sea necesaria para alcanzar un objetivo legítimo de política pública, siempre que la medida: (a) no se aplique de forma que constituya un medio de discriminación arbitraria o injustificable, o una restricción encubierta al comercio; y (b) no imponga restricciones a las transferencias de información mayores a las que se requieren para alcanzar el objetivo."

[21] Acuerdo de comercio entre los EE. UU., México y Canadá, art. 19.11.2. Una medida no cumpliría esta condición si concede un trato diferente a las transferencias de datos únicamente sobre la base de que son transfronterizas de una manera que modifica las condiciones de competencia en detrimento de los proveedores de servicios de otra parte.

[22] Organización Mundial del Comercio (OMC), *Acuerdo General sobre el Comercio de Servicios* (AGCS), 1994, arts. XIV y XIV bis, en https://www.wto.org/english/tratop_e/serv_e/gatsintr_e.htm Véase Acuerdo de libre comercio entre la UE y Nueva Zelanda, arts. 12.5 y 25.1.2.

[23] Véase Acuerdo de libre comercio entre la UE y Nueva Zelanda, arts. 12.5 y 25.1.2.

la Communications Decency Act (CDA) de 1996, una disposición que se ha caracterizado como la "piedra angular de internet" en los EE.UU.[24] La norma otorga amplia inmunidad a los intermediarios en línea frente a responsabilidades derivadas de contenidos generados por terceros.

A la luz de este escenario, y como se ha indicado, el planteamiento de la protección de los datos personales -si bien con un enfoque eminentemente sectorial- está profundamente arraigado en su historia al entender esa privacidad como la protección de la libertad, tal y como la concebía su Bill of Rights. De hecho, en los EE. UU., los agentes privados (responsables y encargados de tratamiento de datos) suelen estar protegidos frente a restricciones en su actividad por la Primera Enmienda. Por lo tanto, las entidades que tratan los datos estarían protegidas frente a posibles exigencias de los interesados o consumidores en relación con el tratamiento de sus datos. Además, en este país, las políticas en torno a la libertad en Internet han buscado continuamente "preservar y ampliar Internet como un espacio abierto y global para la libre expresión, para la organización y la interacción y para el comercio", tal y como confirmó la estrategia de la Casa Blanca sobre inteligencia artificial (IA)[25] bajo la presidencia de Joe Biden, y más recientemente la iniciativa del presidente Donald Trump para prevenir la fragmentación normativa entre Estados en materia de IA.

Mientras que, en virtud de la Primera Enmienda, la libertad de expresión goza de una sólida protección en los EE. UU., la protección de datos se regula de manera fragmentada a través de algunas leyes federales sobre privacidad y de un amplio conjunto de leyes estatales. Estas normas o bien se limitan al sector público, o bien son específicas del tipo de información o del medio a través del cual se difunde (por ejemplo, la información sanitaria, la privacidad de los vídeos o las comunicaciones electrónicas). En este sentido, no existen restricciones generales a la transferencia de datos personales por parte de entidades privadas, y la autorregulación, junto con las denominadas "mejores prácticas", constituye el modelo predominante de protección de la privacidad.

En lo relativo a la protección de los ciudadanos, el legislador estadounidense se centra en las restricciones que permite la Cuarta Enmienda de la Constitución en lo que respecta a la recopilación y al uso de datos personales por parte del gobierno, y no de los actores privados. Este último constituye el enfoque de la normativa europea en términos generales, la cual se encuentra más orientada al cumplimiento normativo (*compliance*). Esta diferencia se ilustra en la ambigüedad existente en EE. UU. respecto de la definición de los roles de responsable y encargado del tratamiento de los datos (agentes privados), frente a la claridad en su definición y alcance en el RGPD[26].

En esta situación normativa, no existe siquiera una definición uniforme y coherente del concepto de datos de carácter personal, tampoco sobre los datos sensibles, más conocidos en Europa, donde están claramente definidos como "categorías especiales de datos personales". Adicionalmente, los datos se consideran un bien susceptible de transacción y, en consecuencia, su exportación no está limitada en absoluto. Por lo general, existe una tendencia hacia una gobernanza liberal, basada en los intereses del mercado, en contraste con la gobernanza socialmente protectora y cimentada en los derechos de los interesados que existe en la UE.

En contraste, los EE. UU. conciben las normas que prohíben la transferencia internacional de datos como limitaciones a la libertad de circulación de la información, que ahoga la competencia y perjudica a las empresas digitales[27]. A pesar de ello, se observa que, desde el final de la Segunda Guerra Mundial, este país atiende de manera fundamental la lógica de la seguridad nacional, el orden público y la soberanía. Empero, en la última década se ha observado un replanteamiento de la cuestión que no parece derivar en un cambio de paradigma rápido y fácil. Esto es especialmente evidente a partir de las reacciones a la sentencia del Tribunal de Justicia de la UE (TJUE) en el asunto Schrems II[28] que, por lo demás, muestran significativas diferencias en los niveles de regulación federal y estatal.

[24] *Communications Decency Act*, 47 U.S.C. § 230 (1996).

[25] Accesible en: The White House, "Fact Sheet: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitments on AI," 26 de julio de 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2024/07/26/fact-sheet-biden-harris-administration-announces-new-ai-actions-and-receives-additional-major-voluntary-commitment-on-ai/. Visitado el 20 de diciembre de 2025.

[26] David Carrizo, "Reflexiones a propósito de la protección de datos en el escenario global digital: El derecho de daños en la litigiosidad internacional," *Revista Boliviana de Derecho*, no. 31 (2021): 451, destacando que estos conceptos son funcionales (tienen por objeto asignar responsabilidades en función del papel real de cada parte en el tratamiento de los datos) y autónomos (deben interpretarse conforme al derecho de la UE).

[27] Otero García-Castrillón. "Protección…", 47.

[28] Tribunal de Justicia de la Unión Europea, asunto C-311/18, *Data Protection Commissioner v Facebook Ireland Ltd y Maximillian Schrems*, sentencia de 16 de julio de 2020 (Schrems II).

Asimismo, algunos estados, como California, Colorado, Connecticut y Virginia cuentan formalmente con idénticos derechos a los garantizados por el RGPD, lo que alcanza un grado de protección de datos de carácter personal similar al europeo. Con todo, la regulación federal es fragmentaria; es decir, el marco regulador está orientado al comercio, por lo que no tiene un principio de prohibición y las escasas normas que existen sobre protección de datos personales son sectoriales y no exhaustivas. En la actualidad, se plantea la adopción de una normativa general a nivel federal que incluya los mismos siete principios y los seis derechos de los interesados que se contemplan en el RGPD. En todo caso, se advierte que el derecho a restringir el tratamiento de datos personales y el derecho a no ser objeto de la toma de decisiones automatizadas y de elaboración de perfiles solo se incluirá parcialmente.

Como es sabido, fue el caso Microsoft contra la Oficina Federal de Investigaciones (FBI por sus siglas en inglés) el que detonó el cambio de la Clarifying Lawful Overseas Use of Data Act (CLOUD Act)[29] para darle carácter extraterritorial. Ahora bien, este cambio fue posible dado que la CLOUD Act se modificó para permitir a las fuerzas de seguridad solicitar directamente (y obtener) datos tratados por proveedores de servicios con sede en el extranjero[30]. En tal sentido, esto ha derivado en el actual panorama de decisiones de adecuación (más conocidos como escudos de privacidad) débiles que se han ido implementando y también invalidando en hasta tres ocasiones desde 2015 hasta la fecha, con la notable supervivencia de su última versión a su paso por el TJUE en septiembre de 2025.

En el caso de referencia, Microsoft se negó a facilitar al FBI la información almacenada en sus servidores de Irlanda. Aunque Microsoft consiguió evitar el acceso a la información, el gobierno de los EE. UU. actuó rápido para cambiar la redacción del CLOUD Act y evitar que ese caso pudiera replicarse en el futuro. El objetivo estadounidense era claro: garantizar el acceso ilimitado por parte de las autoridades estadounidenses a la información almacenada por empresas norteamericanas en el extranjero. Hasta la fecha, esta es la única normativa estadounidense no estrictamente sectorial con carácter extraterritorial. En ese sentido, el derecho estadounidense lleva una década sorteando internamente un conflicto de intereses indudablemente grave en la legislación sobre privacidad. Este conflicto consiste en la confrontación entre la antigua lógica de la seguridad nacional, el orden público y la soberanía en torno a las que se desarrolló el discurso de la privacidad desde el final de la Segunda Guerra Mundial, frente al replanteamiento de la última década. En efecto, esta última visión está más enfocada hacia la explotación extensiva de las oportunidades económicas basadas en la economía del dato.

Algunas leyes estatales en la materia, como las de California, Colorado, Connecticut y Virginia que se han referido, persiguen una aplicación extraterritorial similar a la prevista a nivel federal en el CLOUD Act. No obstante, en estos cuatro Estados los mecanismos de transferencia desarrollados presentan un ámbito de aplicación acotado a nivel sectorial. Esta limitación viene definida por las propias normas estatales, las cuales son de aplicación exclusiva a determinados sectores de actividad, como el sanitario, el financiero o el educativo, entre otros. Dicha restricción sectorial imposibilita su aplicación transversal.

Una aplicación transversal permitiría maximizar el valor añadido de estos mecanismos en la economía digital, especialmente si se considera que las fronteras que delimitan unos sectores de otros son cada vez más difusas, lo que genera una demanda creciente de transversalidad para evitar que tales instrumentos resulten obsoletos. Un ejemplo de ello es la creciente incursión de los proveedores tradicionales de servicios tecnológicos en sectores regulados, como el financiero. Cuando un mecanismo de transferencia solo es aplicable de forma estricta a un sector determinado, cualquier iniciativa de negocio que implique una incursión comercial transectorial podría quedar fuera de su ámbito de aplicación. Esta situación generaría una desventaja competitiva para el libre flujo de información que los competidores más innovadores buscan desarrollar.[31] Se trata de una problemática inherente a la economía del dato y de un debate ya clásica en el seno de la OMC sobre la clasificación de productos y servicios digitales mixtos dentro del ámbito de aplicación del GATT o del GATS.

Asimismo, las mencionadas leyes estatales incorporan algunos requisitos puntuales de almacenamiento local o *data localisation*. De tal modo, este tipo de exigencias, habituales en los

[29] Accesible en: *Clarifying Lawful Overseas Use of Data Act* (CLOUD Act), S.2383, 115th Cong. (2018). Visitado el 27 de diciembre de 2025.

[30] En teoría la actual redacción de la *CLOUD Act* simplemente elimina los conflictos de leyes y no afecta a la jurisdicción sobre los proveedores extranjeros. Éstos están estrictamente limitados por el requisito de jurisdicción personal contenido en la Cláusula de *Due Process* de la Quinta Enmienda de la Constitución.

[31] Los productos de seguros comercializados por Amazon Marketplace son un buen ejemplo de esta situación pues su clasificación dentro de la normativa aseguradora los lastra.

sistemas asiáticos, implican restricciones totales o parciales a la transferencia internacional de datos y derivan de posturas proteccionistas en materia de privacidad.

Un aspecto positivo de estas leyes estatales, en comparación con la perspectiva europea, es la adopción de un enfoque de rendición de cuentas a posteriori o *light touch*, alineado con las recomendaciones de la OCDE. Así, dichas recomendaciones se centran en esquemas de colaboración con la industria y entre reguladores, así como en la corregulación y la autorregulación mediante códigos de conducta, con el fin de facilitar la interacción entre proveedores y usuarios.

Este planteamiento contrasta con las normativas europeas en la materia, que optan por enfoques de responsabilidad proactiva en la rendición de cuentas. En este sentido, el enfoque europeo aún presenta un amplio margen de desarrollo en materia de colaboración con la industria y autorregulación. Si bien el RGPD contempla cierta regulación sobre códigos de conducta, especialmente en relación con las transferencias internacionales de datos, se trata de un ámbito que, hasta el momento, apenas se ha desarrollado.

En septiembre de 2021, el Servicio de investigación del Congreso de EE. UU. publicó un informe sobre las opciones para facilitar la vuelta a la normalidad de las transferencias internacionales entre los EE.UU. y la UE. El informe en cuestión mencionaba expresamente, entre otras alternativas, la creación de una Ley Federal de Privacidad. Sin embargo, esta iniciativa ha quedado estancada. El motivo es evidente: la cuestión parecía haberse resuelto con la Transatlantic Data Privacy Framework[32], al menos a nivel político y temporalmente. Una posible Ley Federal de Privacidad era un gran argumento de negociación con la UE. En consecuencia, una legislación federal habría facilitado la uniformidad normativa entre los diferentes Estados de Norteamérica. Ahora bien, esto no solamente habría atajado la problemática de la fragmentación. Asimismo, habría facilitado que la UE pudiera llevar a cabo un reconocimiento de adecuación *standard* de los EE. UU., en lugar de continuar con el sistema de escudos de privacidad.

Sin embargo, con el reconocimiento del Transatlantic Data Privacy Framework y su reciente resistencia a la embestida del activista Max Schrems en la UE, los EE. UU. han ganado tiempo y pueden permitirse despriorizar la elaboración de dicha norma, al considerar que su principal motivación para plantearla había sido solventar el desencuentro con la UE tras la caída del Privacy Shield. No obstante, el activista Max Schrems ha recurrido a la Sentencia desfavorable a sus pretensiones de septiembre de 2025 sobre la validez de este instrumento ante el TJUE, por lo que se espera un futuro pronunciamiento al respecto a partir de 2026[33], que promete revolucionar de nuevo el panorama transatlántico, e impulsar a EE. UU. a retomar la idea de una Ley Federal de Privacidad.

A la luz de este escenario, el futuro de la regulación de la privacidad y la protección de datos en los EE. UU no parece estar claro. Si bien, cabe intuir, a tenor de las propuestas legislativas -Protecting Americans' Data from Foreign Adversaries Act[34] y la Executive Order 14117-, que ese porvenir promete la incursión de EE. UU en el terreno de las barreras al comercio basadas en motivos de seguridad nacional. Por lo tanto, un cambio de paradigma trascendería las medidas de vigilancia existentes que han sido detonantes del vaivén transatlántico, es decir, un estado cíclico de *fragile trust*.

En esta línea de frágil confianza, en 2021, se halló que un estudio realizado a petición del Comité de Libertades Civiles del Parlamento Europeo destacaba la escasa probabilidad de que las normativas de privacidad estatales, ni la federal lleguen a ofrecer una protección esencialmente equivalente a

[32] Accesible en: "Marco de privacidad de datos UE-EE. UU.", Data Framework Program [DFP] https://www.dataprivacyframework.gov/EU-US-Framework. Visitado el 20 de diciembre de 2025.

[33] Con ocasión de la publicación del *Transatlantic Privacy Framework*, el activista Max Schrems, responsable de llevar ante el TJUE los dos Escudos de privacidad anteriores, ya manifestó que desde su ONG Noyb "*Tenemos varias opciones de impugnación ya en el cajón, aunque estamos hartos de este ping-pong jurídico. Actualmente esperamos que esto vuelva al Tribunal de Justicia a principios del año que viene." A la espera de esta decisión, el TJUE resolvió en septiembre de 2025 el caso T-553/23, Latombe contra Comisión. Aunque se trata de un caso de alcance más limitado, el Tribunal de Justicia no suspendió el nuevo acuerdo considerando que las pequeñas mejoras de la Comisión han sido suficientes. No obstante, Schrems ha declarado no descartar que el TJUE resuelva en otro sentido en un caso de mayor alcance. Al fin y al cabo, durante los últimos 23 años, todos los acuerdos entre la UE y EE. UU. han sido declarados inválidos con carácter retroactivo, haciendo ilegales todas las transferencias de datos realizadas por las empresas en el pasado; parece que ahora vamos a añadir otros dos años de este ping-pong*". *En*: NOYB – European Center for Digital Rights, "La Comisión Europea da un tercer asalto ante el TJUE a las transferencias de datos entre la UE y EE. UU.," 10 de agosto de 2023, en https://noyb.eu/es/european-commission-gives-eu-us-data-transfers-third-round-cjeu. Visitado el 23 de diciembre de 2025.

[34] Este proyecto de ley incluiría los mismos siete principios que el RGPD a nivel federal y seis derechos de los interesados reconocidos en el RGPD, aunque el derecho de limitación de datos de carácter personal y el derecho a no ser objeto de decisiones automatizadas y a la elaboración de perfiles sólo se incluirían parcialmente.

la del RGPD a medio plazo[35], independientemente del acercamiento temporal de posturas derivado de la Transatlantic Data Privacy Framework[36].

1.2. El Camino de Santiago digital: la tecnología al servicio del ser humano

El Camino de Santiago constituye una red histórica de rutas de peregrinación cristianas que convergen en la Catedral de Santiago de Compostela, donde se consideran enterrados los restos del apóstol Santiago, y que ha atraído a peregrinos desde la Edad Media por motivos religiosos, culturales y personales.

No se trata de una única ruta, sino de un entramado de senderos que atraviesa diversos países de la Europa continental. En tanto red de itinerarios recorridos por las motivaciones referidas, el Camino de Santiago puede entenderse como una representación de una UE legislativamente fragmentada[37] y orientada por un compromiso ideológico con una economía digital centrada en el ser humano, con marcados componentes culturales.

En términos estrictamente regulatorios, la normativa europea ha sido la más influyente a nivel global y ha moldeado legislación de todas las regiones del mundo en mayor o menor medida. El término "efecto Bruselas[38]" se acuñó en referencia a este fenómeno.

La UE reconoce que los productos y servicios innovadores desarrollados por las grandes empresas tecnológicas generan beneficios sustanciales para las personas y las sociedades, motivo por el cual considera deseable promover su desarrollo. Esta valoración positiva convive con profundas preocupaciones en torno a la deriva estructural de la economía digital, caracterizada por una creciente concentración de poder económico y político en un número reducido de plataformas globales. Desde la perspectiva de la UE, una economía excesivamente concentrada facilita que determinadas empresas abusen de su posición dominante, restrinjan la competencia y perjudiquen tanto a competidores como a consumidores.

A partir de estas preocupaciones, la UE ha impulsado una intensa agenda regulatoria durante la última década. Como resultado, este proceso ha dado lugar a un entramado normativo que penaliza determinados modelos de negocio de las grandes empresas tecnológicas y refuerza las salvaguardias de los derechos fundamentales en el entorno digital.

Por otro lado, la protección de datos de carácter personal se proclamó como derecho fundamental en la Carta de los Derechos Fundamentales de la UE[39] (CDFUE, art. 8); asimismo, se consideró parte del contenido del derecho a la vida privada y familiar en el Convenio Europeo de Derechos Humanos y Libertades Fundamentales (CEDH, art. 8) sobre el que existe una abundante jurisprudencia del Tribunal Europeo de Derechos Humanos[40] (TEDH)[41].

Si bien la confianza y certeza jurídica en el tratamiento de datos se encuentran en la base reguladora de todo el entorno europeo, la adopción de un reglamento (el RGPD) resultaba fundamental. Por consiguiente, al establecer un cuerpo normativo común en todos los países de la UE, se permite la igualdad de condiciones necesarias para un mercado único digital[42]. En este sentido, la UE reconoce

[35] Parlamento Europeo, *Exchanges of Personal Data after the Schrems II Judgment* (Bruselas: Parlamento Europeo, 2021), https://www.europarl.europa.eu/RegData/etudes/STUD/2021/694678/IPOL_STU(2021)694678_EN.pd. Visitado el 20 de diciembre de 2025.

[36] U.S. Department of Commerce, "EU–U.S. Data Privacy Framework."

[37] Debido a su mercado único inconcluso.

[38] Concepto entendido como la capacidad unilateral de la UE para regular los mercados mundiales de forma no coercitiva, es decir, apoyándose únicamente de su capacidad de presión comercial. Eso se materializa estableciendo normas en diferentes materias como competencia, protección del medio ambiente o seguridad alimentaria. Estándares que sus colaboradores comerciales buscan voluntariamente incorporar en la medida de lo posible a sus legislaciones para garantizar la continuidad de las relaciones comerciales con la UE.

[39] Así como en algunas constituciones nacionales, como en el caso de la española (art. 18).

[40] La doctrina del TEDH ha establecido que el respeto a la vida privada y familiar resulta infringido cuando se accede o se utilizan datos de un individuo sin su consentimiento, salvo si este uso está amparado por la ley y, además, resulta necesario y proporcional al objetivo perseguido, y se ha garantizado al interesado la oportunidad de obtener revisión judicial de la actuación.

[41] Otero García-Castrillón. "Protección…", 38.

[42] Carrizo, "Reflexiones…", 462.

la dimensión económica del dato, por lo que el TJUE interpreta los mecanismos previstos para velar por el mantenimiento del estándar comunitario en las exportaciones de datos[43] y, en consecuencia, aborda su aplicación extraterritorial.

En el asunto *Salemink*[44], el abogado general afirmó que "a efectos de la UE, el territorio de los Estados miembros es el área (no necesariamente territorial, en el sentido espacial o geográfico) de ejercicio de las competencias de la Unión", lo que calificó la conexión entre el ejercicio de la soberanía y un territorio físico más próxima de una verdad contingente que necesaria[45]. Tras la Sentencia *Schrems I*[46], el supervisor europeo de protección de datos indicó que la normativa europea sobre transferencias internacionales de datos de carácter personal se basa en "un grado razonable de pragmatismo con el fin de permitir la interacción con otras partes del mundo". No obstante, se evidencia que algunos autores consideran que esta regulación no busca un equilibrio razonable entre los estándares propios y los de terceros países, sino la afirmación unilateral de los valores de la UE. Por tal motivo, es poco realista esperar soluciones internacionales sin que los demás Estados estén dispuestos a efectuar concesiones[47].

De hecho, numerosos países europeos, entre ellos España, no implementaron normas de privacidad hasta la etapa de negociación de la Directiva europea de protección de datos (DPD) en 1995. A partir de la DPD es cuando la protección de datos se convirtió en un requerimiento esencial para acceder al comercio interior que ofrece la UE. Otro ejemplo llamativo consiste en el caso de Polonia, que en 1997 (siete años antes de convertirse en Estado miembro) ya declaraba en su norma nacional de privacidad que se sometía a la DPD. En consecuencia, resulta evidente que el acceso a la UE requería conciliar con el *acquis* que, entre otros, incluía la normativa de privacidad. Igualmente, los propios reportes sobre el acceso a la UE demuestran un cambio a lo largo del tiempo en la visión de la UE sobre el rol de la protección de datos en la integración europea. Al respecto, se observa que originalmente el capítulo en el que se incorporaban las cuestiones de privacidad solía ser el de "Freedom to provide services". En consonancia con las últimas adhesiones, como la de Croacia, se evidencia un cambio de paradigma con su incorporación en la sección de derechos fundamentales[48].

En ese orden de ideas, plantear la norma de privacidad a todos los territorios de la UE ha sido un aspecto crítico del proyecto original de la creación del Mercado Digital Único. Ahora bien, la adopción de normativas de privacidad no ha sido el único objetivo europeo, sino también asegurar que las existentes en cada uno de los países que se han adherido cumplan con los estándares europeos.

Fuera de su territorio, la UE también ha utilizado su poder negociador en los acuerdos de libre comercio para persuadir a sus *partners* comerciales hacia su propia visión de la privacidad. Por consiguiente, es notable la progresiva incorporación en las negociaciones y acuerdos de libre comercio con países terceros, de referencias e incluso provisiones completas en materia de privacidad y protección de datos. Un buen ejemplo es el art. 30 del acuerdo de libre comercio entre la UE y Chile[49], que se encuadra dentro del bloque de cooperación económica y que se refiere a este aspecto de la privacidad como "facilitador del comercio".

En efecto, las decisiones de adecuación de la Comisión Europea cumplen un rol fundamental en la propagación de la perspectiva de la UE sobre la privacidad. Aunque las decisiones de adecuación funcionan en paralelo a los acuerdos comerciales entre las partes, dependen de estos acuerdos comerciales. Además, se presenta la particularidad de que las decisiones de adecuación se

[43] Otero García-Castrillón. "Protección…", 41.

[44] Conclusiones del Abogado General Pedro Cruz Villalón, presentadas el 8 de septiembre de 2011, asunto C-347/10, *A. Salemink v Raad van bestuur van het Uitvoeringsinstituut Werknemersverzekeringen*, ECLI:EU:C:2011:562, párrs. 54–57.

[45] Taylor Mistale, "Limits That Public International Law Poses on the European Union Safeguarding the Fundamental Right to Data Protection Extraterritorially," en T*ransatlantic Jurisdictional Conflicts in Data Protection Law: Fundamental Rights, Privacy and Extraterritoriality*, ed. Taylor Mistale (Cambridge: Cambridge University Press, 2023), 70.

[46] Tribunal de Justicia de la Unión Europea, asunto C-362/14, *Maximillian Schrems v Data Protection Commissioner*, sentencia de 6 de octubre de 2015 (*Schrems I*).

[47] Christopher Kuner, "Reality and Illusion in EU Data Transfer Regulation Post Schrems", *German Law Journal*, 18, n°. 4 (2017): 917.

[48] Diario Oficial de la Unión Europea (DOUE), L 112, 24 de abril de 2012. Visitado el 27 de diciembre de 2025.

[49] Unión Europea y República de Chile, *Agreement Establishing an Association between the European Community and Its Member States, of the One Part, and the Republic of Chile, of the Other Part*, 2002, https://eur-lex.europa.eu/resource.html?uri=cellar:f83a503c-fa20-4b3a-9535-f1074175eaf0.0004.02/doc_2&format=pdf. Visitado el 28 de diciembre de 2025.

reconocen generalmente de forma unilateral por parte de la UE tras la formalización de dichos acuerdos. De esta forma, se evidencia que, al negociar un acuerdo comercial con la UE, la contraparte solo tiene la capacidad de influir en los términos del propio acuerdo comercial en cuestión. De tal modo, la UE[50] siempre mantendrá la discrecionalidad y unilateralidad con respecto a la decisión de adecuación. Un ejemplo de lo anterior radica en el acuerdo de libre comercio entre la UE y el Reino Unido, donde no se estableció ninguna limitación temporal relevante. Ahora bien, se expuso que la decisión de adecuación que la UE reconoció posteriormente al Reino Unido incluía una limitación temporal de cuatro años. En consecuencia, independientemente de los términos del acuerdo comercial, se retenía esa capacidad unilateral de no renovar la decisión de adecuación en cuestión en 2025. En cualquier caso, la renovación ha tenido lugar satisfactoriamente por un periodo de seis años más.

A la luz de este contexto, queda patente que el efecto Bruselas está totalmente vigente en materia de privacidad, no solo por la extraterritorialidad del RGPD. De hecho, la privacidad y la protección de datos son probablemente de los ejemplos más prominentes de dicho efecto.

Sin embargo, para mantener su posición negociadora privilegiada, la UE aún tiene pendiente la tarea de equilibrar los conceptos de protección de datos y de proteccionismo de los datos, una dicotomía que sus acuerdos de libre comercio buscan abordar fundamentalmente mediante la incorporación de prohibiciones a la localización de datos, la facilitación de los flujos internacionales y la asunción de compromisos concretos, como los relativos a la minimización de datos. En cualquier caso, el acercamiento normativo resulta complejo, pues solo podría alcanzarse de manera parcial.

Incluso en aquellos casos aislados en los que dicho acercamiento fuera total desde un punto de vista formal -como ha ocurrido con la normativa de protección de datos de Andorra o en el caso de Brasil, que acaba de obtener el reconocimiento referido tras una década de espera-, ello no implica necesariamente una implementación idéntica. La mera adopción de una legislación prácticamente equivalente al RGPD no garantiza una interpretación consistente en la práctica, dadas las diferencias culturales respecto de la UE. De hecho, en la propia UE persisten múltiples divergencias entre los estándares de aplicación de las normativas nacionales y el RGPD, pese al alto grado de integración que este último pretendía alcanzar y sin perjuicio de su jerarquía normativa superior[51]. En efecto, Estados como Alemania -con su particular cruzada por el *cloud* nacional para el sector educativo- evidencian una tendencia al localismo que perjudica significativamente la consolidación del proyecto de Mercado Digital Único europeo.

La UE tiene en el contexto del Digital Ómnibus la oportunidad de reafirmar su objetivo que no es otro que garantizar que los estándares regulatorios europeos se conviertan en estándares de alcance global. Esto es, estándares que no sean meramente seguidos en el plano internacional, sino universalmente respetados, en la medida en que están dotados de la credibilidad que únicamente puede proporcionar un pragmatismo responsable.

1.3. La ruta de la seda digital: la tecnología al servicio del Estado

La Ruta de la Seda constituye uno de los pilares históricos más influyentes en la conformación de la identidad comercial y geopolítica de China. Desde la Antigüedad, este conjunto de rutas terrestres y marítimas articula un espacio de intercambio económico, tecnológico y cultural que permite a China proyectar su influencia a lo largo de Asia, Oriente Próximo, África y Europa. En la actualidad, la evolución del ecosistema digital chino y, en particular, la concepción jurídica del comercio digital y del tratamiento de los datos personales muestran una continuidad estructural con este legado histórico: China concibe el flujo de información como un recurso estratégico para asegurar la estabilidad interna, el desarrollo económico y la proyección internacional. En consecuencia, se debe explorar la comparativa entre ambos fenómenos —la Ruta de la Seda y el régimen chino del comercio digital— con el fin de identificar los elementos históricos, jurídicos y culturales que configuran la particular visión china sobre los servicios digitales basados en datos.

La Ruta de la Seda no constituye únicamente un corredor comercial, sino también un sistema de gobernanza. La Corte imperial establece estrictos mecanismos de control sobre las mercancías, la diplomacia y la movilidad, con el objetivo de garantizar la seguridad de las fronteras y regular

[50] Esta unilateralidad no solamente correspondería a la UE, sino a cualquier otro Estado que incorpore los mecanismos de decisiones de adecuación a su legislación. Con todo, en la actualidad ningún Estado ha optado por esta vía tras un reconocimiento de adecuación de la UE. En su lugar, la práctica es replicar ese reconocimiento por parte del Estado en cuestión para beneficiarse del intercambio comercial con la UE.

[51] Merece la pena destacar las constantes tensiones entre normas nacionales y autoridades de protección de datos europeas en cuestiones como la regulación del consentimiento para cookies, en materia de legitimación del tratamiento de datos biométricos o de la monitorización en el ámbito laboral.

la entrada y salida de conocimientos tecnológicos considerados estratégicos. Esta tradición de control estatal sobre los flujos comerciales y transacciones resulta clave para comprender la aproximación contemporánea de China al derecho del entorno digital. En contraste con los modelos liberalizantes de los EE.UU. o los enfoques garantistas de la UE, China adopta un esquema centralizado y normativamente robusto, en el que el Estado asume un papel preponderante en la regulación del comercio digital y en la gobernanza de los datos personales.

En el contexto actual, China estructura un marco jurídico orientado a reforzar su soberanía digital, en el que destacan la *Cybersecurity Law* (2017), la *Data Security Law* (2021) y la *Personal Information Protection Law* (PIPL, 2021). Aunque esta última presenta paralelismos formales con el RGPD de la UE, su lógica subyacente difiere sustancialmente: mientras que en Europa el foco se sitúa en la protección del individuo frente a los poderes públicos y privados, en China el énfasis recae en la protección del Estado y en la preservación del orden social. La información personal se concibe, de manera simultánea, como un recurso económico y como un activo estratégico cuya circulación debe alinearse con los objetivos nacionales de la misma forma que lo hacía el comercio de mercancías que transitaba por la Ruta de la Seda.

El paralelismo con la Ruta de la Seda resulta evidente, dado que, del mismo modo que en la Antigüedad el Imperio chino regulaba los intercambios para asegurar la estabilidad interna y la proyección internacional, en la actualidad China controla los flujos de datos con el fin de evitar riesgos geopolíticos, fomentar la innovación autónoma y garantizar la seguridad nacional. En el marco digital contemporáneo, la información se convierte en la nueva mercancía estratégica. Así como la seda y la pólvora desempeñan un papel central en el comercio imperial, los datos adquieren hoy una función clave en la competitividad de las empresas chinas y en la expansión global de sus plataformas digitales.

Asimismo, la Ruta de la Seda genera asimetrías de poder, en la medida en que China garantiza el acceso al comercio a aquellos territorios que aceptan sus normas y estructuras diplomáticas. En la esfera digital contemporánea se observa una dinámica similar: mediante la Iniciativa de la Franja y la Ruta Digital (*Digital Silk Road*), China exporta infraestructuras tecnológicas, estándares de conectividad y modelos de gobernanza digital que refuerzan su influencia global en ámbitos como las telecomunicaciones, la inteligencia artificial y la gestión de plataformas. En efecto, este proceso reproduce la lógica histórica de creación de espacios interdependientes bajo liderazgo chino.

No obstante, existen diferencias significativas entre ambos contextos. Mientras que la Ruta de la Seda ofrece un espacio de intercambio relativamente abierto, aunque sometido a supervisión imperial, el modelo digital chino contemporáneo prioriza la seguridad nacional por encima de la libertad individual o de la libre circulación de la información. La visión china del comercio digital se inscribe en una concepción más amplia de la "soberanía digital", en la que el control de las plataformas, la infraestructura y los datos resulta esencial para preservar la estabilidad política y la autosuficiencia tecnológica. Esta orientación contrasta con la dinámica transnacional y descentralizada del comercio que caracteriza, en gran medida, a la Ruta de la Seda tradicional.

El modelo regulatorio digital de China se fundamenta, en consecuencia, en un principio estructural: la primacía del control político del Estado y del Partido Comunista de China (PCCh) en la gobernanza del entorno digital. Desde principios del siglo XXI y, de manera más acentuada, bajo la presidencia de Xi Jinping, la estrategia digital china deja de concebir la tecnología como un espacio de autonomía social y pasa a integrarla explícitamente como un instrumento de cohesión política, estabilidad social y legitimidad estatal. Esta orientación se encuentra ampliamente documentada tanto en fuentes oficiales del gobierno chino como en la literatura académica comparada sobre autoritarismo digital[52].

Simultáneamente, China ha demostrado que un modelo de regulación fuertemente estatal y centralizado puede coexistir con una industria tecnológica dinámica, innovadora y de alcance global. Empresas como Alibaba, Tencent, ByteDance, Huawei o Xiaomi se han convertido en actores fundamentales para el crecimiento económico del país y para su ascenso geopolítico, situación que ha reforzado ante la ciudadanía la legitimidad del modelo de desarrollo liderado por el PCCh[53].

En este contexto, el modelo chino se configura como un contrapunto normativo frente a los enfoques occidentales dominantes previamente referidos: el modelo estadounidense orientado al mercado y el europeo centrado en la protección de derechos fundamentales. De tal modo, China

[52] Rogier Creemers, "China's Conception of Cyber Sovereignty: Rhetoric and Realization," en *Governing Cyberspace: Behavior, Power, and Diplomacy*, eds. Dennis Broeders y Bibi van den Berg (Lanham, MD: Rowman & Littlefield, 2020); Rebecca MacKinnon, *Consent of the Networked*: The Worldwide Struggle for Internet Freedom (New York: Basic Books, 2012).

[53] Adam Segal, *When China Rules the Web: Technology in Service of the State* (New York: Council on Foreign Relations, 2018); Gérard Roland y Nancy Qian, "The Evolution of China's Development Strategy," NBER Working Paper no. 29343 (Cambridge, MA: National Bureau of Economic Research, 2021), https://doi.org/10.3386/w29343

posee una larga tradición de restricción informativa y control de las comunicaciones. Desde la era imperial, las élites gobernantes temieron que la circulación sin control de ideas políticas pudiera generar contestación, un patrón que ha persistido en las distintas etapas del gobierno chino[54]. Tras la disponibilidad de acceso público a internet en 1995, el gobierno instauró el conocido "Gran Cortafuegos" (*Great Firewall*), una infraestructura técnico-regulatoria que combina filtrado, bloqueo de sitios *web* y mecanismos administrativos de censura.

Informes institucionales como los del Freedom House de 2022 y estudios académicos documentan cómo esta estructura limita el acceso a periódicos en línea y redes sociales[55]. En consecuencia, los ciudadanos solo pueden acceder mediante redes privadas virtuales (VPN por sus siglas en inglés), cuya utilización también está sujeta a regulaciones estrictas del Estado. Desde 2013, con la llegada de Xi Jinping al gobierno, el enfoque gubernamental hacia el ecosistema digital se ha radicalizado. Xi Jinping ha afirmado repetidamente que internet puede contener "energía negativa" capaz de socavar la "estabilidad social", razón por la cual el Estado debe "purificar" el espacio digital.

Bajo este marco conceptual, el gobierno ha aprobado tres leyes clave que institucionalizan el autoritarismo digital: a) la Ley de Seguridad Nacional de 2015, que reafirma la potestad del Estado para prevenir y sancionar cualquier actividad que amenace la unidad nacional, la soberanía o la estabilidad política; b) la Ley de Ciberseguridad previamente referida, que obliga a individuos y organizaciones a utilizar Internet, pero desde el respeto por el orden público y las "normas morales socialistas", además, establece los requisitos de localización de datos y controles sobre la infraestructura crítica; y c) la Ley de Seguridad de los Datos previamente referida, que restringe la transferencia internacional de datos y confiere al Estado autoridad para acceder y procesar información en nombre de la seguridad nacional. En consecuencia, estas normas[56] consolidan un marco en el que la tecnología queda jurídicamente subordinada al Estado.

En el plano comercial, la entrada de China en la OMC en 2001 propulsó sus relaciones comerciales con la UE y EEUU. China es parte del Regional Comprehensive Economic Partnership Agreement (RCEP), que contiene un capítulo específico sobre comercio electrónico con disposiciones relevantes sobre protección de la información personal, transferencias transfronterizas y localización de datos. También cuenta con acuerdos de libre comercio (TLC) bilaterales con disposiciones o capítulos de comercio electrónico con referencias a privacidad o protección de datos. En general, China tiene una red extensa de TLC bilaterales (por ejemplo, con Australia, Nueva Zelanda, Suiza, Islandia, Pakistán, Chile, Perú, Costa Rica, Georgia, Mauricio, Camboya, y acuerdos de alcance especial con Hong Kong y Macao, entre otros). En muchos de ellos, especialmente los más recientes o actualizados, aparecen capítulos de comercio electrónico, pero no todos incluyen reglas explícitas sobre flujos transfronterizos de datos, limitaciones de localización o un estándar detallado de privacidad comparable a otros modelos destacados en la región. Entre aquellos que sí incorporan reglas explícitas en este sentido destaca el TCL entre China y Corea del Sur. Este TLC incorpora un capítulo de comercio electrónico que contempla, de forma expresa, la protección de la información personal en el entorno del comercio electrónico en términos de cooperación o estímulo normativo. Hoy el comercio digital entre la UE y China continua al alza, pero está crecientemente condicionado por cuestiones regulatorias, geopolíticas y de soberanía tecnológica.

Retomando la cuestión sobre la subordinación tecnológica al Estado, la literatura contemporánea sobre autoritarismos digitales destaca que China no aplica una censura absoluta, sino una censura calibrada (*optimal censorship*) que permite cierto margen de expresión pública evitando así la coordinación de acciones colectivas[57]. En tal sentido, un informe de la Escuela Central del Partido de 2016 -institución clave en la formación del funcionariado del PCCh- explicita que un entorno informativo parcialmente abierto permite al gobierno detectar precozmente descontento social, anticipar problemas y "resolver eficazmente incidentes de opinión pública". Bajo este sistema de censura poroso, se advierte que parte de la población accede ocasionalmente al contenido sensible; empero, este nunca alcanza la masa crítica, lo que evita la movilización colectiva sin generar un rechazo social equivalente al que produciría una censura total.

Por su parte, China ha integrado tecnologías avanzadas, particularmente algoritmos de IA, sistemas de reconocimiento facial y mega-bases de datos, con el fin de construir uno de los sistemas de

[54] Elizabeth J. Perry, "Cultural Governance in Contemporary China," en *Routledge Handbook of Chinese Governance*, ed. Chris Ogden (London: Routledge, 2020).

[55] Gary King, Pan Jennifer y Roberts, Margaret E, "How Censorship in China Allows Government Criticism but Silences Collective Expression", *American Political Science Review*, 107 n°2 (2013): 326–343.

[56] Accesibles en las bases oficiales de legislación del Comité Permanente de la Asamblea Popular Nacional.

[57] Margaret E. Roberts, *Censored: Distraction and Diversion Inside China's Great Firewall* (Princeton: Princeton University Press, 2018).

vigilancia más extensos del mundo[58]. Cabe señalar que el país alberga 8 de las 10 ciudades más vigiladas del planeta[59], cuyo monitoreo se legitima oficialmente como herramienta para mantener la seguridad pública y la armonía social. No obstante, numerosos estudios académicos la han caracterizado como un elemento central del autoritarismo digital chino[60].

El sistema de crédito social, inspirado en diversos proyectos piloto y consolidado en documentos oficiales desde 2014, combina datos provenientes de registros policiales y judiciales, información fiscal y sanitaria, comportamiento financiero y actividad en línea. La doctrina ha evidenciado consistentemente[61] que este sistema pretende modelar comportamientos sociales a través de incentivos y sanciones, y se ha calificado internacionalmente como "orwelliano". Sin perjuicio de lo anterior, encuestas internas sugieren que un sector relevante de la población lo percibe como un mecanismo útil para aumentar la seguridad y reducir fraudes[62].

El modelo chino combina elementos propios del autoritarismo digital con rasgos inspirados en los modelos normativos occidentales. China ha manifestado preocupación por el uso excesivo y abusivo de datos personales por parte de las empresas privadas. En este contexto, la Ley de Protección de la Información Personal (*PIPL*, 2021) –también mencionada previamente– incorpora principios similares a los del RGPD de la UE. Entre estos principios compartidos se encuentran la limitación de la finalidad, la minimización de datos, el consentimiento informado, la prohibición de la discriminación algorítmica y los derechos de acceso, rectificación y supresión.

Asimismo, durante sus primeras décadas de expansión digital, China adoptaba un enfoque extremadamente laxo respecto de las empresas tecnológicas nacionales, lo que favorecía su crecimiento acelerado. Esta estrategia remite al "tecnoliberalismo" estadounidense de las décadas de 1990 y 2000. En este sentido, el origen del ecosistema de capital riesgo en China resulta inseparable de la inversión, la asesoría legal y la cultura empresarial importadas desde Silicon Valley[63]. Por ende, los vehículos societarios, incluidos los *offshore* en las Islas Caimán, utilizados para captar inversión extranjera y cotizar en mercados internacionales, se diseñaron por despachos de abogados estadounidenses.

Adicionalmente, la evolución del modelo digital chino no puede comprenderse sin considerar su política industrial proactiva, orientada a convertir a China en una superpotencia tecnológica autosuficiente. De tal modo, esta estrategia se inscribe dentro de un marco más amplio de tecno-nacionalismo, el cual se comprende como la convicción de que el liderazgo tecnológico constituye un requisito indispensable para la seguridad nacional, el desarrollo económico y la proyección geopolítica[64].

En ese orden de ideas, el Gobierno chino ha impulsado la innovación mediante subsidios estatales masivos, inversiones públicas directas, apoyo a empresas de propiedad estatal y mecanismos de planificación estratégica. De tal modo, estudios de la OCDE[65] y del Center for Strategic and International Studies (CSIS) documentan que este país destina cientos de miles de millones de euros a sectores considerados críticos[66].

Por su parte, el presidente Xi Jinping ha afirmado que la autosuficiencia tecnológica es condición indispensable para una "seguridad nacional integral"; en efecto, una noción que aparece

[58] Paul Mozur, "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras", *New York Times*, 8 de julio de 2018, https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html; James Leibold, "Surveillance in China's Xinjiang Region: Ethnic Sorting, Coercion, and Inducement", Journal of Contemporary China 29, no. 121 (2019): 46–60.

[59] "The World's Most Monitored Cities", *Comparitech*, 2019, Comparitech.com

[60] Xiao Qiang, "The Road to Digital Unfreedom: President Xi's Surveillance State", *Journal of Democracy* 30, no. 1 (2019): 53-67.

[61] Wei Dai, "Social Credit: The China Story, Part XIX", *The China Project*, 2019; Yawei Chen, Darrell M. West, y Xiaoqian Sun. "How China's Social Credit System Currently Works: Evidence from Shanghai and Credit China Data". Brookings Institution, 2021.

[62] Genia Kostka, "China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval", *New Media & Society* 21 no. 7(2019): 1565–1593.

[63] Sebastian Mallaby, *The Power Law: Venture Capital and the Making of the New Future* (New York: Penguin Press, 2022).

[64] Segal, "When China Rules...,".

[65] Organisation for Economic Co-operation and Development (OECD), *OECD Science, Technology and Innovation Outlook 2021: Times of Crisis and Opportunity* (Paris: OECD Publishing, 2021).

[66] IA, computación cuántica, robótica avanzada, semiconductores y telecomunicaciones de nueva generación (5G y 6G).

reiteradamente en documentos oficiales del PCCh.

Lanzado en 2015, *Made in China 2025* (MIC2025) es un plan industrial a diez años destinado a transformar a China en líder de la manufactura avanzada. En esa medida, el programa persigue reducir la dependencia de proveedores extranjeros, desarrollar capacidades nacionales en sectores de alto valor tecnológico, fomentar la exportación de productos tecnológicos chinos y apoyar la adquisición de empresas extranjeras con conocimientos especializados. Los subsidios estimados asociados a este programa superan los 300 000 millones de euros[67].

A pesar del rápido avance tecnológico, China mantiene una dependencia significativa respecto a insumos extranjeros clave, especialmente semiconductores. Por ejemplo, en 2019, el país generó aproximadamente el 60 % de la demanda global, pero solo produjo el 13 % de la oferta mundial[68]. En tal sentido, la respuesta oficial ha sido la de establecer un objetivo de autosuficiencia del 70 % para 2025, reforzado por medidas normativas y subsidios. Por lo tanto, se observa que la guerra tecnológica con los EE.UU. -incluyendo restricciones de las dos administraciones Trump y la Biden a la exportación de chips avanzados y equipos de litografía a China- ha intensificado esta estrategia.

Tras la crisis financiera global de 2008, el PCCh integró la innovación tecnológica como pilar de su legitimidad política y como instrumento de modernización económica. En tal sentido, numerosos estudios[69] plantean que la narrativa estatal vincula directamente el progreso tecnológico con el "rejuvenecimiento nacional", un concepto central en la ideología de Xi Jinping.

La estrategia china de desarrollo tecnológico se articula también a través de restricciones sistemáticas al acceso de empresas extranjeras, mecanismos de transferencia forzosa de tecnología y un marco de proteccionismo digital altamente sofisticado. Este proteccionismo digital opera bajo dos lógicas complementarias: por un lado, garantizar el control estatal sobre los flujos de información y, por otro, asegurar la seguridad política y el desarrollo económico.

Durante las décadas de 2000 y 2010, las grandes empresas tecnológicas chinas mantienen una relación simbiótica con el Estado, caracterizada por el apoyo regulatorio a cambio de cooperación en materia de vigilancia y censura. No obstante, este "pacto tácito" se ha visto profundamente alterado. Con posterioridad, la ofensiva regulatoria de China se ha dirigido principalmente a las empresas de software (plataformas digitales, *fintech*, comercio electrónico y juegos en línea), mientras que los sectores de *hardware* estratégico, como los semiconductores, las telecomunicaciones avanzadas o los vehículos eléctricos, quedan relativamente exentos de sanciones.

Esta selectividad pone de manifiesto que el objetivo estatal no consiste en frenar el desarrollo tecnológico, sino en reorientarlo hacia sectores considerados estratégicos para la seguridad nacional y en reducir el poder de intermediación informacional de las grandes plataformas[70].

Ahora bien, una diferencia estructural con Occidente es que ninguna empresa china ha desafiado públicamente al Estado. Asimismo, mientras que Google o Meta han litigado contra autoridades estadounidenses o europeas, empresas como Alibaba o Tencent han respondido a las sanciones millonarias recibidas con declaraciones de aceptación pública de sanciones, compromisos de cumplimiento normativo e incluso reorganizaciones internas alineadas con las prioridades estatales. En consecuencia, este comportamiento refleja tanto el diseño institucional autoritario como los fuertes incentivos económicos de alineación con el PCCh[71].

Así, el nuevo pacto tecnológico implica que el Estado controla el poder estructural del sector digital, y las plataformas aceptan su subordinación y contribuyen a objetivos de redistribución y estabilidad. A cambio, el Estado continúa apoyando sectores estratégicos (hardware, IA, semiconductores); aún así, no tolera concentración de poder que amenace al PCCh.

A pesar de su aparente coherencia, el modelo presenta paradojas significativas. En ese sentido, China demuestra que un ecosistema innovador puede surgir sin libertades políticas amplias, lo que

[67] Jost Wübbeke, Meissner Mirjam, Zenglein Max J., Ives Jaqueline y Conrad, Björn, *Made in China 2025: The Making of a High-Tech Superpower and Consequences for Industrial Countries* (Berlin: Mercator Institute for China Studies [MERICS], 2016).

[68] Semiconductor Industry Association, *State of the U.S. Semiconductor Industry 2020* (Washington, DC: Semiconductor Industry Association, 2020).

[69] Yongnian Zheng y Lance L. P. Gore, *Technological Innovation and China's Reform: The Dynamics of Change* (London: Routledge, 2022); Elizabeth C. Economy, *The Third Revolution: Xi Jinping and the New Chinese State* (New York: Oxford University Press, 2018).

[70] Segal, *When China Rules the Web*; Scott, *China's Uneven High-Tech Drive*.

[71] Francis Fukuyama, "What Kind of Regime Does China Have?" *Journal of Democracy* 31, no. 1 (2020): 5–20.

contradice la tesis liberal según la cual la innovación depende de un entorno abierto[72]. Sin embargo, quizás debe cuestionarse si esto sería sostenible solo en el corto y medio plazo.

De tal modo, China ha exportado sistemas de vigilancia y gestión de datos a numerosos países (especialmente en África, América Latina, Asia Central y Oriente Medio) a través de proyectos como *Safe Cities* de Huawei[73]. En contextos autoritarios, estos sistemas se utilizan para fortalecer capacidades de control político. Por consiguiente, este país aspira a que su modelo sea una opción atractiva para países con estructuras políticas centralizadas o con débiles instituciones democráticas. Su objetivo –en consonancia con el de la Ruta de la Seda– es exclusivamente introducir sus productos y servicios en dichos países, sin anhelar la exportación de sus cultura y sus valores al mundo.

7. Conclusiones

Desde 1973, prácticamente todos los Estados del mundo han promulgado leyes de protección de datos y privacidad a un ritmo medio de tres nuevas normas nacionales por año. En la actualidad, se encuentran en vigor más de 200 leyes nacionales en esta materia. No obstante, este ritmo legislativo acelerado apenas logra hacer frente a la velocidad creciente con la que evoluciona la tecnología y al consiguiente desarrollo de las oportunidades de negocio internacional.

Puede afirmarse que, en determinados ámbitos, las transacciones digitales superan ampliamente a las analógicas[74]. Así, el comercio digital se ha convertido en el tejido invisible de la economía moderna y de la vida cotidiana. En este contexto, parece evidente que una normativa nacional tradicional por sí sola no puede garantizar un nivel adecuado de protección de los derechos a la protección de datos y a la privacidad en las actividades de tratamiento que incluyen flujos transfronterizos de datos.

De entrada, el comercio y la privacidad constituyen materias dispares que deben mantenerse conceptualmente separadas, puesto que no resulta admisible someter un derecho humano a consideraciones económicas. No obstante, en lo que respecta al derecho a la protección de datos de carácter personal, aun cuando se reconoce como un derecho fundamental, dicha separación no resulta necesariamente evidente en todas las jurisdicciones. Esta situación se observa incluso en las dos principales jurisdicciones en esta materia dentro del marco de las operaciones comerciales —EE. UU. y la UE—. Ambas potencias protagonizan un desarrollo normativo creciente y dinámico en foros comerciales bilaterales y regionales, lo que puede abrir la puerta a ciertos procesos de armonización, sin que ello implique la desaparición de un entorno normativo fragmentado.

De tal modo, EE. UU. se posiciona como el primer exportador mundial de servicios digitales. La Ruta 66, inaugurada en 1926 y convertida en un símbolo cultural estadounidense, ha funcionado históricamente como un eje de movilidad, comercio y proyección identitaria. Esto representa la libertad individual, la expansión económica y el espíritu emprendedor que caracterizan la narrativa del progreso en ese país. De manera análoga, en la actualidad, el ecosistema digital estadounidense constituye una de las principales manifestaciones contemporáneas de ese mismo imaginario: un espacio de innovación acelerada, escasa intervención regulatoria y predominio del mercado como motor organizador. La Ruta 66 y el comercio digital estadounidense comparten, en este sentido, un rasgo definitorio: la centralidad del individuo como agente autónomo.

En el contexto de la Ruta 66, el viajero decidía su recorrido, asumía los riesgos y aprovechaba las oportunidades comerciales que surgían en su trayecto. En el entorno digital contemporáneo, el usuario se conceptualiza de manera similar, como un sujeto libre que decide qué servicios utilizar, qué datos compartir y con qué plataformas interactuar. Sin embargo, esta percepción omite la profunda asimetría informacional que caracteriza al comercio digital. Mientras que los viajeros de la Ruta 66 interactuaban con negocios locales en un entorno económico relativamente equilibrado, el usuario digital actual se enfrenta a empresas cuya capacidad de recopilación y análisis de datos supera ampliamente su entendimiento y su control individual.

Asimismo, este paralelismo se extiende a la forma en que EE. UU. conciben la regulación. La Ruta

[72] Yasheng Huang, *The Rise and Fall of the EAST: Examination, Autocracy, Stability, and Technology* (New Haven: Yale University Press, 2021).

[73] Mozur, "Inside China's Dystopian Dreams".

[74] Por ejemplo, mientras EE. UU. tiene un gran déficit comercial en bienes, tiene un superávit enorme en servicios, especialmente digitales (Déficit en bienes: > USD 900.000 millones, Superávit en servicios: ~ USD 250.000 millones, Superávit digital: ~ USD 240.000 millones). Es decir, los servicios digitales sostienen buena parte del equilibrio comercial de EE. UU.

66 operó durante décadas como un espacio en el que una regulación estatal mínima permitió un florecimiento económico basado en la iniciativa privada. El derecho digital estadounidense reproduce esta tradición, ya que, en lugar de imponer un marco normativo general comparable al de la UE, el país opta por un mosaico de leyes sectoriales (como la *HIPAA* en materia de datos de salud o la *COPPA* respecto de los menores) que fragmentan la protección de los datos personales. En este enfoque, el mercado, más que el Estado, actúa como principal regulador del flujo de información. Si bien este modelo resulta coherente con el *ethos* representado por la Ruta 66 y favorece la competencia y la innovación, sitúa al usuario en una posición de vulnerabilidad estructural.

Del mismo modo, la Ruta 66 funciona como símbolo de movilidad económica y social, al conectar pequeñas poblaciones con grandes mercados. De manera similar, el comercio digital permite que empresas pequeñas y emprendedores accedan a mercados globales mediante servicios basados en datos, posicionándose en un ecosistema de competencia ampliada. Ahora bien, así como la Ruta 66 termina drenando actividad económica hacia grandes conglomerados de autopistas y centros urbanos, el mercado digital estadounidense tiende a favorecer la concentración del poder económico en grandes plataformas tecnológicas, lo que genera nuevas jerarquías que cuestionan el ideal de igualdad de oportunidades.

Desde esta perspectiva, el análisis comparativo entre la Ruta 66 y la concepción estadounidense del derecho del comercio digital pone de manifiesto una continuidad cultural profunda. Ambos configuran espacios de movilidad —física o digital— sustentados en la libertad individual, la mínima intervención estatal y la primacía del mercado como estructura organizadora. Ahora bien, esta analogía permite identificar, al mismo tiempo, las tensiones inherentes a dicho modelo, entre las que destacan la creciente asimetría entre usuarios y empresas, la fragmentación normativa y la concentración del poder económico. Mientras la Ruta 66 simboliza un pasado de oportunidades abiertas, el ecosistema digital estadounidense representa un presente en el que dichas oportunidades coexisten con desafíos legales y éticos que exigen una reflexión más profunda sobre la naturaleza de los datos personales en una sociedad tecnológicamente avanzada.

En contraposición con la Ruta 66, desde la Edad Media el Camino de Santiago constituye un eje de movilidad, intercambio cultural y cooperación entre los territorios europeos. Las rutas jacobeas no solo conectan ciudades y reinos, sino que configuran una red estructurada que garantiza la seguridad, la hospitalidad y la atención al peregrino. Esta red incluye hospitales, albergues y sistemas de señalización que permiten un tránsito relativamente seguro y homogéneo a través de distintas regiones. En este contexto, el peregrino no se desplaza únicamente como individuo aislado, sino que lo hace dentro de un marco normativo y cultural que lo protege y lo integra en una comunidad más amplia. Un proceso similar se observa en la evolución del derecho europeo del comercio digital. Frente al crecimiento exponencial de los servicios digitales y a la centralidad de los datos personales como activo económico, la UE desarrolla un modelo regulatorio orientado a garantizar la protección del usuario y a asegurar un funcionamiento armonizado del mercado digital. Normativas como el RGPD, la DSA o la DMA buscan construir un espacio digital europeo en el que la persona permanece en el centro y en el que la libre circulación de servicios se equilibra con elevados niveles de garantías jurídicas.

En este sentido, el paralelismo más evidente entre ambos fenómenos se encuentra en su función protectora. Así como las rutas jacobeas establecen normas y estructuras destinadas a proteger al peregrino frente a abusos, peligros o situaciones de desamparo, el modelo europeo de regulación digital pretende salvaguardar al usuario frente a los riesgos derivados del tratamiento masivo de datos personales, las asimetrías de poder entre plataformas y consumidores, la opacidad algorítmica o las prácticas comerciales invasivas. Tanto en la peregrinación medieval como en el entorno digital contemporáneo, Europa se concibe a sí misma como garante de una travesía segura.

Además, ambos contextos comparten una clara vocación transfronteriza. El Camino de Santiago ha sido históricamente una red supranacional, capaz de conectar territorios diversos bajo un imaginario común. De manera análoga, la regulación digital europea se concibe para operar más allá de las fronteras estatales, al articular un mercado único de servicios digitales que requiere reglas uniformes para funcionar de manera adecuada. En este sentido, la existencia del denominado "efecto Bruselas" —entendido como la tendencia global a adoptar estándares europeos en materia de privacidad y regulación digital— refuerza la idea de que Europa, al igual que ocurrió con la construcción simbólica del Camino, continúa proyectando modelos normativos más allá de su propio territorio.

En esta línea, la UE, además del modelo de las decisiones de adecuación, ha exportado ampliamente el mecanismo de las cláusulas contractuales tipo a más de 80 Estados, incluidos algunos tan remotos como Arabia Saudí. Con todo, existe una notable disparidad entre los modelos de cláusulas elaborados por dichos Estados terceros, así como entre aquellos facilitados por organizaciones regionales como el Consejo de Europa, la Red Iberoamericana de Protección de Datos o la Asociación de Naciones de Asia Sudoriental (ASEAN). En consecuencia, puede afirmarse que este mecanismo se ha exportado fundamentalmente como concepto, más que como un modelo

normativo homogéneo, relegando su implementación en la práctica.
Para la seguridad y continuidad de las transacciones[75] resulta fundamental la compatibilidad entre los diferentes modelos de cláusulas contractuales tipo. Por lo tanto, la interoperabilidad es necesaria para evitar situaciones en las que los actores (responsables de tratamiento de datos) deben recurrir a distintas cláusulas en función de la transferencia de datos que vayan a realizar (así ocurre, por ejemplo, entre la UE y Asia).

En contraposición con los modelos anteriores -la Ruta 66 y el Camino de Santiago-, la Ruta de la Seda es comparable al enfoque chino ante la privacidad. Ambos sistemas comparten una estructura conceptual: tanto la Ruta de la Seda como la arquitectura digital contemporánea china son mecanismos para consolidar poder, asegurar la estabilidad interna y proyectar influencia internacional. En ese sentido, China entiende los servicios digitales basados en datos personales no solo desde una dimensión económica, sino como elementos esenciales de un nuevo orden global en el que la información reemplaza a las mercancías tradicionales como principal recurso estratégico.

La comparación entre la Ruta de la Seda y la concepción china del derecho del comercio digital pone de relieve una continuidad histórica en la manera en que China entiende la circulación de bienes (materiales o digitales) y su regulación. La gobernanza digital china contemporánea se sustenta en principios que evocan los mecanismos imperiales de control de los flujos comerciales, mediante un enfoque centralizado orientado a la seguridad nacional y a la proyección comercial internacional. En el siglo XXI, los datos personales se configuran como la nueva materia prima de un orden global en transformación y China, en consonancia con la tradición de la Ruta de la Seda, procura asegurar su posición como nodo central en este amplio sistema de intercambios. Por la Ruta de la Seda no transitaban peregrinos ni emprendedores en busca de un futuro mejor, transitaban mercaderes que exportaban sus mercancías al mundo. Del mismo modo, China –en contraste con EE.UU. y la UE– no impregna su estrategia digital de valores ni espiritualidad. Sus objetivos son estrictamente económicos.

Comprender las razones que subyacen a la regulación del entorno digital permite anticipar tanto el impacto de dicha regulación en el mercado como su viabilidad y su capacidad de adaptación en el entorno transfronterizo.

A partir de este escenario, la coexistencia de tres modelos regulatorios diferenciados (EE. UU., UE y China) está dando lugar a una creciente fragmentación global, a menudo descrita como *splinternet*, caracterizada por regímenes jurídicos incompatibles, estándares tecnológicos divergentes y un aumento de los controles geopolíticos sobre los flujos de datos.

Las iniciativas bilaterales y multilaterales podrían, paulatinamente, facilitar la aproximación hacia una regulación equilibrada entre la protección de los datos personales y la flexibilidad para la materialización de los flujos transfronterizos de estos. Al respecto, la OMC continúa negociando sobre "comercio electrónico"[76], mientras que la OCDE, el Consejo de Europa e incluso el Grupo de los Siete (G7) podrían aprovechar su experiencia negociadora en esta materia con el fin de apaciguar las tensiones en torno a la extraterritorialidad de la normativa estadounidense en materia de vigilancia gubernamental. Por el contrario, las tensiones derivadas de la percepción europea sobre la afectación negativa de dicha normativa a los derechos e intereses de los europeos –en determinadas circunstancias– derivan en su incompatibilidad con los términos en que se diseñaron las transferencias internacionales de datos en el RGPD.

En este marco, el modelo chino podría resultar menos problemático al ser coherente con su arquitectura política autoritaria, al priorizar la estabilidad y el control por encima de la libertad individual. La combinación de control estatal, planificación industrial y dinámicas de mercado regulado ha permitido a China consolidarse como una superpotencia tecnológica, desafiando la narrativa liberal que asocia de manera directa la libertad política con el progreso tecnológico. No obstante, las tensiones entre el control estatal y el dinamismo empresarial plantean interrogantes sobre la sostenibilidad de este modelo, especialmente en aquellos sectores que dependen de la innovación disruptiva. Asimismo, la proyección internacional del modelo chino –con sus recientes esfuerzos por flexibilizarse– está reconfigurando el orden digital global, al generar un escenario de competencia normativa con los EE.UU. y la UE. En ese sentido, el futuro del modelo depende de su capacidad para equilibrar seguridad, crecimiento e innovación, en un entorno marcado por una presión geopolítica creciente y una competencia tecnológica acelerada. Por su parte, informes del Banco Mundial de 2022 advierten que una regulación excesivamente punitiva puede reducir

[75] Véase Organisation for Economic Co-operation and Development (OECD), *Moving Forward on Data Free Flow with Trust: New Evidence and Analysis of Business Experiences*, OECD Digital Economy Papers, no. 353 (Paris: OECD Publishing, 2023), 15.

[76] Mozur, "Inside China's Dystopian Dreams".

la productividad en sectores dinámicos. No obstante, también resulta posible concebir sinergias entre los distintos esfuerzos multilaterales que permitan avanzar hacia una arquitectura global de flujos transfronterizos de datos más coherente.

Ante este panorama, EE. UU. y la UE han impulsado estrategias destinadas a frenar la proliferación del modelo chino. Por un lado, EE. UU. aplica la doctrina de la "reciprocidad digital", mediante restricciones a la exportación de tecnologías y la imposición de sanciones a empresas chinas. Por otro, la UE avanza en la construcción de su autonomía digital, fundamentada en la protección de los derechos fundamentales. Ambos actores conciben el modelo chino como un contrapeso normativo que cuestiona los pilares del orden digital liberal. En este sentido, los próximos años pondrán a prueba la voluntad de cooperación internacional en el ámbito de la regulación del comercio digital –y en particular de la economía del dato–. Hasta entonces, el escenario se caracteriza por una situación de *fragile trust*.

Bibliografía

Carrizo, David. "Reflexiones a propósito de la protección de datos en el escenario global digital: El derecho de daños en la litigiosidad internacional." *Revista Boliviana de Derecho*, no. 31 (2021): 445–472.

Chen, Yawei, Darrell M. West y Xiaoqian Sun. "How China's Social Credit System Currently Works: Evidence from Shanghai and Credit China Data." Brookings Institution, 2021.

Comisión Europea. "EU–US Data Privacy Framework: Questions and Answers on the Adequacy Decision." 13 de diciembre de 2022. https://ec.europa.eu/commission/presscorner/detail/de/qanda_22_7632.

Comisión Europea. "Servicios – estadísticas." Acceso el 8 de marzo de 2026. https://trade.ec.europa.eu/access-to-markets/en/content/services-statistics.

Comité Europeo de Protección de Datos (EDPB). *Recomendaciones 02/2020 sobre las garantías esenciales europeas para medidas de vigilancia*. Adoptadas el 10 de noviembre de 2020. https://www.edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-022020-european-essential-guarantees_es.

Creemers, Rogier. "China's Conception of Cyber Sovereignty: Rhetoric and Realization." En *Governing Cyberspace: Behavior, Power, and Diplomacy*, editado por Dennis Broeders y Bibi van den Berg. Lanham, MD: Rowman & Littlefield, 2020.

Cruz Villalón, Pedro. *Conclusiones del Abogado General en el asunto C-347/10, A. Salemink v Raad van bestuur van het Uitvoeringsinstituut Werknemersverzekeringen*. 8 de septiembre de 2011. ECLI:EU:C:2011:562.

Dai, Wei. "Social Credit: The China Story, Part XIX." *The China Project*. 2019.

Economy, Elizabeth C. *The Third Revolution: Xi Jinping and the New Chinese State*. New York: Oxford University Press, 2018.

Fukuyama, Francis. "What Kind of Regime Does China Have?" *Journal of Democracy* 31, no. 1 (2020): 5–20.

Huang, Yasheng. *The Rise and Fall of the EAST: Examination, Autocracy, Stability, and Technology*. New Haven: Yale University Press, 2021.

Jacques Delors Institute. *Europe in the World: Mapping the EU's Digital Trade—A Global Leader Hidden in Plain Sight?* París: Jacques Delors Institute, 2023.

Kennedy, Scott. *China's Uneven High-Tech Drive: Implications for the United States*. Washington, DC: Center for Strategic and International Studies, 2020.

King, Gary, Jennifer Pan y Margaret E. Roberts. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107, no. 2 (2013): 326–343.

Kostka, Genia. "China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval." *New Media & Society* 21, no. 7 (2019): 1565–1593.

Kuner, Christopher. "Reality and Illusion in EU Data Transfer Regulation Post Schrems." *German Law Journal* 18, no. 4 (2017): 881–918.

Leibold, James. "Surveillance in China's Xinjiang Region: Ethnic Sorting, Coercion, and Inducement." *Journal of Contemporary China* 29, no. 121 (2019): 46–60.

MacKinnon, Rebecca. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York: Basic Books, 2012.

Mallaby, Sebastian. *The Power Law: Venture Capital and the Making of the New Future*. New York: Penguin Press, 2022.

Mistale, Taylor. "Limits That Public International Law Poses on the European Union Safeguarding the Fundamental Right to Data Protection Extraterritorially." En *Transatlantic Jurisdictional Conflicts in Data Protection Law: Fundamental Rights, Privacy and Extraterritoriality*, editado por Taylor Mistale. Cambridge: Cambridge University Press, 2023.

Mozur, Paul. "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras." *New York Times*. 8 de julio de 2018. https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html.

NOYB – European Center for Digital Rights. "La Comisión Europea da un tercer asalto ante el TJUE a las transferencias de datos entre la UE y EE. UU." 10 de agosto de 2023. https://noyb.eu/es/european-commission-gives-eu-us-data-transfers-third-round-cjeu.

Organisation for Economic Co-operation and Development (OECD). *OECD Science, Technology and Innovation Outlook 2021: Times of Crisis and Opportunity*. Paris: OECD Publishing, 2021.

Organisation for Economic Co-operation and Development (OECD). "Declaration on Government Access to Personal Data Held by Private Sector Entities." *OECD Legal Instruments*. 13 de mayo de 2023. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0487.

Organisation for Economic Co-operation and Development (OECD). *Moving Forward on Data Free Flow with Trust: New Evidence and Analysis of Business Experiences*. OECD Digital Economy Papers, no. 353. Paris: OECD Publishing, 2023.

Organización Mundial del Comercio (OMC). *Acuerdo General sobre el Comercio de Servicios (AGCS)*. 1994.

Otero García-Castrillón, Carmen. *Protección de datos en la economía digital: Una aproximación desde la regulación del comercio internacional*. Pamplona: Thomson Reuters Aranzadi, 2021.

Parlamento Europeo. *Exchanges of Personal Data after the Schrems II Judgment*. Bruselas: Parlamento Europeo, 2021.

Parlamento Europeo y Consejo de la Unión Europea. *Reglamento (UE) 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento General de Protección de Datos)*. DOUE L 119, 4 de mayo de 2016.

Perry, Elizabeth J. "Cultural Governance in Contemporary China." En *Routledge Handbook of Chinese Governance*, editado por Chris Ogden. London: Routledge, 2020.

Qiang, Xiao. "The Road to Digital Unfreedom: President Xi's Surveillance State." *Journal of Democracy* 30, no. 1 (2019): 53–67.

Roberts, Margaret E. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton: Princeton University Press, 2018.

Roland, Gérard, y Nancy Qian. "The Evolution of China's Development Strategy." NBER Working Paper no. 29343. Cambridge, MA: National Bureau of Economic Research, 2021. https://doi.org/10.3386/w29343.

Schwartz, Paul M., y Karl-Nikolaus Peifer. "Structuring International Data Privacy Law." *International Data Privacy Law* 9, no. 1 (2019): 7–21.

Segal, Adam. *When China Rules the Web: Technology in Service of the State*. New York: Council on Foreign Relations Press, 2018.

Semiconductor Industry Association. *State of the U.S. Semiconductor Industry 2020*. Washington, DC: Semiconductor Industry Association, 2020.

The White House. "Fact Sheet: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitments on AI." 26 de julio de 2024.

The Transatlantic Economy 2021: Digital Acceleration. The Transatlantic Economy. 2021. https://transatlanticrelations.org/wp-content/uploads/2021/03/TA-economy-2021_CH4.pdf.

Tribunal de Justicia de la Unión Europea. *Data Protection Commissioner v Facebook Ireland Ltd y Maximillian Schrems*. Asunto C-311/18. Sentencia de 16 de julio de 2020 (*Schrems II*).

Tribunal de Justicia de la Unión Europea. *Maximillian Schrems v Data Protection Commissioner*. Asunto C-362/14. Sentencia de 6 de octubre de 2015 (*Schrems I*).

U.S. Department of Commerce. "EU–U.S. Data Privacy Framework." *Data Privacy Framework Program*. https://www.dataprivacyframework.gov/EU-US-Framework.

United States. *Communications Decency Act*. 47 U.S.C. § 230 (1996).

United States Congress. *Clarifying Lawful Overseas Use of Data Act (CLOUD Act)*. S.2383. 115th Cong. 2018.

Unión Europea. *Diario Oficial de la Unión Europea*. L 112. 24 de abril de 2012.

Unión Europea y República de Chile. *Agreement Establishing an Association between the European Community and Its Member States, of the One Part, and the Republic of Chile, of the Other Part*. 2002.

Wübbeke, Jost, Mirjam Meissner, Max J. Zenglein, Jaqueline Ives y Björn Conrad. *Made in China 2025: The Making of a High-Tech Superpower and Consequences for Industrial Countries*. Berlin: Mercator Institute for China Studies (MERICS), 2016.

Zheng, Yongnian, y Lance L. P. Gore. *Technological Innovation and China's Reform: The Dynamics of Change*. London: Routledge, 2022.

Tutela laboral ante la brecha digital: retos jurídicos ante la desigualdad tecnológica[1]

Labour Protection Against the Digital Gap: Legal Challenges to Technological Inequality

Manuel Jesús Garnica Corbacho
Investigador en formación
Universidad de Cádiz

Resumen:
En la digitalización del trabajo, la brecha digital emerge como una nueva forma de desigualdad en el mercado laboral español, diferenciando entre brechas de acceso, uso y aprovechamiento de las tecnologías. En consonancia con la desigualdad tecnológica laboral, se examinan sus manifestaciones en el acceso al empleo, la segmentación de ocupaciones y el teletrabajo, con especial incidencia en las desigualdades de género, edad, territorio y clase. Sobre esta base, se revisa el marco normativo internacional, europeo y español (Carta Social Europea, Estrategia de la Década Digital, Reglamento de IA, ET, LO 3/2018, Ley 10/2021, Ley 3/2023 y negociación colectiva) y se formulan propuestas de tutela centradas en la igualdad de oportunidades digitales, la transparencia algorítmica y la protección reforzada de colectivos vulnerables.

Palabras clave:
Brecha digital laboral; desigualdad tecnológica; teletrabajo; derechos digitales; inclusión digital.

Abstract:
In the digitalisation of employment, the digital divide emerges as a novel form of inequality within the Spanish labour market, distinguishing between gaps in access to, usage of, and effective exploitation of technologies. In line with the concept of technological inequality at work, the article examines its manifestations in access to employment, occupational segmentation, and teleworking arrangements, with particular emphasis on disparities arising from gender, age, territorial, and class-based factors. Upon this foundation, the international, European, and Spanish regulatory framework is reviewed (including the European Social Charter, Digital Decade Strategy, AI Regulation, Workers' Statute, Organic Law 3/2018, Law 10/2021, Law 3/2023, and collective bargaining) and remedial proposals are advanced, centring on equal digital opportunities, algorithmic transparency, and enhanced safeguards for vulnerable groups.

Keywords:
Labour digital divide; technological inequality; teleworking; digital rights; digital inclusion.

[1] El autor agradece el apoyo financiero del Programa Operativo FEDER Andalucía 2021-2027 y de la Consejería de Universidad, Investigación e Innovación de la Junta de Andalucía (Proyecto FEDER – UCA-2024-A2-15 "La eficacia de los derechos fundamentales en el ámbito laboral: nuevos retos para la práctica jurídica").

Summary:

1. *Introduction.* ***2.*** *Conceptual framework: digital divide and technological inequality at work.* ***2.1.*** *The multidimensional concept of the digital divide.* ***2.2.*** *Technological inequality in employment: definition and dimensions.* ***3.*** *Regulatory framework: labour protection against the digital divide.* ***3.1.*** *International and European context of protection.* ***3.1.1.*** *European Social Charter.* ***3.1.2.*** *Digital Decade Strategy.* ***3.1.3.*** *Artificial Intelligence Regulation.* ***3.2.*** *Spanish labour and equality regulations.* ***3.2.1.*** *Workers' Statute and equality protection.* ***3.2.2.*** *Law 10/2021 on remote work.* ***3.2.3.*** *Organic Law on Data Protection.* ***3.2.4.*** *Employment Law 3/2023.* ***3.2.5.*** *The role of collective bargaining.* ***4.*** *Manifestations of the digital divide in the Spanish labour market.* ***4.1.*** *Digital divide in access to employment and labour segmentation.* ***4.2.*** *Telework, disconnection and new inequalities.* ***5.*** *Protection instruments and reform proposals.* ***5.1.*** *Strengthening equal digital opportunities in employment.* ***5.2.*** *Enhancing transparency and governance of algorithmic systems.* ***5.3.*** *Specific guarantees for vulnerable groups.* ***6.*** *Conclusions.*

1. Introducción

La transformación digital de la economía y la sociedad constituye uno de los grandes retos del Derecho, con implicaciones profundas para el mundo del trabajo. La pandemia del Covid-19 aceleró exponencialmente los procesos de digitalización que anteriormente se desarrollaban de forma gradual, como el teletrabajo masivo, la automatización acelerada, la vigilancia o gestión algorítmica, plataformas digitales o la propia formación a distancia. Sin embargo, esta aceleración, lamentablemente, no se ha producido de manera uniforme[2]. La Estrategia Década Digital de la Unión Europea ha reconocido precisamente la brecha digital como uno de los retos más importantes que la sociedad actual presenta[3]. Este documento, entre sus objetivos, trata de "*colmar la brecha digital*" y reducirla, otorgando "*acceso a las tecnologías y datos digitales en condiciones abiertas, accesibles y justas*".

El concepto de brecha digital ha evolucionado desde su definición inicial, centrada en el acceso a la tecnología (brecha de acceso), hacia dimensiones más actuales con efectos sobre capacidades digitales (brecha de uso) y de beneficios (brecha de aprovechamiento). En el contexto laboral, esta evaluación es particularmente relevante, porque la brecha digital, es decir, trabajadores sin acceso a las tecnologías, que desconocen cómo funcionan o no la aprovechan lo suficiente, se traduce directamente en desigualdad de acceso al empleo, condiciones de trabajo, de oportunidades en la carrera profesional y de exposición a nuevos riesgos, como la algoritmización, automatización o sesgos discriminatorios. La profesora Ordiales Hurtado la define como "*las diferencias en el acceso y uso de las tecnologías digitales*" presentando un carácter multidimensional[4].

El ordenamiento jurídico español ha ido incorporando, de forma gradual, algunos elementos de tutela y protección frente a los riesgos de la digitalización en el ámbito del trabajo, en particular, la Ley Orgánica de Protección de Datos (LOPDGDD) con énfasis en derechos digitales laborales, la

[2] Lucía Aragüez Valenzuela. "Desafíos de la digitalización de las relaciones laborales: algoritmos digitales, robotización y trabajo a distancia", *e-Revista Internacional de la Protección Social*, no. 1, (2022), p. 9

[3] Unión Europea. *Década Digital de Europa. Decisión (UE) 2022/2481 del Parlamento Europeo y del Consejo de 14 de diciembre de 2022 por la que se establece el programa estratégico de la Década Digital para 2030 (Texto pertinente a efectos del EEE)*. Diario Oficial de la Unión Europea, 19 de diciembre de 2022, L 323/4.

[4] Inmaculada Ordiales Hurtado. "Brechas digitales en España y su relación con el mercado de trabajo". *Hipatia. Observatorio de las Ocupaciones*. Recuperado de https://www.sepe.es/HomeSepe/es/que-es-observatorio/Hipatia/cuadernos-mercado-trabajo/revista-cuadernos-mercado-trabajo/detalle-articulo.html?detail=/revista/Cuarta-revoluci-n-industrial-y-su-impacto-en-el-mercado-laboral-y-la-formaci-n/brechasdigitalesenespaaysurelacionconelmercadodetrabajo

Ley 10/2021 de trabajo a distancia, que viene a intentar democratizar el teletrabajo, o la Ley 3/2023 de Empleo, que enfatiza la digitalización inclusiva. Esto implica la adopción de unas medidas hacia la supervisión humana de la Inteligencia Artificial (IA), transparencia algorítmica o regulación del trabajo en plataformas digitales. Son instrumentos que, aunque valiosos, no han sido pensados de forma sistemática para paliar el desafío de la desigualdad tecnológica, pero que se configuran como elementos imprescindibles en su erradicación. Por ello resulta preciso analizar este reto desde el prisma de un nuevo modelo de desigualdad laboral.

2. Marco conceptual: brecha digital y desigualdad tecnológica en el trabajo

2.1. El concepto multidimensional de brecha digital

El término brecha digital tuvo su origen en Francia, entre los años 70 y 80, a raíz del proyecto Minitel, que tenía como intención la digitalización de las guías telefónicas, donde la empresa estudió si los terminales se debían proporcionar, sin coste alguno, a los usuarios de la red telefónica, o si tendría que abonarse alguna contraprestación. En la diferencia entre las personas que podían o no pagar dicho precio estaba el origen de la brecha digital[5]. Se trata, por lo tanto, de la desigualdad entre unas personas y otras de acceder a las herramientas TICs, cuestión que en Derecho del Trabajo resultará esencial para el acceso y/o mantenimiento a ciertos empleos, en especial en la actual era de la digitalización que ha transformado las relaciones laborales.

La profesora Olarte Encabo ha entendido, ciertamente, que la brecha digital es el reflejo de unas desigualdades sociales preexistentes, que así mismo demuestran cómo las medidas adoptadas, es decir, políticas públicas de inversión tecnológica y promoción de acceso digital, "*no son suficientes*" y que se precisan "*medidas más amplias e integrales que ataque la causa originaria de la desigualdad, la causa de la pobreza de estos colectivos*", incluyendo medidas de acción positiva, más allá de medidas digitales indiscriminadas[6].

Precisamente, podemos hablar de tres dimensiones derivadas, la primera brecha, sobre acceso material, una segunda sobre capacidades y destrezas, y la tercera sobre usos y beneficiarios. La primera brecha, de acceso material, hará referencia a la disponibilidad de una infraestructura tecnológica, dispositivos y conectividad, es decir, la posibilidad o no de que un trabajador disponga de tales dispositivos o herramientas TICs. Esta primera es una de las brechas más bajas que estudiamos, pues estudios como la Encuesta sobre Equipamiento y uso de las TIC en los Hogares del INE ha revelado que cerca de 16 millones de hogares en España ya disponían de acceso a internet en 2022, lo que corresponde a un 95,9% del total de hogares, y unos 14 millones tenían dispositivos digitales, en torno al 83% de la población. En 2025, los datos han mejorado todavía más, con el 97,4% de hogares con conexión de banda ancha.

La segunda brecha trata las capacidades y destrezas, es decir, el uso de las competencias digitales que posee la población. Una vez la población tiene acceso a las tecnologías, debe tener conocimientos que les permita usar las herramientas, comprender los interfaces, navegar por las plataformas o inclusive resolver los problemas técnicos que se planteen. El INE, en su consulta sobre utilización de productos TIC por las personas entre 16 y 74 años, ha mostrado que un 19,1% alega dominar competencias digitales básicas, un 31,4% de habilidades bajas, un 41,1% de habilidades avanzadas y solo un 1,6% de los encuestados declaran no poseer ninguna habilidad. En este sentido, la UE a través del Programa Europa Digital encabeza un fuerte compromiso e inversión para la capacitación, perfeccionamiento y reciclaje profesional con la finalidad de la mejora de las capacidades digitales de los trabajadores europeos. No obstante, el Derecho debe estar en la protección de los más vulnerables, aunque represente porcentajes mínimos de la realidad social.

Finalmente, la tercera de las brechas indica que no todas las personas que acceden a la tecnología logran los mismos beneficios, tratándose de una brecha de aprovechamiento. Estudios como el de la Fundació Ferrer y Guàrdia de 2022, concluyen que, a menor edad, mayor nivel competencial, aumentando el aprovechamiento, y en especial, a mayor nivel de ingresos, también mayor aprovechamiento . Esta brecha refleja frecuentemente factores como el idioma, la alfabetización general, la confianza en la tecnología o el entorno social. Un mayor aprovechamiento[7]de las

[5] Ana María Martín Romero. "La brecha digital generacional", *Temas Laborales*, no. 151, (2020), p. 79.

[6] Sofía Olarte Encabo. "Brecha digital, pobreza y exclusión social", *Temas Laborales*, no. 136, (2017), p. 293.

[7] Fundació Ferrera i Guàrdia. *La brecha digital en España: conocimiento clave para la promoción de la inclusión digital.* (2022).

tecnologías permite un mayor control de las herramientas digitales en determinados empleos, mientras que el que no la tiene, queda excluido del mercado laboral. Por lo tanto, la brecha de tercer nivel viene establecida por un uso productivo y eficiente de internet[8].

En el contexto del trabajo, estas tres dimensiones o brechas, la de acceso, uso y aprovechamiento, operan de forma conjunta. Por ejemplo, un trabajador sin acceso a su ordenador personal en casa quedará excluido de ofertas de teletrabajo. Otro con acceso, pero sin competencias en herramientas específicas, quedará también excluido en determinados procesos de selección. Un tercero, con acceso y ciertas competencias, pero sin dominio de plataformas complejas como Excel, verá limitadas sus oportunidades de promoción. Ante esto, la inclusión digital debe emerger como un auténtico derecho universal, con su traslado al ámbito digital, que garantice una verdadera integración de la población en la era de la digitalización[9].

2.2. Desigualdad tecnológica laboral: definición y dimensiones

La desigualdad tecnológica en el empleo podemos definirla como el conjunto de situaciones de desfavorecimiento estructural que experimentan determinados trabajadores o colectivos laborales como consecuencia de la digitalización de procesos, acceso desigual a capacidades digitales, y/o exposición a sistemas de gestión algorítmica, que generan efectos discriminatorios, excluyentes o degradantes en el acceso, mantenimiento o calidad del empleo, sin necesidad de que exista intención discriminatoria explícita.

Esta definición es deliberadamente amplia porque captura un fenómeno que trasciende la discriminación clásica. Una mujer no es discriminada por razón de sexo en sentido estricto si queda fuera del teletrabajo porque no tiene acceso a banda ancha de calidad (efecto de pobreza digital), pero sí experimenta una desigualdad tecnológica cuando, por defectos en la capacitación en materia de teletrabajo ofrecida por la empresa, tiene sobrecarga de cuidados que la penaliza en permanencia laboral en comparación con hombres que sí reciben esa capacitación[10]. Concretamente, el profesor Jiménez Vargas se refiere a la brecha tecnológica como la distinción entre el número de hombres y mujeres que usan internet, cuyas causas se encuentran en la educación, así como en "*roles que han relegado a las mujeres a trabajos menos cualificados y poco remunerados que en gran medida están relacionados con tareas domésticas y de cuidados*".

La desigualdad tecnológica en el empleo viene analizándose desde varias dimensiones. En primer lugar, el acceso desigual al empleo digital, pues la digitalización de los procesos de selección, por medio de plataformas y algoritmos, excluyen a trabajadores sin competencias digitales y se crea segmentación laboral. Por otro lado, el teletrabajo es desigual, porque consolida las desigualdades y roles de género, con la sobrecarga de cuidados, por la edad y sus dificultades técnicas, por cuestiones territoriales (y la consecuente falta de conectividad rural) y de clase, con trabajadores precarios sin un derecho al teletrabajo de calidad. También, existen riesgos ante la discriminación algorítmica, a través de los sesgos algorítmicos derivados de las variables proxy y la codificación de los algoritmos[11]. Esta discriminación amplifica la vulnerabilidad de colectivos vulnerables como las personas con discapacidad, de edad avanzada o mujeres, entre otros[12].

3. Marco normativo: tutela laboral frente a la brecha

3.1. Contexto internacional y europeo de protección

3.1.1. Carta Social Europea

La Carta Social Europea (CSE), revisada en 1996, constituye uno de los tratados fundamentales de garantía de derechos sociales de Europa. En su articulado reconoce el derecho a trabajar y a ganarse la vida con libertad, junto a la tutela contra la pobreza y exclusión social, derecho a la seguridad social, vivienda, protección social o asistencia social. Este instrumento del Consejo

[8] Ana María Martín Romero. "La brecha digital generacional", *Temas Laborales*, no. 151, (2020), p. 79.

[9] Sofía Olarte Encabo. "Brecha digital, pobreza y exclusión social", *Temas Laborales*, no. 136, (2017), p. 293.

[10] Fundació Ferrera i Guàrdia. *La brecha digital en España: conocimiento clave para la promoción de la inclusión digital.* (2022).

[11] Karen Hao. "Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo", *MIT Technology Review*, (2019).

[12] Mike Mullane. "La eliminación de los sesgos en los algoritmos", *Revista UNE*, no. 11, (2021), p. 3.

de Europa, ratificado por España, puede interpretarse con el conjunto del ordenamiento jurídico español, así como del Derecho de la Unión Europea.

En el Derecho de la Unión Europea, el Pilar Europeo de Derechos Sociales reconoce la igualdad de oportunidades y de acceso al mercado de trabajo, que posibilidad el derecho de toda persona a "*una educación, formación y aprendizaje permanente inclusivos y de calidad, a fin de mantener y adquirir capacidades que les permitan participar plenamente en la sociedad y gestionar con éxito las transiciones en el mercado laboral*", una mención que cobra especial relevancia en el actual contexto de la digitalización y brecha digital en el empleo, implicando el derecho de todo trabajador a recibir una formación que le recapacite y lo forme, también en la nueva era de la industria 4.0. Se trata pues de un acceso a la formación y a la obtención de competencias, como derecho social fundamental, en contextos de transformación digital.

3.1.2. Estrategia de la Década Digital

La Estrategia de la Década Digital o Década Digital Europea es un programa estratégico de la UE que tiene entre sus objetivos (art. 3) el "*reforzar la resiliencia colectiva de los Estados miembros y colmar la brecha digital*" junto a "*promover la implantación y el uso de las capacidades digitales con vistas a reducir la brecha digital geográfica y otorgar acceso a las tecnologías y datos digitales en condiciones abiertas, accesibles y justas*", con la finalidad de obtener una innovación tecnológica en las empresas europeas, desarrollando un ecosistema global y sostenibles de las infraestructuras digitales.

Para tales propósitos, existen algunos componentes interdependientes esenciales para la consecución, como la digitalización de los servicios públicos (administración, justicia o sanidad), la transformación digital de la economía o las infraestructuras digitales, mediante la conectividad de banda ancha de ultra-alta velocidad universal[13].

3.1.3. Reglamento de Inteligencia Artificial

El 20 de octubre de 2020, el Parlamento Europeo aprobó una Resolución con "*recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas*", un documento que, en consonancia con el nuevo Reglamento 2024/1689 sobre IA, va a constituir un marco de referencia basado en el principio de la dignidad humana, así como la igual autonomía de todos los ciudadanos y el respeto a los derechos fundamentales, en especial, el derecho a la no discriminación. En este sentido, todo sistema de IA debe respetar la Carta de Derechos Fundamentales de la UE (CDFUE) y el acervo comunitario, como la Directiva 2000/78/CE sobre igualdad en el empleo que recoge la prohibición de las conductas discriminatorias, directas como indirectas, en el empleo, donde encaja la brecha digital.

El RIA viene a establecer el conjunto de normas armonizadas en materia de IA y contiene una exhaustiva regulación basada en el riesgo, donde se clasifican los sistemas en cuatro categorías, el riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo. Esto implica un método que va modulando las obligaciones y los requisitos exigibles en base a su posible impacto, en los derechos fundamentales[14].

En cuanto a los efectos que tiene sobre la brecha digital, el Reglamento, a lo largo de su articulado, y previamente en sus considerandos, recoge el principio de no discriminación, junto a otros, como la supervisión humana, transparencia, diversidad, rendición de cuentas o bienestar social, y trata de, conforme a las Directrices éticas para una IA fiable de 2019, contribuir "*al diseño de una IA coherente, fiable y centrada en el ser humano*" (Considerando 27). En particular, define la diversidad, no discriminación y equidad como la obligación de que los sistemas de IA se desarrollen y utilicen "*de un modo que incluya a diversos agentes*" y que promueva "*la igualdad de acceso, la igualdad de género y la diversidad cultural*" mientras que se evitan los efectos discriminatorios y los sesgos. El Considerando 28 vuelve a poner el énfasis en el respeto a los valores y derechos fundamentales de la UE, entre ellos, la no discriminación, protección de datos o intimidad[15]. Este especial interés de la UE en torno a la protección de los derechos fundamentales va a resultar decisorio, clasificando como de alto riesgo todo sistema con efectos adversos en los derechos protegidos por la CDFUE.

[13] Juan Miguel Márquez Fernández. "El desarrollo digital de España en el marco de la Estrategia Digital Europea", *ICE, Revista de Economía*, no. 938, (2025), p. 18.

[14] Inés Delgado López "El impacto de los sistemas algorítmicos en los procesos de selección de personal. Análisis jurídico-laboral a la luz del nuevo Reglamento europeo en materia de inteligencia artificial", *Revista de Trabajo y Seguridad Social. CEF*, no. 483, (2024), p. 53.

[15] Jesús Mercader Uguina y Belén Velasco Pardo "El impacto laboral del reglamento de inteligencia artificial", *Revista Derecho Social y Empresa*, no. 23, (2025), p. 111.

Entre ellos, el Considerando 48 vuelve a nombrar la no discriminación o el derecho a la educación. En el ámbito laboral, conforme al Considerando 57, deben clasificarse como de alto riesgo los sistemas que se utilicen en el empleo, gestión de trabajadores y acceso al autoempleo, concretamente, para la "*contratación y la selección de personal, para la toma de decisiones que afecten a las condiciones de las relaciones de índole laboral, la promoción y la rescisión de relaciones contractuales de índole laboral*" y así se concreta en el Anexo III y el art. 6.2.

3.2. Normativa española laboral y de igualdad

3.2.1. Estatuto de los Trabajadores y tutela de la igualdad

El Estatuto de los Trabajadores es la piedra angular sobre la que se asienta la tutela del derecho al empleo en igualdad de condiciones. El art. 17 entronca el principio de no discriminación en las relaciones laborales, siendo nulos y sin efecto "*los preceptos reglamentarios, las cláusulas de los convenios colectivos, los pactos individuales y las decisiones unilaterales del empresario que den lugar en el empleo*" e igual que sobre la jornada, retribución u otras condiciones, ante situaciones discriminatorias de forma directa o indirecta, y que sean desfavorables "*por razón de edad o discapacidad o a situaciones de discriminación directa o indirecta por razón de sexo, origen, incluido el racial o étnico, estado civil, condición social, religión o convicciones, ideas políticas, orientación e identidad sexual, expresión de género, características sexuales, adhesión o no a sindicatos y a sus acuerdos, vínculos de parentesco con personas pertenecientes a o relacionadas con la empresa y lengua dentro del Estado español*". Igualmente, el art. 4.2 c) reconoce el derecho de las personas trabajadoras "*a no ser discriminadas directa o indirectamente para el empleo*" por los mismos motivos.

La discriminación puede ser directa o indirecta[16]. Directa cuando, de forma general, una persona sufre un trato menos favorable que otra que se encuentra en la misma o análoga situación, por el motivo que sea, sexual, orientación, identidad, origen racial, discapacidad, estado civil, etc. Será indirecta la discriminación provocada por una situación que, aparentemente neutra, ocasiona una desventaja particular a una persona. Autores como Rubio Sánchez la definen como una "discriminación camuflada" la cual "no se puede comprobar directamente y por lo general es más difícil de probar"[17]. A estos efectos, es importante la aplicación, en complemento con el ET, de normas como la LO 3/2007 de Igualdad Efectiva de Mujeres y Hombres, o la LO 15/2022 de garantía integral de la libertad sexual, ambas con responsabilidades específicas ante la no discriminación por razón de sexo[18].

3.2.2. Ley 10/2021 de Trabajo a Distancia

La Ley 10/2021, de 9 de julio, de trabajo a distancia, constituye un avance regulatorio importante, especialmente post-pandemia, con el incremento del número de personas trabajando a distancia[19]. El texto normativo establece un marco relevante sobre los derechos de los teletrabajadores y las obligaciones empresariales.

Sobre los derechos, recogidos en el Capítulo III, se trata de garantizar de forma específica algunos derechos ya integrados en el ET, como el derecho a la formación (art. 9) que implica que "*las empresas deberán adoptar las medidas necesarias para garantizar la participación efectiva en las acciones formativas de las personas que trabajan a distancia*", es decir, dotar de formación a los trabajadores para que conozcan cómo funciona el teletrabajo y puedan desempeñarlo, reduciendo la brecha digital, tanto de uso como de aprovechamiento. Especialmente, la empresa tendrá la obligación de garantizar una "*formación necesaria para el adecuado desarrollo de su actividad tanto en el momento de formalizar el acuerdo" así como "cuando se produzcan cambios en los medios o tecnologías utilizadas*". Del mismo modo, el derecho a la promoción profesional (art. 10) en los mismos términos que los trabajadores presenciales.

El trabajo a distancia debe ser recompensado, y así lo indica el derecho a la dotación suficiente y mantenimiento de medios, equipos y herramientas, garantizando la reducción de la brecha digital de acceso, puesto que las personas trabajadoras pueden solicitar la dotación, al igual que el

[16] Eduardo Caamaño Rojo. "La discriminación laboral indirecta", *Revista de derecho (Valdivia)*, Vol. 12, no. 2, (2001), p. 70.

[17] Francisco Rubio Sánchez, F. "Discriminación indirecta" en *Un decenio de jurisprudencia laboral sobre la Ley de igualdad entre mujeres y hombres*, ed. por Carmen Sánchez Trigueros, (2018), p. 201.

[18] Fernando Elorza Guerrero. "Despido y prueba de la discriminación indirecta por razón de sexo", *Temas Laborales*, no. 103, (2010), p. 251.

[19] Cristina Ayala del Pino. "La nueva regulación del trabajo a distancia no es la panacea", *Anuario Jurídico y Económico Escurialense*, LV, (2022), p. 117.

mantenimiento adecuado para "*los medios, equipos y herramientas necesarios para el desarrollo de la actividad*". Inclusive, este art. 11, indica que se deberá de atender de forma precisa a los trabajadores en caso de dificultades técnicas, tratándose de una medida para evadir la brecha digital de uso o inclusive de aprovechamiento. Estos derechos, junto a la protección de privacidad, desconexión digital, los gastos de conexión, supervisión de carga o participación en la vida laboral de la empresa conforman un verdadero programa de tutela y protección ante las nuevas formas de trabajo.

Lamentablemente, existen deficiencias que afectan a la brecha digital, como la formación obligatoria, frente al actual modelo potestativo, así como discriminaciones invisibles por discapacidad, y es que a pesar de que se establece la accesibilidad en el trabajo, no hay una obligación explícita de adaptación tecnológica, con los costes que supone, a través de softwares de accesibilidad u otras herramientas de asistencia. Esto se une a la brecha de género en teletrabajo, que queda sin abordar. Aunque en la teoría el teletrabajo puede mejorar la conciliación, numerosos estudios demuestran que ante la inexistencia de redistribución de cuidados y de formación específica, se amplifica la histórica desigualdad de género, recayendo sobre las mujeres el rol presencial de cuidados y casa[20]. La Radiografía del Teletrabajo en España, reconoce que el teletrabajo es más frecuente entre mujeres (con un 17% de las ocupadas) frente a los hombres (un 14%)[21].

3.2.3. Ley Orgánica de Protección de Datos

La Ley Orgánica 3/2018 de Protección de Datos Personales y Garantía de Derechos Digitales (LOPDGDD) y el Reglamento General de Protección de Datos (RGPD, aplicable directamente) establecen derechos digitales de los trabajadores respecto a sus datos personales: derecho de acceso, rectificación, supresión, portabilidad, y oposición a decisiones basadas únicamente en tratamiento automatizado (artículo 22 RGPD, artículo 18 LOPD)[22].

Los principales derechos en el ámbito laboral quedan recogidos en los arts. 87 y ss. El art. 87 trata el derecho a la intimidad y uso de dispositivos digitales en el ámbito laboral, el art. 88 sobre desconexión digital, art. 89 sobre intimidad frente a la videovigilancia y grabación de sonidos en el lugar de trabajo, el art. 90 sobre derecho a la intimidad ante uso de sistemas de geolocalización o el art. 91 de derechos digitales en la negociación colectiva. Conforma una base protectora de tutela basada en la previa información al trabajador, el respeto a sus derechos fundamentales, como intimidad o privacidad, y la consagración jurisprudencial posterior del principio de proporcionalidad aplicado ante estos fenómenos laborales de digitalización. En lo que respecta a la brecha digital, el art. 80 trata el derecho a la neutralidad de internet, por el que los proveedores deben proporcionar una oferta transparente de servicios "*sin discriminación por motivos técnicos o económicos*" así como el art. 81 sobre acceso universal, independientemente de la "*condición personal, social, económica o geográfica*" y que debe ser además "*asequible, de calidad y no discriminatorio*". En todo caso, el acceso a internet, de hombres y mujeres, tendrá que superar tanto la brecha de género, "*en el ámbito personal como laboral*" y la "*brecha generacional mediante acciones dirigidas a la formación y el acceso a las personas mayores*". Inclusive, garantizando un acceso efectivo a entornos rurales y en condiciones de igualdad a personas con necesidades especiales[23].

3.2.4. Ley 3/2023 de Empleo

Ley 3/2023, de 28 de febrero, de Empleo recoge reformas importantes sobre digitalización inclusiva del empleo. Precisamente, reconoce la transformación digital como eje transversal de políticas de empleo[24]. En el preámbulo, indica que "*la mejora de las habilidades de comunicación oral y escrita, así como de manejo útil de herramientas digitales y tecnológicas se configuran por esta ley como competencias básicas para la empleabilidad*". También se enfatiza la inclusión digital en el acceso a empleo, especialmente para colectivos vulnerables. Se refleja la "evitación de discriminaciones y estereotipos de cualquier índole" en el diseño como implementación de las políticas. La profesora Garrido Pérez recuerda que la norma, entre otros propósitos, responde a las "nuevas realidades

[20] María Belén Fernández Collados. "¿Es el teletrabajo una fórmula de conciliación de la vida personal, familiar y laboral?". *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, Vol. 10, no. 1, (2022), p. 200.

[21] Infojobs. *V Radiografía del Teletrabajo en España*. Septiembre 2025. Recuperado de https://s36300.pcdn.co/wp-content/uploads/2025/09/Informe_teletrabajo_2025-2.pdf

[22] Raquel Aguilera Izquierdo. "El derecho a la protección de datos en el ámbito laboral. Los sistemas de videovigilancia y geolocalización", *Revista de Trabajo y Seguridad Social. CEF*, no. 442, (2020), p. 97.

[23] Tamara Álvarez Robles. "El derecho de acceso universal a internet en el marco normativo español: presente y futuro", *Derecho Digital e Innovación. Digital Law and Innovation Review*, no. 7, (2020).

[24] Jaime Román Lemos. "La digitalización en la ley de empleo: empleo, colocación e intermediación", *Trabajo, Persona, Derecho, Mercado*, Monográfico (2023), p. 183.

sociales, económicas y tecnológicas"[25].

El art. 13, sobre el Plan Anual para el Fomento del Empleo Digno, incluye, obligatoriamente, en su eje 2 sobre formación, recuerda que se deberá tener en cuenta la brecha digital, "garantizando la atención presencial a la población que la padece". El art. 33, de sistemas de formación en el trabajo, declara como finalidad la mejora de las competencias profesionales, en especial las digitales y de sostenibilidad, "que inciden en su desarrollo profesional y personal", mejorando la empleabilidad, con empleo de calidad. Además, deberá "acompañar los procesos de transformación digital y ecológica y favorecer la cohesión social y territorial, así como la igualdad de género".
En cuanto a las competencias básicas para la empleabilidad, art. 38, vuelve a reconocer el manejo y aprovechamiento de las herramientas digitales y tecnológicas, "asegurándose la plena accesibilidad y la no discriminación en el uso de dichas herramientas" conformando una competencia transversal en todas las actividades de empleabilidad. Todo esto se une con la prohibición de discriminación, art. 39.

3.2.5. El alcance de la negociación colectiva

Los convenios colectivos españoles apenas han abordado de manera sistemática la brecha digital laboral. Algunos convenios sectoriales recientes (estiba, banca digital, química) incluyen cláusulas sobre formación en competencias digitales, información y consulta sobre introducción de nuevas tecnologías, límites a vigilancia y monitorización algorítmica o garantías de accesibilidad para personas con discapacidad en sistemas digitales.

Sin embargo, la cobertura es fragmentaria y débil. En muchos sectores (retail, hostelería, servicios a domicilio), donde la brecha digital afecta gravemente a trabajadores, los convenios son prácticamente ausentes sobre digitalización. De hecho, organizaciones como UGT advierte que "*la IA tiende a perpetuar y amplificar roles de género en el ámbito laboral" con un "acentuadísimo sesgo machista, vinculado a las profesiones de prestigio y alta cualificación siendo siempre para un hombre*"[26].

Por ejemplo, el Convenio Colectivo de la Banca (código de convenio 99000585011981), en su importante art. 80, que trata los derechos digitales, reconoce la desconexión digital y laboral, el derecho a la intimidad y uso de dispositivos digitales, como derecho a la inteligencia artificial (derecho a no ser objeto de decisiones automatizadas, derecho a la no discriminación o derecho de información) así como el derecho a la educación digital. Este último, especialmente relevante, obliga a las empresas a formar al personal, en "*competencias y habilidades digitales necesarias para afrontar la transformación digital y facilitar así su reconversión digital*" adaptando los nuevos puestos y evitando y erradicando "*las brechas digitales y garantizar su empleabilidad*". Todo ello, con el compromiso de los trabajadores de participar en estas acciones[27].

Por su parte, el Convenio de la Estiba del año 2022 (código de convenio 99012545011993)[28] tiene importantes disposiciones en lo que a automatización y digitalización se refiere, donde las partes "*consideran necesario realizar todos los esfuerzos precisos para evitar, minimizar o solidarizar los efectos negativos sobre la calidad y garantía del empleo*" además de un "*proceso de diálogo tendente a analizar las consecuencias en el empleo del proyecto empresarial*" con especial observancia a "*las consecuencias de la nueva tecnología y/o automatización*". Igualmente, cuando se implante nuevas tecnologías se deberá "*permitir el desarrollo de los procesos formativos del personal por cualquiera de las modalidades de importación previstas*" en un claro ánimo de cualificación profesional frente a la desigualdad tecnológica.

El Convenio Colectivo general de la Industria Química (código de convenio 99004235011981)[29] en

[25] Eva Garrido Pérez. "La Ley 3/2023 de Empleo: alcance y dimensiones estratégicas a favor de la empleabilidad", *Trabajo, Persona, Derecho, Mercado*, Monográfico (2024), p. 38.

[26] Unión General de Trabajadoras y Trabajadores. *UGT advierte que la IA tiende a perpetuar y amplificar roles de género en el ámbito laboral*. Recuperado de https://www.ugt.es/ugt-advierte-de-que-la-ia-tiende-perpetuar-y-amplificar-roles-de-genero-en-el-ambito-laboral

[27] Ministerio de Trabajo y Economía Social. Resolución de 20 de diciembre de 2024, de la Dirección General de Trabajo, por la que se registra y publica el XXV Convenio colectivo del sector de la banca (código de convenio 99000585011981). BOE núm. 1, de 1 de enero de 2025.

[28] Ministerio de Trabajo y Economía Social. Resolución de 4 de mayo de 2022, de la Dirección General de Trabajo, por la que se registra y publica el V Acuerdo para la regulación de las relaciones laborales en el sector de la estiba portuaria (código de convenio 99012545011993), BOE núm. 118, de 18 de mayo de 2022.

[29] Ministerio de Trabajo y Economía Social. Resolución de 6 de febrero de 2025, de la Dirección General de Trabajo, por la que se registra y publica el XXI Convenio colectivo general de la industria química (código de convenio 99004235011981). BOE núm. 41, de 17 de febrero de 2025.

su art. 10 recoge que cuando se "*introduzcan nuevas tecnologías*" que supongan modificaciones sustanciales de las condiciones de trabajo o periodo de formación o adaptación superior a un mes, "*se deberán comunicar las mismas con carácter previo a los representantes de las personas trabajadoras en el plazo suficiente para poder analizar y prever sus consecuencias*" en lo que respecta al "*empleo, salud laboral, formación y organización del trabajo*", igualmente, se facilitarán la "*formación adecuada y precisa para el desarrollo de su nueva función*".

Estas manifestaciones en convenios colectivos demuestran cómo la negociación y el diálogo social constituye un factor fundamental frente al reto de la digitalización, donde las empresas no quedan exclusivamente sujetas a lo establecido en el marco legal de la LOPDGDD y del Estatuto de los Trabajadores. En la actualidad, el V Acuerdo para el Empleo y la Negociación Colectiva (AENC) apunta un peso relevante a la transición tecnológica digital, para anticiparse a los cambios y evitar el "*impacto negativo en el empleo y garantizar el cumplimiento del derecho legal recogido en el artículo 64 del Estatuto de los Trabajadores*", ajustándose a los principios de control humano, transparencia y seguridad. El Capítulo V, en lo que concierne a la formación y cualificación profesional, se ve "fundamental la adaptación permanente de las plantillas a las nuevas realidades a las que se enfrenta el mundo del trabajo", es decir, establecer mecanismos concretos de formación para capacitar ante los retos de la digitalización, particularmente se indica la necesidad de capacitar con necesarias actualizaciones, "*marcadas, entre otras, por la transición digital y ecológica y el envejecimiento de la población trabajadora*"[30].

De cara al futuro, creemos interesante mencionar la propuesta de CCOO y UGT del VI AENC donde se reconoce cómo la transición digital necesita una "*implementación justa y ordenada que tenga en cuenta el impacto sobre el empleo*" además de "*necesidades formativas de las personas trabajadoras para adaptarse a los cambios y nuevas herramientas*", es decir, instrumentos potentes y útiles para revertir la brecha digital[31]. Esta visión sitúa las necesidades formativas no como una opción voluntaria, sino como un pilar estratégico para la empleabilidad, donde la capacitación en nuevas herramientas debe realizarse necesariamente dentro de la jornada laboral, mediante itinerarios de "*upskilling*" y "*reskilling*" que garanticen que ningún colectivo quede rezagado por razones de edad o cualificación.

4. Manifestaciones de la brecha digital en el mercado laboral español

4.1. Brecha digital de acceso al empleo y segmentación laboral

La automatización y digitalización de procesos laborales están transformando la demanda de empleo en España. Según el VI Informe sobre Desigualdad en España 2024 de Fundación Alternativas, en la última década ha habido una importante destrucción del empleo poco cualificado, especialmente en manufactura, logística manual, y servicios administrativos, donde automatización y digitalización han reemplazado trabajadores. No obstante, favorecerá el crecimiento de empleo en ocupaciones digitales, como en programación, análisis de datos o especialistas en ciberseguridad[32]. Otra cuestión es la creación de empleo precario digitalizado, es decir, plataformas de reparto, trabajo remoto freelance, servicios a demanda, que ofrecen flexibilidad pero bajo protección, bajos salarios, sin garantías, y que sigue sin una respuesta legislativa más allá de la presunción de laboralidad de los riders recogidas en el Disposición Adicional 23ª del ET.

El resultado del informe es concluyente, una segmentación laboral creciente: trabajadores con competencias digitales de nivel alto acceden a empleo estable, bien remunerado, con autonomía; trabajadores sin competencias digitales quedan rezagados en desempleo, inactividad, o empleo precario temporal. Esta segmentación tiende a perpetuarse porque trabajadores precarios tienen menos capacidad de acceder a formación (tiempo, dinero), mientras que trabajadores estables pueden invertir continuamente en upskilling[33]. La segmentación laboral y brecha digital, en este sentido, ya se viene anunciando desde hace tiempo, como reconocían Miren Barrenechea o Antonio

[30] Ministerio de Trabajo y Economía Social. Resolución de 19 de mayo de 2023, de la Dirección General de Trabajo, por la que se registra y publica el V Acuerdo para el Empleo y la Negociación Colectiva (código de convenio 99100015092012). BOE núm. 129, de 31 de mayo de 2023.

[31] Unión General de Trabajadoras y Trabajadores. *UGT y CCOO presentan sus propuestas para el VI AENC*. 2026. Recuperado de https://www.ugt.es/ugt-y-ccoo-presentan-sus-propuestas-para-el-vi-aenc

[32] Fundación alternativas. *VI Informe sobre la Desigualdad en España 2024. Los efectos de las transiciones demográfica, climática y digital en la desigualdad*. no. 6, (2024). Recuperado de https://fundacionalternativas.org/wp-content/uploads/2024/05/IDES_2024-3.pdf

[33] El upskilling hace referencia al proceso de adquirir nuevas competencias, de forma especial en el ámbito de la tecnología. Surge en las relaciones laborales como un elemento esencial para fortalecer la competitividad y la cualificación profesional.

Cardona en 2003[34].

El factor territorial empeora la brecha digital. En zonas rurales, los datos de brecha son aún mayores, causados por el envejecimiento de la población, éxodo rural y la falta de una conexión de calidad que no favorece la inclusión digital del territorio. Muchas zonas de España siguen sin conexión a internet de banda ancha, con una "*escasez de infraestructuras de telecomunicaciones*" que causa una "*brecha visible que lleva a sus habitantes a ir un paso por detrás del resto de los españoles en las principales tendencias de consumo de internet*"[35]. El Observatorio ASTEO en 2024 concluyó que al menos 7 de cada 10 usuarios con un perfil de internet medio o básico, consideraban necesaria una formación en habilidades digitales para usar internet de manera efectiva, especialmente las mujeres, en un 83% y los mayores de 55 años, en un 90%[36].

Esta brecha de acceso no solo se manifiesta en la posesión de herramientas físicas, sino en la "*alfabetización algorítmica*", porque la IA ya no es una tecnología de futuro, sino que sus manifestaciones están inmersas en la infraestructura social, también laboral, resultan necesario "*formar a ciudadanos críticos capaces de entender los sesgos algorítmicos, de exigir justicia en los sistemas automatizados y, sobre todo, de participar activamente en el diseño ético de esta sociedad*"[37]. Un ejemplo de cómo la IA afecta a las relaciones laborales es su actuación como filtro invisible en sistemas de selección de personal, que puede incluir filtros que penalizan a quienes no dominan los códigos digitales, así como sistemas de reconocimiento facial, análisis de currículums o test de proporcionalidad, provocando una barrera de entrada difícilmente franqueable para perfiles con baja cualificación tecnológica[38].

En febrero de 2026 ha sido publicado el informe "*El impacto de las nuevas tecnologías en la desigualdad salarial en España*" impulsado por el Observatorio Social de la Fundación "la Caixa" y que ha dado interesantes resultados[39]. El principal hallazgo de la investigación ha sido que la automatización, ese proceso de aplicación de la tecnología para ejecutar tareas sin intervención humana, contribuye al incremento de la desigualdad salarial, donde sin el cambio tecnológico, entorno al 10% de los trabajadores con salarios más elevados habrían acumulado una porción salarial un 4% menor, mientras que los trabajadores de segmentos inferiores habrían aumentado su participación. El estudio determina que "*el impacto adverso de la automatización es sistemáticamente mayor para los trabajadores con niveles educativos bajos*" al mismo tiempo que los trabajadores con estudios superiores son los menos perjudicados, manifestándose claramente cómo la digitalización en el trabajo mantiene las desigualdades y brechas salariales. En lo que a IA se refiere, su implementación aumenta los salarios de los trabajadores más cualificados, con un impacto que constata "*el aumento de la desigualdad salarial*" respecto de trabajadores de cualificaciones menores. Como conclusión, el estudio indica que "*el cambio tecnológico de las dos últimas décadas ha modificado la estructura salarial en España*" con un impacto negativo principalmente en "*la parte media de la distribución salarial y en los trabajadores menos cualificados*".

4.2. Teletrabajo, desconexión y nuevas desigualdades

El teletrabajo en España se ha consolidado tras la pandemia, pero sigue siendo minoritario y con fuerte sesgo social y de género: alrededor de un 14-15% de las personas ocupadas teletrabaja, cifra muy por debajo de la media europea, con más del 24%[40], y con mayor incidencia entre mujeres y trabajadores con más cualificación. Esta realidad revela una desigualdad tecnológica

[34] Miren Barrenechea Ayesta y Antonio Cardona Rodriguez. "La brecha digital como fuente de nuevas desigualdades en el mercado de trabajo", *Economistas*, no. 95, (2003).

[35] European Anti-Poverty Network. *Estudio: Brecha digital, rural y de género*, (2022), p. 31. Recuperado de https://www.eapn.es/ARCHIVO/documentos/documentos/1672316546_eapn_estudio-brecha-rural_271222.pdf

[36] Observatorio de Asteo Red Neutra. *III Informe del Observatorio Astro "Conectividad rural frente al Reto Demográfico: internet en áreas con menos de 10.000 habitantes*". (2024).

[37] Natividad Cobarrubias, Humberto Rodríguez, Yevgeni Martínez, y Eduardo Delsordo. "De la lectoescritura a la inteligencia artificial: hacia la alfabetización crítica en la sociedad algorítmica", *Revista Digital de Tecnologías Informáticas y Sistemas*, vol. 9, no. 1, 2025, p. 53.

[38] Anna Ginés i Fabrellas. "Inteligencia artificial en la fase precontractual de la relación laboral: selección de personas" en *Tratado sobre Inteligencia Artificial y Relaciones de Trabajo*, ed. Miguel Rodríguez-Piñero Royo, Aranzadi, 2025.

[39] Fundación La Caixa. *Observatorio Social. "El impacto de las nuevas tecnologías en la desigualdad salarial en España"*, (2026). Recuperado de https://elobservatoriosocial.fundacionlacaixa.org/es/-/impacto-nuevas-tecnologias-desigualdad-salarial-espana

[40] El upskilling hace referencia al proceso de adquirir nuevas competencias, de forma especial en el ámbito de la tecnología. Surge en las relaciones laborales como un elemento esencial para fortalecer la competitividad y la cualificación profesional.

doble: quien no dispone de competencias y medios digitales queda excluido de puestos que permiten teletrabajar, y al mismo tiempo el teletrabajo refuerza privilegios preexistentes (como los empleos estables, cualificados, urbanos) frente a sectores más precarios que continúan anclados al trabajo presencial, especialmente en zonas rurales. Se trata claramente de la conocida brecha de habitabilidad y equipamiento, puesto que no bastaría con poseer las competencias digitales sino también de disponer de espacios físicos adecuados y de conexión estable, costes que la Ley 10/2021 obliga a sufragar a la empresa pero que en la práctica suelen ser asumidos de forma encubierta por el trabajador.

Un estudio de Randstad Research sobre teletrabajo en España reconoció que el 41,5% de las viviendas españolas no estaban adecuadas para el trabajo, afectando negativamente a la productividad de los profesionales y a las posibilidades de conciliar[41]. Esta situación genera una clara discriminación, especialmente por condición económica, donde aquellos con menores ingresos teletrabajan en condiciones con mayor precariedad ergonómica y ambiental, que a largo plazo deriva en nuevos riesgos para la salud física y mental que no afectan por igual a todas las escalas salariales. Como una clara consecuencia podemos nombrar los riesgos psicosociales, aquellas alteraciones que afectan a la salud psicológica de las personas trabajadoras y que se debe a causas ambientales del entorno laboral o sus condiciones, incluyendo la ansiedad, depresión, estrés, falta de concentración o *burnout*[42]. El Tribunal Supremo en su Sentencia Nº 760/2025, de 10 de septiembre (ECLI:ES:TS:2025:3940), resolvió que los teletrabajadores no tenían derecho a que la empresa les dote de sillas ergonómicas en relación a la prevención de riesgos laborales. La resolución fue bastante relevante ya que discutió si existe o no una vulneración del principio de igualdad en el teletrabajo por la no dotación de este medio. El Tribunal entendió que el principio de igualdad de trato y oportunidades y no discriminación que se contempla en el art. 4.1 de la LTD, exige "*que las personas trabajadoras que desarrollan su actividad en la modalidad de teletrabajo gocen de los mismos derechos que las personas trabajadoras que prestan servicios presencialmente*", sin embargo, aunque la LTD incorpore diversas menciones a la dotación de medios, no hay una mención específica que reconozca el derecho genérico del teletrabajador a la silla ergonómica, ya que el legislador remite a la negociación colectiva el establecimiento de la dotación de medios, o bien al acuerdo individual de teletrabajo[43]. Por ello, se concluye que no se aprecia la infracción de dicho artículo.

En clave de género, los estudios muestran que el teletrabajo es ligeramente más frecuente entre mujeres, pero sin redistribución de cuidados y sin una organización del tiempo corresponsable, esta modalidad intensifica la doble jornada y puede traducirse en un techo de cristal digital, donde las mujeres teletrabajan más, pero promocionan menos[44]. Además, la edad, el territorio y la situación socioeconómica introducen nuevas brechas: las personas mayores y residentes en zonas rurales afrontan más dificultades de conectividad, equipamiento y competencias, de modo que el teletrabajo de calidad se concentra en núcleos urbanos y en perfiles jóvenes y formados. En cierto modo, el teletrabajo puede suponer una "resistencia" ante el modelo rígido actual de la laboralidad. En zonas rurales, el III Informe del Observatorio ASTEO, al que anteriormente nos referimos, indicó que 9 de cada 10 residentes de estas zonas veían positivo el teletrabajo para reactivar los pequeños municipios.

El derecho a la desconexión digital, reconocido en la LOPDGDD y desarrollado por la Ley 10/2021 de trabajo a distancia y el art. 20 bis ET, pretende equilibrar el tiempo de trabajo con el descanso, pero en la práctica su efectividad es desigual, especialmente en entornos donde se normaliza la disponibilidad permanente en línea[45]. En contextos de brecha digital, la desconexión puede convertirse paradójicamente en un privilegio de quienes ocupan posiciones fuertes en la empresa, mientras que trabajadores más vulnerables aceptan la hiperconexión por miedo a perder oportunidades o a ser sustituidos por perfiles más digitales.

Específicamente, la hiperconexión es uno de los retos más serios al que se enfrenta la desconexión

[41] Miren Barrenechea Ayesta y Antonio Cardona Rodriguez. "La brecha digital como fuente de nuevas desigualdades en el mercado de trabajo", *Economistas*, no. 95, (2003).

[42] European Anti-Poverty Network. *Estudio: Brecha digital, rural y de género*, (2022), p. 31. Recuperado de https://www.eapn.es/ARCHIVO/documentos/documentos/1672316546_eapn_estudio-brecha-rural_271222.pdf

[43] Observatorio de Asteo Red Neutra. *III Informe del Observatorio Astro "Conectividad rural frente al Reto Demográfico: internet en áreas con menos de 10.000 habitantes"*. (2024).

[44] Natividad Cobarrubias, Humberto Rodríguez, Yevgeni Martínez, y Eduardo Delsordo. "De la lectoescritura a la inteligencia artificial: hacia la alfabetización crítica en la sociedad algorítmica", *Revista Digital de Tecnologías Informáticas y Sistemas*, vol. 9, no. 1, 2025, p. 53.

[45] Julio Mendoza Huillca. "El trabajo remoto, la desconexión digital y el tecno estrés". *Ciencia Latina, Revista Multidisciplinar,* Vol. 6, no. 4, (2022), p. 1144.

digital. Esta hiperconectividad se ha visto reflejada incluso por la Organización Internacional del Trabajo, la cual en el Informe "*Trabajar en cualquier momento y en cualquier lugar: consecuencias en el ámbito laboral*" de 2019 realizado junto a Eurofound, concluyeron que la introducción de nuevas TICs provoca una ampliación injustificada de la jornada laboral sin el respeto debido a los descansos que laboralmente están establecidos, desencadenando en nuevos riesgos psicosociales[46]. Se trata de un "*escenario en el que, acabada la jornada diaria, los asuntos del trabajo siguen a la persona trabajadora hasta su casa*", un deber que inclusive puede estar "*implícito en la cultura organizativa de la empresa*" o bien en la autoimposición del trabajador[47]. Algunos autores vienen a llamar este fenómeno en lo vulgarmente conocido como "*esclavitud digital moderna*" donde se vulnera uno de los más antiguos derechos del trabajador, como es el del tiempo de descanso entre jornadas[48].

5. Instrumentos de tutela y propuestas de reforma

El ordenamiento español ha ido construyendo un mosaico de garantías frente a los riesgos de la digitalización (ET, LO 3/2018, Ley 10/2021, Ley 3/2023, o la propia negociación colectiva), pero estas normas no se han diseñado aún desde una lógica unitaria de lucha contra la desigualdad tecnológica. Es necesario reorientar estos instrumentos para que operen como un auténtico vector de igualdad de oportunidades digitales en el empleo, con obligaciones empresariales claras, políticas públicas activas y un papel reforzado de la negociación colectiva.

5.1. Reforzar la igualdad de oportunidades digitales en el empleo

La primera línea de actuación pasa por convertir las competencias digitales básicas en un verdadero derecho subjetivo ligado a la empleabilidad, superando la actual formulación programática de la Ley 3/2023. Esto implica que los servicios públicos de empleo y las políticas activas incorporen itinerarios formativos digitales obligatorios y adaptados a colectivos con mayores carencias (personas desempleadas de larga duración, mayores de 55 años, mujeres con responsabilidades de cuidado o población rural), garantizando siempre modalidades presenciales para quienes sufren más intensamente la brecha digital[49].

En el plano de la empresa, la Ley 10/2021 debería reforzar la obligatoriedad de la formación digital vinculada al puesto, imponiendo obligaciones específicas de actualización cuando se introduzcan nuevas herramientas, plataformas o sistemas de gestión. La negociación colectiva puede traducir este mandato en planes de capacitación digital sectoriales con garantías de tiempo de formación dentro de la jornada, prioridad para trabajadores más vulnerables y certificación de las competencias adquiridas, evitando que el "*upskilling*" se convierta en una carga individual y no en una responsabilidad organizativa.

5.2. Fortalecer la transparencia y gobernanza de sistemas algorítmicos

La regulación europea de la inteligencia artificial y la nueva Directiva 2024/2831 sobre trabajo en plataformas han introducido obligaciones de transparencia, supervisión humana y rendición de cuentas en la gestión algorítmica, especialmente en contextos de alto riesgo como la selección de personal o la organización del trabajo[50]. Sin embargo, estas obligaciones deben aterrizar en el Derecho interno laboral, articulando derechos concretos de información, explicación y revisión de

[46] Eurofound y OIT. *Working anytime, anywhere: The effects on the world of work*. (2019). Recuperado de https://www.ilo.org/publications/working-anytime-anywhere-effects-world-work

[47] Sarai Rodríguez González. "El riesgo de la conectividad permanente: un reto para la consecución de un tiempo de trabajo decente". p. 5, en *Health at work, ageing and environmental effects on future social security and labour law systems* ed. por Nuno Cerejeira Namora (ed. lit.), Lourdes Mella Méndez (ed. lit.), Duarte Abrunhosa e Sousa (ed. lit.), Gonçalo Cerejeira Namora (ed. lit.), Eduardo Castro Marques (ed. lit.), ISBN 9781527514010, Cambridge Scholars Publishing, Reino Unido, (2018).

[48] Talita Correa Gomez Cardim. "De la hiperxonexión del trabajador a la esclavitud digital: riesgos psicosociales y desafíos de la conciliación entre tiempo de trabajo y vida privada". *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, Vol. 11, no. 1, (2023), p. 428.

[49] José Climent Rodríguez. "Análisis de la empleabilidad y la orientación laboral en el contexto de las políticas activas de empleo y su concreción actual en la Ley 3/2023 de 28 de febrero, de empleo". *Trabajo, Persona, Derecho, Mercado*, monográfico (2023), p. 90.

[50] Alana Engelmann. "Algorithmic transparency as a fundamental right in the democratic rule of law", *Brazilian Journal of Law, Technology and Innovation*, Vol. 1, no 2, (2023), p. 174.

decisiones automatizadas para todos los trabajadores, no solo en plataformas digitales[51].

Resulta preciso reconocer el derecho de la representación legal de los trabajadores a acceder a información sustantiva sobre el funcionamiento de los sistemas algorítmicos que inciden en contratación, promoción, evaluación del rendimiento o despido, más allá de fórmulas genéricas en el art. 64.4 d) del ET[52]. La negociación colectiva debería incorporar cláusulas de transparencia algorítmica, límites a la monitorización digital, obligación de evaluaciones de impacto en igualdad y no discriminación, así como canales para auditorías independientes que detecten sesgos y mecanismos correctores accesibles y comprensibles para la plantilla[53].

5.3. Garantías específicas para colectivos vulnerables

La brecha digital converge con factores de género, edad, discapacidad, origen y territorio, por lo que las medidas generales resultan insuficientes si no se acompañan de garantías reforzadas para estos colectivos. La Ley 3/2023 y la legislación en materia de igualdad deben desarrollar un ejercicio de coordinación para exigir planes de inclusión digital laboral que combinen formación adaptada, ajustes razonables tecnológicos, apoyo económico para equipamiento y conectividad, y canales de orientación personalizados.

En el caso de personas con discapacidad, la regulación sobre trabajo a distancia y accesibilidad digital ha de incorporar una obligación expresa de proporcionar tecnologías de apoyo (por ejemplo, lectores de pantalla, software de reconocimiento de voz o interfaces accesibles) y de verificar su compatibilidad con los sistemas corporativos. La accesibilidad digital universal es imprescindible para erradicar la brecha digital en estas personas y fomentar su inclusión laboral.

En España, el número de personas con discapacidad en edad de trabajar es superior al 6% donde las políticas activas tanto del Estado como de las empresas en materia de accesibilidad se convierten en una de las más relevantes oportunidades, como las adaptaciones en los software[54]. Todo ello debe además analizarse desde el nuevo prisma implantado por medio de la Ley 2/2025, de 29 de abril, que modificó el art. 49.1 del ET. La nueva articulación prohíbe la extinción automática del contrato laboral en caso de sufrir una incapacidad permanente (IP), o lo que es lo mismo, convertirse en una persona con discapacidad, resultando obligatorio, una vez declarada dicha IP, adaptar el puesto de trabajo por medio de los ajustes razonables, salvo que suponga una "*carga excesiva para la empresa*" o no exista un puesto vacante y disponible acorde con el perfil de la nueva situación de la persona trabajadora, o bien el propio trabajador rechace dicho cambio[55].

Por su parte, la acción pública en zonas rurales debe priorizar la extensión de la banda ancha de alta velocidad, la creación de centros de recursos digitales comunitarios y la oferta de formación presencial y gratuita, de modo que el código postal deje de ser un factor determinante de exclusión tecnológica en el empleo.

Los trabajadores en edad próxima a la jubilación son también uno de los colectivos más vulnerables frente a la brecha digital, con menores capacitaciones en lo que a competencias digitales se refiere[56]. El profesor Álvarez del Cuvillo convenientemente defiende que "*en las sociedades modernas del capitalismo avanzado, la edad sigue siendo una circunstancia socialmente relevante, que genera clasificaciones sociales muy significativas*"[57]. Sin duda, es un grupo que posee menores

[51] de Zárate Alcarazo, L. O. "Explicabilidad (de la inteligencia artificial)" Eunomía. Revista en Cultura de la Legalidad, 22, 2022, p. 333.

[52] Adrián Todolí Signes. "Cambios normativos en la digitalización del trabajo: comentario a la Ley Rider y los derechos de información sobre los algoritmos". *IUSLabor*, no. 2, (2021), p. 43.

[53] Juana María Serrano García. "La humanización de la gestión algorítmica de la directiva de trabajo en plataformas", *Briefs Asociación Española de Derecho del Trabajo y de la Seguridad Social*, no. 100, (2024).

[54] Mercedes Herrero de la Fuente, Begoña Miguel San Emeterio, y Javier Sierra Sánchez. "Digital Skills and Technological Accessibility as Challenges for the Labour Market Insertion of People with Disabilities in the Audiovisual Sector". *UCJC Business and society review*, (2022).

[55] Manuel Jesús Garnica Corbacho. "La extinción de la relación laboral por incapacidad permanente a raíz de la Ley 2/2025: ámbito de aplicación y retos". *Revista General de Derecho del Trabajo y de la Seguridad Social*, no. 71, (2025), p. 407.

[56] Pompeyo Gabriel Ortega Lozano. "Transición digital y trabajadores en edad próxima a la jubilación: brecha y desigualdades tecnológicas". *Revista de Derecho de la Seguridad Social, Laborum*, no. 6, (2024), p. 183.

[57] Antonio Álvarez del Cuvillo. "La identificación y refutación de la discriminación por edad a partir del concepto de discriminación grupal", en *Trabajador, edad y pensiones de jubilación*, Comunicaciones del XXXIV Congreso Nacional de la Asociación Española de Derecho del Trabajo y de la Seguridad Social, (2024), p. 564.

posibilidades de reincorporarse al mercado laboral una vez se encuentran fuera, teniendo, habitualmente, un nivel de estudios bajo, siendo la edad un factor importante de vulnerabilidad, determinante en la desigualdad tecnológica laboral[58]. Frente a esta brecha generacional, la tutela laboral debe enfocarse en ejes como el reciclaje o "reskilling", donde la empresa tiene que garantizar acciones dirigidas específicamente a la formación de los trabajadores en competencias, en este caso, digitales, adaptada a los ritmos y necesidades de aprendizaje del sector[59].

6. Conclusiones

La brecha digital en el ámbito laboral no se reduce a una mera cuestión de acceso a dispositivos o conexión a internet, sino que configura una nueva forma de desigualdad estructural que condiciona el acceso al empleo, la estabilidad, la promoción profesional y, en definitiva, la calidad del trabajo. Sus manifestaciones en España, desde la segmentación entre empleos digitales y no digitales hasta las desigualdades que genera el teletrabajo y la gestión algorítmica, muestran cómo las transformaciones tecnológicas tienden a amplificar vulnerabilidades preexistentes si no se acompañan de políticas de tutela adecuadas.

El marco normativo internacional, europeo y español ofrece una base relevante para afrontar estos retos, pero todavía fragmentaria y reactiva, más orientada a mitigar riesgos que a garantizar una verdadera inclusión digital del trabajo. La integración de la igualdad de oportunidades digitales como principio transversal del Derecho del Trabajo exige reforzar el derecho a la formación digital, consolidar la transparencia y gobernanza algorítmica, y desplegar garantías específicas para aquellos colectivos que más sufren la desigualdad tecnológica. Es decir, hay una clara insuficiencia legal que repercute negativamente en el mercado laboral español, persistiendo la brecha digital y la desigualdad tecnológica, especialmente en colectivos más vulnerables como personas con discapacidad, mujeres, mayores, personas LGTBI+ o raciales, así como en zonas rurales y de la periferia.

Desde esta perspectiva, la tutela laboral frente a la brecha digital debe concebirse como una política de derechos fundamentales, en la que el acceso, uso y aprovechamiento de las tecnologías en el empleo se distribuyan de forma equitativa y compatible con la dignidad, la privacidad y la no discriminación de las personas trabajadoras. Solo así la digitalización dejará de ser un vector de exclusión y se convertirá en una herramienta de cohesión social y de democratización real de las oportunidades en el mercado de trabajo. Para la consecución de estos propósitos, empresas, sindicatos, trabajadores y el propio Estado deben comprometerse a establecer mecanismos de participación, transparencia, y sobre todo acciones de capacitación, actualización y formación para garantizar un nivel de competencia digital suficiente a las necesidades del nuevo mercado, junto a la necesaria reducción de desigualdades, salariales como de oportunidades, poniendo el foco en la vulnerabilidad y en consonancia con los deberes del Estado Social.

Bibliografía

Álvarez del Cuvillo, Antonio. "La ley integral para la igualdad: un frágil puente entre el Derecho Europeo y la Constitución". *Temas Laborales* no. 165 (2022).

Aragüez Valenzuela, Lucía. "Desafíos de la digitalización de las relaciones laborales: algoritmos digitales, robotización y trabajo a distancia". *e-Revista Internacional de la Protección Social* 7, no. 1 (2022).

Balladares Burgos, Jorge. "Principios y valores para una ética digital". *Oxímora, Revista Internacional de Ética y Política* no. 23 (2023)

Braqui, Margarita. "Radiografía de un experimento sociológico". En *Feminismo 2.0*, editado por Luis Santos, 27-63. Barcelona: Ínsula, 2020

Castaño Collado, Cecilia. *La segunda brecha digital*. Madrid: Ediciones Cátedra, 2008.

Cerrillo i Martínez, Agustí y Clara Isabel Velasco Rico. "La transparencia algorítmica". *Working Papers Digitapia, Universitat Oberta de Catalunya* no. 3 (2025).

[58] Ana María Martín Romero. "La edad como factor de vulnerabilidad en las relaciones laborales" (Tesis doctoral, Universidad de Vigo, 2022).

[59] Ramona Diana Leon. "Employees' reskilling and upskilling for industry 5.0: Selecting the best professional development programmes", *Technology in Society*, no. 75, (2023), p. 2.

De las Heras García, Aránzazu. "La fiscalidad de la dotación de medios y compensación de gastos en el teletrabajo: entre la omisión de regulación estatal y la respuesta foral vasca". *Lan Harremanak* no. 54 (2025).

Delgado López, Inés. "El impacto de los sistemas algorítmicos en los procesos de selección de personal. Análisis jurídico-laboral a la luz del nuevo Reglamento europeo en materia de inteligencia artificial". *Revista de Trabajo y Seguridad Social. CEF* no. 483 (2024).

Eduardo Haz-Gómez, Gabriel López-Martínez y Salvador Manzanera-Román. "La exclusión digital como una forma de exclusión social: una revisión crítica del concepto de brecha digital". *Studia Humanitatis Journal* 4, no. 1 (2024).

Fernández Alles, José Joaquín. "La jurisprudencia constitucional sobre el principio de no discriminación por razón de discapacidad". *Revista de Estudios Fronterizos del Estrecho de Gibraltar* no. 10 (2022).

García-Izquierdo, Antonio León y García-Izquierdo, Mariano. "Discriminación, igualdad de oportunidades en el empleo y selección de personal en España". *Revista de Psicología del Trabajo y de las Organizaciones* 23, no. 1 (2007).

Iturmendi Rubia, José Miguel. "La discriminación algorítmica y su impacto en la dignidad de las personas y los derechos humanos. Especial referencia a los inmigrantes". *Deusto Journal of Human Rights* no. 12 (2023).

Lombardero Rodil, Luis. *Trabajar en la Era Digital: Tecnología y Competencias para la Transformación Digital*. Córdoba: Editorial Almuzara, 2015.

López Ahumada, Eduardo. "El desarrollo del principio europeo de transparencia algorítmica en el trabajo en plataformas digitales". *Briefs AEDTSS, Asociación Española de Derecho del Trabajo y de la Seguridad Social* no. 69 (2025).

Macías García, María del Carmen. "El Plan Nacional de Competencias Digitales en España y su repercusión en la población activa". *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo* 9, no. 4 (2021).

Martín Romero, Ana María. "La brecha digital generacional". *Temas Laborales* no. 151 (2020).

Martínez-Carrión, José Miguel, Pérez-Castroviejo, Pedro María, Puche Gil, Javier y Ramón-Muñoz, Josep María. "La brecha rural-urbana de la estatura y el nivel de vida al comienzo de la industrialización española". *Historia Social* (2014).

Mercader Uguina, Jesús y Velasco Pardo, Belén. "El impacto laboral del reglamento de inteligencia artificial". *Revista Derecho Social y Empresa* no. 23 (2025).

Olarte Encabo, Sofía. "Brecha digital, pobreza y exclusión social". *Temas Laborales* no. 136 (2017).

Ordiales Hurtado, Inmaculada. "Brechas digitales en España y su relación con el mercado de trabajo". *Cuadernos del Mercado de Trabajo* no. 9 (2023).

Ortiz de Zárate Alcarazo, Lucía. "Explicabilidad (de la inteligencia artificial)". *Eunomía. Revista en Cultura de la Legalidad* no. 22 (2022).

Ribes Moreno, M. Isabel. "Marina Mercante, derechos de conciliación y negociación colectiva, ¿es posible conciliar a bordo de un buque?". *Temas Laborales* no. 132 (2016).

After all, what are Regulatory Sandboxes? Key characteristics and a working definition

Después de todo, ¿qué son los sandboxes regulatorios? Características clave y una definición de trabajo

Thiago Guimaraes Moraes
Vrije Universiteit Brussel (VUB) / Universidade de Brasília (UnB)

Resumen:

Este artículo examina cómo se definen los sandboxes regulatorios en la literatura académica y en documentos de política pública, con el objetivo de clarificar las principales características de esta herramienta regulatoria, que ha sido adoptada de manera creciente por formuladores de políticas. El análisis se basa en una revisión cualitativa de más de 40 publicaciones académicas y fuentes de política pública que, en conjunto, examinan más de 100 sandboxes regulatorios, abarcando una amplia variedad de sectores y jurisdicciones, tanto en Europa como fuera de ella. El artículo se estructura en dos partes principales. En primer lugar, revisa y compara las definiciones existentes de sandboxes regulatorios e identifica siete elementos recurrentes. En segundo lugar, examina críticamente cada uno de estos elementos a la luz de los debates académicos y de política pública sobre experimentación regulatoria y gobernanza de la innovación. A partir de este análisis, el artículo propone una definición de los sandboxes regulatorios que refleja los enfoques predominantes en la literatura y destaca sus principales características funcionales.

Palabras clave:

Sandboxes regulatorios; aprendizaje regulatorio; flexibilidad regulatoria; orientación regulatoria; gobernanza colaborativa.

Sumario:

Abstract:

This paper examines how regulatory sandboxes are defined in academic literature and policy documents, with the aim of clarifying the key characteristics of this regulatory tool, which has been increasingly adopted by policymakers. The assessment is performed on basis of a qualitative review of more than 40 academic publications and policy sources that together assessed more than 100 regulatory sandboxes, covering a broad range of sectors and jurisdictions, within Europe and beyond. The paper is structured in two main parts. First, it reviews and compares existing definitions of regulatory sandboxes and identifies seven recurring elements. Second, it critically examines each of these elements in light of academic and policy discussions on regulatory experimentation and innovation governance. Based on this analysis, the paper proposes a working definition of regulatory sandboxes that reflects prevailing approaches in the literature and highlights their core functional features.

Keywords:

Regulatory sandboxes; regulatory learning; regulatory flexibility; regulatory guidance; collaborative governance.

Summary:

1. Introduction

The term sandbox has been historically used in the information technology ecosystem to describe isolated, controlled environments where products or software can be tested securely[1]. ICT industry drew inspiration from the analogy of children's playgrounds, where sandboxes are seen as spaces for safe play[2]. This concept has also been adopted by the social sciences to describe experimental settings in various fields, such as sustainable urban development[3] or education[4]. However, these applications do not necessarily involve legal oversight or regulatory involvement and therefore are not considered regulatory sandboxes.

By contrast, regulatory sandboxes – a term increasingly employed in legal and policy contexts - also evoke the idea of a safe and controlled environment, but with a broader meaning of "safety." On one hand, they are intended to provide innovators with a protected space where regulatory requirements are relaxed to encourage experimentation[5]. On the other hand, they must also ensure safety for third parties, such as safeguarding consumer protection - as addressed by financial sandboxes[6] or addressing broader societal concerns, including health, safety, and the protection of fundamental rights - as emphasized by the EU AI Act.[7]

Due to the increasing adoption of regulatory sandboxes, it becomes increasingly important to reflect on how this concept is defined and characterised, in order to support their appropriate design and operation. To this end, this paper conducted a qualitative review of more than 40 academic papers and policy documents from the European Union and multilateral organisations such as the Organization for Economic Cooperation and Development – OECD, and the World Bank. The initial corpus comprised highly cited papers available through major academic databases - Springer, ProQuest, and Web of Science - and was subsequently expanded through a snowballing technique, whereby references in the selected texts led to further relevant materials. The reviewed literature includes one systematic literature review on regulatory sandboxes[8], and sector-specific studies in the financial sector[9], the energy sector[10], data protection[11] and AI-related sandboxes[12].

[1] Sharmista Appaya et al., *Global Experiences from Regulatory Sandboxes*, Fintech Note | No. 8 (World Bank, 2020), 5; Hannah Ruschemeier, "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation," in *Bridging the Gap Between AI and Reality*, ed. Bernhard Steffen (Springer Nature Switzerland, 2024), 321, https://doi.org/10.1007/978-3-031-73741-1_20.

[2] Francesco Longo et al., "Value-Oriented and Ethical Technology Engineering in Industry 5.0: A Human-Centric Perspective for the Design of the Factory of the Future," *Applied Sciences* 10, no. 12 (2020): 25, 12, https://doi.org/10.3390/app10124182.

[3] Ruschemeier, "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation."

[4] Sergio Sánchez et al., "Augmented Reality Sandbox: A Platform for Educative Experiences," November 2, 2016, 602, https://doi.org/10.1145/3012430.3012580

[5] L. Bromberg et al., "Fintech Sandboxes: Achieving a Balance between Regulation and Innovation," *Journal of Banking and Finance Law and Practice* 28, no. 4 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090844.

[6] Hilary J. Allen, "Regulatory Sandboxes," *George Washington Law Review* 87, no. 3 (2019): 626, https://doi.org/10.2139/ssrn.3056993; Jonathan Everhart, "The Fintech Sandbox: An Overview of Regulatory Sandbox Regimes," *Southern Journal of Business and Ethics* 12 (2020): 64–73.

[7] See Article 57(6), 57(11) and Recital 140 of the AI Act, all of which refer to risk management of fundamental rights, health and safety within sandbox environments.

[8] Nova Sky, "Systematic Review of Regulatory Sandboxes : Implications for the European Union's Artificial Intelligence Act" (Diplomityö, N. Sky, 2024), https://oulurepo.oulu.fi/handle/10024/50797.

[9] Appaya et al., *Global Experiences from Regulatory Sandboxes*; European Securities and Markets Authority et al., *FinTech: Regulatory Sandboxes and Innovation Hubs* (Brussels, 2018), https://www.esma.europa.eu/sites/default/files/library/jc_2018_74_joint_report_on_regulatory_sandboxes_and_innovation_hubs.pdf; Dirk Zetzsche et al., "Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation," *Fordham Journal of Corporate & Financial Law* 23, no. 1 (2017): 31.

[10] F. Gangale et al., *Making Energy Regulation Fit for Purpose: State of Play of Regulatory Experimentation in the EU : Insights from Running Regulatory Sandboxes*. (European Commission - Joint Research Centre, 2023), DOI.org (CSL JSON), https://data.europa.eu/doi/10.2760/32253.

[11] Business at OECD, *Regulatory Sandboxes for Privacy - Analytical Report.Pdf* (BIAC, 2020), https://25159535.fs1.hubspotusercontent-eu1.net/hubfs/25159535/website/documents/pdf/Digital%20Economy/Regulatory%20Sandboxes%20for%20Privacy%20-%20Analytical%20Report.pdf; Nathan Genicot, *From Blueprint to Reality: Implementing AI Regulatory Sandboxes under the AI Act* (FARI, 2024).

[12] Organization for Economic Cooperation and Development, *Regulatory Sandboxes in Artificial Intelligence* (OECD, 2023), https://doi.org/10.1787/8f80a0e6-en

Altogether, these works analyse more than 100 regulatory sandboxes in all regions of the globe, and some of those references cover regulatory sandbox cases from 2015, while the most recent report assessed has been published in 2026. Through cross-referencing and comparative analysis, the review enabled the identification of recurring definitions, core features, and emerging classifications, as well as key points of debate surrounding the role of regulatory sandboxes in governance and innovation policy.

Thus, the current paper is structured into two main sections. First, seven key characteristics of regulatory sandboxes are extracted from the definitions identified in the reviewed literature. Second, these core features are critically examined in light of the broader scholarly and policy discussions on regulatory experimentation and innovation governance.

2. Defining Regulatory Sandboxes

As the concept progressed into the regulatory landscape, different definitions have been proposed by a variety of entities, from the private sector, the public sector, academia and civil society. In Table 1, a non-exhaustive list of definitions for regulatory sandbox is presented as identified in literature. The table is provided in chronological order.

Table 1: Proposed definitions of "regulatory sandbox" found in the reviewed literature

Source	Definition
Zetzsche et al.[13]	"In finance, a regulatory sandbox refers to a regulatory "safe space" for innovative financial institutions and activities underpinned by technology. At the most basic level, the sandbox creates an environment for businesses to test products with less risk of being "punished" by the regulator for non-compliance. In return, regulators require applicants to incorporate appropriate safeguards to insulate the market from risks of their innovative business."
European Financial Supervisory Authorities[14]	"Schemes to enable firms to test, pursuant to a specific testing plan agreed and monitored by a dedicated function of the competent authority, innovative financial products, financial services or business models."
Council of the European Union[15]	"Concrete frameworks that, by providing a structured context for experimentation, enable, where appropriate, in a real world environment, the testing of innovative technologies, products, services or approaches - especially in the context of digitalization - for a limited time and in a limited part of a sector or area under regulatory supervision, ensuring that appropriate safeguards are in place."
World Bank Group – WBG[16]	"[A] controlled, time-bound, live testing environment, which may feature regulatory waivers at regulators' discretion. The testing environment may involve limits or parameters within which firms must operate."
OECD[17]	"A limited form of regulatory waiver or flexibility for firms, enabling them to test new business models with reduced regulatory requirement."
Business at OECD – BIAC[18]	"A controlled environment wherein for some predetermined period of time and for a defined use case, a close collaboration between firms and a regulator enables firms to test new data uses, technologies and applications while receiving regulatory guidance."
Didenko[19]	A standing (or at least long-term) regulatory strategy that facilitates the development of innovative technology-driven solutions in the financial sector and (i) involves actual on-market testing involving real customers, (ii) is conducted under regulatory supervision and (iii) provides limited exceptions from the otherwise applicable regulatory framework (but regardless of specific regulatory tools used to administer such exceptions)."

[13] Zetzsche et al., "Regulating a Revolution."

[14] European Securities and Markets Authority et al., *FinTech: Regulatory Sandboxes and Innovation Hubs*.

[15] Council of the European Union, *Regulatory Sandboxes and Experimentation Clauses as Tools for Better Regulation: Council Adopts Conclusions* (EU Council, 2020)

[16] Appaya et al., *Global Experiences from Regulatory Sandboxes*

[17] Angela Attrey et al., *The Role of Sandboxes in Promoting Flexibility and Innovation in the Digital Age*, OECD Going Digital Toolkit Notes (OECD, 2020), https://doi.org/10.1787/cdf5ed45-en.

[18] Business at OECD, *Regulatory Sandboxes for Privacy - Analytical Report.Pdf*

The Future Society[20]	"Time-bound programmes administered by authorities. They enable entrepreneurs to test new products and services in the market, but with increased oversight from authorities".
European Parliament Research Service[21]	"[R]egulatory tools allowing businesses to test and experiment with new and innovative products, services or businesses under supervision of a regulator for a limited period of time."
Agência Nacional de Proteção de Dados – ANPD[22]	"A regulatory sandbox is a collaborative experimentation between the regulator, the regulated entity and other stakeholders with the aim to test innovations against the regulatory framework by adopting a structured methodology."
European Commission[23]	"Schemes that enable firms to test innovations in a controlled real-world environment, under a specific plan developed and monitored by a competent authority" and which are "usually organised on a case-by-case basis, include a temporary loosening of applicable rules, and feature safeguards to preserve overarching regulatory objectives, such as safety and consumer protection."
Undheim et al.[24]	"A regulatory sandbox is an example of a 'soft law' mechanism in emerging technologies, introduced in highly regulated industries such as finance and energy, or related to specific spheres or regulations, such as AI or GDPR, with the goal of promoting responsible innovation/and or competition, addressing regulatory barriers to innovation and advancing regulatory learning".
Ranchordás and Vinci[25]	"Collaborative regulatory instruments where regulators interact closely with a selected group of market actors (usually startups) to create a safe testbed to understand how to best regulate new types of services or products."
The Datasphere Initiative[26]	"Controlled learning environment designed for structured experimentation under defined governance frameworks, timeframes, and built-in safeguards to support iterative, multi-actor collaboration and evidence-based decision-making"

While the plethora of definitions may reveal some different interpretations of this concept, it also allows to identify points of convergence. One widely shared element is the idea of regulatory sandboxes as *controlled environments for the temporary testing of innovations*. This notion brings together three core characteristics: (i) they are *innovation-driven* – as several definitions associate them with the testing of new services or products, or new business models; (ii) they operate within a *controlled regulatory framework*, implying the presence of a structured methodology (which also highlights the relevance on the leading role of regulators on setting up and monitoring these schemes)[27]; and (iii) they are *time-limited*. These features closely align with traditional experimental

[19] Anton Didenko, "A Better Model for Australia's Enhanced Fintech Sandbox," *University of New South Wales Law Journal* 44, no. 3 (2021).

[20] The Future Society, *Sandboxes without the Quicksand: Making EU AI Sandboxing Work for Regulators, Entrepreneurs and Society* (TFS, 2022).

[21] Tambiana Madiega and Anne Louise Van de Pol, *Artificial Intelligence Act and Regulatory Sandboxes* (European Parliament Research Service, 2022).

[22] Agência Nacional de Proteção de Dados, *Consulta à Sociedade - Sandbox Regulatório de IA* (ANPD, 2023), https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/anpd-sandbox-regulatorio-consulta-bilingue.pdf.

[23] European Commission, *Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final* (European Commission, 2023), https://data.consilium.europa.eu/doc/document/ST-12199-2023-INIT/en/pdf

[24] Kristin Undheim et al., "True Uncertainty and Ethical AI: Regulatory Sandboxes as a Policy Tool for Moral Imagination," *AI and Ethics* 3, no. 3 (2023): 997–1002, https://doi.org/10.1007/s43681-022-00240-x

[25] Sofia Ranchordás and Valeria Vinci, "Regulatory Sandboxes and Innovation-Friendly Regulation: Between Collaboration and Capture," *Italian Journal of Public Law*, March 14, 2024, https://www.ijpl.eu/regulatory-sandboxes-and-innovation-friendly-regulation-between-collaboration-and-capture/

[26] The Datasphere Initiative, *Sandboxes for DPI* (The Datasphere Initiative, 2026), https://www.thedatasphere.org/wp-content/uploads/2026/02/Report-Sandboxes-for-DPI-Datasphere-Intiative-2026.pdf

[27] European Commission, Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final, 15. "A Better Model for Australia's Enhanced Fintech Sandbox." The Future Society Sandboxes without the Quicksand: Making EU AI Sandboxing Work for Regulators, Entrepreneurs and Society.

approaches, supporting the classification of regulatory sandboxes as a form of regulatory experimentation tool[28].

Another key element present across several definitions is (iv) *regulatory flexibility*. Once again, the central role of the regulator is emphasized, due to its legitimate role to apply flexible rules, within the limits of existing law. A clear example of this is the use of temporary regulatory waivers, explicitly mentioned in the OECD's definition. However, most definitions do not explicitly refer to waivers. For instance, the European Commission refers instead to the "loosening of applicable rules," a broader and more ambiguous concept, while Undheim et al. mention that regulatory sandboxes address "regulatory barriers to innovation."

The fifth element is (v) *regulatory guidance*, explicitly mentioned by BIAC. This concept emphasizes the role of the regulator in informing and supporting innovators on how to better address legal obligations. It connects to the "soft law" approach, where administrative bodies provide non-binding recommendations - typically through guidelines – but under an interactive, experimental setting[29]. Undheim et al.[30] explicitly associate regulatory sandboxes with soft law. However, regulatory sandboxes differ from typical soft law tools: besides guidance, they enable temporary relaxation of binding rules, a feature soft law is not known for.

Ranchordás and Vinci highlight the notion of (vi) *regulatory learning*, defining regulatory sandboxes as mechanisms to help regulators understand how best to regulate new types of services and products. This function is primarily directed at the regulator, who gains insight into how abstract legal norms apply in specific socio-technical contexts[31]. Nonetheless, this learning process also benefits regulatees, as both parties can gather practical insights, identify good practices, and generate lessons that guide future legal and policy development[32]. This aligns with the "test and learn" philosophy, which suggests that the best way to evaluate innovative ideas is through experimentation[33]. Moreover, this approach can extend to evaluating regulatory effectiveness itself through a process known as evidence-based learning[34].

Finally, several definitions imply that regulatory sandboxes are (vii) *collaboration oriented*. This highlights an expectation of close cooperation between regulators and innovators, with the former adopting a supportive rather than enforcement-focused role during experimentation. The objective is not to audit participants but to jointly explore and discuss the regulatory landscape. Therefore, trust is critical to this collaborative model, enabling transparency and the open sharing of information - particularly in areas of emerging risk[35]. At the same time, such trust must be balanced with accountability and independent oversight to prevent regulatory capture and ensure that sandbox experimentation serves the public interest[36].

In most definitions, the collaborative aspect focus narrowly on the interaction between regulators and businesses, overlooking scenarios in which the public sector itself may be a source of innovation, or yet failing to acknowledge the potential contributions of other stakeholders within innovation ecosystems, such as academia and civil society. A notable exception is the definition proposed by the Brazilian Data Protection Authority (ANPD), which describes regulatory sandboxes as "collaborative experimentation between the regulator, the regulated entity, and other stakeholders." The term "regulated entity" is deliberately broad, opening the possibility for experimentation within the public sector. In fact, France's *Commission nationale de l'informatique et des libertés*, held one

[28] Sofia Ranchordás, "Experimental Regulations and Regulatory Sandboxes – Law Without Order?," *Law and Method*, 2021, https://www.boomportaal.nl/tijdschrift/LaM/lawandmethod-D-21-00012.

[29] Walter G. Johnson, "Caught in Quicksand? Compliance and Legitimacy Challenges in Using Regulatory Sandboxes to Manage Emerging Technologies," *Regulation & Governance* 17, no. 3 (2023): 709–25, https://doi.org/10.1111/rego.12487.

[30] Undheim et al., "True Uncertainty and Ethical AI," 998.

[31] Jonathan McCarthy, "From Childish Things: The Evolving Sandbox Approach in the EU's Regulation of Financial Technology," *Law, Innovation and Technology* 15, no. 1 (2023): 1–24, https://doi.org/10.1080/17579961.2023.2184131.

[32] European Commission, *Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final*; Sofia Ranchordás, "Experimental Regulations for AI: Sandboxes for Morals and Mores," *Morals & Machines* 1, no. 1 (2021): 86.

[33] Nils O. Fonstad and Jeanne W. Ross, "Learning How to Test and Learn | *MIT CISR," MIT Center for Information Systems Research*, February 15, 2018, https://cisr.mit.edu/publication/2018_0201_TestAndLearn_FonstadRoss

[34] Business at OECD, *Regulatory Sandboxes for Privacy - Analytical Report.Pdf*.

[35] Johnson, "Caught in Quicksand?"

[36] Ranchordás and Vinci, "Regulatory sandboxes and innovation-friendly regulation."

sandbox edition which targeted public sector innovation[37]. The inclusion of "other stakeholders" in ANPD's definition also signals an openness to wider participation, although it doesn't make clear who they could be.

3. Regulatory Sandboxes core characteristics

Fostering innovation is often cited as the primary goal of regulatory sandboxes and is widely understood as a shared objective between regulators and innovator[38]. Regulatory sandboxes are portrayed as a potential solution to the longstanding regulation versus innovation conundrum, where regulation is frequently perceived as a barrier to technological development[39]. However, this claiming is highly debatable. The State plays a fundamental role in enabling innovation when providing legal certainty and actively supporting innovation through well-designed policies and incentives. In this context, regulatory sandboxes can be viewed as one of these key policy instruments. Nevertheless, relying solely on innovation to justify the establishment of regulatory sandboxes - without a proper plan for their operationalization - can introduce significant risks, including harm to competition and, more critically, to individuals and groups impacted by the technologies being tested[40].

Moreover, defining "innovation" is a challenge itself. As seen in the proposed definitions for regulatory sandboxes, innovation is often associated with the introduction of "new products and services" or "new business models." Yet, these notions offer little practical guidance on how to assess the innovative character of a given technology. The OECD's *Oslo Manual defines innovation as "a new or improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)."*[41] The manual also provides a useful framework for measuring business innovation. However, in practice, regulators rarely possess the necessary resources to conduct a thorough assessment of "innovativeness." Instead, project selection decisions are often made subjectively by the teams managing the sandbox[42].

It is important to consider that innovations tested in regulatory sandboxes may involve different levels of technology maturity. In Norway, Datatilsynet's privacy sandbox has supported projects in different stages of development – ranging from initiatives still in the conceptual design phase[43], to those testing high-fidelity prototypes under simulated or real-world conditions[44]. In Singapore, the

[37] Commission Nationale de l'Informatique et des Libertés, "'Sandbox': CNIL Launches Call for Projects on Artificial Intelligence in Public Services," CNIL, July 28, 2023, https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services.

[38] Davide Baldini and Kate Francis, "AI Regulatory Sandboxes between the AI Act and the GDPR," *CEUR Workshop Proceedings* 3731 (January 2024), https://ceur-ws.org/Vol-3731/paper07.pdf; Fabio Seferi, "A Comparative Analysis of Regulatory Sandboxes from Selected Use Cases : Insights from Recurring Operational Practices," in *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers f or Stakeholders* (Cybersecurity National Lab, 2025), Zotero, https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/; Sky, "Systematic Review of Regulatory Sandboxes"; Katerina Yordanova, "The Shifting Sands of Regulatory Sandboxes for AI," *CiTiP Blog*, July 18, 2019, https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/.

[39] Sofia Ranchordás, "Innovation Experimentalism in the Age of the Sharing Economy," *Lewis & Law Clark Review* 19 (2015), Zotero, https://law.lclark.edu/live/files/21702-lcb194art1ranchordaspdf.

[40] Anastasiia Dardykina, "Is There Enough Sand in the Sandbox?," SSRN Scholarly Paper no. 5126862 (Social Science Research Network, October 1, 2024), https://doi.org/10.2139/ssrn.5126862; Erik Longo and Filippo Bagni, "From Legal Experimentation to Regulatory Sandboxes: The EU's Pioneering Approach to Digital Innovation and Regulation," in *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers for Stakeholders* (Cybersecurity National Lab, 2025), https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/; Ruschemeier, "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation."

[41] Organization for Economic Cooperation and Development, *Oslo Manual 2018 - Guidelines for Collecting, Reporting, and Using Data on Innovation* (OECD, 2018)

[42] McCarthy, "From Childish Things," 14.

[43] See for instance: Datatilsynet, "Ruter: On Track with Artificial Intelligence," Datatilsynet, 2023, https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/ruter-on-track-with-artificial-intelligence/; Datatilsynet, "Simplifai and NVE: Digital Employee," Datatilsynet, 2023, https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/simplifai-and-nve-digital-employee/.

[44] Some examples: Datatilsynet, "The AVT Project: Activity Data for Assessment and Adaptation," Datatilsynet, 2022, https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/the-avt-project-activity-data-for-assessment-and-adaptation/; Datatilsynet, "NAV: Prediction of the Development of Sick Leave," Datatilsynet, 2022, https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/nav-prediction-of-the-development-of-sick-leave/.

Infocomm Media Development Authority – IMDA, has established a Privacy-Enhancing Technology (PET) Sandbox, where most cases focus on testing proof of concepts within a shared data infrastructure[45]. By its turn, the UK Financial Conduct Authority - FCA, requires a certain degree of market readiness as an eligibility criteria for assessing its regulatory sandbox[46], but offers the Digital Sandbox as an alternative tool for firms into earlier stages of development, where they can assess data sets to build and experiment prototype solutions[47].

The EU Digital Europe EUSAiR project, which has been supporting the implementation of AI regulatory sandboxes across the EU, has suggested adopting the Technology Readiness Level (TRL) scale to assess the technology maturity[48]. This scale has been adopted by the European Commission since 2020. TRL 1-2 represent those technologies in the conceptual and planning stages of development; TRL 3-7 progress to functional prototypes, including proofs of concept (TRL 3), laboratory validation (TRL 4), real environment validation (TRL 5), relevant environment validation (TRL 6), and prototype validation (TRL 7). Finally, TRL 8-9 represent the final stages of development, where technology is mature enough to be applied in real-world contexts using real-life data and interacting with end users.

According to EUSAiR, a regulatory sandbox may engage with technologies at different TRLs. earlier TRLs (2-3) may primarily involve regulatory guidance, clarification of compliance expectations and preliminary risk assessments. Later stages would be necessary for testing in real world conditions (at least TRL 4) and for preparing conformity assessments (at least TRL 5). In the context of the AI Act, regulators may prefer to establish testing environments focusing on innovations at more advanced stages of maturity, since this facilitates access to market after the sandbox exit. However, doing so also requires more robust safeguards, since under these circumstances, end-users, consumers and other rightsholders may already be impacted by the technologies being tested, even before their formal market deployment.

Given the inherently subjective nature of innovation, fostering it should not be the sole purpose of establishing a regulatory sandbox. Ranchordás[49] highlights that innovation is, by its very nature, a risky endeavour, characterized by uncertainty and potential societal impacts, both positive and negative. It is therefore the regulator's role to mitigate - or, in extreme cases, prevent - developments that could harm society. When properly established, regulatory sandboxes can contribute to the lawful, sustainable, and ethical development of technologies such as AI systems[50].

3.2. Controlled regulatory framework (experimental approach)

The second key characteristic of regulatory sandboxes is their experimental nature. Sandboxes are one of several experimentation tools adopted in recent years to assess emerging technologies across both technical and legal dimensions. As testing environments, they must offer a structured methodology that reinforces their role as a "safe space", on a twofold perspective: for the innovator, the sandbox should reduce the risk of sanctions when exploring certain uncharted areas; for society, it must identify and address potential risks to individuals and groups[51].

Even within the financial sector, authorities have incorporated considerations of consumer protection into both the entry criteria and the testing processes conducted within sandboxes[52] . For instance, the UK FCA - widely recognized as a pioneer in the development of regulatory sandboxes - has emphasized both the potential benefits to consumers and the need for risk management to

[45] Infocomm Media Development Authority, "Privacy Enhancing Technology Sandboxes," Infocomm Media Development Authority, accessed March 9, 2026, https://www.imda.gov.sg/how-we-can-help/data-innovation/privacy-enhancing-technology-sandboxes.

[46] Financial Conduct Authority, "Regulatory Sandbox," FCA, United Kingdom, 2026, https://www.fca.org.uk/firms/innovation/regulatory-sandbox

[47] Financial Conduct Authority, "Digital Sandbox," FCA, 2025, https://www.fca.org.uk/firms/innovation/digital-sandbox.

[48] EUSAiR, *The EU AI Ecosystem and AI Regulatory Sandboxes: Potential Synergies (Preliminary Assessment)* (EUSAiR, 2025), https://eusair-project.eu/app/uploads/2025/08/EUSAiR_Report_InnovationEcosystem.pdf.

[49] Ranchordás, "Innovation Experimentalism in the Age of the Sharing Economy."

[50] Baldini and Francis, "AI Regulatory Sandboxes between the AI Act and the GDPR."

[51] Genicot, *From Blueprint to Reality*, 10.

[52] Everhart, "The fintech sandbox."

ensure their protection as integral parts of its sandbox methodology[53]. However, the concept of the "consumer" in the financial sector tends to be narrowly defined, typically referring to retail clients, investors and depositors[54]. When considering the diverse contexts of AI applications, societal protection needs a broader scope, to address the rights of stakeholders who may fall outside consumer legal frameworks. Take for instance, data subject rights that encompasses situations such as employment relationships and state-citizens interactions regarding public services[55].

Understanding regulatory sandboxes as controlled frameworks highlights the crucial role of the regulator on coordinating the sandbox. This marks a significant distinction from other experimentation tools such as testbeds and living labs, where regulator involvement is optional and, when present, typically not in a leading role[56]. By contrast, in a regulatory sandbox, the regulator plays a pivotal role in establishing and supervising the framework in which innovations are developed and tested: it is responsible for setting the parameters of the sandbox, overseeing the experimentation process, evaluating its implementation, and assessing the outcomes. This leadership role is also essential to enabling three additional core features of regulatory sandboxes: regulatory flexibility, regulatory guidance, and regulatory learning, which are discussed in further sections.

3.3. Time-limited

The temporary nature of regulatory sandboxes is also understood as one of their key features. Testing periods typically range 6 months to 2 years,[57] although there are exceptional cases in which they have operated for 5 or even 20 years.[58] Seferi notes that testing durations longer than 12 months usually happen in sectors which involve larger-scale trials and infrastructure assessments, such as telecommunications (12 to 24 months), energy (24 to 48 months), and transportation (12 to 60 months)[59]. A regulatory sandbox needs to be time-bound, as it imposes limits on the duration and resources that regulators dedicate to a specific innovation[60]. This helps mitigate the risk of preferential treatment toward a narrow group of regulatees and confines the period during which sandbox participants may operate under lighter regulatory conditions, limiting potential risks of regulatory capture[61].

3.4. Regulatory flexibility

The relaxation of rules is a core characteristic of regulatory sandboxes, designed to foster an innovation-friendly environment and offered as an incentive for innovators to engage in testing in collaboration with regulators[62]. This flexibility is arguably the most controversial feature of sandboxes, as it can introduce risks for stakeholders affected by the technologies being tested, and may create imbalances between those operating within the sandbox and those outside it.

Contrary to common sense, regulatory flexibility does not mean exemptions from liability. Since the launch of its regulatory sandbox, the FCA has never exempted participants from liability for harm

[53] Allen, "Regulatory Sandboxes."

[54] Zetzsche et al., "Regulating a Revolution," 7.

[55] Natali Helberger et al., "The Perfect Match? A Closer Look at the Relationship between Eu Consumer Law and Data Protection Law," *Common Market Law Review* 54, no. 5 (2017), https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2017118.pdf.

[56] European Commission, *Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final*.

[57] 73 of 99 fintech sandbox projects assessed by the World Bank had their testing periods ranging from 6 months to 2 years, with almost half of those (33) with 1 year duration. See Appaya et al., *Global Experiences from Regulatory Sandboxes*, 22.

[58] While Sky came to a similar conclusion as the World Bank, she also identified cases in which testing last 5 to 20 years. See: Sky, "Systematic Review of Regulatory Sandboxes," 32.

[59] "A Comparative Analysis of Regulatory Sandboxes from Selected Use Cases : Insights from Recurring Operational Practices," 157.

[60] Allen, "Regulatory Sandboxes," 638.

[61] Ranchordás and Vinci, "Regulatory sandboxes and innovation-friendly regulation."137.

[62] Bromberg et al., "Fintech Sandboxes: Achieving a Balance between Regulation and Innovation"; Dardykina, "Is There Enough Sand in the Sandbox?"; Ruth Plato-Shinar and Andrew Godwin, "Regulatory Cooperation in AI Sandboxes: Insights from Fintech," *SSRN Scholarly Paper* no. 5199887 (Social Science Research Network, April 1, 2025), https://doi.org/10.2139/ssrn.5199887.

caused during the testing phase[63]. The WBG study also reveals that most jurisdictions explicitly require applicants to demonstrate sufficient resources to cover potential liabilities, particularly for customer-related harms[64]. These findings suggest that regulatory sandboxes are not mechanisms to evade responsibility or legal consequences for third-party harms. In the EU, Regulations enacting provisions on regulatory sandboxes state that these environments should not result on liability exemptions neither affect the supervisory and corrective powers of competent authorities[65].

Rather, the regulatory leeway provided by regulatory sandboxes typically involves a temporary easing of specific regulatory requirements. Based on the reviewed literature, three common methods have been identified. The first one is the issuance of "*no-enforcement letters*" (or "no-action letters"), whereby regulators commit to refraining from enforcement actions under certain conditions[66]. These commitments are usually tailored to each case[67]. However, offering such letters entails a degree of risk for the regulator: if it later revokes its position, the letter could be used as evidence in legal proceedings. Moreover, going back on such commitments could damage the regulator's credibility and jeopardize future participation in the sandbox. In any case, while these letters can serve as strong incentives for innovators, they may also weaken the effectiveness of legal protections intended to safeguard third-party interests[68]. As such, their use must be carefully weighed to avoid undermining public trust and regulatory legitimacy.

A second option involves *restricted authorization*, such as *temporary licenses*, which allow innovators to engage in certain activities within the sandbox[69]. This approach is particularly suited to legal frameworks based on authorization regimes, which are common in sectors like finance, energy, telecommunications, and healthcare[70]. However, other legal regimes, such as data protection, do not rely on a licensing model. Instead, data protection law adopts a risk-based approach, where entities (e.g., data controllers and processors) assess their own compliance by conducting risk management evaluations[71]. These assessments ensure that data processing activities comply with legal requirements, with obligations tailored to the risk level involved.

Third, instead of a temporary licence, regulators may establish specific legal exemptions (i.e. legal derogations), allowing innovators to temporarily deviate from certain legal obligations during the testing phase[72]. These exemptions are typically tailored to the needs of the experimentation process and vary depending on what regulators wish to facilitate. For instance, the Monetary Authority of Singapore (MAS) offers flexible rules in standardization processes for sandbox participation but still enforces strict requirements on customer confidentiality and financial handling; while Canada's Ontario Securities Commission has granted relief on audit and certain disclosure and reporting requirements, but only for services prioritized in their sandbox[73]. These exemptions are generally narrower than temporary licenses, as they waive specific legal duties based on the regulator's risk appetite within the sandbox[74]. Blanket exemptions, however, are highly criticized

[63] Financial Conduct Authority, *Regulatory Sandbox Guide* (FCA, 2022), https://www.fca.org.uk/publication/fca/fca-regulatory-sandbox-guide.pdf.

[64] Appaya et al., *Global Experiences from Regulatory Sandboxes*, 25.

[65] See Article 12(4) and (5) of the Interoperable Europe Act, Article 33(2) of the Cyber Resilience Act, Article 33(4) and (6) of the Net-Zero Industry Act and Article 57(11) and (12) of the EU AI Act.

[66] Thomas Buocz et al., "Regulatory Sandboxes in the AI Act: Reconciling Innovation and Safety?," *Law, Innovation and Technology* 15, no. 2 (2023): 363, https://doi.org/10.1080/17579961.2023.2245678.

[67] Zetzsche et al., "Regulating a Revolution."

[68] Buocz et al., "Regulatory Sandboxes in the AI Act."

[69] Sky, "Systematic Review of Regulatory Sandboxes," 39.

[70] Thiago Moraes, "Regulatory Sandboxes as Tools for Ethical and Responsible Innovation of Artificial Intelligence and Their Synergies with Responsive Regulation," in *The Quest for AI Sovereignty, Transparency and Accountability*, ed. Luca Belli and Walter B. Gaspar (Springer Nature Switzerland, 2025), https://doi.org/10.1007/978-3-032-02762-7_18.

[71] Raphaël Gellert, "Meta Regulation in Data Protection Law: The Risk-Based Approach," in *The Risk-Based Approach to Data Protection*, ed. Raphaël Gellert (Oxford University Press, 2020), 21, https://doi.org/10.1093/oso/9780198837718.003.0006.

[72] Buocz et al., "Regulatory Sandboxes in the AI Act," 363; Giuseppe Mobilio and Matteo Giannelli, "Legal Basis for Regulatory Sandboxes: Key Aspects for a Coherent Theoretical and Practical Framework," in *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers f or Stakeholders* (Cybersecurity National Lab, 2025), 40, Zotero, https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/.

[73] Zetzsche et al., "Regulating a Revolution," 36–37.

[74] Sky, "Systematic Review of Regulatory Sandboxes," 34.

by academics due to the potential risks they introduce, and regulatory sandboxes should not be treated as "regulatory holidays"[75]: the Australian Securities and Investments Commission (ASIC) approach has faced significant criticism and has even been dismissed by some academics as a proper regulatory sandbox.[76] A 2018 report by the European Supervisory Authorities for the EU financial sector concluded that, at the time, fintech sandboxes in the region still required firms to obtain a licence before beginning testing.[77] Flexibility was granted in how supervisory powers were exercised: occasionally through measures like no-enforcement letters, or through proportionality levers related to national rules implementing EU Directives (e.g. where those rules exceeded the EU's minimum requirements), or otherwise where those levers were already embedded in EU legislation.

Two final considerations must be made regarding the flexibility of regulatory sandboxes. First, this leeway requires a statutory basis for its legitimacy,[78] and to be in accordance with the legality principle under the rule of law[79]. These are usually experimental clauses in legislation, which permit temporary deviations from specific legal requirements[80]. These clauses also provide the legal framework for regulators to initiate a sandbox program[81]. In the EU AI Act, Article 57 outlines the general rules for AI regulatory sandboxes and can be understood as the experimental clause which legitimizes these tools to operate and govern how regulators should experiment within these controlled environments. Additional rules may complement the AI Act's provisions, including the Commission's implementing acts, national legislation and government regulations or decisions.[82] Aside from flexibility, these statutory bases also play essential roles for defining scope (legal and technical), type of settings (such as laboratory or real-world conditions), powers and tasks of the regulators in this context[83]. Second, it is important to note that derogatory provisions in a specific sector or regulation do not automatically imply derogations across all applicable legal frameworks[84]. This is particularly relevant in contexts where technologies being tested, such as AI, may impact multiple legal regimes simultaneously, and it is debatable if an experimentation clause can limit supervisory powers of multiple legal domains at once.[85]

3.5. Regulatory guidance

Participating in a regulatory sandbox can offer innovators a unique opportunity to engage in direct dialogue with regulators and receive guidance on regulatory expectations and how to meet legal obligations[86], and some regulatory sandboxes environments have been framed as "regulatory

[75] Allen, "Regulatory Sandboxes"; Didenko, "A Better Model for Australia's Enhanced Fintech Sandbox."

[76] Bromberg et al. argue that ASIC's blanket exemption is inconsistent with the logic that underpins regulatory sandboxes, since this kind of leeway doesn't provide safe spaces designed by the regulator to prevent (or mitigate) consumer harm. See: Bromberg et al., "Fintech Sandboxes: Achieving a Balance between Regulation and Innovation," 17.

[77] European Securities and Markets Authority et al., *FinTech: Regulatory Sandboxes and Innovation Hubs*.

[78] Buocz et al., "Regulatory Sandboxes in the AI Act"; Longo and Bagni, "From Legal Experimentation to Regulatory Sandboxes: The EU's Pioneering Approach to Digital Innovation and Regulation"; Mobilio and Giannelli, "Legal Basis for Regulatory Sandboxes: Key Aspects for a Coherent Theoretical and Practical Framework."

[79] Stefan Philipsen et al., "Legal Enclaves as a Test Environment for Innovative Products: Toward Legally Resilient Experimentation Policies," Regulation & Governance 15, no. 4 (2021): 1135, https://doi.org/10.1111/rego.12375.

[80] Genicot, *From Blueprint to Reality*, 36.

[81] Johnson, "Caught in Quicksand?," 712.

[82] Article 58 of the AI Act requires the Commission to enact implementing acts on further rules of EU AI regulatory sandboxes.

[83] Mobilio and Giannelli, "Legal Basis for Regulatory Sandboxes: Key Aspects for a Coherent Theoretical and Practical Framework."

[84] Esther C. Van Der Waal et al., "Participatory Experimentation with Energy Law: Digging in a 'Regulatory Sandbox' for Local Energy Initiatives in the Netherlands," *Energies* 13, no. 2 (2020): 10, https://doi.org/10.3390/en13020458.

[85] Nathan Genicot and Thiago Guimaraes Moraes, "Exploring the Boundaries of AI Regulatory Sandboxes under the AI Act: Flexibility and Real-World Testing," *Cambridge Forum on AI: Law and Governance* 1 (January 2025): 8, https://doi.org/10.1017/cfl.2025.10013.

[86] Ahmad Alaassar et al., "Exploring How Social Interactions Influence Regulators and Innovators: The Case of Regulatory Sandboxes," *Technological Forecasting and Social Change* 160 (November 2020): 120257, https://doi.org/10.1016/j.techfore.2020.120257; Sofia Ranchordás, "Experimental Lawmaking in the EU: Regulatory Sandboxes," preprint, October 22, 2021, https://doi.org/10.2139/ssrn.3963810; Jon Truby et al., "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications," *European Journal of Risk Regulation* 13, no. 2 (2022): 270–94, https://doi.org/10.1017/err.2021.52;

dialogues"[87]. When assessing the UK fintech ecosystem, Fahy found that the most common motivation for companies to join the FCA regulatory sandbox was to "make sure they were following the law."[88] The sandbox was perceived as an expedited route to market entry, offering a form of legal shelter for product testing and commercialisation. Similarly, Alaassar et al.[89] conducted interviews with regulators and firms and observed a shared understanding that the former could support sandbox participants by identifying regulatory gaps and offering early-stage guidance.

The pursuit of legal certainty is thus widely recognised as a central motivation for firms to participate in regulatory sandboxes. However, while sandboxes may enhance legal certainty for participating firms, they may simultaneously reduce it for those outside the sandbox.[90] In the absence of mechanisms to disseminate lessons learned, competitors may perceive the regulatory flexibility and tailor-made guidance granted to sandbox participants as unfair, potentially leading to legal disputes. Additionally, without adequate risk management, regulatory experimentation may cause severe interferences with the fundamental rights of individuals, and there is also a risk that products exiting the sandbox could be misinterpreted by the public as having been officially endorsed by authorities[91].

To mitigate these risks, Ranchordás argues that regulators should be transparent about the experimental nature of the sandbox[92]. This includes clearly communicating the scope and duration of the experimentation, the evaluation criteria, and the mechanisms for exit. Furthermore, citizens must be able to understand the societal benefits of these experimental regimes.

3.6. Regulatory learning

While regulatory guidance is often seen as the main incentive for innovators to join a sandbox, the primary benefit for regulators lies in gaining a deeper understanding of how new technologies function and how they interact with existing legal frameworks[93]. Baldini and Francis argue that regulatory sandboxes offer regulators the opportunity to acquire technical insights on emerging technologies and their novel applications well before they reach the market, by establishing a new channel for evidence-based regulatory learning[94]. Similarly, Ranchordás affirms that experimental legal instruments contribute to evidence-based lawmaking and promote the continuous reassessment and adaptation of regulation[95]. This links sandboxes to anticipatory regulation, which develops legal responses iteratively to address the "pacing problem": the challenge that technological change outpaces the ability of regulatory systems to respond appropriately[96].

However, the effectiveness of this regulatory learning process is highly dependent on the design and integration of the sandbox in a broader strategy. Without a structured methodology, sandboxes

[86] [cont.] Katerina Yordanova and Natalie Bertels, "Regulating AI: Challenges and the Way Forward Through Regulatory Sandboxes," in *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, ed. Henrique Sousa Antunes et al. (Springer International Publishing, 2024), https://doi.org/10.1007/978-3-031-41264-6_23.

[87] Appaya et al., *Global Experiences from Regulatory Sandboxes*; Genicot, *From Blueprint to Reality*; Ranchordás and Vinci, "Regulatory sandboxes and innovation-friendly regulation.""; Yordanova and Bertels, "Regulating AI."

[88] "Regulator Reputation and Stakeholder Participation: A Case Study of the UK's Regulatory Sandbox for Fintech," *European Journal of Risk Regulation* 13, no. 1 (2022): 138–57, https://doi.org/10.1017/err.2021.44.

[89] Alaassar et al., "Exploring How Social Interactions Influence Regulators and Innovators."

[90] Philipsen et al., "Legal Enclaves as a Test Environment for Innovative Products," 1135.

[91] Truby et al., "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications," 279.

[92] Ranchordás, "Experimental Regulations and Regulatory Sandboxes – Law Without Order?," 18.

[93] Ana Paula Gonzalez Torres and Nitin Sawhney, "Role of Regulatory Sandboxes and MLOps for AI-Enabled Public Sector Services," *The Review of Socionetwork Strategies* 17, no. 2 (2023): 297–318, https://doi.org/10.1007/s12626-023-00146-y; Johnson, "Caught in Quicksand?"; Longo and Bagni, "From Legal Experimentation to Regulatory Sandboxes: The EU's Pioneering Approach to Digital Innovation and Regulation"; Ruschemeier, "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation"; Sky, "Systematic Review of Regulatory Sandboxes"; Yordanova and Bertels, "Regulating AI."

[94] Baldini and Francis, "AI Regulatory Sandboxes between the AI Act and the GDPR," 5.

[95] Ranchordás, "Experimental Regulations for AI," 23.

[96] Emily Leckenby et al., "The Sandbox Approach and Its Potential for Use in Health Technology Assessment: A Literature Review," *Applied Health Economics and Health Policy* 19, no. 6 (2021): 857–69; Gary E. Marchant, "Addressing the Pacing Problem," in *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, ed. Gary E. Marchant et al. (Springer Netherlands, 2011), https://doi.org/10.1007/978-94-007-1356-7_13.

may have a limited learning value with a low outreach of its findings[97]. Therefore, sandboxes must be properly integrated into regulatory and policymaking strategies. If lessons learned are not systematically gathered, shared, and considered for regulatory reform - whether through amendments to existing norms or the creation of new legal frameworks - the sandbox's potential as a tool for evidence-based learning is undermined[98]. A good practice in this regard is the publication of evaluation reports. For example, the UK ICO publishes both individual project exit reports and broader insights reports[99]. These documents contribute to sharing of lessons learned and also work as soft law instruments, providing guidance to the wider community of regulatees beyond the sandbox participants.

Alaassar et al. claim that unwillingness to explore regulatory changes could undermine the potential of sandboxes to work as innovation-friendly regulatory tools[100]. Treating sandboxes merely as "conformity assessment" mechanisms limits their transformative potential and risks creating competition imbalances by offering regulatory guidance exclusively to sandbox participants. On the other hand, evidence-based learning on sandboxes should not become a pathway at deregulation or at reducing protection standards. Rather, it should aim to foster the development of adaptive and appropriate regulatory frameworks that bridge the gap between technological innovation and legal oversight - without compromising fundamental rights, safety, and other societal protections[101].

3.7. Collaborative governance

Last, but not least, regulatory sandboxes are set up as collaborative environments[102]. This means that interactions between regulators and regulatees happen on the common understanding that enforcement is not the order of the day. Of course, for this to be effective all parties engaged in the sandbox must be able to trust each other[103]. Trust is typically built through open, honest, and transparent dialogue: Regulators must demonstrate a genuine interest in identifying and addressing regulatory challenges, rather than focusing on punitive measures[104]. At the same time, safeguards must be in place to prevent regulatory capture and ensure that collaboration remains balanced and accountable. Regulatees must understand that if they stop engaging in good faith, disregard regulatory recommendations, or fail to address identified risks, the experimentation may be swiftly suspended, and in serious cases, enforcement action or liability claims may follow[105].

This collaborative nature is another reason why sometimes regulatory sandboxes are framed as "regulatory dialogues"[106]. Nevertheless, in many cases, periodic reports have become the primary means of communication between regulators and regulatees[107]. While important, relying solely on written reports risks missing out on richer forms of engagement - such as meetings, workshops, and direct interactions - that can significantly enhance evidence-based regulatory learning. The emphasis on dialogue in regulatory sandboxes also requires managing expectations, as innovators may mistakenly assume that regulatory sandboxes always include access to technical infrastructure or financial support. However, this is often not the case. Genicot assessed several privacy sandboxes and found that only Singapore provided financial support or access to data

[97] Ranchordás, "Experimental Regulations and Regulatory Sandboxes – Law Without Order?"

[98] Allen, "Regulatory Sandboxes," 640.

[99] Information Commissioner's Office, "Previous Participants," ICO, November 19, 2024, https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/previous-participants/.

[100] Alaassar et al., "Exploring How Social Interactions Influence Regulators and Innovators," 11.

[101] Ruschemeier, "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation," 323.

[102] Dardykina, "Is There Enough Sand in the Sandbox?"; Plato-Shinar and Godwin, "Regulatory Cooperation in AI Sandboxes"; Sky, "Systematic Review of Regulatory Sandboxes"; Truby et al., "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications."

[103] Alaassar et al., "Exploring How Social Interactions Influence Regulators and Innovators"; Johnson, "Caught in Quicksand?"

[104] Johnson, "Caught in Quicksand?"

[105] RanchRanchordás and Vinci, "Regulatory sandboxes and innovation-friendly regulation."

[106] Appaya et al., *Global Experiences from Regulatory Sandboxes*; Genicot, *From Blueprint to Reality*; Ranchordás and Vinci, "Regulatory sandboxes and innovation-friendly regulation."; Yordanova and Bertels, "Regulating AI."

[107] Sky, "Systematic Review of Regulatory Sandboxes," 41.

infrastructure.[108] Similarly, the WBG's report on fintech sandboxes highlighted their role in fostering continuous dialogue but was unclear on the extent to which they offer shared infrastructure. Also, the UK FCA explicitly states participants must meet certain resource requirements before joining its sandbox, implying that financial or technical support is not provided[109]. Finally, an overview presented by the European Commission on regulatory sandboxes in the energy sector found that only Lithuania's regulatory sandbox offered funding[110].

This does not mean, however, that regulatory sandboxes cannot offer financial benefits. Fahy notes that companies often associate participation in a regulator's sandbox with an enhanced corporate reputation, which can increase their attractiveness to investors[111]. This factor can be particularly attractive to SMEs and startups[112]. McCarthy's shows that participation in fintech sandboxes has led to an average 15% increase in capital raised - especially among smaller and younger firms[113]. Likewise, the FCA reported that 40% of companies that successfully completed its 2015 sandbox program secured investment during of afterward testing[114]. Regulatory sandboxes can also appeal to small businesses and startups, which may lack resources but benefit from facilitated engagement with regulators[115]. Participation is generally free of charge, and sandboxes may be explicitly designed to support small and medium-sized enterprises (SMEs) and startups[116]. In the context of the AI Act, there is an unique opportunity to integrate EU's regulatory sandboxes with the AI innovation ecosystem it is developing through its policies.[117]

A common outcome of this collaborative process is exit reports, usually drafted either by the regulator or by the participating organisation, with the final content mutually agreed upon by both parties[118]. This document generally summarises the objectives, testing process, regulatory considerations, and key findings of the sandbox participation, and often serves as proof that the entity has successfully completed the sandbox programme. Many privacy sandboxes have publicly disclosed such reports to share lessons learned and support broader regulatory learning among stakeholders[119]. In license-based regulated sectors, such as finance, energy or telecommunications, sandbox participation may also convert temporary authorisations granted during testing into permanent permissions[120]. Under the EU AI Act, in addition to an exit report, sandbox participants

[108] Genicot assessed 8 privacy regulatory sandboxes from Brazil, Colombia, Denmark, France, Luxembourg, Norway, Singapore and the UK. See: Genicot, *From Blueprint to Reality*.

[109] Financial Conduct Authority, *Regulatory Sandbox Guide*.

[110] The Commission overview included energy-sector regulatory sandboxes of the following Member States: Belgium, Denmark, Spain, France, Italy, Hungary, Lithuania, Netherlands, Portugal and Sweden. See: European Commission, *Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final*, 121.

[111] Fahy, "Regulator Reputation and Stakeholder Participation," 145.

[112] Antonella Zarra, "Operationalizing AI Regulatory Sandboxes: A Look at the Incentives for Participating Start-Ups and SMEs beyond Compliance," in *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers for Stakeholders* (Cybersecurity National Lab, 2025), https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/.

[113] McCarthy, "From Childish Things," 10.

[114] Financial Conduct Authority, *Regulatory Sandbox Lessons Learned Report* (FCA, 2017), https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf.

[115] Ranchordás, "Experimental Regulations and Regulatory Sandboxes – Law Without Order?"; Zarra, "Operationalizing AI Regulatory Sandboxes: A Look at the Incentives for Participating Start-Ups and SMEs beyond Compliance."

[116] Genicot, *From Blueprint to Reality*; Appaya et al., *Global Experiences from Regulatory Sandboxes*.

[117] Examples include the European Digital Innovation Hubs (EDIHs), Test and Experimentation Facilities (TEFs) AI Factories and data spaces. See: Genicot, *From Blueprint to Reality*.

[118] Genicot, *From Blueprint to Reality*; Fabio Seferi, *A Working Experimentation Model for Cyber Resilience Regulatory Sandboxes*, 2025, https://ceur-ws.org/Vol-3962/paper67.pdf; Sky, "Systematic Review of Regulatory Sandboxes."

[119] Examples include privacy sandboxes from Norway's Datatilsynet (Datatilsynet, "Sandbox - Reports," 2023, https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/reports/.), France's CNIL (Commission Nationale de l'Informatique et des Libertés, "Sandbox," 2025, https://www.cnil.fr/en/tag/Sandbox.), UK's ICO (Information Commissioner's Office, "Previous Participants.") and Singapore's IMDA (Infocomm Media Development Authority and Personal Data Protection Commission, "Privacy Enhancing Technology Sandboxes," Infocomm Media Development Authority, 2022, https://www.imda.gov.sg/how-we-can-help/data-innovation/privacy-enhancing-technology-sandboxes.).

[120] Moraes, "Regulatory Sandboxes as Tools for Ethical and Responsible Innovation of Artificial Intelligence and Their Synergies with Responsive Regulation."

may request written documentation of the activities carried out within the sandbox to support the demonstration of compliance during subsequent conformity assessment procedures[121]. This documentation may contribute to a presumption of compliance with certain regulatory requirements; however, it should not be interpreted as proof of full compliance, as sandbox experimentation may not cover all regulatory requirements under the AI Act.

Beyond regulator–regulatee interaction, regulatory sandboxes also offer a valuable opportunity for collaboration between regulators - particularly in contexts where technologies intersect multiple legal frameworks, whether across sectors at the national level or between jurisdictions[122]. Data-driven technologies are inherently cross-cutting in nature, which necessitates coordinated regulatory approaches that present both opportunities and challenges. For example, fintech solutions have long been recognized for their broad impact across multiple areas of public policy. As a result, it has been recommended they adopted a multidisciplinary supervisory cooperation and knowledge sharing on issues such as competition, fraud, anti-money laundering, cybersecurity, consumer protection, and data protection[123]. In the UK, the financial and data protection authorities – the FCA and the ICO – have been collaborating to facilitate coordinated regulatory support for companies operating across both domains. This cooperation raises firms confidence that they are getting aligned guidance, regardless of which sandbox environment the participate in[124]. The EU AI Act also points into that direction, explicitly stating that national competent authorities should cooperate with other relevant authorities, whenever appropriate, such as when testing involves the processing of personal data or falls within the remit of other sectoral regulators[125].

Regulatory cooperation can promote greater harmonisation and reduce the risk of forum shopping - where innovators choose to enter sectors or jurisdictions perceived as more flexible or "innovation-friendly", creating competition or other regulatory imbalances[126]. One strategy to mitigate this risk is the creation of one-stop shops for sandbox participation, such as the arrangements of fintech sandboxes of the UK and Hong Kong[127]. A similar model was proposed for the Dutch model on AI regulatory sandbox: AI developers and providers would submit queries through a single-entry point, which would screen the requests and refer them either to a central coordination team or to the relevant authority[128]. Only questions deemed appropriate for sandbox experimentation would proceed, allowing innovators and regulators to jointly develop a sandbox plan.

At the cross-border level, there are still relatively few examples, though initiatives like the ASEAN Financial Innovation Network stand out[129]. The EU AI Act may stimulate more cross-border experimentation by explicitly allowing for the joint establishment of regulatory sandboxes.[130] Also, the Interoperable Europe Act requires Union entities and bodies to operate interoperability regulatory sandboxes, with the specific goal of supporting cross-border interoperability of trans-European digital public services[131]. Cross-jurisdiction initiatives should mitigate risks of forum shopping[132].

On the other hand, multi-sector and cross-border sandboxes introduce complexity and can

[121] EU AI Act, Article 57(7)

[122] Alaassar et al., "Exploring How Social Interactions Influence Regulators and Innovators"; John W. Harris, "NC Regulatory Sandbox Act: Encouraging Innovation despite Missing Some Opportunities," *North Carolina Banking Institute* 26 (2022): 301.

[123] Radostina Parenti, *Regulatory Sandboxes and Innovation Hubs for FinTech: Impact on Innovation, Financial Stability and Supervisory Convergence* (European Parliament, 2020), 10, Zotero, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652752/IPOL_ST U(2020)652752_EN.pdf.

[124] Nikil Rathi, "Tech, Trust and Teamwork: How the FCA and ICO Are Helping Innovation Take Off," FCA, Financial Conduct Authority, May 28, 2025, https://www.fca.org.uk/news/blogs/tech-trust-and-teamwork.

[125] EU AI Act, Article 57(4) c/c 57(10).

[126] Zarra, "Operationalizing AI Regulatory Sandboxes: A Look at the Incentives for Participating Start-Ups and SMEs beyond Compliance."

[127] Plato-Shinar and Godwin, "Regulatory Cooperation in AI Sandboxes," 13.

[128] Autoriteit Persoonsgegevens, *Proposal Dutch Regulatory Sandbox* (Autoriteit Persoonsgegevens, 2025).

[129] Appaya et al., Global Experiences from Regulatory Sandboxes, 34.

[130] Article 57(1) and Recital 138 of the AI Act.

[131] Articles 3(14) c/c 11 of the Interoperable Europe Act.

[132] Dardykina, "Is There Enough Sand in the Sandbox?"; Madiega and Van de Pol, *Artificial Intelligence Act and Regulatory Sandboxes*.

increase the burden on regulators - particularly where participation is frequently required. For example, Article 57(10) of the AI Act states that whenever personal data processing is involved, national data protection authorities (DPAs) must be involved in sandbox activities. This could place significant strain on DPAs, many of which are already under-resourced. Similarly, Article 12 of the Interoperable Europe Act requires DPAs (among other regulators) to be engaged in interoperability regulatory sandboxes. Effective engagement in sandboxes requires regulators to possess not only legal expertise (which is expected) but also a basic level of technical understanding to enable meaningful dialogue with innovators[133]. If innovators perceive that regulators are not sufficiently prepared to engage in such discussions, they may be discouraged from participating[134]. Despite these challenges, the EU financial authorities have reported that these initiatives have strengthened cooperation with other regulators, including those outside the financial domain, such as DPAs[135].

Finally, given the collaborative nature of regulatory sandboxes, they appear to offer a suitable environment for engaging a more diverse set of actors. In this regard, the OECD[136] identifies multistakeholder cooperation as a key element for the success of AI regulatory sandboxes. Such cooperation should involve not only coordination among different public authorities but also active engagement with market participants, academic experts, and civil society representatives. Similarly, the World Economic Forum[137] emphasizes that a multistakeholder approach enables collaborative assessment of the socio-economic impacts of emerging technologies such as AI. Nevertheless, existing sandbox experiences suggest that this broader multistakeholder approach remains largely underdeveloped. Beyond the interaction between regulators and regulatees, it is often unclear whether and how other stakeholders, particularly from civil society, are meaningfully involved. If their participation is limited to responding to surveys or providing data, civil society actors are relegated to the role of test subjects rather than active contributors. This raises an important reflection, whether civil society can be meaningfully engaged into these experimentation environments.

Participation of impacted stakeholders in AI and data governance is receiving growing attention, driven by calls for stronger democratic participation and participation equity: these actors should participate more directly in the governance of the technologies that impact them, since this improves general democratic values of plurality, participation, equal representation, and non-discrimination.[138] In this context, approaches such as value sensitive design[139] and participatory design[140] could bring an added value. They promote several stakeholder engagement methods throughout the development of technological artefacts, such as co-design workshops for value elicitation and system requirement specification, surveys and interviews to gather user feedback and iterative development processes that support critical design thinking.

Furthermore, establishing multistakeholder advisory bodies can help embed the perspectives of rights-holders within experimentation environments. This approach has been promoted by the Brazilian DPA[141], the Norwegian DPA[142], and by sandbox initiatives led by civil society organisations (e.g.

[133] Plato-Shinar and Godwin, "Regulatory Cooperation in AI Sandboxes," 9.

[134] Johnson, "Caught in Quicksand?," 718.

[135] European Securities and Markets Authority et al., *Update on the Functioning of Innovation Facilitators – Innovation Hubs and Regulatory Sandboxes* (2023).

[136] Organization for Economic Cooperation and Development, *Regulatory Sandboxes in Artificial Intelligence*.

[137] World Economic Forum, *AI Governance Alliance: Briefing Paper Series 2024* (WEF, 2024), https://www3.weforum.org/docs/WEF_AI_Governance_Alliance_Briefing_Pape r_Series_2024.pdf.

[138] Margot E. Kaminski and Gianclaudio Malgieri, "Impacted Stakeholder Participation in AI and Data Governance," SSRN Scholarly Paper no. 4836460 (Social Science Research Network, May 21, 2024), https://papers.ssrn.com/abstract=4836460.

[139] Batya Friedman et al., "Value Sensitive Design and Information Systems," in *Early Engagement and New Technologies: Opening up the Laboratory*, ed. Neelke Doorn et al. (Springer Netherlands, 2013), https://doi.org/10.1007/978-94-007-7844-3_4.

[140] Susanne Bødker et al., *Participatory Design*, Synthesis Lectures on Human-Centered Informatics (Springer International Publishing, 2022), https://doi.org/10.1007/978-3-031-02235-7.

[141] Agência Nacional de Proteção de Dados, *Consulta à Sociedade - Sandbox Regulatório de IA*.

[142] Tom E. Markussen and Arild Opheim, *Evaluation of the Norwegian Data Protection Authority's Regulatory Sandbox for Artificial Intelligence* (Datatilsynet, 2023), https://www.datatilsynet.no/contentassets/41e268e72f7c48d6b0a177156a815c5b/agenda-kaupang-evaluation-sandbox_english_ao.pdf.

COR Sandbox[143], and IFOW Responsible AI Sandbox[144]). Such mechanisms can establish structured channels for dialogue and oversight, so that experimentation processes remain attentive to societal impacts and fundamental rights protection.

4. Conclusion

Based on the analysis conducted and the key elements identified in the reviewed literature, this paper proposes the following working definition of regulatory sandboxes:

"Regulatory sandboxes are controlled, time-limited experimentation environments that enable coordinating entities to engage with innovators and other actors on testing innovative products, services, or business models, within existing legal regimes, often relying on incentives such as regulatory flexibility and regulatory guidance, to support mutual learning, and foster collaboration among engaged stakeholders."

While no single definition can fully capture the diversity of regulatory sandbox models currently in operation, this paper argues that the proposed definition reflects both how they are commonly understood in the literature and how they ought to be conceived for responsible governance practice. In particular, it emphasises that regulatory sandboxes should not merely facilitate innovation through regulatory relief. Rather, they should work as collaborative spaces oriented towards learning, and dialogue between regulators, innovators, and other relevant stakeholders.

In Table 2, three different examples of regulatory sandbox environments are presented to illustrate how these key characteristics can be identified: the UK FCA fintech sandbox[145], the first known regulator to establish a regulatory sandbox framework; the Norwegian Datatilsynet privacy sandbox[146], which has focused on trustworthy AI in the context of data protection from 2021 to 2025; and the Spanish *Secretaría de Estado de Digitalización y Inteligencia Artificial* (SEDIA) AI sandbo[147], a pilot initiative designed to create lessons for the forthcoming AI Regulatory Sandboxes mandated under the EU AI Act[148].

Table 2: The key characteristics of a Regulatory Sandbox in some specific sandbox environments

Core characteristic	FCA Regulatory fintech sandbox (UK)	Datatilsynet Regulatory privacy sandbox (Norway)	SEDIA Pilot AI Sandbox (Spain)
Innovation-driven	Supports testing of innovative financial products, services, or business models in fintech and related sectors.	Foster privacy-enhancing innovation and responsible data governance, including AI systems	Pilot to test AI systems and implement requirements of the upcoming EU AI Act.
Controlled framework	Firms test innovations in a supervised environment with regulatory oversight and consumer protection safeguards.	Selected organisations experiment with AI projects under close supervision of the DPA	Pilot projects are conducted within a structured testing environment supervised by the Spanish State Secretary on Digitalization and AI (SEDIA).
Time-limited	Testing activities last up to 6 months; Application to exit last up to 24 months	Yearly based cohorts	Testing activities last around 12 months, after participants selection

[143] CORSandbox, "CORSandbox," Sandbox, 2026, https://www.corsandbox.org

[144] Institute for the Future of Work, "Responsible AI Sandbox - IFOW," United Kingdom, 2026, https://www.ifow.org/landing-page/sandbox.

[145] Financial Conduct Authority, "Regulatory Sandbox."

[146] Datatilsynet, "Regulatory Privacy Sandbox," Datatilsynet, accessed July 20, 2024, https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/.

[147] Secretaria de Estado de Digitalización e Inteligencia Artificial - SEDIA, "Sandbox IA - Home," 2025, https://avance.digital.gob.es/sandbox-IA/Paginas/sandbox-IA.aspx.

[148] European Commission, "First Regulatory Sandbox on Artificial Intelligence Presented | Shaping Europe's Digital Future," Brussels, 2022, https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented.

Core characteristic	FCA Regulatory fintech sandbox (UK)	Datatilsynet Regulatory privacy sandbox (Norway)	SEDIA Pilot AI Sandbox (Spain)
Regulatory flexibility	Firms may benefit from tools such as restricted authorisation, waivers, or no-enforcement letters.	Focuses less on waivers and more on interpreting GDPR obligations during experimentation.	Allows experimentation with AI Act compliance processes before full regulatory application.
Regulatory guidance	One of the available tools is individual guidance to help firms understand regulatory requirements during testing.	Provides free regulatory guidance and advisory support to all participating organisations.	Participants receive guidance on implementing AI Act obligations and conformity assessment processes.
Regulatory learning	Insights from sandbox tests support FCA policies on emerging financial technologies.	Lessons from projects on data protection and AI governance good practices are shared in public reports.	Outcomes will provide evidence for supporting future AI executive acts and assess the impact of the AI Act.
Collaborative governance	Interaction between firms, regulators, and sometimes consumers in real-market tests.	Cooperation between DPA, participant organisations, and members of an advisory board. Some projects also engaged with end-users and affected stakeholders such as hospital staff and patients; or school staff, students and legal guardians.	Collaboration between the Spanish government, AI providers, and a panel of experts to develop AI regulatory implementation practices.

Importantly, this definition highlights the broader societal role of regulatory sandboxes. Beyond benefits for participating innovators and authorities, sandboxes should also contribute to the public interest by fostering responsible innovation, anticipating risks to fundamental rights, and guiding future regulatory development. When appropriately designed and implemented, regulatory sandboxes can serve as sites for inclusive experimentation, where diverse perspectives are integrated and safeguards for affected third parties - such as consumers, end-users, and other rightsholders - are meaningfully considered. Framing them as learning-oriented and participatory governance tools can support more transparent, and socially responsive approaches to regulatory experimentation in sandbox environments.

Use of Generative AI

This paper used generative AI in a limited manner for spelling, translation and minor stylistic refinements.

References

Agência Nacional de Proteção de Dados. *Consulta à Sociedade - Sandbox Regulatório de IA*. ANPD, 2023. https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/anpd-sandbox-regulatorio-consulta-bilingue.pdf.

Alaassar, Ahmad, Anne-Laure Mention, and Tor Helge Aas. "Exploring How Social Interactions Influence Regulators and Innovators: The Case of Regulatory Sandboxes." *Technological Forecasting and Social Change* 160 (November 2020): 120257. https://doi.org/10.1016/j.techfore.2020.120257.

Allen, Hilary J. "Regulatory Sandboxes." *George Washington Law Review* 87, no. 3 (2019): 579–645. https://doi.org/10.2139/ssrn.3056993.

Appaya, Sharmista, Helen Luskin Gradstein, and Mahjabeen N. Haji. *Global Experiences from Regulatory Sandboxes*. Fintech Note | No. 8. World Bank, 2020. http://documents.worldbank.org/curated/en/912001605241080935/Global-Experiences-from-Regulatory-Sandboxes.

Autoriteit Persoonsgegevens. *Proposal Dutch Regulatory Sandbox*. Autoriteit Persoonsgegevens, 2025.

Baldini, Davide, and Kate Francis. "AI Regulatory Sandboxes between the AI Act and the GDPR." *CEUR Workshop Proceedings* 3731 (January 2024). https://ceur-ws.org/Vol-3731/paper07.pdf.

Bødker, Susanne, Christian Dindler, Ole S. Iversen, and Rachel C. Smith. *Participatory Design*. Synthesis Lectures on Human-Centered Informatics. Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-02235-7.

Bromberg, L., A. Godwin, and I. Ramsay. "Fintech Sandboxes: Achieving a Balance between Regulation and Innovation." *Journal of Banking and Finance Law and Practice* 28, no. 4 (2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090844.

Buocz, Thomas, Sebastian Pfotenhauer, and Iris Eisenberger. "Regulatory Sandboxes in the AI Act: Reconciling Innovation and Safety?" *Law, Innovation and Technology* 15, no. 2 (2023): 357–89. https://doi.org/10.1080/17579961.2023.2245678.

Business at OECD. *Regulatory Sandboxes for Privacy - Analytical Report.Pdf*. BIAC, 2020. https://25159535.fs1.hubspotusercontent-eu1.net/hubfs/25159535/website/documents/pdf/Digital%20Economy/Regulatory%20Sandboxes%20for%20Privacy%20-%20Analytical%20Report.pdf.

Commission Nationale de l'Informatique et des Libertés. "Sandbox." 2025. https://www.cnil.fr/en/tag/Sandbox.

Commission Nationale de l'Informatique et des Libertés. "'Sandbox': CNIL Launches Call for Projects on Artificial Intelligence in Public Services." CNIL, July 28, 2023. https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services.

CORSandbox. "CORSandbox." Sandbox, 2026. https://www.corsandbox.org.

Council of the European Union. *Regulatory Sandboxes and Experimentation Clauses as Tools for Better Regulation: Council Adopts Conclusions*. EU Council, 2020.

Dardykina, Anastasiia. "Is There Enough Sand in the Sandbox?" SSRN Scholarly Paper No. 5126862. Social Science Research Network, October 1, 2024.

Datatilsynet. "NAV: Prediction of the Development of Sick Leave." Datatilsynet, 2022. https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/nav-prediction-of-the-development-of-sick-leave/.

Datatilsynet. "Regulatory Privacy Sandbox." Datatilsynet. Accessed July 20, 2024. https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/.

Datatilsynet. "Ruter: On Track with Artificial Intelligence." Datatilsynet, 2023. https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/ruter-on-track-with-artificial-intelligence/.

Datatilsynet. "Sandbox - Reports." 2023. https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/reports/.

Datatilsynet. "Simplifai and NVE: Digital Employee." Datatilsynet, 2023. https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/simplifai-and-nve-digital-employee/.

Datatilsynet. "The AVT Project: Activity Data for Assessment and Adaptation." Datatilsynet, 2022. https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/reports/the-avt-project-activity-data-for-assessment-and-adaptation/.

Didenko, Anton. "A Better Model for Australia's Enhanced Fintech Sandbox." *University of New South Wales Law Journal* 44, no. 3 (2021). https://classic.austlii.edu.au/au/journals/UNSWLawJl/2021/38.html.

European Commission. "First Regulatory Sandbox on Artificial Intelligence Presented | Shaping Europe's Digital Future." Brussels, 2022. https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented.

European Commission. *Regulatory Learning in the EU. Guidance on Regulatory Sandboxes, Testbeds, and Living Labs in the EU, with a Focus Section on Energy - (Commission Staff Working Document) - SWD(2023) 277 Final*. European Commission, 2023. https://data.consilium.europa.eu/doc/document/ST-12199-2023-INIT/en/pdf.

EUSAiR. *The EU AI Ecosystem and AI Regulatory Sandboxes: Potential Synergies (Preliminary Assessment)*. EUSAiR, 2025. https://eusair-project.eu/app/uploads/2025/08/EUSAiR_Report_InnovationEcosystem.pdf.

Everhart, Jonathan. "The Fintech Sandbox: An Overview of Regulatory Sandbox Regimes." *Southern Journal of Business and Ethics* 12 (2020): 64–73.

Fahy, Lauren. "Regulator Reputation and Stakeholder Participation: A Case Study of the UK's Regulatory Sandbox for Fintech." *European Journal of Risk Regulation* 13, no. 1 (2022): 138–57. https://doi.org/10.1017/err.2021.44.

Financial Conduct Authority. "Digital Sandbox." FCA, 2025. https://www.fca.org.uk/firms/innovation/digital-sandbox.

Financial Conduct Authority. "Regulatory Sandbox." FCA, United Kingdom, 2026. https://www.fca.org.uk/firms/innovation/regulatory-sandbox.

Financial Conduct Authority. *Regulatory Sandbox Guide*. FCA, 2022. https://www.fca.org.uk/publication/fca/fca-regulatory-sandbox-guide.pdf.

Financial Conduct Authority. *Regulatory Sandbox Lessons Learned Report*. FCA, 2017. https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf.

Fonstad, Nils O., and Jeanne W. Ross. "Learning How to Test and Learn | *MIT CISR." MIT Center for Information Systems Research*, February 15, 2018. https://cisr.mit.edu/publication/2018_0201_TestAndLearn_FonstadRoss.

Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Huldtgren. "Value Sensitive Design and Information Systems." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman. Springer Netherlands, 2013. https://doi.org/10.1007/978-94-007-7844-3_4.

Gangale, F., A. Mengolini, L. Covrig, S. Chondrogiannis, and R. Shortall. *Making Energy Regulation Fit for Purpose: State of Play of Regulatory Experimentation in the EU : Insights from Running Regulatory Sandboxes* . European Commission - Joint Research Centre, 2023. DOI.org (CSL JSON). https://data.europa.eu/doi/10.2760/32253.

Gellert, Raphaël. "Meta Regulation in Data Protection Law: The Risk-Based Approach." In *The Risk-Based Approach to Data Protection*, edited by Raphaël Gellert. Oxford University Press, 2020. https://doi.org/10.1093/oso/9780198837718.003.0006.

Genicot, Nathan, and Thiago Guimaraes Moraes. "Exploring the Boundaries of AI Regulatory Sandboxes under the AI Act: Flexibility and Real-World Testing." *Cambridge Forum on AI: Law and Governance* 1 (January 2025): e36. https://doi.org/10.1017/cfl.2025.10013.

Gonzalez Torres, Ana Paula, and Nitin Sawhney. "Role of Regulatory Sandboxes and MLOps for AI-Enabled Public Sector Services." *The Review of Socionetwork Strategies* 17, no. 2 (2023): 297–318. https://doi.org/10.1007/s12626-023-00146-y.

Harris, John W. "NC Regulatory Sandbox Act: Encouraging Innovation despite Missing Some Opportunities." *North Carolina Banking Institute* 26 (2022): 301.

Helberger, Natali, Frederik Zuiderveen Borgesius, and Agustin Reyna. "The Perfect Match? A Closer Look at the Relationship between Eu Consumer Law and Data Protection Law." *Common Market Law Review* 54, no. 5 (2017). https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2017118.pdf.

Infocomm Media Development Authority. "Privacy Enhancing Technology Sandboxes." Infocomm Media Development Authority. Accessed March 9, 2026. https://www.imda.gov.sg/how-we-can-help/data-innovation/privacy-enhancing-technology-sandboxes.

Infocomm Media Development Authority and Personal Data Protection Commission. "Privacy Enhancing Technology Sandboxes." Infocomm Media Development Authority, 2022. https://www.imda.gov.sg/how-we-can-help/data-innovation/privacy-enhancing-technology-sandboxes.

Information Commissioner's Office. "Previous Participants." ICO, November 19, 2024. https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/previous-participants/.

Institute for the Future of Work. "Responsible AI Sandbox - IFOW." United Kingdom, 2026. https://www.ifow.org/landing-page/sandbox.

Johnson, Walter G. "Caught in Quicksand? Compliance and Legitimacy Challenges in Using Regulatory Sandboxes to Manage Emerging Technologies." *Regulation & Governance* 17, no. 3 (2023): 709–25. https://doi.org/10.1111/rego.12487.

Kaminski, Margot E., and Gianclaudio Malgieri. "Impacted Stakeholder Participation in AI and Data Governance." SSRN Scholarly Paper No. 4836460. Social Science Research Network, May 21, 2024. https://papers.ssrn.com/abstract=4836460.

Leckenby, Emily, Dalia Dawoud, Jacoline Bouvy, and Páll Jónsson. "The Sandbox Approach and Its Potential for Use in Health Technology Assessment: A Literature Review." *Applied Health Economics and Health Policy* 19, no. 6 (2021): 857–6

Longo, Erik, and Filippo Bagni. "From Legal Experimentation to Regulatory Sandboxes: The EU's Pioneering Approach to Digital Innovation and Regulation." In *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers for Stakeholders*. Cybersecurity National Lab, 2025. https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/

Longo, Francesco, Antonio Padovano, and Steven Umbrello. "Value-Oriented and Ethical Technology Engineering in Industry 5.0: A Human-Centric Perspective for the Design of the Factory of the Future." *Applied Sciences* 10, no. 12 (2020): 12. https://doi.org/10.3390/app10124182.

Madiega, Tambiana, and Anne Louise Van de Pol. *Artificial Intelligence Act and Regulatory Sandboxes*. European Parliament Research Service, 2022.

Marchant, Gary E. "Addressing the Pacing Problem." In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, edited by Gary E. Marchant, Braden R. Allenby, and Joseph R. Herkert. Springer Netherlands, 2011. https://doi.org/10.1007/978-94-007-1356-7_13.

Markussen, Tom E., and Arild Opheim. *Evaluation of the Norwegian Data Protection Authority's Regulatory Sandbox for Artificial Intelligence*. Datatilsynet, 2023. https://www.datatilsynet.no/contentassets/41e268e72f7c48d6b0a177156a815c5b/agenda-kaupang-evaluation-sandbox_english_ao.pdf

McCarthy, Jonathan. "From Childish Things: The Evolving Sandbox Approach in the EU's Regulation of Financial Technology." *Law, Innovation and Technology* 15, no. 1 (2023): 1–24. https://doi.org/10.1080/17579961.2023.2184131.

Mobilio, Giuseppe, and Matteo Giannelli. "Legal Basis for Regulatory Sandboxes: Key Aspects for a Coherent Theoretical and Practical Framework." In *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers f or Stakeholders*. Cybersecurity National Lab, 2025. Zotero. https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/.

Moraes, Thiago. "Regulatory Sandboxes as Tools for Ethical and Responsible Innovation of Artificial Intelligence and Their Synergies with Responsive Regulation." In *The Quest for AI Sovereignty, Transparency and Accountability*, edited by Luca Belli and Walter B. Gaspar. Springer Nature Switzerland, 2025. https://doi.org/10.1007/978-3-032-02762-7_18.

Organization for Cooperation and Economic Development. *Regulatory Sandboxes in Artificial Intelligence*. OECD, 2023. https://doi.org/10.1787/8f80a0e6-en.

Organization for Economic Cooperation and Development. *Oslo Manual 2018 - Guidelines for Collecting, Reporting, and Using Data on Innovation*. OECD, 2018. https://www.oecd.org/en/publications/oslo-manual-2018_9789264304604-en.html.

Parenti, Radostina. *Regulatory Sandboxes and Innovation Hubs for FinTech: Impact on Innovation, Financial Stability and Supervisory Convergence*. European Parliament, 2020. Zotero. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652752/IPOL_ST_U(2020)652752_EN.pdf.

Philipsen, Stefan, Evert F. Stamhuis, and Martin de Jong. "Legal Enclaves as a Test Environment for Innovative Products: Toward Legally Resilient Experimentation Policies." *Regulation & Governance* 15, no. 4 (2021): 1128–43. https://doi.org/10.1111/rego.12375

Plato-Shinar, Ruth, and Andrew Godwin. "Regulatory Cooperation in AI Sandboxes: Insights from Fintech." SSRN Scholarly Paper No. 5199887. Social Science Research Network, April 1, 2025. https://doi.org/10.2139/ssrn.5199887

Ranchordás, Sofia. "Experimental Lawmaking in the EU: Regulatory Sandboxes." Preprint, October 22, 2021. https://doi.org/10.2139/ssrn.3963810

Ranchordás, Sofia. "Experimental Regulations and Regulatory Sandboxes – Law Without Order?" *Law and Method*, 2021. https://www.boomportaal.nl/tijdschrift/LaM/lawandmethod-D-21-00012.

Ranchordás, Sofia. "Innovation Experimentalism in the Age of the Sharing Economy." *Lewis & Law Clark Review* 19 (2015). Zotero. https://law.lclark.edu/live/files/21702-lcb194art1ranchordaspdf.

Ranchordás, Sofía, Vinci, Valeria, 2024. Regulatory sandboxes and innovation-friendly regulation: Between collaboration and capture. Italian Journal of Public Law 1. https://doi.org/10.2139/ssrn.4696442

Rathi, Nikil. "Tech, Trust and Teamwork: How the FCA and ICO Are Helping Innovation Take Off." FCA, Financial Conduct Authority, May 28, 2025. https://www.fca.org.uk/news/blogs/tech-trust-and-teamwork.

Ruschemeier, Hannah. "Thinking Outside the Box? Regulatory Sandboxes as a Tool for AI Regulation." In *Bridging the Gap Between AI and Reality*, edited by Bernhard Steffen. Springer Nature Switzerland, 2024. https://doi.org/10.1007/978-3-031-73741-1_20.

Sánchez, Sergio, Laura Delgado, Miguel Gimeno-González, Teresa Martín García, Fernando Almaraz-Menendez, and Camilo Ruiz. "Augmented Reality Sandbox: A Platform for Educative Experiences." November 2, 2016, 602. https://doi.org/10.1145/3012430.3012580.

Secretaría de Estado de Digitalización e Inteligencia Artificial - SEDIA. "Sandbox IA - Home." 2025. https://avance.digital.gob.es/sandbox-IA/Paginas/sandbox-IA.aspx.

Seferi, Fabio. "A Comparative Analysis of Regulatory Sandboxes from Selected Use Cases : Insights from Recurring Operational Practices." In *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers f or Stakeholders*. Cybersecurity National Lab, 2025. Zotero. https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/.

Seferi, Fabio. *A Working Experimentation Model for Cyber Resilience Regulatory Sandboxes*. 2025. https://ceur-ws.org/Vol-3962/paper67.pdf.

Sky, Nova. "Systematic Review of Regulatory Sandboxes : Implications for the European Union's Artificial Intelligence Act." Diplomityö, N. Sky, 2024. https://oulurepo.oulu.fi/handle/10024/50797.

The Datasphere Initiative. *Sandboxes for DPI*. The Datasphere Initiative, 2026. https://www.thedatasphere.org/wp-content/uploads/2026/02/Report-Sandboxes-for-DPI-Datasphere-Intiative-2026.pdf.

Truby, Jon, Rafael Dean Brown, Imad Antoine Ibrahim, and Oriol Caudevilla Parellada. "A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications." *European Journal of Risk Regulation* 13, no. 2 (2022): 270–94. https://doi.org/10.1017/err.2021.52.

Undheim, Kristin, Truls Erikson, and Bram Timmermans. "True Uncertainty and Ethical AI: Regulatory Sandboxes as a Policy Tool for Moral Imagination." AI and Ethics 3, no. 3 (2023): 997–1002. https://doi.org/10.1007/s43681-022-00240-x.

Van Der Waal, Esther C., Alexandra M. Das, and Tineke Van Der Schoor. "Participatory Experimentation with Energy Law: Digging in a 'Regulatory Sandbox' for Local Energy Initiatives in the Netherlands." *Energies* 13, no. 2 (2020): 458. https://doi.org/10.3390/en13020458.

World Economic Forum. *AI Governance Alliance: Briefing Paper Series 2024*. WEF, 2024. https://www3.weforum.org/docs/WEF_AI_Governance_Alliance_Briefing_Pape r_Series_2024.pdf.

Yordanova, Katerina. "The Shifting Sands of Regulatory Sandboxes for AI." *CiTiP Blog*, July 18, 2019. https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/.

Yordanova, Katerina, and Natalie Bertels. "Regulating AI: Challenges and the Way Forward Through Regulatory Sandboxes." In *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, edited by Henrique Sousa Antunes, Pedro Miguel Freitas, Arlindo L. Oliveira, Clara Martins Pereira, Elsa Vaz de Sequeira, and Luís Barreto Xavier. Springer International Publishing, 2024. https://doi.org/10.1007/978-3-031-41264-6_23.

Zarra, Antonella. "Operationalizing AI Regulatory Sandboxes: A Look at the Incentives for Participating Start-Ups and SMEs beyond Compliance." In *Regulatory Sandboxes for AI and Cybersecurity. Questions and Answers for Stakeholders*. Cybersecurity National Lab, 2025. https://cybersecnatlab.it/white-paper-on-regulatory-sandboxes/.

Zetzsche, D., Buckley, R., Barberis, J., Arner, D., 2017. Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. Fordham Journal of Corporate & Financial Law 23, 31.

Los datos genéticos: un estudio crítico de su pasado, presente y futuro[1]

Genetic Data: A Critical Study of Its Past, Present, and Future

Mikel Recuero
Universidad del País Vasco (EHU)

Resumen:

Los avances en las capacidades de computación y la reducción drástica de los costes de secuenciación han permitido que, en unas pocas décadas, el tratamiento de datos genéticos se haya ido proyectando a una pluralidad de disciplinas y finalidades otrora inimaginables. No es de extrañar, pues, que este potencial de los datos genéticos, junto con los posibles riesgos derivados de su tratamiento, hayan sido objeto de reconocimiento legal expreso no solo por una multiplicidad de instrumentos sectoriales, también, por las propias normas de protección de datos. El presente artículo tiene por objeto llevar a cabo un recorrido por el pasado, presente y futuro de los datos genéticos, planteando una lectura crítica de su actual configuración legal, sus certezas e interrogantes; y anticipando algunas de las discusiones jurídico-científicas más relevantes.

Palabras clave:

Datos genéticos; datos genómicos; categorías especiales de datos; ADN; datos sensibles.

Sumario:

Abstract:

Progress in computing and data processing capabilities, coupled with massive reductions in sequencing costs, have led to the expansion of genetic data processing into a multitude of areas and applications that were inconceivable just a few decades ago. It is therefore not surprising that the potential of genetic data, together with the possible risks arising from its processing, are the subject of express legal recognition not only in a multitude of sectoral instruments, but also in data protection regulations. This paper provides an insight into the past, present and future of genetic data by critically examining its actual legal status, and anticipating some of the key legal and scientific controversies and discussions.

Keywords:

Genetic data; genomic data; special categories of data; DNA; sensitive data.

Summary:

[1] Este trabajo se ha realizado en el marco del proyecto IMPaCT-Data 2 – PDI (Exp. PMP24/00024) que ha sido financiado por el Instituto de Salud Carlos III (ISCIII) y co-financiado por la Unión Europea – NextGenerationEU. El autor quiere agradecer asimismo el apoyo institucional prestado por el Grupo de Investigación en Ciencias Sociales y Jurídicas aplicadas a las Nuevas Tecnociencias (GI-CISJANT) de la Universidad del País Vasco (EHU), con número de referencia IT 1541-22.

1. Introducción y conceptos esenciales

Los avances en las capacidades de computación y la reducción drástica de los costes de secuenciación han permitido que, en unas pocas décadas, el tratamiento de datos genéticos se haya ido proyectando a una pluralidad de disciplinas y finalidades otrora inimaginables. No es más que la punta del iceberg de un conocimiento genético que no dejará de incrementarse al abrigo de un incesante progreso científico y tecnológico. Sin ánimo de exhaustividad, este exponencial incremento en las posibilidades y potencial de los datos genéticos puede constatarse en cuatro grandes áreas. En primer lugar, en la prestación de la asistencia sanitaria, con aplicaciones tales como la medicina personalizada de precisión, para ofrecer un diagnóstico o un tratamiento adaptados a la composición genética de cada paciente; incluyendo la farmacogenómica, que puede contribuir a anticipar la respuesta de cada individuo ante un fármaco o producto sanitario. En segundo lugar, en el campo de la medicina legal y forense, por ejemplo, para la identificación de cadáveres o en el marco de la investigación criminal. En tercer lugar, en el desarrollo de productos y la prestación de servicios directamente a las personas consumidoras, tales como pruebas genéticas dirigidas al público general (conocidas en inglés como *direct-to-consumer* o «DTC») y ofertadas mayormente a través de plataformas en línea. Finalmente, en el ámbito de la investigación científica, propiciando el surgimiento de áreas tan prolíficas como la investigación genética y la genómica.

No es de extrañar, pues, que este potencial de los datos genéticos, junto con los posibles riesgos derivados de su tratamiento, hayan sido objeto de reconocimiento legal expreso no solo por una multiplicidad de instrumentos sectoriales, también, por las normas de protección de datos. Los datos genéticos conforman hoy en el Reglamento (UE) 2016/679 General de Protección de Datos («RGPD»)[2], una categoría especial dotada de entidad y autonomía propias, si bien no siempre fue así. La convivencia entre la normativa sectorial y las disposiciones en materia de protección de datos no siempre ha sido (ni es) pacífica debido a la coexistencia de fórmulas legales dispares e incluso contradictorias. Una relación que bien puede terminar tornándose todavía más controversial a medida que avance el estado del arte científico-técnico y ello permita ahondar en el conocimiento de las funciones —muchas, hoy en día ignotas— del ADN y los genes.

Ahora bien, al aludir a las nociones de datos genéticos, ADN, gen o genoma, se están manejando una serie de conceptos y fundamentos propiamente extrajurídicos, que a menudo resultan complejos de comprender para las y los profesionales de la privacidad y la protección de datos. Es por ello por lo que, como prefacio al estudio de la categoría de datos genéticos de carácter personal, procede realizar una breve aproximación a algunos de estos conceptos esenciales. En gran medida, porque muchos de los actuales debates jurídicos encuentran su origen y razón de ser en discusiones científicas subyacentes, siendo así que los datos genéticos constituyen un paradigma inigualable de cómo se han abordado legalmente algunos de los más disruptivos avances científicos y tecnológicos.

El ácido desoxirribonucleico (ADN) es la molécula que contiene toda la información genética de un ser vivo y que se encuentra distribuida en fragmentos o cromosomas. La estructura del ADN, tras su descubrimiento por Watson y Crick, se compone de dos cadenas helicoidales enrolladas alrededor de un mismo eje. Cada una de las cadenas o hilos se compone, a su vez, de una sucesión secuencial de cuatro bases nucleótidas: adenina (A), timina (T), citosina (C) y guanina (G).

El gen, procedente del griego *génos* (generación, linaje) es la sucesión específica de bases del ADN que constituye la unidad funcional para la transmisión de la información hereditaria. Los genes contienen las instrucciones sobre la composición biológica de un ser vivo y determinan, en gran medida, su nacimiento, sus características y su desarrollo. No obstante, no todo el ADN está integrado por genes. Tan solo una parte del ADN, llamada «ADN codificante», contiene información sobre la herencia, por lo que existen muchas secuencias de bases nucleótidas que cumplen funciones no vinculadas directamente con la transmisión de la herencia o cuya función resulta, simplemente, desconocida. A estas últimas se les denomina «ADN no codificante» y presentan otros usos tales como la identificación forense o la medicina legal, aunque es probable que su utilidad siga incrementándose a medida que se ahonde en el descubrimiento científico de su significación y potenciales funciones.

Por su parte, se denomina genoma, procedente del sufijo griego *-oma* (conjunto, masa) a todo el conjunto de ADN de un organismo, incluidos sus genes. Los seres humanos comparten aproximadamente el 99,9% de la secuencia de ADN y su orden. Por lo tanto, solo el 0,1% restante es diferente, lo que es sin embargo suficiente para la individualización de las personas y su distinción

[2] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. DO L119 04/5/2016, pp. 1-88.

de los demás individuos de su especie. Esta ínfima variación en el orden de la secuencia genética determina las características físicas y biológicas que hacen a una persona diferente de otra y puede influir en su salud y en la predisposición genética a ciertas enfermedades. El estudio de estas variaciones puede realizarse, por un lado, mediante test o análisis específicamente dirigidos a genes o regiones particulares y, por otro, por medio de la secuenciación del genoma completo. Claro está, esto último representa el análisis más detallado del ADN de una persona y, en consecuencia, el tratamiento de una mayor cantidad y entidad de datos.

Así pues, los anteriores conceptos resultan fundamentales a la hora de trazar una clara línea distintiva entre datos genéticos y datos genómicos que es, a la postre, trascendental a efectos de la protección de datos personales:

- i) Datos genéticos. La genética consiste en el estudio de los genes, es decir, de los factores hereditarios, sus funciones y sus efectos[3]. Por lo tanto, los datos genéticos son *stricto sensu* aquellos que se refieren a la información genética, es decir, al código o secuencia genética como tal. Representan el elemento inmaterial de los genes (la información), en contraposición al elemento material (la base física, la molécula de ADN), siendo el primero de ellos el relevante a efectos de la protección de datos de carácter personal[4].

- ii) Datos genómicos. La genómica consiste en el estudio del genoma de una persona, incluyendo las interacciones entre genes y factores ambientales[5]. Por consiguiente, los datos genómicos son *lato sensu* aquellos que se refieren a la información genómica, es decir, al conjunto del ADN, al genoma completo, y a sus interacciones con el entorno y con otros factores socioeconómicos, ambientales o conductuales. De este modo, la categoría de datos genómicos evoca una amplitud que no siempre es fácil de delimitar, siquiera en el plano científico, razón que lleva a algunos autores a sostener que los datos genómicos son «intrínsecamente *Big Data*»[6].

La anterior distinción resulta, como se anticipaba, muy pertinente a efectos legales, en la medida en que las normas de protección de datos (como es el caso del vigente RGPD) regulan y se refieren al tratamiento de datos genéticos (incluyendo los fenotípicos y los genotípicos[7]), pero no así a los datos genómicos como tal. Ahora bien, como se verá seguidamente, esta opción legislativa no ha significado que el tratamiento de datos genómicos haya quedado despojado de una protección legal reforzada, pues si bien solo una parte de los datos genómicos puede subsumirse en la categoría especial de datos genéticos, nada impide que las restantes informaciones que sean objeto de tratamiento puedan llegar a acomodarse en otras categorías especiales existentes, en particular, en la categoría amplia (omnicomprensiva) de datos relativos a la salud.

El presente artículo tiene por objeto llevar a cabo un recorrido por el pasado, presente y futuro de los datos genéticos en tanto categoría especial de datos personales, planteando una lectura crítica de su actual configuración legal y anticipando algunas de las discusiones jurídico-científicas más relevantes. A tales efectos, partiendo de los conceptos y fundamentos básicos arriba enunciados, se realizará un breve exordio relativo a la sensibilidad de los datos genéticos, incluyendo sus características singulares, como argumentos para su protección legal reforzada (epígrafe 2). Tras ello, se emprenderá un recorrido crítico desde los orígenes normativos de esta categoría especial de datos (epígrafe 3) a su actual configuración jurídica en el RGPD, sus certezas e interrogantes (epígrafe 4). Finalmente, se esbozan algunas de las perspectivas normativas del futuro de los datos genéticos, junto con una serie de consideraciones finales (epígrafe 5).

[3] Francis Collins et al., "A vision for the future of genomics research", Nature 422 (2003):839. DOI https://doi.org/10.1038/nature01626

[4] Pilar Nicolás, La protección jurídica de los datos genéticos de carácter personal (Granada: Comares, 2006), 4-5. Collins et al., "A vision for the future of genomics research", 839.

[5] Collins et al., "A vision for the future of genomics research", 839.

[6] Paul Quinn y Liam Quinn, "Big genetic data and its big data protection challenges", Computer Law & Security Review 34, no.5 (2018):1000-1018. DOI https://doi.org/10.1016/j.clsr.2018.05.028

[7] Los datos genéticos pueden contener, además, información sobre un determinado código o composición genética (genotipo) e información observable sobre la concreta significación de ese código o composición (fenotipo).

2. La singularidad y la sensibilidad de los datos genéticos como argumentos para su protección legal reforzada

La particular naturaleza y funcionamiento de los genes confieren *per se* a la información genética una condición singular que ha sido invocada tradicionalmente como fundamento del carácter sensible de los datos genéticos y, eventualmente, como un poderoso argumento para su protección legal reforzada. En líneas generales, son cinco los rasgos que convierten a los datos genéticos en una categoría de datos singular:

- i) Son informaciones únicas. Definen e individualizan a una persona, distinguiéndola de otros individuos de su especie. Si bien la mayor parte del genoma es compartido, las diferencias aparentemente insignificantes convierten a cada individuo en un ser genéticamente irrepetible. Son, por ende, reflejo de la individualidad de la persona y dan cuenta de su identidad genética única y singular[8].

- ii) Son informaciones generacionales. Pese a ser únicos y singulares, en la medida en que gran parte del genoma es compartido, los datos genéticos pueden revelar vínculos e informaciones sobre la familia biológica, comprendida la descendencia, e incluso sobre el grupo o comunidad al que pertenezca la persona en cuestión[9]. Asimismo, por su carácter hereditario, pueden revelar información sobre los ancestros o las generaciones venideras.

- iii) Son informaciones estructurales, permanentes e inalterables. La información genética no varía normalmente (a salvo de mutaciones) a lo largo de la vida de la persona. Asimismo, dado que el ADN está presente en todas las células del organismo, su soporte es indestructible: acompaña a la persona a lo largo de toda su vida e incluso tras su muerte[10].

- iv) Tienen capacidades predictivas. Permiten predecir enfermedades antes incluso de la aparición de síntomas o, cuanto menos, indicar la susceptibilidad o predisposición genética de una persona con respecto a la enfermedad[11]. En cierto modo, sin perjuicio de la evidencia existente y creciente en relación con los orígenes multifactoriales, los datos genéticos contienen «una suerte de diario futuro de cada individuo que describe, con mayor o menor probabilidad o certeza, una parte importante de su porvenir, de su pasado y de su presente»[12].

- v) Tienen un carácter no voluntario. No escogemos nuestros genes. De esta forma, los datos genéticos y la información que pudieran revelar no dependen de la voluntad del individuo, pues habrán sido transmitidos por sus progenitores sin que el portador llegue a ser tan siquiera conocedor de su concreta significación[13].

Los anteriores cinco rasgos caracterizadores podrían ser argumentos más que suficientes para otorgar a los datos genéticos de carácter personal la condición de sensibles y, en consecuencia, una protección legal reforzada en lo que respecta a su tratamiento. Sin embargo, la sensibilidad tiende a constituir un atributo subjetivo y relativo, de modo que la naturaleza intrínsecamente singular de la información genética no debería aducirse como razón jurídica incontrovertible para justificar el carácter sensible de los datos genéticos en todo caso. Dicho de otro modo, el argumento legislativo para una protección legal reforzada de los datos genéticos de carácter personal no ha sido solamente el de la singularidad[14]. Como sucede con el resto de las categorías

[8] Santiago Grisolía. "Introducción científica", en El Derecho ante el proyecto genoma humano, ed. Fundación BBV, vol.1 (Bilbao: Fundación BBV, 1994), 35.

[9] Art. 4, letra a) de la Declaración Internacional sobre los Datos Genéticos Humanos de la UNESCO.

[10] Ayday, Erman et al., "Whole Genome Sequencing: Revolutionary Medicine or Privacy Nightmare?", Computer 48, no.2 (2015): 63. DOI

[11] Art. 4, letra a) de la Declaración Internacional sobre los Datos Genéticos Humanos de la UNESCO.

[12] George Annas, Leonard Glantz y Patricia Roche, "Drafting the Genetic Privacy Act: Science, Policy, and Practical Considerations", Journal of Law, Medicine & Ethics 23, no.4, (1995):360-366. DOI https://doi.org/10.1111/j.1748-720x.1995.tb01378.x

[13] Carlos María Romeo Casabona, Los genes y sus leyes: el derecho ante el genoma humano (Granada: Comares, 2002), 63.

[14] Como apunta JOVE, "más allá de si es por naturaleza o por contexto (...) el fundamento de las categorías especiales reside en su potencial para generar un mayor riesgo de afectación de los derechos y libertades fundamentales". Daniel Jove Villares, La protección de lo sensible, o cuando la naturaleza del dato no lo es todo (Valencia: Tirant lo Blanch, 2023), 313.

especiales de datos, el argumento de peso no es otro que el grado de afectación que el tratamiento de datos (genéticos) de carácter personal pueda desencadenar sobre los derechos y libertades fundamentales de las personas interesadas[15].

En concreto, el tratamiento de datos genéticos de carácter personal tiene el potencial de afectar a los derechos fundamentales a la intimidad y a la vida privada. Pueden revelar una enorme cantidad y variedad de informaciones pertenecientes a la esfera más íntima y privada de una persona: aspectos relativos a la salud presente, pasada y futura de la persona o de sus familiares, origen étnico o racial e incluso aspectos relacionados con la vida sexual o reproductiva[16]. Es más, el Tribunal Europeo de Derechos Humanos (en adelante, «TEDH») ha llegado a proclamar que la injerencia en el derecho a la vida privada no solo se produciría en el momento de la recogida y tratamiento de los datos genéticos, también, en el instante mismo de la extracción y retención de las muestras biológicas o los perfiles de ADN[17].

Por otra parte, cabe referirse sucintamente al derecho a la no discriminación, en este caso, por razones fundadas en características genéticas. La discriminación genética tiene lugar cuando el carácter predictivo de la información genética se invoca para negar un tratamiento igualitario a una persona y su familia o a todo un colectivo o comunidad[18]. En este sentido, las prácticas discriminatorias pueden manifestarse en una pluralidad de ámbitos (laboral, asistencial, financiero, asegurador, etc.) y podrían acaecer, asimismo, en dos niveles: el individual (v.g. un test genético que identifica la susceptibilidad a una enfermedad y que se emplea para denegar la suscripción de un seguro) o el colectivo (v.g. una investigación que determina que un grupo étnico presenta una mayor predisposición genética a ciertas enfermedades y ello termina por estigmatizar a sus integrantes)[19]. Por todo ello, se ha venido reclamando no solo su protección a nivel *iusfudamental*, en clave de injerencias transversales en el derecho a la no discriminación, también por medio de una normativa específica en la que la protección de datos juega un papel destacado[20].

Por último, no cabe olvidar el tratamiento de datos genéticos también puede afectar al núcleo esencial de la dignidad humana: un individuo no debe quedar reducido a un resultado de la interpretación de sus características genéticas, garantizándose el pleno respeto al carácter único de cada persona y a la diversidad genética humana. Ello implica un rechazo frontal a las teorías del determinismo y reduccionismo genéticos[21].

En pocas palabras, es la combinación de sus características singulares, junto con la constatación de los riesgos derivados de su tratamiento para los derechos y libertades descritos, la que ha justificado una protección legal reforzada de los datos genéticos de carácter personal en los términos que seguidamente se expondrán. No obstante, este reconocimiento no se produjo con tanta prontitud como el de otras categorías especiales de datos personales, tales como los datos relativos a la salud. Es más, la respuesta legislativa inicial fue, precisamente, la de asimilar los datos genéticos a una suerte de subtipo de datos relativos a la salud, lo que llevó a cuestionarse si los

[15] Las normas de protección de datos europeas, con el RGPD a la cabeza, han venido despojándose del recurso a la «sensibilidad» en tanto fundamento exclusivo para el reconocimiento legal de categorías de datos merecedoras de una especial protección. En consecuencia, el régimen vigente reposa, como cabe extractar del Considerando 51 del RGPD, en los riesgos (y, en su caso, en el grado de afectación) que los datos personales pueden tener en relación con los derechos y libertades fundamentales.

[16] Sentencia del TEDH del 4 de diciembre de 2008, Asunto Marper v. Reino Unido, §72 y 76.

[17] Vid. entre otras y por todas: Sentencia del TEDH de 16 de febrero de 2000, Asunto Amann v. Suiza (§69); y Sentencia del TEDH de 4 de diciembre de 2008, Asunto Marper v. Reino Unido, §68-76.

[18] Sostiene a este respecto HENDRIKS que la discriminación genética es diferente al resto de formas de discriminación, puesto que tiene lugar en el presente en base a un riesgo o susceptibilidad futuros y que no tienen por qué llegar a materializarse. Aart Hendriks, "Genetic discrimination: How to Anticipate Predictable Problems?, European Journal of Health Law 9, no.2 (2002):87. DOI https://doi.org/10.1163/157180902400821039

[19] Jalayne Arias et al, "Trust, vulnerable populations, and genetic data sharing", Journal of Law and the Biosciences 2, no.3 (2016):747–753. DOI https://doi.org/10.1093/jlb/lsv044

[20] Algunos países han aprobado leyes específicas contra la discriminación genética, siendo la Genetic Information Nondiscrimination Act (GINA) estadounidense uno de los máximos exponentes. En cambio, en Europa, no existen leyes específicas contra la discriminación genética, por un lado, debido a que se encuentra prohibida en el plano iusfundamental y, por otro, dado que la protección de datos ha operado en este terreno de manera efectiva, inclusive evitando la materialización de los riesgos aludidos.

[21] El reduccionismo genético propugna reducir la interpretación o explicación universal de la vida y comportamiento humanos a las características genéticas y a la información derivada de aquellas. Por su parte, muy vinculado con el anterior, el determinismo genético implica la aceptación de que toda la vida humana y su desarrollo futuro quedarían de un modo u otro determinado por los genes.

datos genéticos eran realmente merecedores o no de un reconocimiento jurídico particular con respecto al resto de informaciones relevantes o relacionadas con la salud[22].

3. Los orígenes del dato genético de carácter personal: una heterogeneidad de fórmulas y respuestas

Los datos genéticos integran hoy en el RGPD una categoría especial dotada de entidad y autonomía propias, pero no siempre fue así: la situación precedente se caracterizó por una coexistencia de fórmulas legales de lo más dispares e incluso contradictorias. En términos estrictos de protección de datos, originalmente, ni el Convenio 108 del Consejo de Europa[23], ni la Directiva de Protección de Datos[24] incorporaron mención alguna a los datos genéticos de carácter personal. En esta misma línea cabe situar a la Ley Orgánica 15/1999[25] española, si bien con posterioridad en su reglamento de desarrollo[26] sí se llegó a subsumir de manera expresa la información genética dentro de la categoría de datos relacionados con la salud (art. 5.1, letra g).

En consecuencia, las primeras categorizaciones y definiciones de los datos genéticos fueron desarrolladas a nivel sectorial e inicialmente en recomendaciones o declaraciones jurídicamente no vinculantes[27]. Más adelante, coincidiendo con el final del Proyecto del Genoma Humano y los avances científico-técnicos, vieron la luz ciertos instrumentos, ahora sí vinculantes. Destacan, a tales efectos, el Protocolo Adicional al Convenio de Oviedo, relativo a los análisis genéticos[28]; o nuestra Ley de Investigación Biomédica (en adelante «LIB»)[29], pionera en elevar a derecho positivo no solo una concreta definición legal de datos genéticos, también, un régimen particular para el tratamiento de datos genéticos de carácter personal. De todos, el Convenio de Oviedo tuvo una influencia decisiva en la actual categorización y definición de los datos genéticos en las normas de protección de datos, hasta el punto en que el RGPD o el Convenio 108+ del Consejo de Europa han llegado a replicar en gran medida sus definiciones.

Así pues, salvando ciertas excepciones a nivel sectorial, las normas de protección de datos europeas no incluyeron inicialmente una categoría especial de datos genéticos, ausencia que fue en parte mitigada, como decimos, por medio de una subsunción en la categoría expansiva de datos relacionados con la salud. Sin embargo, tal solución implicaba la necesidad de vincular en todo momento el tratamiento de datos genéticos con informaciones relativas a la salud física o mental de una persona o, en su caso, con informaciones que pudieran revelar su origen étnico o racial[30]. Esto expelía del régimen de protección reforzada aquellos tratamientos de datos genéticos que no revelasen o pudieran revelar tales informaciones; por no hablar de que, en términos médicos y científicos, sigue siendo enormemente problemático trazar una clara línea divisoria entre las informaciones genéticas que puedan o no ser relevantes para la salud de las personas.

[22] Planteamiento defendido por una parte de la literatura especializada de la época, entre otros y por todos: George Annas, "Privacy Rules for DNA Databanks: Protecting Coded Future Diaries", JAMA 270, no.19 (1993):2346-2350. DOI https://doi.org/10.1001/jama.1993.03510190102034

[23] Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal (Convenio 108). Estrasburgo, 28 de enero de 1981. «BOE» núm. 274, 15/11/1985.

[24] Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. DO L281 23/11/1995, pp. 31–50.

[25] Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. «BOE» núm.298, 14/12/1999.

[26] Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999. «BOE» núm.17, 19/01/2008.

[27] Cabe destacar a estos efectos la Recomendación R97(5) del Consejo de Europa, de 13 de febrero de 1997, sobre Protección de Datos Médicos que, aunque asimilaba los datos genéticos a los datos médicos, llegó a prever un cierto régimen de tratamiento diferenciado para los primeros. Otro ejemplo es la Declaración Universal de la UNESCO sobre los Datos Genéticos Humanos, que reconocía que la información genética forma parte del acervo general de los datos médicos, proclamando su sometimiento a dichos principios reguladores, pero destacando la singularidad de los datos genéticos.

[28] Protocolo Adicional al Convenio para la protección de los derechos humanos y la dignidad del ser humano con respecto a las aplicaciones de la biología y la medicina (Convenio de Oviedo), sobre análisis genéticos con fines de salud. Estrasburgo, 27/11/2008.

[29] Ley 14/2007, de 3 de julio, de Investigación biomédica. «BOE» núm. 159, 04/07/2007.

[30] Grupo de Trabajo del Artículo 29 (GT29), Documento de Trabajo sobre Datos Genéticos, 5; Documento de Trabajo sobre el tratamiento de datos relativos a la salud en Historias Clínicas, 7.

Sea como fuere, antes de la llegada del RGPD los datos genéticos no gozaban de un reconocimiento autónomo como categoría especial de datos personales. Adicionalmente, las diversas fórmulas de definición empleadas por la normativa que precedió al Reglamento evidenciaron la presencia de enormes disparidades en atención a los siguientes dos criterios fundamentales:

- i) Criterio relacional. Por una parte, si la categoría de datos genéticos ha de abarcar, en sentido estricto, las informaciones relativas a las características hereditarias o, de modo sumativo, las informaciones relativas a las características hereditarias y adquiridas. En este sentido, cabe anticipar que el RGPD aboga ostensiblemente por la segunda de las opciones.

- ii) Criterio originario. Por otra parte, en cuanto al origen de los datos genéticos, si se opta por una concepción amplia (toda información genética con independencia de su procedencia) o una más rigurosa (información genética obtenida de un análisis genético u otros análisis científicos). Aunque este último criterio sigue generando controversia, a nuestro entender, el RGPD aboga por la aproximación más rigurosa.

En consecuencia, el Reglamento ve la luz en un momento decisivo en lo concerniente al tratamiento de datos genéticos de carácter personal. En primer lugar, debiendo optarse entre su inclusión como una categoría autónoma de datos especialmente protegidos o, en su caso, continuar con su tradicional subsunción o asimilación a la categoría de datos relativos a la salud. En segundo lugar, de abogar efectivamente por la primera de las opciones, debiendo escogerse qué criterios concretos seguir a la hora de delimitar jurídicamente la definición y alcance de la categoría.

4. Los datos genéticos en el Reglamento General de Protección de Datos: certezas e incertidumbres

La llegada del RGPD puso fin a la primera de las incertidumbres, a saber, la ausencia de una categoría especial de datos genéticos cuyo tratamiento desencadenase la aplicación de un régimen legal de protección reforzada. En efecto, el Reglamento incluyó de manera expresa a los datos genéticos en el listado del art. 9.1, de modo que ya no es indispensable demostrar caso por caso la existencia de un vínculo entre la información genética y la salud pasada, presente y futura de la persona. Ahora bien, esta decisión legislativa también conllevó asumir que los datos genéticos, a pesar de sus mencionadas características singulares, no resultaban merecedores de una mayor protección legal que los datos relativos a la salud o los datos biométricos. En caso de existir alguna diferencia sustancial en las condiciones relativas a su tratamiento, éstas habrían de ser introducidas por los Estados miembros en virtud de la cláusula del art. 9.4.

Así pues, resuelta la primera de las cuestiones, la más controvertida en la actualidad no es otra que la fórmula empleada por el RGPD para delimitar el concepto y el alcance jurídicos de los datos genéticos, y que sigue siendo fuente de dudas e incertidumbres. La definición actualmente contenida en el art. 4(13) reza como sigue:

> «datos personales relativos a las características genéticas *heredadas* o *adquiridas* de una persona física que proporcionen una *información única* sobre la *fisiología o la salud de esa persona*, obtenidos en particular del *análisis de una muestra biológica* de tal persona».

Del fragmento arriba reproducido, junto con las valiosas precisiones del Considerando 34, cabe deducir que, para hablar de datos genéticos en términos del RGPD, deberán estar cumulativamente presentes los siguientes cuatro elementos:

- i) Elemento relacional y de contenido. Exige que los datos personales se encuentren relacionados con las características genéticas heredadas o adquiridas y que, concomitantemente, proporcionen información sobre la fisiología o la salud.

- ii) Elemento personal. Los datos deben referirse a la persona física de cuya muestra biológica se derive la información en cuestión, excluyéndose otros sujetos potencialmente afectados por el tratamiento. Por ejemplo, la familia, el grupo biológico o la comunidad a que pertenezcan.

- iii) Elemento de identificabilidad. Los datos genéticos deben proporcionar una información única sobre la persona, en el sentido de que deben permitir su identificación unívoca.

- iv) Elemento originario. Aunque su concreta significación se encuentra todavía cuestionada, exigiría que los datos personales procedan únicamente del análisis de una muestra biológica.

↘ En los próximos apartados se procede a desglosar con mayor detalle cada uno de los componentes arriba enunciados, llevando a cabo un estudio crítico en el que se plantean algunas de las principales dudas e incertidumbres existentes.

4.1. Elemento relacional y de contenido: características genéticas heredadas o adquiridas que proporcionen información sobre la fisiología o la salud

Tal y como se anticipaba en el epígrafe tercero, el RGPD toma como punto de partida la aproximación sumativa en lo que se refiere al criterio relacional: los datos genéticos comprenden no solo las características genéticas hereditarias, también las adquiridas. Esto genera ciertas disfunciones a la hora de trabajar con la normativa sectorial, por ejemplo, con la LIB, cuyas disposiciones son de aplicación exclusivamente a las características genéticas correspondientes a la línea germinal, mientras que las del RGPD tanto a las germinales como las somáticas, incluyendo aquellas mutaciones que pudieran acaecer a lo largo de la vida de la persona[31]. Ahora bien, parece que en ambos casos quedarían expelidas de la categoría, sobre la base del conocimiento existente en la actualidad, los datos personales relativos a las regiones del ADN no codificante, pues en sentido estricto no contienen información sobre los genes ni sobre sus características[32]. Naturalmente, esta exclusión no significa que tales datos queden desprovistos de la protección reforzada del RGPD, habida cuenta de que podrán subsumirse en otras categorías especiales como los datos relativos a la salud o, de ser empleadas con fines de identificación, en la de datos biométricos.

Por último, este elemento relacional se encuentra inextricablemente unido a uno de contenido: las características genéticas (heredadas o adquiridas) habrán de proporcionar información sobre la fisiología o la salud de la persona. A la postre, esto denota una postura expansiva: ya no es estrictamente necesario que los datos genéticos provean información sobre la salud de una persona, pues el régimen de protección reforzada también será objeto de aplicación cuando la información proporcionada se refiere a su fisiología. En cambio, esta opción excluye aquellos tratamientos de que pudieran derivarse otro tipo de informaciones, por ejemplo, las relativas al origen étnico o racial de la persona[33]. Tal matiz ha sido deliberadamente introducido a los efectos de preservar una cierta autonomía a la categoría de datos relacionados con el origen étnico o racial, sin perjuicio de las dificultades que a menudo produce su efectiva distinción en la práctica.

4.2. Elemento personal: la persona física

La información genética, pese a ser única o singular respecto de una persona física, tiene un carácter generacional: puede revelar al mismo tiempo características o circunstancias relativas a su familia biológica o, incluso, un grupo social relacionado pero ajeno al propio círculo familiar (el grupo biológico)[34]. A pesar de ello, el RGPD proclama que los datos genéticos cuyo tratamiento debe gozar de una protección reforzada serán solamente los relativos al sujeto fuente de la muestra biológica (la persona física de cuya muestra biológica se derive la información en cuestión).

Esta opción legislativa ha sido muy criticada por la doctrina y los operadores jurídicos, al entenderse que excluye de esa protección reforzada a otras personas o sujetos potencialmente afectados por el tratamiento de los datos personales. En efecto, hasta el propio Grupo de Trabajo del Artículo 29 («GT29») ha llegado a admitir la posibilidad de que los datos genéticos sean «compartidos» con la familia o el grupo biológico, sugiriendo dos alternativas para su conciliación en la normativa: i) extender la condición de «interesado» a las familias y/o grupos biológicos; o ii) reconocer ciertos derechos propios y diferenciados de aquellos de las personas interesadas (una suerte de derechos *sui generis*)[35]. En particular, la segunda de las propuestas ha suscitado un gran interés académico,

[31] Pilar Nicolás, et al., Informe sobre implicaciones legales para el desarrollo de un modelo de gestión de datos genéticos, (Barcelona: Fundació TICSALUT, 2021), 8, https://ticsalutsocial.cat/wp-content/uploads/2021/11/INFORME-PROYECTO-DATOS-GENETICOS.pdf

[32] Carlos María Romeo Casabona. "Datos genéticos (Comentario al artículo 4.13 RGPD)", en Comentario al Reglamento General de Protección de Datos, ed. por Antonio Troncoso Reigada (Cizur Menor: Aranzadi, 2021), 701.

[33] Iñigo de Miguel Beriain y Ricardo de Lorenzo y Aparici, Claves Prácticas. Datos genéticos y datos relativos a la salud (Madrid: Lefebvre, 2020), 24. También, en este mismo sentido: Verónica Miño-Vásquez, "The Protection of Genetic Data under the General Data Protection Regulation: A European view", Datenschutz und Datensicherheit 43 (2019):158. DOI https://doi.org/10.1007/s11623-019-1083-9

[34] GT29, Documento de Trabajo sobre Datos Genéticos, 9.

[35] Ídem, 7.

habiendo sido explorada por algunos autores, que abogan por la creación de «categorías genéticas» vinculadas con el sujeto fuente de la muestra o, incluso, por la introducción de dos nuevos grupos de interesados (los sujetos o interesados «primarios» y los «secundarios»)[36].

Por más que las anteriores críticas pudieran resultar acertadas, lo cierto es que tanto la literalidad de la definición de datos genéticos del RGPD, como la incuestionable fijación de la norma para con la persona interesada en tanto sujeto individual, comportan que ni la familia ni el grupo biológico pueden erigirse como titulares compartidos de derechos sobre los datos genéticos de carácter personal. Al menos, no en el vigente sistema de reconocimiento de derechos y asignación de responsabilidades con respecto al tratamiento de datos personales. Por ejemplo, aunque llegase a reconocerse en el Reglamento un derecho *sui generis* de acceso a la información genética por parte de la familia biológica, tal facultad difícilmente podría ser ejercida de manera efectiva si los sujetos integrantes de dicho núcleo familiar no hubieran sido previamente informados de que el tratamiento de datos estaba teniendo lugar[37].

En definitiva, todo parece indicar que el RGPD solo tiene en cuenta a estos afectos a la persona física de cuya muestra biológica se derive la información en cuestión, lo que no ha de significar que los intereses de terceras personas o grupos vayan a quedar jurídicamente desatendidos. Existen otras vías legales y judiciales para dirimir los eventuales conflictos que pudieran acaecer de modo que, llegados a este punto, es preciso asumir que ni el derecho ni la normativa de protección de datos parecen ser la vía idónea[38] para proporcionar unos criterios generalmente aplicables a la miríada de escenarios —jurídicamente complejos y heterogéneos— que pueden sucederse en el tratamiento de datos genéticos. Hoy día la opción más practicable no es otra que la realización evaluaciones individualizadas de cada supuesto de tratamiento en atención a las circunstancias concurrentes y con plena observancia de la legislación sectorial relevante.

4.3. Elemento de identificabilidad: información única sobre una persona física

Si bien el tercer componente de la definición de datos genéticos ha pasado a menudo desapercibido por la vehemencia de las demás discusiones, lo cierto es que la frase «que proporcionen una información *única* sobre la fisiología o la salud de esa persona» (con énfasis en «única») no deja de ser, cuanto menos, enigmática. Máxime, toda vez que el RGPD no parece esclarecerlo en ningún momento, ni siquiera en sus considerandos. Tal ambigüedad puede traducirse, a nuestro entender, en dos posibles interpretaciones.

Una primera consistente en entender esa unicidad como una referencia al carácter individual y excluyente de la información. En otras palabras, se estaría matizando el elemento personal de la definición, de modo que solamente se considerarían datos genéticos aquellos que aportasen información (únicamente) sobre la persona física en cuestión, excluyendo, con ello, la información relativa a la familia o el grupo biológico. Sin perjuicio de que, como se ha visto, el RGPD no reconozca derechos sobre los datos genéticos a la familia o grupo biológico, esta lectura se antoja difícilmente admisible. El reconocimiento legal del sujeto fuente como núcleo principal del elemento personal no ha de significar que los datos genéticos no puedan aportar información sobre la familia o grupo biológico, pues tal aseveración sería científicamente inexacta debido al carácter heredable de la información genética. De concederse, el legislador habría asumido una discordancia evidente entre la caracterización jurídica y científica de los datos genéticos.

Así, la segunda —y más razonable— interpretación consiste en concebir la unicidad de la información como un componente de identificabilidad de la persona física en cuestión. Es decir, para poder hablar de datos genéticos de carácter personal, el tratamiento deberá aportar información única en el sentido de que permita la identificación unívoca de la persona. Este y no otro es, a nuestro entender, el significado conferido a la expresión y ello por dos razones fundamentales. Primera, porque la información sobre las características genéticas —especialmente, el genoma completo— a menudo constituye un identificador único e inherente a cada persona física. Es más, el propio RGPD llega a mencionar de manera expresa en su art. 4(1), en tanto identificador susceptible

[36] Vid. respectivamente: Dara Hallinan y Paul de Hert. "Genetic Classes and Genetic Categories: Protecting Genetic Groups Through Data Protection Law", en Group Privacy, ed. por Luciano Floridi y Bart van der Sloot (Springer, 2016), 177; Mark Taylor, Genetic data and the law: a critical perspective on privacy protection (Cambridge: Cambridge University, 2012), 105.

[37] Taner Kuru, "Genetic Data: The Achilles' Heel of the GDPR?", European Data Protection Law Review 7, no.45 (2021):51. DOI https://doi.org/10.21552/edpl/2021/1/8

[38] En este mismo sentido vid.: Verónica Miño-Vásquez, "The Protection of Genetic Data under the General Data Protection Regulation: A European view", 156.

de determinar (directa o indirectamente) la identidad de una persona: «uno o varios elementos propios de la identidad genética de dicha persona». Segunda, porque el otro caso excepcional en que el Reglamento emplea una técnica legislativa semejante es a los efectos de proclamar, en su art. 4(14), qué se entiende por un tratamiento de datos biométricos: que «permitan o confirmen la identificación *única* de dicha persona». Como cabe apreciar, las similitudes resultan evidentes, en buena parte, porque ambas categorías y definiciones fueron introducidas *ex novo* al mismo tiempo por el RGPD.

Por lo demás, una exigencia explícita de la unicidad de la información como parte de la definición de datos genéticos *de carácter personal*, creemos, resulta jurídicamente desafortunada. Lo que determina la identificabilidad de una persona física no es la unicidad del atributo o de la pieza de información en cuestión, sino la existencia de una probabilidad razonable de identificación, en atención a los factores objetivos y al contexto, incluyendo su posible combinación con otros conjuntos de datos. Una interpretación literal de este precepto implicaría que solo ameriten situarse en el núcleo de la protección legal reforzada aquellos tratamientos de datos relativos a las características genéticas hereditarias o adquiridas que, además de ser de carácter personal, puedan servir para identificar de manera unívoca a la persona. Dicha opción legislativa no parece coherente a la luz de un enfoque garantista para con los derechos y libertades fundamentales de las personas interesadas, pues excluiría de tal categorización especial aquellas informaciones que, indirectamente o de manera no tan unívoca, pudieran identificar al sujeto de los datos genéticos. Ciertamente, no todos los datos relativos a las características genéticas son iguales en términos de identificabilidad ni facultan —en cualquier escenario y con total independencia del contexto del tratamiento— la identificación unívoca de una persona. Algo que ha sido admitido por el propio Comité Europeo de Protección de Datos («CEPD») para quien la identificabilidad de los datos genéticos: «sigue siendo una cuestión por resolver»[39].

Consecuentemente, no todo dato genético equivale automáticamente a un identificador único de la persona ni todos son iguales en términos de calidad y representatividad[40]. Incluso, cabe cuestionarse si la información sobre las características genéticas pueda constituir, inherentemente, por sí misma, un identificador directo, sin una atención debida al contexto o al entorno de tratamiento; siendo así que lo que realmente aumenta la probabilidad de identificación no es tanto la unicidad de esa pieza de información, sino su vinculabilidad[41]. Es decir, la posibilidad de relacionar alguno de los marcadores o características genéticas con otros conjuntos o bases de datos. Esto último convertiría a esa pieza de información genética en un identificador indirecto, pero no en una información inherente o directamente identificativa[42].

En definitiva, no todos los datos genéticos van a ser susceptibles de identificar de manera unívoca a una persona, pues ello depende, como se ha dicho, de la probabilidad razonable de que dicha identificación tenga lugar, a la luz de todos los factores objetivos y contextuales del tratamiento en cuestión. Así, debida atención merecen, entre otras, la disponibilidad de otras fuentes o bases de datos, las medidas técnicas y organizativas implementadas, los marcos de gobernanza o los mecanismos institucionales de supervisión y control sobre el acceso. La exigencia de unicidad de la información en tanto elemento definitorio de la categoría especial de datos genéticos no debe prejuzgar, en un sentido u otro, el resultado arrojado por el test de identificabilidad del Considerando 26. Además, de interpretarse de manera estricta, operará como un criterio rígido y poco garantista, dejando fuera del régimen de la protección legal reforzada a una multiplicidad de tratamientos de datos (potencialmente) relativos a las características genéticas.

4.4. Elemento originario: provenientes del análisis de una muestra biológica

Para finalizar, los tratamientos de datos personales que cumplan con los tres elementos anteriores

[39] «Todavía falta por demostrar qué combinación de medidas técnicas y organizativas es susceptible de ser empleada para remover los datos genéticos del ámbito material de aplicación del RGPD». Vid. CEPD, Documento de aclaraciones relativas a la aplicación coherente del RGPD en investigación en salud, 12.

[40] Por ejemplo, no todos los datos genéticos van a ser tan detallados como un análisis del genoma completo (WGS). Como concluyen LOWRANCE y COLLINS, resulta complicado establecer un estándar o nivel de identificabilidad aplicable a la generalidad de datos genéticos: «how much is sufficient for identifying highly depends on the region and extent of genome covered, the density of mapping, the rarity of variants, the degree of linkage disequilibrium and other factors». William Lowrance y Francis Collins, "Identifiability in genomic research", Science 317, no.5838 (2007):601. DOI https://doi.org/10.1126/science.1147699

[41] Colin Mitchell, et al., The GDPR and genomic data, 38 (Cambridge: PHG Foundation, 2020), https://www.phgfoundation.org/wp-content/uploads/2023/10/gdpr-and-genomic-data-report.pdf

[42] Mahsa Shabani y Luca Marelli, "Re-identifiability of genomic data and the GDPR", EMBO Reports 20, no.6 (2019). DOI https://doi.org/10.15252/embr.201948316

solamente podrán subsumirse en la categoría especial de datos genéticos si han sido «obtenidos, en particular, del análisis de una muestra biológica de tal persona». En este punto, la redacción del art. 4(13) del RGPD resulta algo confusa, no quedando claro si ese «obtenidos en particular» pretende ser una expresión dotada de carácter meramente ejemplificativo o, al contrario, determina de una manera excluyente la fuente o la procedencia de los datos genéticos. Por tal motivo, este último elemento de la definición ha sido habitualmente interpretado de dos formas: i) los datos genéticos pueden ser obtenidos, entre otros, del análisis de una muestra biológica; o ii) los datos genéticos deben obtenerse necesariamente del análisis de una muestra biológica.

La primera de las lecturas tendría su argumento en una interpretación algo forzada de la literalidad del Considerando 34 del Reglamento: «o del análisis de cualquier otro elemento que permita obtener información equivalente». En este sentido, lo relevante sería la obtención de la información en cuestión y no el método o la fuente, de modo que el legislador estaría indicando cuál es el método típicamente utilizado para la obtención de los datos genéticos, pero no el único existente, pues de lo contrario se habría optado por emplear otras expresiones más contundentes como «solo» o «exclusivamente»[43].

Ahora bien, por sorprendente que pudiera parecer, es la segunda y más restrictiva de las interpretaciones la que, a nuestro juicio, ha sido deliberadamente escogida por el legislador europeo. Dicho de otro modo, el elemento originario de la definición debe estar siempre cumulativamente presente para hablar de dato genético en el sentido del RGPD: los datos genéticos que ameritan una protección legal reforzada son aquellos obtenidos del análisis de una muestra biológica. Existen dos motivos de peso que refuerzan, a nuestro juicio, esta postura:

- i) Una traducción desafortunada de la definición de datos genéticos. La literalidad del art. 4(13) del RGPD ha sido deficientemente traducida al español, ya que ha omitido una parte clave del texto original en inglés («*and which result*») que, de haber sido incorporada de manera adecuada, indicaría indubitadamente el carácter cumulativo de este último componente («y que se obtengan»)[44]. Si dicho argumento no fuera suficiente, la lectura del Considerando 34 resulta todavía más esclarecedora, puesto que emplea la expresión «en particular» no como un ejemplo de la diversidad de fuentes posibles, sino como una enumeración *numerus apertus* de los métodos más extendidos para el análisis de las muestras biológicas. Por lo tanto, la enumeración que el Reglamento deja abierta es la de los métodos de análisis de la muestra biológica (análisis cromosómico, del ADN, del ARN o de «cualquier otro elemento que permita obtener información equivalente»).

- ii) Se toma como referencia para la definición y caracterización contenida en el Convenio de Oviedo. En particular, en su Protocolo Adicional sobre análisis genéticos con fines de salud, cuyo ámbito de aplicación queda de por sí circunscrito al análisis de «muestras biológicas de origen humano». Es más, el Considerando 34 del RGPD reproduce cuasiliteralmente la definición del art. 2.3 del Protocolo.

En consecuencia, a nuestro juicio el RGPD adopta una aproximación rigurosa en lo que concierne al elemento originario de los datos genéticos: para ser considerados datos especialmente protegidos, deben obtenerse necesariamente del análisis de una muestra biológica humana. Con ello, quedan excluidas de esta categoría especial de datos personales aquellas informaciones relativas a las características genéticas, pero que hubieran sido obtenidas de otras fuentes. Por ejemplo, los datos genealógicos recogidos mediante cuestionarios, los datos ambientales o conductuales que permitan extraer conclusiones sobre la composición genética de una persona[45], o incluso aquellos obtenidos a partir de un examen físico, la historia clínica o un análisis de determinaciones bioquímicas que no necesariamente consistan en el material biológico[46].

Aunque más restrictiva —y menos garantista—, esta opción parece haber sido deliberadamente escogida, tal vez, en un intento de diferenciar la categoría especial de datos genéticos de otras afines. Otra posibilidad es que, con ello, se pretendiera ofrecer una mayor seguridad a responsables

[43] Iñigo de Miguel Beriain y Daniel Jove, "Is it possible to place limits on the self-determination of your own genetic data?", BioLaw Journal 1S (2021):10. DOI https://doi.org/10.15168/2284-4503-782

[44] La versión en español del RGPD ha omitido la conjunción «y» («obtenidos en particular del análisis de la muestra biológica de tal persona»). Puede observarse su falta de correspondencia con la versión original en inglés («and which result, in particular, from an analysis of a biological sample from the natural person in question»), por lo que la traducción correcta sería la siguiente: «y que se obtengan, en particular, del análisis de la muestra biológica de tal persona».

[45] Gauthier Chassang, "The impact of the EU general data protection regulation on scientific research", Ecancermedicalscience 11, no.709 (2017):5. DOI https://doi.org/10.3332/ecancer.2017.709

[46] Aitziber Emaldi, "Protección de datos personales en el ámbito sanitario y de investigación biomédica: una visión europea", Actualidad Jurídica Iberoamericana 14 (2021), 732.

y encargados del tratamiento, de modo que el alcance jurídico de la categoría especial de datos genéticos no quedase permanentemente cuestionado tal y como sucede con aquella expansiva de los datos relativos a la salud. No obstante, al margen de las apuntadas disfunciones interpretativas y aplicativas, no creemos que esta decisión haya conllevado unas repercusiones críticas en la práctica en lo que se refiere a una posible desprotección de las personas físicas, pues nada impide subsumir, caso por caso y atendiendo a factores contextuales y/o finalistas, aquellos tratamientos de datos eventualmente excluidos de la categoría de datos genéticos en otras íntimamente relacionadas, tales como los datos relativos a la salud, los datos biométricos si se tratan con fines de identificación y/o los datos que revelen el origen étnico o racial.

5. Consideraciones finales y el futuro de los datos genéticos de carácter personal

Como se ha visto en las líneas precedentes, si bien el RGPD resolvió algunas de las incertidumbres existentes con la introducción de una nueva categoría especial de datos genéticos, en cambio, la fórmula legal escogida para la definición y delimitación de dicha categoría especial, más que de certezas, ha sido una fuente de constantes discusiones y controversias; algunas ya enfrentadas a lo largo del presente artículo.

Así las cosas, como colofón, es preciso cuestionarse cuál será el futuro de los datos genéticos de carácter personal, en especial, a la luz de ciertas novedades legislativas e incluso reformas de calado que bien pudieran alterar el *statu quo*.

5.1. El Reglamento del Espacio Europeo de Datos de Salud

El Reglamento (UE) 2025/327 relativo al Espacio Europeo de Datos de Salud (en adelante, «REEDS»)[47] está llamado a ser una de las normas más relevantes para el sector en los próximos años. Si bien un examen pormenorizado del citado texto excede del objeto del presente artículo, cabe señalar que el REEDS va un paso más allá del RGPD en lo que a su ámbito de aplicación se refiere, al acuñar una nueva categoría de «datos de salud electrónicos personales» en la que habrán de entenderse comprendidos tanto los datos relativos a la salud como los datos genéticos que se traten en formato electrónico (art. 1.2, letra a). Por lo tanto, las disposiciones del REEDS también resultarán de aplicación al tratamiento de datos genéticos (personales o no, pues no puede obviarse que este Reglamento expande su alcance más allá de la mera protección de datos), al quedar comprendidos dentro de esa nueva noción omnicomprensiva de datos de salud electrónicos.

Además, a diferencia de la categorización legal dada por el RGPD, en la que se maneja un criterio originario estricto (procedentes del análisis de una muestra biológica de la persona física) el REEDS deja claro en todo momento que el origen o la fuente de los datos puede ser enormemente diverso[48]. Más aún, el art. 51.1 del REEDS establece un extenso catálogo de las categorías mínimas de datos de salud electrónicos que deben hacerse disponibles para el uso secundario, entre las que se citan expresamente los «datos genómicos, epigenómicos y genómicos humanos» (letra f). Por lo tanto, no solo no se estaría exigiendo que los datos genéticos procedan estrictamente del análisis de una muestra biológica, adicionalmente, se estaría llevando a cabo una alusión explícita a los datos genómicos que, recordemos, se encuentran excluidos de la categoría especial de datos genéticos del RGPD.

Finalmente, otro posible foco de conflictos con el RGPD y con la normativa sectorial precedente podría estar en que, si bien este listado de categorías mínimas del REEDS alude a «datos genómicos, epigenómicos y genómicos *humanos*», no aclara si dichos datos, para entenderse efectivamente comprendidos dentro de la categoría de datos de salud electrónicos, deben limitarse a aquellos que proporcionen información sobre la salud humana o, en su caso, también sobre la fisiología de la persona[49].

[47] Reglamento (UE) 2025/327 del Parlamento Europeo y del Consejo, de 11 de febrero de 2025, relativo al Espacio Europeo de Datos de Salud, y por el que se modifican la Directiva 2011/24/UE y el Reglamento (UE) 2024/2847. «DOUE» núm.327, de 05/03/2025.

[48] Vid. en este sentido los Considerandos 55 y 56 del REEDS.

[49] Mikel Recuero, "El uso secundario de datos de salud electrónicos: el futuro Reglamento del Espacio Europeo de Datos de Salud y su interacción con la protección de datos personales", InDret 2 (2024):532. DOI https://doi.org/10.31009/indret.2024.i2.13

5.2. La propuesta de Reglamento «Ómnibus Digital»

La otra gran novedad legislativa que ha copado gran parte del debate sobre el futuro de la protección de datos no es otra que la Propuesta de Reglamento relativa a la simplificación del marco legislativo digital («Ómnibus Digital»)[50] que, en el momento en que se escriben estas líneas, se encuentra todavía en estadios iniciales de negociación. La Propuesta persigue propiciar un marco regulador optimizado, coherente y cohesionado para el tratamiento y la disponibilidad de los datos en la UE, lo que incluye la introducción de modificaciones de calado en el RGPD. Por ejemplo, en lo que concierne a las definiciones de datos personales e investigación científica, el recurso al interés legítimo para el desarrollo de Sistemas de IA o la inclusión de nuevas circunstancias para levantar el veto al tratamiento de categorías especiales de datos.

Al margen del evidente impacto que dichas enmiendas puedan tener en distintas áreas relacionadas con el tratamiento de datos genéticos, lo cierto es que el Ómnibus Digital parece que no afectará de modo inmediato a la manera en que se define y delimita jurídicamente esta categoría especial de datos personales. De hecho, hasta el controversial texto filtrado de la propuesta inicial llegó a reconocer explícitamente que «la protección reforzada de los datos genéticos y biométricos debería permanecer intacta debido a sus características únicas y específicas»[51]. Contrasta esto último con la polémica suscitada en torno a una posible restricción del alcance tradicionalmente expansivo de la categoría de datos relativos a la salud, de la que se llegó a proponer que únicamente diera cobijo a informaciones que revelen directamente el estado de salud de una persona física[52]. La ausencia de una tentativa legislativa equivalente para con los datos genéticos puede explicarse, a nuestro entender, por dos razones.

Primera, porque el regulador ha sido en todo momento plenamente consciente de la singularidad de los datos genéticos de carácter personal y de los posibles riesgos dimanantes de su tratamiento indiscriminado en beneficio de la competitividad europea; una toma de conciencia que, según parece, no fue igual de prolija en el caso de los datos relativos a la salud.

Segunda —tal y como el presente artículo ha terminado por poner de manifiesto— porque la categoría especial de datos genéticos del RGPD es, en efecto, mucho más restrictiva de lo que *a priori* pudiera pensarse, de modo que no sería procedente acotarla legalmente todavía más.

Bibliografía

Annas, George, Leonard Glantz y Patricia Roche. "Drafting the Genetic Privacy Act: Science, Policy, and Practical Considerations". *Journal of Law, Medicine & Ethics* 23, no.4, (1995):360-366. DOI 10.1111/j.1748-720X.1995.tb01378.x

Annas, George. "Privacy Rules for DNA Databanks: Protecting Coded Future Diaries". *JAMA* 270, no.19 (1993):2346-2350. DOI 10.1001/jama.1993.03510190102034

Arias, Jalayne, Genevieve Pham-Kanter, Rosa González y Eric Campbell. "Trust, vulnerable populations, and genetic data sharing". *Journal of Law and the Biosciences* 2, no.3 (2016):747–753. DOI 10.1093/jlb/lsv044

Ayday, Erman, Emiliano De Cristofaro, Jean-Pierre Hubaux, Gene Tsudik. "Whole Genome Sequencing: Revolutionary Medicine or Privacy Nightmare?". *Computer* 48, no.2 (2015): 58-66. DOI 10.1109/MC.2015.59

Chassang, Gauthier. "The impact of the EU general data protection regulation on scientific research". *Ecancermedicalscience* 11, no.709 (2017). DOI 10.3332/ecancer.2017.709

Collins, Francis, Eric Green, Alan Guttmacher y Mark Guyer. "A vision for the future of genomics research". *Nature* 422 (2003): 835–847. DOI: 10.1038/nature01626

De Miguel Beriain y Daniel Jove. "Is it possible to place limits on the self-determination of your own genetic data?". *BioLaw Journal* 1S (2021): 209-222. DOI 10.15168/2284-4503-782

[50] Propuesta de Reglamento del Parlamento Europeo y del Consejo relativa a la simplificación del marco legislativo digital. COM(2025)837 final. Bruselas, 19.11.2025.

[51] Vid. Considerando 26 del texto inicial filtrado de la Propuesta. Disponible en: https://cdn.netzpolitik.org/wp-upload/2025/11/EU-Kommission-Digital-Omnibus-A-Data-Act-und-DSGVO.pdf

[52] Art. 2.1, letra b) del texto inicial filtrado.

De Miguel Beriain, Iñigo y Ricardo de Lorenzo y Aparici. *Claves Prácticas. Datos genéticos y datos relativos a la salud*. Madrid: Lefebvre, 2020.

Emaldi, Aitizber. "Protección de datos personales en el ámbito sanitario y de investigación biomédica: una visión europea". *Actualidad Jurídica Iberoamericana* 14 (2021):718-747.

Grisolía, Santiago. "Introducción científica". En *El Derecho ante el proyecto genoma humano*, editado por Fundación BBV, vol. 1, 33-40. Bilbao: Fundación BBV, 1994.

Hallinan, Dara y Paul de Hert. "Genetic Classes and Genetic Categories: Protecting Genetic Groups Through Data Protection Law". En *Group Privacy*, editado por Luciano Floridi y Bart van der Sloot, 175-196. Springer, 2016.

Hendriks, Aart. "Genetic discrimination: How to Anticipate Predictable Problems?. *European Journal of Health Law* 9, no.2 (2002): 87-92. DOI 10.1163/157180902400821039

Kuru, Taner. "Genetic Data: The Achilles' Heel of the GDPR?". *European Data Protection Law Review* 7, no.45 (2021):45-58. DOI 10.21552/edpl/2021/1/8

Lowrance, William y Francis Collins. "Identifiability in genomic research". *Science* 317, no.5838 (2007):600-602. DOI 10.1126/science.1147699

Miño-Vásquez, Verónica. "The Protection of Genetic Data under the General Data Protection Regulation: A European view". *Datenschutz und Datensicherheit* 43 (2019): 154-158. DOI 10.1007/s11623-019-1083-9

Mitchell, Colin, Johan Ordish, Emma Johnson y Alison Hall. The GDPR and genomic data. Cambridge: PHG Foundation, 2020. https://www.phgfoundation.org/wp-content/uploads/2023/10/gdpr-and-genomic-data-report.pdf

Nicolás, Pilar, Carlos María Romeo Casabona, Iñigo de Miguel Beriain y Guillermo Lazcoz. *Informe sobre implicaciones legales para el desarrollo de un modelo de gestión de datos genéticos*. Barcelona: Fundació TICSALUT, 2021. https://ticsalutsocial.cat/wp-content/uploads/2021/11/INFORME-PROYECTO-DATOS-GENETICOS.pdf

Nicolás, Pilar. *La protección jurídica de los datos genéticos de carácter personal*. Granada: Comares, 2006.

Quinn, Paul y Liam Quinn. "Big genetic data and its big data protection challenges", *Computer Law & Security Review* 34, no.5 (2018): 1000-1018. DOI 10.1016/j.clsr.2018.05.028

Recuero, Mikel. "El uso secundario de datos de salud electrónicos: el futuro Reglamento del Espacio Europeo de Datos de Salud y su interacción con la protección de datos personales". *InDret* 2 (2024):525-551. DOI 10.31009/InDret.2024.i2.13

Romeo Casabona, Carlos María. "Datos genéticos (Comentario al artículo 4.13 RGPD)", en *Comentario al Reglamento General de Protección de Datos*, editado. por Antonio Troncoso Reigada, 697-707. Cizur Menor: Aranzadi, 2021.

Los genes y sus leyes: el derecho ante el genoma humano. Granada: Comares, 2002.

Shabani, Mahsa y Luca Marelli. "Re-identifiability of genomic data and the GDPR". *EMBO Reports* 20, no.6 (2019). DOI 10.15252/embr.20194831

Taylor, Mark. *Genetic data and the law: a critical perspective on privacy protection*. Cambridge: Cambridge University, 2012.

Inteligencia Artificial en la Educación Superior y el derecho fundamental a la protección de datos personales: ¿innovación pedagógica o amenaza a la privacidad?

Artificial Intelligence in Higher Education and the Fundamental Right to Data Protection: Pedagogical Innovation or a Threat to Privacy?

Carolina López Medina
Abogada colegiada ICAM
especializada en protección de datos personales

Resumen:

Este artículo analiza y reflexiona sobre la integración de la Inteligencia Artificial (IA) en la Educación Superior (ES) desde una perspectiva jurídica y pedagógica. Examina la tensión dialéctica entre la innovación pedagógica y la amenaza a la privacidad; y destaca tanto su potencial innovador y ventajoso como los riesgos asociados a la vulneración de derechos. Se defiende una postura proactiva y garantista (la integración de la IA es un imperativo social que exige una gobernanza ética y corresponsable), así como se invita a reflexionar sobre la transformación del rol docente hacia un modelo tecnopedagógico capaz de armonizar conocimiento, pedagogía y tecnologías. En un ecosistema donde la IA redefine las trayectorias académicas, la protección de datos se erige como el escudo esencial para preservar la dignidad humana y el rigor académico. Adicionalmente, subraya la necesidad de fomentar un uso ético, responsable y jurídicamente garantista de la IA. Pone de relieve la relevancia de la alfabetización digital como condición indispensable para una integración segura y eficaz, apoyándose en marcos internacionales. Finalmente, se analizan políticas institucionales y buenas prácticas que evidencian la conveniencia de establecer marcos normativos comunes, sistemas de supervisión humana y enfoques de privacidad desde el diseño y por defecto.

Palabras clave:

Alfabetización digital; derechos fundamentales; educación; inteligencia artificial; protección de datos personales.

Abstract:

This article analyses and reflects on the integration of Artificial Intelligence (AI) in Higher Education (HE) from both a legal and pedagogical perspective. It examines the dialectical tension between pedagogical innovation and the threat to privacy, highlighting both its innovative and advantageous potential and the risks associated with potential infringements of fundamental rights. It advocates a proactive and rights-protective approach, arguing that the integration of AI constitutes a social imperative that requires ethical and co-responsible governance. It further invites reflection on the transformation of the teaching role towards a techno-pedagogical model capable of harmonising knowledge, pedagogy, and technology. Within an ecosystem in which AI is redefining academic trajectories, data protection emerges as an essential safeguard to preserve human dignity and academic integrity. Additionally, the article emphasises the need to promote an ethical, responsible, and legally compliant use of AI. It underscores the importance of digital literacy as an indispensable condition for safe and effective integration, drawing on international frameworks. Finally, it examines institutional policies and best practices that demonstrate the value of establishing common regulatory frameworks, systems of human oversight, and privacy-by-design and privacy-by-default approaches.

Keywords:

Digital skills; fundamental rights; education; artificial intelligence; personal data protection.

Summary:

1. *Introduction.* ***2.*** *Uses of Artificial Intelligence in Higher Education: the impact of Generative Artificial Intelligence.* ***3.*** *Main challenges of the use of Artificial Intelligence in universities: privacy and data protection.* ***4.*** *Relevance of digital literacy and digital skills in the university context.* ***5.*** *Conclusions.*

1. Introducción

Los términos "Privacidad", "Innovación" y "Tecnología" - que precisamente ponen nombre a esta Revista de la Agencia Española de Protección de Datos (AEPD, en adelante) - poseen una relevancia jurídica y social tal que cada uno, por separado o de forma interrelacionada, podría ser analizado y conformar el tema central de la publicación de un artículo, libro o incluso una tesis. Además, con la posibilidad de aplicarse a diversos ámbitos debido a su naturaleza transversal. No obstante, dentro de las nuevas tecnologías disruptivas, se considera que la que reclama una atención preferente es la relativa a la Inteligencia Artificial (IA, en adelante), fundamentalmente por su rápida expansión e integración, así como por los desafíos que plantea en materia de privacidad y seguridad. De ahí que en el presente artículo pone el foco en la integración de la IA en el contexto educativo de Educación Superior (ES, en adelante) y sus implicaciones respecto del derecho fundamental de protección de datos personales.

La rápida expansión y utilización generalizada de la IA en los últimos años la ha convertido en uno de los temas tendencia de investigación científica y de debate más relevantes en la actualidad, así como también generadores de mayor interés y preocupación social. La IA ha generado tanta exaltación por sus posibilidades como preocupación por sus riesgos, lo que hace urgente crear espacios de reflexión académica.

Si bien la literatura jurídica y pedagógica reciente ha explorado cómo la IA ha transformado el paradigma de la ES, la mayoría de estas aportaciones coinciden en considerar a la IA como un motor de transformación radical que, simultáneamente, plantea severas limitaciones y riesgos para la integridad académica.

La tesis que sostiene esta autora se alinea con una visión propositiva pero cautelosa y garantista, en el sentido de que la IA debe ser entendida como un instrumento a nuestro favor, facilitador, de soporte y apoyo complementario en las funciones y tareas propias de los agentes que intervienen en el ámbito de la ES; siempre que su evolución técnica sea acompañada por un despliegue ético y normativo que garantiza un uso conforme a la legalidad, además de responsable y ético.

La incidencia de la IA en la ES trasciende la mera optimización del proceso de enseñanza-aprendizaje, especialmente con la aparición de la denominada "IA Generativa" (IAG, en adelante) y las metodologías activas; sino que también es relevante por su posible impacto en los derechos y libertades fundamentales de las personas físicas, entre los que nos centraremos en el derecho a la protección de los datos personales y a la educación. En este sentido, la normativa reguladora del sistema universitario en España[1], asigna a la ES los fines esenciales de (i) la educación y formación integral de los estudiantes, (ii) la formación académica y profesional, (iii) la generación, transmisión y difusión del conocimiento, (iv) así como la promoción de la innovación y la preparación para el aprendizaje permanente. Fines que en la sociedad digitalizada y globalizada en la que vivimos deben convivir con algoritmos, predicciones y diversos tratamientos de datos masivos.

A los efectos del presente artículo y para dotar de rigor conceptual a este análisis, interesa hacer una breve referencia a la distinción conceptual entre IA, sistemas de IA, sistema de IA de uso general, modelo de IA de uso general e IAG. Así, se entiende la *IA* puede ser definida como aquella "*disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico*". En otras palabras, constituye un campo de la informática que busca replicar procesos cognitivos

[1] En España, las funciones del sistema universitario se reflejan actualmente en el art. 2 de la Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario (LOSU). Disponible en el siguiente enlace: https://www.boe.es/buscar/act.php?id=BOE-A-2023-7500&p=20240802&tn=1#a2. Más información en el Código de Universidades. Disponible en el siguiente enlace: https://www.boe.es/biblioteca_juridica/codigos/codigo.php?id=133&modo=2&nota=0&tab=2.

humanos en máquinas[2]. De otro lado, la Ciencia y la Cultura (UNESCO) define los *sistemas de IA* como las "*tecnologías de tratamiento de la información que integran modelos y algoritmos y que producen una capacidad de aprendizaje y de realización de tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales. (…) están diseñados para funcionar con diversos grados de autonomía, mediante la modelización y representación del conocimiento y la explotación de datos y el cálculo de correlaciones (…)*"[3]. En similar sentido, se define en el vigente Reglamento europeo de IA (RIA, en adelante)[4].

En cambio, el concepto de *sistema de IA de uso general* es entendido como "*sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA*", según el artículo 3, 66) RIA). Diferente del concepto de *modelo de IA de uso general*, que se define por el RIA en su artículo 3, 63) como: "*modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado*".

Por último, la IAG se identifica por su capacidad de "*generar un contenido a partir de nuestras peticiones*" denominadas generalmente "*prompts*", siendo las más utilizadas de forma cotidiana los *Largue Language Models* (LLM), capaces de "*procesar texto en lenguaje natural para producir respuestas a partir del entrenamiento*" utilizando gran cantidad de datos[5], siendo las herramientas más conocidass *ChatGPT*, *Copilot* de Microsoft o *Gemini* desarrollado por Google[6] .

Distinguidos los anteriores conceptos, resulta evidente que la integración de la IA y su aplicación como IAG en el ámbito educativo universitario no es una mera posibilidad futura, sino que es una realidad presente en la ES en todas sus modalidades. Así, es una realidad fáctica susceptible de constituir una amenaza al derecho fundamental a la protección de datos personales consagrado en el ordenamiento jurídico español en el artículo 18.4 de la Constitución Española de 1978 (CE, en adelante) al expresar: "*La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos*"[7]. Al respecto, interesa citar que corresponde a la AEPD, como autoridad nacional de control, la supervisión del cumplimiento de la normativa europea aplicable reguladora de la materia, fundamentalmente, el Reglamento General de Protección de Datos (RGPD, en adelante)[8] y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD, en adelante)[9].

De forma paralela, la CE reconoce en su artículo 27.1 el derecho fundamental a la educación,

[2] Cruz Gómez, A. *Introducción a la Inteligencia Artificial y algoritmos IFCT155PO* (Almarir Consultores, editado por Formadir, S.L., 2025),13-14.

[3] UNESCO, "Recomendación sobre la ética de la Inteligencia Artificial", (2021). Disponible en el siguiente enlace: https://www.unesco.org/es/legal-affairs/recommendation-ethics-artificial-intelligence.

[4] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828. Disponible en el siguiente enlace: https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079. Establece un marco jurídico uniforme, que recoge una utilización de la IA conforme con los valores de la Unión Europea, centrada en el ser humano, y garantizando un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales. los define en su artículo 3, 1) como "un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales".

[5] Gómez, *Introducción a la Inteligencia Artificial y algoritmos IFCT155PO*,13-14.

[6] El RIA en su Considerando 99 precisa que: "*Los grandes modelos de IA generativa son un ejemplo típico de un modelo de IA de uso general, ya que permiten la generación flexible de contenidos, por ejemplo, en formato de texto, audio, imágenes o vídeo, que pueden adaptarse fácilmente a una amplia gama de tareas diferenciadas*".

[7] Ello, en línea con lo dispuesto en el artículo 8 de la Carta de los Derechos Fundamentales de la Unión Europea y el artículo 16 del Tratado de Funcionamiento de la Unión Europea.

[8] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. Disponible en el siguiente enlace: https://www.boe.es/doue/2016/119/L00001-00088.pdf.

[9] Disponible en el siguiente enlace: https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

reconociendo en sus dos primeros apartados que: "*Todos tienen el derecho a la educación. Se reconoce la libertad de enseñanza" y "La educación tendrá por objeto el pleno desarrollo de la personalidad humana en el respeto a los principios democráticos de convivencia y a los derechos y libertades fundamentales*". Ambos derechos tienen la consideración de "fundamentales" gozando, por tanto, de una protección reforzada; y al regularse en la Sección primera del Capítulo segundo del Título I, "De los derechos y deberes fundamentales" de la CE, entra en aplicación el límite extrínseco general previsto en su artículo 10.1 CE, que prevé de forma literal: "*La dignidad de la persona, los derechos inviolables que le son inherentes, el libre desarrollo de la personalidad, el respeto a la ley y a los derechos de los demás son fundamento del orden político y de la paz social*". De ahí que configuran dos derechos fundamentales que configuran un bloque de protección reforzado que orienten cualquier implementación tecnológica en las aulas, debiendo proteger la dignidad y el libre desarrollo de la persona.

Volviendo a la normativa reguladora de IA, se ha constatado que el RIA hace referencia expresa al derecho a la educación en doce ocasiones, esto es, en varios de sus Considerandos (4, 44, 48, 56, 96), en dos artículos (el 4 cuando se refiere a la Alfabetización en materia de IA y 9 sobre el sistema de gestión de riesgos) y finalmente, en su Anexo III sobre el citado sistema de IA de alto riesgo. En dichos apartados, se pone de relieve que la IA contribuye a generar beneficios diversos en todos los sectores, y en particular, en los ámbitos de la educación y la formación "*(...) al mejorar la predicción, optimizar las operaciones y la asignación de los recursos, y personalizar las soluciones digitales que se encuentran a disposición de la población y las organizaciones*" (Considerando 4 RIA). Además, refleja que "*(...) los sistemas de IA que detectan o deducen las emociones o las intenciones de las personas físicas a partir de sus datos biométricos pueden tener resultados discriminatorios y pueden invadir los derechos y las libertades (...) Por tanto, debe prohibirse la introducción en el mercado, la puesta en servicio y el uso de sistemas de IA destinados a ser utilizados para detectar el estado emocional de las personas en situaciones relacionadas con el lugar de trabajo y el ámbito educativo. (...)*" (Considerando 44 RIA). Es decir, introduce la prohibición para tales finalidades en ámbito educativo en el que manifiesta un desequilibrio de poder existente entre el profesorado y el alumnado, y, por último, considera que es para clasificar un sistema de IA como de alto riesgo, es importante tener en cuenta la magnitud de las consecuencias adversas de un sistema de IA para los derechos fundamentales, entre los que se incluyen específicamente el derecho a la protección de datos y el derecho a la educación (Considerando 48 RIA).

Desde el punto de vista normativo y de protección de datos, el RGPD hace referencia expresa al contexto educativo en su Considerando 132, instando a las autoridades de control a promover actividades de sensibilización específicamente en el contexto educativo: "Entre las actividades de sensibilización del público por parte de las autoridades de control deben incluirse medidas específicas dirigidas a los responsables y los encargados del tratamiento, (...), así como las personas físicas, en particular en el contexto educativo". En el ordenamiento interno, aunque la LOPDGDD no hace referencia expresa a la IA, sí consagra en su artículo 83 el derecho a la educación digital, que obliga al sistema educativo a garantizar la plena inserción del alumnado "en la sociedad digital y el aprendizaje de un consumo responsable y un uso crítico y seguro de los medios digitales y respetuoso con la dignidad humana, la justicia social y la sostenibilidad medioambiental, los valores constitucionales, los derechos fundamentales y, particularmente con el respeto y la garantía de la intimidad personal y familiar y la protección de datos personales (...)". Para ello, impone la incorporación de la competencia digital y la prevención de riesgos asociados al uso inadecuado de las nuevas tecnologías en los currículos, la formación específica del profesorado en estos ámbitos, la inclusión de contenidos sobre uso seguro de medios digitales y derechos en los planes de estudio universitarios, y la incorporación de materias sobre derechos digitales y protección de datos en los procesos selectivos de la Administración Pública.

Partiendo de las anteriores premisas, el objetivo general del presente artículo, titulado "Inteligencia Artificial en la Educación Superior y el derecho fundamental a la protección de datos personales: ¿innovación pedagógica o amenaza a la privacidad?", es examinar de forma analítica y reflexiva -desde una doble perspectiva jurídica y didáctica- el impacto de esta convergencia. Se pretende dar respuesta a la pregunta detonante relativa a cómo podríamos integrar la IA en el contexto educativo de ES como un motor de innovación pedagógica y una herramienta a favor de la sociedad sin que implique una amenaza a la privacidad y a la ética.

La perspectiva que se defiende sostiene que la IA es una herramienta de gran potencial y utilidad en todos los ámbitos y, en particular de la ES, cuya adopción exige un cambio de paradigma en la sociedad en general y en los agentes que participan en ámbito educativo a través de una cultura y alfabetización en IA y en competencias digitales. Resulta imperativo transitar hacia una cultura de competencias digitales que no solo aborde qué herramientas existen y cómo utilizarlas, sino también para qué usos y fines y bajo qué límites legales y éticos. En definitiva, se defiende un modelo de corresponsabilidad y gobernanza ética, donde la supervisión humana y la privacidad desde el diseño y por defecto no sean meros requisitos formales, sino pilares de una integración académica segura, responsable, ética y eficaz.

2. Usos de IA en Educación Superior: el impacto de la IAG

Partiendo de la consideración de que la IA es un arma de doble filo en ámbito educativo de ES, que conlleva múltiples ventajas, pero, paralelamente, genera riesgos y desafíos. Su implementación exige un análisis crítico que pondere, de forma equilibrada, la innovación frente a la preservación de derechos fundamentales. Por un lado, entre las principales ventajas del uso de la IA en este ámbito y para la mejora del proceso enseñanza-aprendizaje, destacan para los docentes, la reducción de carga administrativa o automatización de tareas administrativas como la calificación o control de asistencia; o material de impartición de clase o evaluación, o la posibilidad de realizar un ofrecimiento de un aprendizaje personalizado. Los docentes pueden integrar la IA, por ejemplo, en la evaluación, en la producción de material formativo, con la inclusión de herramientas de IA o incorporación de herramientas de IA en el aula por pos docentes o por las metodologías activas. Para los estudiantes universitarios, resaltan el aumento de su motivación mediante la gamificación y la interactividad, el acceso rápido a la formación, o el acercamiento a la realidad de la práctica profesional actual y adquisición de competencias digitales necesarias. Los estudiantes, pueden aplicarla IA en EL acceso a fuentes de información, utilización de IAG en la redacción de trabajos, trabajos colaborativos o en equipo en línea.

Por otro lado, respecto a los desafíos y riesgos del uso de la IA, se determinan en función de los usos o aplicaciones concretas que se le otorguen a la IA dentro del ES. En este punto, entre los principales usos que la IA puede tener en el sistema de ES, destacan los siguientes dependiendo del tipo de agente interviniente (docentes, estudiantes, institución de ES), por lo que no todos los usos de la IA son susceptibles de generar los mismos desafíos, sino que lo serán en mayor o menor medida y requerirán la toma de unas medidas concretas de mitigación y/o prevención, en función de su aplicación y finalidad en el caso en particular.

Por consiguiente, resulta imperativo desglosar estas aplicaciones bajo una óptica de proporcionalidad y gestión del riesgo, atendiendo a la naturaleza del agente y al impacto potencial sobre los derechos:

- (i) En la docencia: La IA se presenta como un colaborador en la creación de contenidos y en el proceso de enseñanza, pero también en el diseño de actividades y en la evaluación. Entre los posibles riesgos se encuentra el acceso y transmisión de información errónea o fruto de "alucinaciones de la IA", por lo que se exige que el docente mantenga la responsabilidad y el control humano.

- (ii) En el ámbito del alumnado: La IA potencia la autonomía y la motivación, pero su uso plantea interrogantes éticos sobre la integridad académica y el desarrollo de competencias críticas y reflexivas. La respuesta regulatoria debe ser la regulación, el fomento de la transparencia y de una cultura de uso responsable que previene el plagio y garantiza que la tecnología sea un medio para el conocimiento y una herramienta de apoyo o facilitadora del aprendizaje.

- (iii) En la esfera institucional: La adopción de sistemas de IA para la gestión de admisiones de alumnado y revisión de solicitudes, la asignación de becas o la analítica de aprendizaje sitúa a la institución en una posición de garantía. Aquí, el desafío no es solo técnico, sino de transparencia algorítmica, que requiere una auditoría de cumplimiento constante para evitar que sesgos en los datos deriven en decisiones discriminatorias que afectan la equidad en el acceso al sistema educativo.

En esta misma línea, ya el RIA establece en su Considerando 56 que: "*El despliegue de sistemas de IA en el ámbito educativo es importante para fomentar una educación y formación digitales de alta calidad y para que todos los estudiantes y profesores puedan adquirir y compartir las capacidades y competencias digitales necesarias, incluidos la alfabetización mediática, y el pensamiento crítico, (…). No obstante, deben clasificarse como de alto riesgo los sistemas de IA que se utilizan en la educación o la formación profesional, y en particular aquellos que determinan el acceso o la admisión, distribuyen a las personas entre distintas instituciones educativas y de formación profesional o programas de todos los niveles, evalúan los resultados del aprendizaje de las personas, evalúan el nivel apropiado de educación de una persona e influyen sustancialmente en el nivel de educación y formación que las personas recibirán o al que podrán acceder, o supervisan y detectan comportamientos prohibidos de los estudiantes durante las pruebas, ya que pueden decidir la trayectoria formativa y profesional de una persona y, en consecuencia, puede afectar a su capacidad para asegurar su subsistencia. (…)*".

Por ello, como se indicaba, es necesario tener en cuenta que aplicar los sistemas de IA en este ámbito puede tener riesgos e implicaciones no tan positivas, como los reflejados expresamente en el mencionado Considerando, que seguidamente expresa que: "*(…) Cuando no se diseñan y utilizan correctamente, estos sistemas pueden invadir especialmente y violar el derecho a la educación y*

la formación, y el derecho a no sufrir discriminación, además de perpetuar patrones históricos de discriminación (…)".

Ahora bien, se puede considerar que existen dos grandes posicionamientos o posturas en torno a la integración y utilización de la IA en ámbito universitario. Un grupo de autores[10] se muestra favorable y, aunque reconoce posibles limitaciones y riesgos (como la desinformación, la disminución de capacidades de expresión y otras competencias esenciales como la creatividad o la pérdida del pensamiento crítico) - respecto a los que proponen medidas mitigadoras, se centran en sus ventajas y oportunidades, así como en la realidad del contexto actual digitalizado. Como medidas mitigadoras, proponen principalmente (i) su uso ético y responsable, de modo que se garantice el respeto de los principios y valores éticos - asegurando equidad y transparencia; (ii) la alfabetización digital y en materia de IA para fortalecer la competencia digital, (iii) la continua evaluación crítica de la información; o (iv) la homogeneización normativa. Consideran, en este sentido, que "*las universidades pueden potenciar la calidad de la enseñanza, personalizar el proceso de aprendizaje de los estudiantes y preparar a la próxima generación de profesionales para un mundo impulsado por la tecnología*"; concluyendo que "*el uso de la IA para su plena explotación pedagógica está supeditada a una alfabetización digital*"[11]. En esta línea, se propone, además, la elaboración de una normativa de uso de IAG en el ámbito docente, reflexionando que es esencial garantizar un acceso universal y justo a las nuevas tecnologías, y en especial, a la IA.

Otro grupo de autores que se mantienen al margen de todo lo relacionado con la IA – lo que puede ser una postura incluso peligrosa por estar alejada de la realidad en las sociedades, especialmente aquellas desarrolladas tecnológicamente-; o bien, se muestran contrarios a su uso e incorporación, considerando incluso su prohibición para generar "*los mejores resultados en cuanto al aprendizaje, las habilidades y el bienestar de los estudiantes. Además, una prohibición implicará una menor participación de la universidad en resultados y procesos moralmente objetables*"[12].

Desde mi perspectiva, aunque este último posicionamiento es permisible y puede justificarse, en el contexto actual y siguiendo la tendencia de las sociedades globalizadas y cada vez más interconectadas del siglo XXI, sería más adecuada mantener una postura favorable hacia el uso de la IA en favor de las Universidades y los agentes que en ella interactúan, pero siempre con el establecimiento de las oportunas garantías y con la implantación supervisada de medidas mitigadoras y preventivas eficaces, unidas a un seguimiento continuado bajo el principio de mejora continua. Ello, con observancia, en todo caso, de las exigencias de la normativa reguladora de la IA, así como de la ética y la responsabilidad.

En este mismo enfoque se posiciona Navarro Salcedo[13], que recomienda implementar programas de capacitación docente en nuevas tecnologías, incluida y especialmente la IA y, además, estrategias de internacionalización. A este respecto, se traen a colación, en primer lugar, los resultados de un estudio en el que se analizaba cómo la IAG ha revolucionado el contexto universitario, cuyo principal hallazgo es que la formación en IAG no solo mejora su conocimiento por los estudiantes (fomentando su uso consciente, crítico, ético y responsable), sino que, además, conlleva una mejora de los elementos de las asignaturas sobre competencias digitales e IAG, concluyendo la necesidad de adaptar el enfoque sobre la IAG a las disciplinas específicas, especialmente a las menores tecnológicas[14]. En segundo lugar, interesa hacer referencia a los resultados de un estudio comparativo de guías publicadas sobre uso de IA en ámbito universitario español. Concluye que existe heterogeneidad y ausencia de marcos comunes, resalta la importancia de la alfabetización en IA y propone políticas unificadas y coordinadas realizadas por expertos internacionales para evitar diferencias significativas.

[10] Entre los que se citan Pablo Gallego Rodríguez. "El derecho constitucional a la educación y la IA generativa: propuestas para mitigar la exclusión digital educativa". *Estudios De Deusto*, 73, (1), (2025): 183-212. DOI: https://doi.org/10.18543/ed.3329; y Alfonso López-Pulido y José Manuel Sánchez. "La Inteligencia Artificial en contextos educativos: usos éticos de la información y alfabetización digital". *Derecom. Revista Internacional de Derecho de la Comunicación y de las Nuevas Tecnologías*, 38(1), (2025): 3-11.DOI: https://doi.org/10.5209/dere.102342.

[11] Flor M. Mella-Mella, María A. Calatayud y Ángel J. Lucas. "Uso de la IA como Estrategia de y para el Aprendizaje en Educación Superior". *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 24(1) (2026). DOI: https://doi.org/10.15366/reice2026.24.1.007.

[12] Karl de Fine Licht. "Generative Artificial Intelligence in Higher Education: Why the 'Banning Approach' to Student use is Sometimes Morally Justified". *Philos. Technol.* 37, (2024): 113. DOI: https://doi.org/10.1007/s13347-024-00799-9

[13] Gabriel Navarro Salcedo. "Internacionalización y Formación Docente: El Rol de las TIC en la Educación Superior." *Revista De La Escuela De Ciencias De La Educación*. 21 (2026). DOI: https://revistacseducacion.unr.edu.ar/index.php/educacion/article/view/907

[14] Teresa Romeu Fontanillas et al., "Desafíos de la Inteligencia Artificial generativa en educación superior: fomentando su uso crítico en el estudiantado". *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2), (2025): 209–231. DOI: https://doi.org/10.5944/ried.28.2.43535.

En este sentido, se hace referencia a que la AEPD ha publicado en noviembre de 2025 su propia Política general interna para el uso de IAG en procesos administrativos de la AEPD[15], que establece una política general para la implementación, gobernanza y uso responsable de sistemas de inteligencia artificial generativa en el ámbito interno de la AEPD. Su finalidad es reforzar la capacidad tecnológica y organizativa de la Agencia, asegurando una transformación digital segura, ética y plenamente conforme con el marco normativo vigente.

Por su parte, algunas universidades también han elaborado su propia Política, sus Códigos, o, como en el caso de la Universidad Internacional de la Rioja (UNIR), una "*Declaración para un uso ético y responsable de la IA en ES*"[16] para sentar las bases de consenso que guíen a todos los actores (docentes, estudiantes, personal de gestión, investigadores, etc.) en el uso, aplicación, e incluso desarrollo de soluciones basadas en IA. Resulta especialmente interesante que en dicha Declaración de la UNIR prevé que los principios que fundamentan el diseño, desarrollo y aplicación de las soluciones y herramientas de IA son los siguientes: (i) el principio de Contribución social, "*se garantizará el alineamiento entre los intereses propios de proyectos científicos y esfuerzos técnicos relacionados con la IA y aquellos de la sociedad*" (ii) equidad, "*se enfatizará el uso de sistemas IA sin ningún tipo de discriminación de usuario o público objetivo" (…) "se evitarán, por tanto, los sesgos por sectores o cualquier otra categorización*", (iii) capacitación en el uso de sistemas de IA, "*para que sigan prácticas responsables en el uso, distribución, divulgación y producción de las tecnologías y servicios basados en IA, coherentes con las normas éticas del grupo. Asimismo, se fomentará la concienciación sobre los mismos, a cualquier nivel, en cualquier sector del ecosistema educativo. Del mismo modo, se formará en los riesgos específicos del uso de IA Generativa para minimizar los posibles efectos negativos en el aprendizaje, la enseñanza y la evaluación*".

También, se establece (iv) el principio de supervisión humana, "*se mantendrán el uso, configuración e implementación de la IA de forma controlada por personal competente*", (v) sostenibilidad, "*para minimizar el impacto ambiental y consumo energético. Se procura dejar la menor huella ecológica posible y la aplicación del principio DNSH (No Causar Daño Significativo, por sus siglas en inglés)*"; y los principios de (vi) transparencia e identificación, "*se garantizarán la identificación y trazabilidad de los contenidos producidos por IA. Todo sistema de IA generativa desarrollado o utilizado por UNIR indicará claramente su origen artificial y su trazabilidad asociada de manera inequívoca, ya sean imágenes, vídeos, textos o cualquier otro formato o producto*"; y (vii) Principio de Confidencialidad, "*se respetarán la seguridad y privacidad de datos en uso y desarrollo de sistemas IA*", sobre el que se centra el siguiente apartado. Asimismo, esta serie de principios viene supervisada por un sistema de auditoría y seguimiento interno eficaz, adaptado a los requerimientos del RIA, que evalúa periódicamente el nivel de cumplimiento del compromiso que establece el presente manifiesto, y que sugiere medidas correctivas y transparentes en caso de identificar cualquier desviación significativa. En definitiva, y se puede tomar como ejemplo que debería ser generalizado en ámbito universitario, la UNIR, declara su compromiso "*a utilizar la inteligencia artificial de manera ética y responsable, a potenciar el conocimiento, promover la innovación y beneficiar a la sociedad en su conjunto respetando siempre la legislación vigente*".

En virtud de todo lo anterior, desde mi perspectiva y mi experiencia profesional, la IA es una herramienta muy útil que ha de ser vista desde un enfoque positivo en el proceso de enseñanza-aprendizaje en ES. La integración de la IA en la praxis docente no debe ser interpretada como una opción, sino como un imperativo de adecuación a la realidad social y tecnológica. No obstante, esta transición exige un compromiso deontológico y profesional trascendental. La postura institucional y docente no puede ser la del aislamiento o la evitación, sino la de una alfabetización crítica que enseñe a gestionar no solo la potencialidad de estas herramientas de IA, sino también sus riesgos intrínsecos en materia de privacidad y seguridad de datos, así como en la ética.

La integración de la IA en este sector está transformando la forma en que se ejerce la profesión de la docencia y también el rol del alumnado, respecto a lo que no debemos tener miedo ni significa que vaya a desaparecer la función docente o investigadora, pero está claro que están cambiando radicalmente y lo seguirá haciendo durante los próximos años. Esto debe servir para abrir una reflexión mucho más importante sobre a que sí puedan desaparecer aquellos profesionales que no evolucionen junto a los cambios tecnológicos o se adapten a estos de forma proactiva y resiliente. En mi opinión, los profesionales tienen que ser educadores, bajo el modelo tecnopedagógico TPACK (*Technological Pedagogical Content Knowledge*), que defiente que el docente actual debe ser un líder del conocimiento capaz de armonizar las tres áreas de la ciencia, la pedagogía y la tecnología, así como sus intersecciones. Esta transformación exige una actitud de vanguardia en una sociedad globalizada, donde la IA actúa a su favor y no como un sustituto del criterio humano.

[15] Disponible en el siguiente enlace: https://www.aepd.es/documento/politica-iag-aepd.pdf.

[16] Disponible en el siguiente enlace: https://www.unir.net/actualidad/responsabilidad-social-corporativa/declaracion-unir-para-un-uso-etico-de-la-inteligencia-artificial-en-educacion-superior/.

Para finalizar, en línea con lo anterior, se considera que la ES no debe mantenerse al margen de las novedades tecnológicas de la sociedad digital y globalizada e interconectada de la que forma parte, sino que debe evolucionar y transformarse a su ritmo. El papel de los docentes de ES es esencial en la promoción del conocimiento, la integración y el manejo de las aplicaciones de la IA con un uso adecuado, legal, ético y responsable, con garantía de los derechos fundamentales. Resulta esencial conocer y compartir con el alumnado que el uso de la IA también puede conllevar efectos negativos o desafíos, como la amenaza a la privacidad o la seguridad, respecto a los que hay que ir más allá y buscar soluciones y prevenir o minimizar su impacto de forma, proactiva y no reactiva. Solo mediante una estrategia de una formación proactiva y compartida entre docentes y alumnado sobre los efectos adversos de la IA, se podrá asegurar que la innovación tecnológica fortalezca, y no deteriore, la integridad, la calidad y el rigor de la ES.

3. Principales retos y desafíos de la utilización de IA en las universidades: Privacidad y protección de datos

La promoción de competencias digitales y alfabetización digital, tanto dirigida al personal docente e investigador como al alumnado universitario, se considera clave para garantizar una integración de la IA en la ES con garantía y respeto de los derechos fundamentales a la educación y a la protección de datos. En una reflexión profunda sobre el tema, cabe cuestionarnos la pregunta sobre la que se centra el presente artículo: Inteligencia artificial en la educación superior y el derecho fundamental a la protección de datos: ¿innovación pedagógica o amenaza a la privacidad? Aunque la IA promete optimizar aprendizaje, su implementación debe superar una serie de obstáculos críticos, que van desde la brecha digital hasta la ausencia de marcos regulatorios comunes y sólidos que guían su despliegue en la educación.

Siguiendo a Flores Portillo[17] y Pedroso[18], entre los retos destacan la brecha digital y necesidad de alfabetización digital; de garantizar transparencia, privacidad y protección de datos, así como la ética y confiabilidad de los sistemas, junto a los posibles sesgos algorítmicos. Además, la ausencia de marcos regulatorios comunes y éticos solidos para guiar la implementación de la IA y en particular, en ámbito de ES. En particular, se considera esencial el enfoque de privacidad por defecto y desde el diseño o también conocido en inglés como "*privacy by design*" y "*privacy by default*". Al respecto, sabe señalar que el principio de privacidad desde el diseño implica, por un lado, que se debe utilizar un enfoque "*orientado a la gestión del riesgo y de responsabilidad proactiva para establecer estrategias que incorporen la a protección de la privacidad a lo largo de todo el ciclo de vida del objeto*"[19]. Por otro lado, el principio de protección de datos por defecto hace referencia a las "*las elecciones realizadas con respecto a los valores de configuración u opciones de tratamiento fijadas en los sistemas y procedimientos que implementan el tratamiento y que determinan la cantidad de los datos personales recopilados, el alcance de su procesamiento, el período de su conservación y su accesibilidad*". Ello implica que se debe tener en cuenta en el uso de herramientas y soluciones de IA el respeto de la privacidad y protección de datos, por ejemplo, mediante la eliminación de cualquier dato personal tal y como se define en el RGPD, que pudiera identificar o hacer identificable a una persona física, y máxime si son de los datos considerados categorías especiales (anteriormente, sensibles), como los relacionados con salud o ideología política.

Navarro Salcedo[20], mencionado anteriormente, resalta igualmente el desarrollo de habilidades digitales, el principio de mejora continua y la integración efectiva de las tecnologías y la IA; así como la redefinición de estrategias de evaluación basadas en las herramientas digitales para evaluar de manera justa y válida. En esta misma línea concluye la investigación desarrollada por Niño[21] sobre la opinión de los estudiantes respecto a la IA, quienes reportan percepciones positivas,

[17] María de Lourdes Flores, Juan D. Machin-Mastromatteo, y Javier Tarango. "Propuesta metodológica para analizar la aceptación, uso e integración de la inteligencia artificial generativa en instituciones de educación superior". *AULA Revista De Humanidades Y Ciencias Sociales*, 72 (1) (2025).

[18] João Pedroso, "IA, Derechos fundamentales y regulación: Constitucionalismo digital y AI Act". Ponencia presentada en la Facultad de Derecho, Derecho Constitucional de la *Universidad de Jaén*, noviembre, 2025.

[19] AEPD, *Guía de Privacidad desde el Diseño*. Disponible en el siguiente enlace: https://www.aepd.es/guias/guia-privacidad-desde-diseno.pdf. AEPD, *Guía de Privacidad por Defecto*. Disponible en el siguiente enlace: https://www.aepd.es/guias/guia-proteccion-datos-por-defecto.pdf.

[20] Navarro Salcedo. "Internacionalización y Formación Docente…".

[21] Shamaly A. Niño et al., "Inteligencia artificial en la formación universitaria: una revisión de estudios centrados en la opinión de los estudiantes". Formación universitaria, 18(2), (2025): 107-124. DOI: https://dx.doi.org/10.4067/s0718-50062025000200107.

pero insuficiencia de conocimiento y habilidades para su uso, junto a preocupaciones sobre la privacidad y los efectos negativos de su uso excesivo, en particular, en el pensamiento crítico y creativo.

Otra investigación también relevante que se trae a colación y con la que coincido es la de Castedo[22], que analiza los marcos regulatorios del m-learning (aprendizaje móvil), concretando sus implicaciones a la luz de la normativa de protección de datos, y proponiendo un modelo de gobernanza algorítmica y medidas verificables contra la brecha digital para proteger los derechos fundamentales en el contexto universitario. Considera que ante los retos que plantea el m-learning y la inteligencia artificial en la ES, las universidades deben pasar del simple análisis de riesgos a una praxis reguladora efectiva. Desde mi perspectiva, la digitalización solo reforzará la calidad educativa si se asegura un entorno transparente, accesible y protegido, donde la tecnología sea una herramienta pedagógica que no sustituya el juicio académico ni comprometa la equidad o la autonomía de los estudiantes.

También comparto las conclusiones de Cotino Hueso en su publicación sobre La Carta de los Derechos Digitales[23], en la que analiza los avances de la Carta de Derechos Digitales en la relación entre la ciudadanía y la Administración Pública, destacando principios que coinciden plenamente con los desafíos y soluciones propuestas en las fuentes respecto al m-learning y la IA en el ámbito universitario. En sintonía con ambas posturas, se concluye que la legitimidad de la IA y el m-learning en ES depende de la transición de un marco teórico de derechos a una praxis reguladora proactiva y verificable.

El reto principal es lograr que la integración y utilización de la IA se realice de un modo lícito, ético y responsable, en ámbito universitario, institucional y en la sociedad en general. Con garantía de cumplimiento de la normativa de protección de datos materia en la que plantea desafíos por el tratamiento masivo de información, falta de transparencia, la opacidad de ciertos modelos algorítmicos y de otros riesgos de usos adicionales no previstos.La recogida y análisis de datos académicos, informes del alumnado y del personal universitario puede afectar a este derecho fundamental si no se aplican estrictamente los principios del RGPD, en particular la licitud, minimización, limitación de la finalidad, transparencia y responsabilidad proactiva. A ello se suma el riesgo de sesgos algorítmicos, decisiones automatizadas con efectos significativos y posibles transferencias internacionales de datos, que exigen evaluaciones de impacto, supervisión humana efectiva y la incorporación de los enfoques de privacidad desde el diseño y por defecto.

En este contexto, se sostiene que las universidades deben adoptar políticas claras de gobernanza de la IA, reforzar las medidas de seguridad y promover la alfabetización digital, con el fin de garantizar una integración de la IA compatible con los derechos fundamentales y la confianza de la comunidad universitaria. La IA es, sin duda, un aliado para el presente y el futuro de la ES, pero su legitimidad depende de su capacidad para operar dentro de un marco de licitud, ética y responsabilidad. Solo a través de una aplicación estricta de los principios de transparencia y responsabilidad proactiva, así como de la normativa de protección de datos, podremos asegurar que la innovación pedagógica no erosione la confianza de la comunidad universitaria ni vulnere sus derechos más elementales, en particular, a la protección de datos personales. En definitiva, la conclusión final es que la ES debe liderar la digitalización y alfabetización en IA mediante una gobernanza robusta que proteja la autonomía intelectual y la equidad, asegurando que la IA sea un recurso de innovación pedagógica integrado en las aulas para potenciar el pensamiento crítico y analítico sin influir en la libertad de cátedra ni sustituir nunca el juicio académico ni comprometer o amenace los derechos fundamentales de las personas físicas, especialmente la privacidad.

4. Relevancia de la alfabetización digital y las competencias digitales en ámbito universitario

Anteriormente se ha expuesto la relevancia y necesidad de llevar a cabo un uso de la IA en ámbito universitario con un equilibrio entre innovación pedagógica y respecto de los derechos fundamentales.

A tal fin se han citado varias propuestas y recomendaciones a lo largo del mismo, así como medidas deben adoptar las universidades para garantizar el cumplimiento de protección de datos, entre las

[22] Castedo Espeso, A.: "M-learning y protección de datos: retos constitucionales en la Educación Superior", *Revista de Derecho de la UNED (RDUNED)*, nº 36, (2025), 81-100. DOI: 10.5944/rduned.36.2025.47720

[23] Lorenzo Cotino Hueso (ed.): *La Carta de Derechos Digitales*, Tirant lo Blanch, Valencia, (2022), 251 y ss.

que se consideran esenciales promover y dotar de recursos la alfabetización digital[24] y formación en competencias digitales. Esta propuesta educativa podría ser una estrategia que primero se forme en el personal docente e investigador, para que este a su vez forme y potencie las competencias digitales, incluida la IA, en los estudiantes universitarios.

Ante el hecho de que la IA está cada vez más presente en la sociedad, los profesionales de la educación y los sistemas educativos universitarios han de ser proactivos en materia de formación de competencias digitales, incluyendo la IA, para que los estudiantes sean conscientes de los riesgos, y lleven a cabo un uso lícito, seguro, ético y responsable de la IA. Al respecto de esta idea, interesa traer a colación que la UNESCO (en su función de apoyo a los Estados Miembros para que aprovechen el potencial de la IA con miras a la consecución de la Agenda de Educación 2030[25], al tiempo que "vela porque su aplicación en contextos educativos responda a los principios básicos de inclusión y equidad"[26]), en su "Marco de competencias para docentes en materia de IA[27] de 2025, sugiere estrategias de implementación, entre las que destacan la elaboración de políticas y condiciones propicias para el uso de IA en educación, así como el diseño y optimización de programas de capacitación en competencias de IA, para lograr la preparación a los estudiantes para ser ciudadanos responsables y creativos en la era de la IA. También se ha publicado un "Marco de competencias en IA para estudiantes"[28] de la UNESCO publicado en 2025, en aras de ayudar a los docentes con la integración de la IA, describiendo doce competencias en tres niveles de progresión: comprender, aplicar y crear; y pasan por cuatro dimensiones: una forma de pensar centrada en las personas, la ética de la IA, técnicas y aplicaciones de la IA y la IA para el diseño de sistemas de IA.

Por todo ello, se considera clave que en el ámbito de la ES se exija la formación en competencias digitales a los estudiantes universitarios, así como de alfabetización digital para obtener los mayores beneficios de los sistemas de IA con protección de los derechos fundamentales y la ética, así como la seguridad. Esta alfabetización, tal y como refleja el RIA en su Considerando 20, para la toma de decisiones con conocimiento de causa en relación con los sistemas de IA. En cuanto a la alfabetización en materia de IA, continúa expresando que "debe proporcionar a todos los agentes pertinentes de la cadena de valor de la IA los conocimientos necesarios para garantizar el cumplimiento adecuado y la correcta ejecución. Además, la puesta en práctica general de medidas de alfabetización en materia de IA y la introducción de acciones de seguimiento adecuadas podrían contribuir a mejorar las condiciones de trabajo y, en última instancia, sostener la consolidación y la senda de innovación de una IA fiable en la Unión. El Consejo Europeo de Inteligencia Artificial (en lo sucesivo, «Consejo de IA») debe apoyar a la Comisión para promover las herramientas de alfabetización en materia de IA, la sensibilización pública y la comprensión de los beneficios, los riesgos, las salvaguardias, los derechos y las obligaciones en relación con el uso de sistemas de IA. En cooperación con las partes interesadas pertinentes, la Comisión y los Estados miembros deben facilitar la elaboración de códigos de conducta voluntarios para promover la alfabetización en materia de IA entre las personas que se ocupan del desarrollo, el manejo y el uso de la IA".

Desde una perspectiva crítico-jurídica, la alfabetización en IA en ámbito de ES no es un valor añadido, sino un presupuesto de validez. La RIA establece que tanto proveedores como responsables del despliegue (en este caso, las Universidades) deben garantizar que las personas afectadas tengan el conocimiento necesario para comprender las oportunidades y, principalmente, los perjuicios potenciales de la IA. Por tanto, la formación en competencias digitales se configura como el escudo más eficaz contra la vulneración de derechos fundamentales. El éxito de esta transición se medirá por nuestra capacidad de utilizarlos para formar estudiantes y ciudadanos más responsables, informados, libres, éticos, críticos y protegidos en la era digital.

En este contexto, a lo largo del presente artículo se han mencionado diversas propuestas, recomendaciones y medidas que las ES deben adoptar para garantizar el cumplimiento de la normativa en materia de protección de datos y el uso responsable de la IA. Entre dichas medidas,

[24] Definida por el artículo 3, 56) del RIA como las "capacidades, los conocimientos y la comprensión que permiten a los proveedores, responsables del despliegue y demás personas afectadas, teniendo en cuenta sus respectivos derechos y obligaciones en el contexto del presente Reglamento, llevar a cabo un despliegue informado de los sistemas de IA y tomar conciencia de las oportunidades y los riesgos que plantea la IA, así como de los perjuicios que puede causar".

[25] Más información en el siguiente enlace: https://www.educacionfpydeportes.gob.es/biblioteca-central/blog/2024/febrero/agenda-2030-educacion.html.

[26] Recuperado del siguiente enlace: https://www.unesco.org/es/digital-education/artificial-intelligence.

[27] UNESCO, "Marco de competencias para docentes en materia de IA" (2025). Disponible en el siguiente enlace: https://www.unesco.org/es/articles/marco-de-competencias-para-docentes-en-materia-de-ia.

[28] UNESCO, "Marco de competencias en IA para estudiante", (2025). Disponible en el siguiente enlace: https://www.unesco.org/es/articles/marco-de-competencias-para-estudiantes-en-materia-de-ia.

destaca especialmente la promoción de programas estructurados de alfabetización digital y de formación en competencias digitales. Desde una perspectiva institucional, resulta razonable que esta estrategia educativa se implemente de forma progresiva y escalonada. La creciente presencia de la IA en múltiples ámbitos de la vida social, económica y profesional implica que los sistemas de ES deben adoptar una actitud proactiva en relación con la formación en competencias digitales. No se trata únicamente de capacitar técnicamente a los estudiantes para el uso de determinadas herramientas, sino de formar ciudadanos capaces de comprender el impacto social, ético y jurídico de estas tecnologías. En este sentido, la formación universitaria debe incorporar contenidos que permitan identificar los riesgos asociados al uso de sistemas de IA, tales como los sesgos algorítmicos, los problemas de privacidad, la opacidad en la toma de decisiones automatizadas o las posibles vulneraciones de derechos fundamentales.

En definitiva, se defiende la postura que entiende que las competencias digitales y en materia de IA constituyen un elemento esencial para maximizar los beneficios derivados del uso de los sistemas de IA a favor de la sociedad, al tiempo que permiten reducir o minimizar los riesgos asociados a su utilización, como los relativos a las amenazas a la privacidad con formación también en esta materia. En particular, la alfabetización digital permite comprender los mecanismos de funcionamiento de estas tecnologías, evaluar críticamente sus resultados y tomar decisiones informadas, responsables y éticas en relación con su uso. Esta unida a la cultura y formación en privacidad y protección de datos serán la fórmula del éxito de la integración innovadora pedagógica de la IA en la ES con respeto a la privacidad y sin amenazar la vulneración del derecho fundamental a la protección de datos, que permitirá responder favorablemente a la pregunta detonante a la que se refiere el presente artículo.

5. Conclusiones

La rápida expansión de la IA, y en especial de la IAG, está reconfigurando el contexto de la ES, conformando su integración una innovación pedagógica de alto impacto, siendo necesario y positivo aprovechar todo su potencial y sus ventajas, pero siempre con responsabilidad y respeto de la ética y la integridad académica. El uso de la IA de forma ética y responsable es una responsabilidad de todos.

Dentro de las tecnologías disruptivas, la IA reclama una atención preferente debido a su rápida expansión e integración, planteando desafíos críticos en materia de privacidad, seguridad e integridad académica. Se defiende una visión donde la IA actúa como un instrumento facilitador y de apoyo para los agentes educativos, siempre que su evolución técnica esté acompañada por un despliegue ético y normativo que garantice un uso legal y responsable.

La utilización de sistemas de la IA en ámbito universitario puede afectar de manera directa a derechos fundamentales, como el derecho a la protección de datos personales y al derecho a la educación, lo que justifica su consideración como tecnología de alto riesgo en determinados usos. Es fundamental tener en cuenta los principios de protección de datos, y en particular, el enfoque de privacidad por defecto y desde el diseño.

Se evidencia, una nueva forma de entender el proceso enseñanza-aprendizaje en el contexto universitario, respecto de la que se generan riesgos y desafíos que investigar, mitigar y prevenir. La alfabetización digital y la formación en competencias de IA y digitales en general en todos los agentes que intervienen, tanto del profesorado como del alumnado, se configuran como el principal mecanismo preventivo frente a los riesgos de discriminación, pérdida de pensamiento crítico y vulneración de la privacidad. No es solo un valor añadido, sino un presupuesto de validez y el escudo más eficaz contra la vulneración de derechos; es esencial para que tanto docentes como alumnos realicen un uso crítico, ético y seguro de estas herramientas.

Las normativas europeas de protección de datos y de IA proporcionan un marco imprescindible para garantizar una implementación lícita, transparente y responsable de la IA en las universidades. Es preferible adoptar una postura institucional proactiva y regulada frente a enfoques prohibitivos o pasivos, integrando la IA bajo principios de ética, equidad, supervisión humana, transparencia y confidencialidad.

Las instituciones deben adoptar un enfoque de responsabilidad proactiva, integrando la protección de la privacidad en todo el ciclo de vida de los sistemas de IA y configurando las herramientas para que, de forma predeterminada, minimicen el tratamiento de datos personales.

El docente actual debe transformarse en un líder del conocimiento capaz de armonizar el conocimiento, la pedagogía y la tecnología (modelo TPACK), adaptándose de forma proactiva para evitar el aislamiento frente a los cambios tecnológicos.

La adopción de políticas, códigos éticos y sistemas de auditoría —como los promovidos por la AEPD— constituye una buena práctica replicable y necesaria para consolidar una IA segura y fiable en la educación superior . La integración de la IA requiere un modelo de corresponsabilidad basado en principios como la transparencia, la equidad (evitando sesgos algorítmicos) y la supervisión humana efectiva, asegurando que la tecnología no sustituya nunca el juicio académico.

Bibliografía

Castedo Espeso, A.: "M-learning y protección de datos: retos constitucionales en la Educación Superior", *Revista de Derecho de la UNED (RDUNED)*, nº 36, (2025), 81-100. DOI: 10.5944/rduned.36.2025.47720

Cotino Hueso, Lorenzo (ed.): *La Carta de Derechos Digitales*, Tirant lo Blanch, Valencia, (2022).

Cruz Gómez, A. *Introducción a la Inteligencia Artificial y algoritmos IFCT155PO*: Almarir consultores, editado por Formadir, SL, 2025.

Fine Licht, Karl. "Generative Artificial Intelligence in Higher Education: Why the 'Banning Approach' to Student use is Sometimes Morally Justified". *Philos. Technol.* 37, (2024):113. DOI: https://doi.org/10.1007/s13347-024-00799-9.

Flores, María de L., Machin-Mastromatteo, Juan D. y Tarango, Javier. "Propuesta metodológica para analizar la aceptación, uso e integración de la inteligencia artificial generativa en instituciones de educación superior". *AULA Revista De Humanidades Y Ciencias Sociales, 72*(1) (2025).

Gallego Rodríguez, Pablo. "El derecho constitucional a la educación y la IA generativa: propuestas para mitigar la exclusión digital educativa". *Estudios De Deusto*, *73*(1), (2025):183-212. DOI: https://doi.org/10.18543/ed.3329.

López-Pulido Alfonso. y Sánchez José M. "La Inteligencia Artificial en contextos educativos: usos éticos de la información y alfabetización digital. *Derecom". Revista Internacional de Derecho de la Comunicación y de las Nuevas Tecnologías, 38*(1) (2025): 3-11. DOI: https://doi.org/10.5209/dere.102342.

Mella-Mella, Flor M., Calatayud, María A. y Lucas Ángel J. "Uso de la IA como Estrategia de y para el Aprendizaje en Educación Superior". *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 24*(1). (2026). DOI: https://doi.org/10.15366/reice2026.24.1.007.

Navarro Salcedo, Gabriel. "Internacionalización y Formación Docente: El Rol de las TIC en la Educación Superior". *Revista De La Escuela De Ciencias De La Educación, 1* (21) (2026). DOI: https://revistacseducacion.unr.edu.ar/index.php/educacion/article/view/907

Niño, Shamaly A., Castellanos, Juan C., López, Mónica L. y Parra, Karla L. "Inteligencia artificial en la formación universitaria: una revisión de estudios centrados en la opinión de los estudiantes". *Formación universitaria, 18*(2), (2025): 107-124. DOI: https://dx.doi.org/10.4067/s0718-50062025000200107.

Pedroso, João, "IA, Derechos fundamentales y regulación: Constitucionalismo digital y AI Act", Ponencia presentada en la Facultad de Derecho, Derecho Constitucional, Universidad de Jaén, noviembre 2025.

Romeu, Teresa, Romero, Marc., Guitert, Montse. y Baztán, Pablo. "Desafíos de la Inteligencia Artificial generativa en educación superior: fomentando su uso crítico en el estudiantado". *RIED-Revista Iberoamericana de Educación a Distancia, 28*(2), (2025): 209–231. DOI: https://doi.org/10.5944/ried.28.2.43535.

UNESCO, "Recomendación sobre la ética de la Inteligencia Artificial", (2021). Disponible en el siguiente enlace: https://www.unesco.org/es/legal-affairs/recommendation-ethics-artificial-intelligence.

UNESCO, "Marco de competencias para docentes en materia de IA", (2025). Disponible en el siguiente enlace: https://www.unesco.org/es/articles/marco-de-competencias-para-docentes-en-materia-de-ia.

UNESCO, "Marco de competencias en IA para estudiantes", (2025). Disponible en el siguiente enlace: https://www.unesco.org/es/articles/marco-de-competencias-para-estudiantes-en-materia-de-ia.

Gestión de riesgos y protección de datos: la tecnificación normativa ecuatoriana frente al paradigma europeo

Risk Management and Data Protection: Ecuadorian Normative Technification versus the European GDPR Paradigm

Darío Echeverría Muñoz
Abogado, Docente en Derecho Digital

Resumen:

El presente artículo analiza la evolución del marco regulatorio de protección de datos personales en la República del Ecuador, con especial énfasis en la estrategia de implementación técnica desplegada por la Superintendencia de Protección de Datos Personales (SPDP). A través de una metodología analítica-comparativa —que combina análisis documental de fuentes normativas primarias, revisión bibliográfica de literatura especializada y estudio de casos de herramientas regulatorias concretas— se examina el espectro regulatorio que transita desde un modelo garantista basado en derechos, pasando por el enfoque basado en riesgos del Reglamento General de Protección de Datos (RGPD), hasta consolidarse en un modelo de metarregulación. El estudio deconstruye la normativa ecuatoriana y sus guías de gestión de riesgos cuantitativa (FAIR, Monte Carlo), contrastando este enfoque prescriptivo con la flexibilidad operativa del modelo europeo representado por el sistema GESTIONA de la Agencia Española de Protección de Datos (AEPD) y el marco de la CNIL francesa. La investigación concluye que Ecuador apuesta por una objetivación matemática del cumplimiento, transformando obligaciones jurídicas abstractas en requisitos de ingeniería prescriptivos para cerrar la brecha de implementación regional, distanciándose del modelo de confianza europeo. Este enfoque, si bien reduce la incertidumbre jurídica, conlleva riesgos de rigidez normativa y exclusión de actores con recursos limitados.

Palabras clave:

LOPDP; RGPD; Metarregulación; Gestión de Riesgos; Análisis Comparado.

Abstract:

This article analyzes the evolution of the personal data protection regulatory framework in the Republic of Ecuador, with special emphasis on the technical implementation strategy deployed by the Superintendence of Personal Data Protection (SPDP). Through an analytical-comparative methodology — combining documentary analysis of primary normative sources, bibliographic review of specialized literature, and case studies of concrete regulatory tools— the regulatory spectrum is examined, transitioning from a guarantee-oriented rights-based model, through the GDPR's risk-based approach, to consolidate into a meta-regulation model. The study deconstructs the Ecuadorian regulations and quantitative risk management guides (FAIR, Monte Carlo), contrasting this prescriptive approach with the operational flexibility of the European model represented by the GESTIONA system of the Spanish Data Protection Agency (AEPD) and the French CNIL framework. The research concludes that Ecuador is betting on mathematical objectification of compliance, transforming abstract legal obligations into prescriptive engineering requirements to bridge the regional implementation gap, distancing itself from the European trust-based model. While this approach reduces legal uncertainty, it carries risks of normative rigidity and exclusion of actors with limited resources.

Keywords:

LOPDP; GDPR; Meta-regulation; Risk Management; Comparative Analysis.

1. Introducción

La protección de datos personales ha dejado de ser una disciplina jurídica periférica para convertirse en el eje central de la gobernanza digital contemporánea. En la República del Ecuador, esta transformación posee una dimensión profundamente constitucional. El artículo 66, numeral 19 de la Constitución de la República reconoce y garantiza el derecho a la protección de datos de carácter personal, elevando este mandato a un nivel de garantía fundamental.[1]

Sin embargo, la materialización de este derecho ha sufrido una metamorfosis acelerada que merece análisis riguroso. Si bien la Ley Orgánica de Protección de Datos Personales (LOPDP) de 2021 fue influenciada por el denominado «Efecto Bruselas» —fenómeno mediante el cual la Unión Europea extiende su poder regulatorio más allá de sus fronteras territoriales— y el RGPD, la implementación regulatoria liderada por la Superintendencia de Protección de Datos Personales (SPDP) en el periodo 2024-2025, y consolidada con la actualización a la versión 2 de su guía técnica en 2026, marca un hito disruptivo: el paso hacia una metarregulación.

Esta transformación no ocurre en un vacío regional. América Latina ha experimentado una ola de convergencia regulatoria en materia de protección de datos durante la última década. Brasil promulgó su *Lei Geral de Proteção de Dados* (LGPD) en 2018, estableciendo la *Autoridade Nacional de Proteção de Dados* (ANPD) en 2020. Chile actualizó su normativa mediante la Ley 21.096 de 2018, mientras que Colombia ha avanzado significativamente en la aplicación de su Ley 1581 de 2012. Argentina, pionera regional con su Ley 25.326 de 2000, obtuvo en 2019 la declaración de adecuación por parte de la Comisión Europea, reconociendo un nivel de protección esencialmente equivalente al europeo.

En este contexto regional, Ecuador se distingue no por ser el primero en regular, sino por la estrategia de implementación adoptada. A diferencia del modelo europeo, que descansa en la ponderación jurídica y el análisis cualitativo de riesgos bajo el principio de *accountability*, Ecuador ha optado por una estrategia que impone una ingeniería de la privacidad cuantificable. Este nuevo paradigma busca reducir la incertidumbre jurídica mediante la precisión de la ingeniería de datos, integrando estándares internacionales (ISO 27001, ISO 27701, NIST) y modelos matemáticos (FAIR, Monte Carlo) directamente en la normativa secundaria.[2]

La pregunta central que articula esta investigación es: ¿Constituye el modelo ecuatoriano una evolución necesaria del paradigma basado en riesgos del RGPD, o representa una sobre-tecnificación que podría comprometer la flexibilidad y adaptabilidad que caracterizan al enfoque europeo? Para responderla, el artículo adopta una metodología analítica-comparativa basada en:

[1] Asamblea Constituyente de la República del Ecuador, Constitución de la República del Ecuador, Registro Oficial 449 (Quito: 20 de octubre de 2008), art. 66, num. 19.

[2] Superintendencia de Protección de Datos Personales del Ecuador, *Guía de Gestión de Riesgos y Evaluación de Impacto del Tratamiento de Datos Personales – Versión 2* (Quito: SPDP, 2026), p. 12.

- i) Análisis documental de fuentes normativas primarias —textos legales, resoluciones y guías técnicas—
- ii) Revisión sistemática de literatura académica y regulatoria especializada; y
- iii) Estudio comparado de herramientas concretas de supervisión (GESTIONA-AEPD, metodología PIA-CNIL).

Esta triangulación metodológica permite contrastar principios abstractos con implementaciones operativas, superando los límites del análisis puramente dogmático.

2. De la Burocracia Legal a la Ingeniería Regulatoria

Para comprender el modelo ecuatoriano, es necesario desmitificar la idea de que representa una invención normativa absoluta. Más que situarse en una vanguardia inédita, la regulación ecuatoriana se distingue por adoptar un modelo híbrido que fusiona el garantismo jurídico tradicional con la ingeniería de datos aplicada. Esta estrategia operacionaliza los paradigmas existentes de manera prescriptiva y técnicamente específica para suplir carencias institucionales y de cultura de cumplimiento, marcando una diferencia pragmática respecto a jurisdicciones con mayor madurez.

2.1. Evolución de los Modelos de Regulación

La teoría regulatoria contemporánea identifica tres modelos fundamentales que han estructurado históricamente la protección de datos personales, cada uno respondiendo a contextos tecnológicos, políticos y epistemológicos específicos.

2.1.1. Modelo Basado en Derechos (Rights-Based Model)

El enfoque histórico fundacional, anclado en la dignidad humana y los derechos fundamentales, concibe la protección de datos como un derecho autónomo derivado del derecho a la intimidad pero con sustantividad propia. Este modelo encuentra su expresión paradigmática en la Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales.[3]

El considerando 1 del RGPD establece que "[l]a protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental."[4] De manera análoga, el artículo 66, numeral 19, de la Constitución ecuatoriana reconoce y garantiza este derecho fundamental.[5] La LOPDP, en su Capítulo III (arts. 12-23), despliega un amplio catálogo de derechos que incluyen información, acceso, rectificación, actualización, eliminación, oposición, portabilidad, y el derecho a no ser objeto de decisiones basadas únicamente en valoraciones automatizadas.[6]

Sin embargo, el modelo basado en derechos no escala adecuadamente con el riesgo tecnológico y depende excesivamente de la capacidad de ejercicio del titular, quien frecuentemente carece de los conocimientos técnicos, recursos o información necesaria para activar efectivamente sus facultades frente a organizaciones con asimetrías de poder significativas.

2.1.2. Modelo Basado en Riesgos (Risk-Based Model y Accountability)

El segundo estadio evolutivo representa un cambio de paradigma: del enfoque reactivo centrado en el titular hacia un modelo proactivo centrado en el responsable del tratamiento. El artículo 5.2 del RGPD establece que "[e]l responsable del tratamiento será responsable del cumplimiento

[3] Parlamento Europeo y Consejo de la Unión Europea, "Directiva 95/46/CE relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos", *Diario Oficial de las Comunidades Europeas* L 281 (1995).

[4] Parlamento Europeo y Consejo de la Unión Europea, "Reglamento (UE) 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD)", *Diario Oficial de la Unión Europea* L 119 (2016), considerando 1.

[5] Asamblea *Constituyente, Constitución de la República del Ecuador*, art. 66, num. 19.

[6] Asamblea Nacional del Ecuador, *Ley Orgánica de Protección de Datos Personales*, Registro Oficial Suplemento 459 (Quito: 26 de mayo de 2021), arts. 12-23.

de lo dispuesto en el apartado 1 y capaz de demostrarlo (responsabilidad proactiva)."[7] Esta responsabilidad proactiva se operacionaliza mediante la gestión de riesgos: el responsable debe evaluar continuamente los riesgos y adoptar medidas técnicas y organizativas apropiadas.[8]

Un ejemplo paradigmático de la maduración de este enfoque es el Marco de Ciberseguridad del NIST (NIST CSF 2.0), actualizado en febrero de 2024, que elevó la «Gobernanza» (GOVERN) como función central y transversal.[9] La función GOVERN abarca seis categorías: Contexto Organizacional (GV.OC), Estrategia de Gestión de Riesgos (GV.RM), Roles, Responsabilidades y Autoridades (GV.RR), Política (GV.PO), Supervisión (GV.OV), y Gestión de Riesgos de la Cadena de Suministro (GV.SC).[10]

En Europa, la AEPD desarrolló el sistema GESTIONA[11], herramienta digital que guía a las organizaciones en la realización de análisis de riesgos mediante metodologías estructuradas pero flexibles, manteniendo un significativo grado de discrecionalidad metodológica que permite a las organizaciones adaptar sus sistemas a sus circunstancias específicas.

2.1.3. Modelo Basado en Estándares (Standards-Based Model)

Como respuesta directa a la ambigüedad inherente de los principios legales abstractos, surge un tercer estadio evolutivo: el modelo basado en estándares técnicos internacionales. Este enfoque recurre a normas desarrolladas por organismos como ISO/IEC, NIST o CIS para establecer métricas objetivas, procedimientos de ingeniería auditables y niveles de madurez verificables.

Por ejemplo, el mandato legal de «garantizar la confidencialidad de los datos» se traduce en controles concretos: Control ISO 27001:2022 A.8.24 (uso de criptografía), Control NIST SP 800-53 Rev. 5 SC-8 (protección de integridad en tránsito), y CIS Control 3 (protección de datos mediante cifrado en reposo y en tránsito). Estos controles son medibles, auditables y permiten la certificación por terceros independientes.

Sus ventajas —objetividad, interoperabilidad, mejora continua y transferencia de conocimiento— coexisten con limitaciones igualmente relevantes: rigidez ante innovaciones disruptivas, costos prohibitivos para organizaciones pequeñas, riesgo de tecnocracia que privilegie el cumplimiento formal sobre la protección efectiva, y dependencia regulatoria de estándares anglosajones. Ecuador, consciente de estas tensiones, construye un modelo híbrido que preserva el marco garantista de la LOPDP mientras instrumentaliza el cumplimiento mediante estándares técnicos.

2.2. Tipologías Regulatorias

Para caracterizar con precisión la estrategia ecuatoriana, es fundamental distinguir las tipologías de intervención estatal que la componen: heterorregulación, autorregulación y metarregulación.

2.2.1. Heterorregulación

La heterorregulación representa la intervención estatal directa mediante normas de comando y control. En Ecuador, el fundamento heterorregulatorio parte del reconocimiento constitucional del artículo 66, numeral 19. Ejemplos paradigmáticos en la LOPDP incluyen: la prohibición de tratar datos sensibles sin consentimiento explícito (art. 26),[12] la prohibición de transferencias internacionales a países sin nivel adecuado de protección (art. 35),[13] y el régimen sancionatorio con multas de hasta el 5% de los ingresos anuales (art. 56).[14]

---

[7] Parlamento Europeo y Consejo de la Unión Europea, "Reglamento (UE) 2016/679", *Diario Oficial de la Unión Europea* L 119 (2016), art. 5.2.

[8] Parlamento Europeo, "RGPD", art. 32.1.

[9] National Institute of Standards and Technology, *Framework for Improving Critical Infrastructure Cybersecurity, Version 2.0* (Gaithersburg, MD: NIST, 2024), https://doi.org/10.6028/NIST.CSWP.29

[10] Ibid.

[11] Agencia Española de Protección de Datos, "La AEPD presenta su herramienta Gestiona como ayuda para realizar análisis de riesgos y evaluaciones de impacto en la protección de datos", nota de prensa, AEPD, accedido el 29 de diciembre de 2025, https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/la-aepd-presenta-su-herramienta-gestiona-como-ayuda-para

[12] Asamblea Nacional, LOPDP, art. 26.

[13] Asamblea Nacional, LOPDP, art. 35.

[14] Asamblea Nacional, LOPDP, art. 56.

2.2.2. Autorregulación

El RGPD europeo fomenta explícitamente la autorregulación a través de códigos de conducta sectoriales (art. 40), supervisión del cumplimiento por organismos acreditados (art. 41), y mecanismos de certificación (arts. 42-43).[15] Ecuador, en contraste, integra la autorregulación como un mecanismo subordinado de cumplimiento técnico. Los artículos 52-53 de la LOPDP y los artículos 68-70 del Reglamento General[16] permiten códigos de conducta sectoriales, pero los sujetan explícitamente a verificación y aprobación por la SPDP, lo que evita el riesgo de captura regulatoria o disminución encubierta de estándares.[17]

2.2.3. Metarregulación

La innovación distintiva de Ecuador radica en la implementación sistemática de la metarregulación: el Estado no dicta cada micro-conducta operativa, sino que supervisa los sistemas de gestión interna que las organizaciones diseñan para cumplir con los fines públicos establecidos en la ley. Mientras Europa mantiene una metarregulación abierta y metodológicamente flexible, Ecuador apuesta por una metarregulación técnicamente prescriptiva que exige metodologías cuantitativas específicas (FAIR, Monte Carlo) para tratamientos de alto riesgo, jerarquía metodológica obligatoria, documentación de *rationales* con evidencia trazable, y la integración de marcos internacionales (ISO 27001/27701, NIST SP 800-53).

Esta estrategia responde a un diagnóstico pragmático: en contextos institucionales débiles, la flexibilidad puede devenir en arbitrariedad o simulación. Al prescribir métodos técnicos, Ecuador reduce la incertidumbre epistémica del cumplimiento, facilitando tanto la implementación por parte de los responsables como la verificación por la autoridad. El riesgo —sobre-prescripción generadora de rigidez y costos desproporcionados para organizaciones pequeñas— es reconocido expresamente por la propia guía de la SPDP.

2.3. Gobernanza Dual: Responsable, CISO y DPO

El paradigma híbrido de metarregulación requiere una estructura de roles claramente diferenciada. El Responsable del Tratamiento —conforme al artículo 24 del RGPD[18] y los artículos 4, 10 y 40 de la LOPDP—[19] ostenta la responsabilidad última de garantizar el cumplimiento y asignar los recursos necesarios para ello.

El Oficial de Seguridad de la Información (CISO) actúa como encargado técnico: diseña, implementa y supervisa los controles de seguridad conforme a los estándares prescritos (ISO 27001, NIST CSF),[20] sin asumir por ello la responsabilidad jurídica de garantía de derechos fundamentales, que permanece en la esfera del Responsable.

El Delegado de Protección de Datos (DPO), regulado en los artículos 37-39 del RGPD[21] y los artículos 47-51 de la LOPDP,[22] opera con independencia funcional garantizada: el artículo 38.3 del RGPD establece que no puede recibir instrucciones sobre el desempeño de sus funciones.[23] Su función es de segunda línea de defensa: valida que los controles implementados por el CISO y las decisiones del Responsable se alineen con la normativa de protección de datos.

Esta estructura tripartita (Responsable-CISO-DPO) es crítica en el modelo ecuatoriano: el Responsable decide e invierte, el CISO ejecuta técnicamente con base en estándares prescritos,

[15] Parlamento Europeo, "RGPD", arts. 40-43.

[16] Presidencia de la República del Ecuador, *Reglamento General a la Ley Orgánica de Protección de Datos Personales*, Decreto Ejecutivo 904, Registro Oficial Tercer Suplemento 435 (Quito: 13 de noviembre de 2023), arts. 68-70.

[17] Asamblea Nacional, LOPDP, art. 52.

[18] Parlamento Europeo, "RGPD", art. 24.

[19] Asamblea Nacional, LOPDP, arts. 4, 10 y 40.

[20] International Organization for Standardization, *ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements* (Ginebra: ISO, 2022), Anexo A.

[21] Parlamento Europeo, "RGPD", arts. 37-39.

[22] Asamblea Nacional, LOPDP, arts. 47-51.

[23] Parlamento Europeo, "RGPD", art. 38.3.

y el DPO supervisa el cumplimiento normativo documentando evidencia para demostrar *accountability*. La confusión de roles entre el CISO y el DPO —uno de los errores más frecuentes en la implementación práctica— genera dilución de responsabilidades que la SPDP considera sancionable.

3. El Núcleo Operativo de la Metarregulación

La Resolución No. SPDP-SPD-2025-0003-R, que aprueba la Guía de Gestión de Riesgos y Evaluación de Impacto del Tratamiento de Datos Personales (cuya actualización a la versión 2 en 2026 reafirma esta postura prescriptiva), constituye el instrumento operativo central del modelo ecuatoriano. Esta guía materializa prescriptivamente el artículo 40 de la LOPDP —que obliga a implementar medidas de seguridad adaptadas al riesgo— y el artículo 42, que regula la Evaluación de Impacto en la Protección de Datos (EIPD). El análisis comparativo frente al sistema GESTIONA de la AEPD y las metodologías de la CNIL revela convergencias estructurales y divergencias metodológicas que ilustran la apuesta ecuatoriana por la cuantificación como vehículo de certeza jurídica.

3.1. Taxonomía de la Incertidumbre

Una de las contribuciones conceptuales más sofisticadas de la guía ecuatoriana es la distinción explícita entre dos tipos fundamentales de incertidumbre que deben gestionarse diferenciadamente.[24]

La incertidumbre aleatoria (*aleatoric uncertainty*) se refiere a la variabilidad intrínseca e irreducible de los eventos futuros, originada en la naturaleza estocástica de ciertos fenómenos: desastres naturales, fallas aleatorias de hardware, comportamiento humano impredecible, ataques oportunistas no dirigidos. Su gestión se centra en el modelado probabilístico basado en datos históricos, la cuantificación de impactos y la transferencia de riesgos mediante seguros cibernéticos.

La incertidumbre epistemológica, en contraste, es *reducible*: deriva de la falta de conocimiento sobre el propio sistema —desconocimiento del inventario de datos, ignorancia sobre vulnerabilidades, ausencia de inteligencia de amenazas, incomprensión de flujos de datos—. La guía ecuatoriana interpreta el artículo 40 de la LOPDP[25] como una obligación de reducir activamente esta incertidumbre mediante auditorías técnicas, mapeo de procesos y análisis forense, incluso la versión 2 de la guía ecuatoriana es tajante al respecto, estableciendo como objetivo primordial reducir esta incertidumbre epistemológica frente a escenarios de amenaza. Ignorarla se considera negligencia sancionable. Esta distinción no aparece explícitamente en las guías europeas (GESTIONA-AEPD, metodología PIA-CNIL), aunque ambas operan implícitamente bajo el mismo presupuesto.

3.2. Metodologías de Análisis

La guía ecuatoriana impone una jerarquía metodológica que vincula el nivel de riesgo del tratamiento con el tipo de análisis requerido, contrastando marcadamente con la flexibilidad europea.

3.2.1. El Problema del «Ruido» en Evaluaciones Cualitativas

Las evaluaciones cualitativas basadas en juicios subjetivos presentan limitaciones severas que la guía de la SPDP reconoce explícitamente: ambigüedad semántica (¿qué significa «probabilidad alta»?), compresión de rangos en matrices 3×3, sesgo de anclaje hacia eventos recientes y manipulabilidad de las clasificaciones. Para combatir este ruido decisional, la guía exige la documentación obligatoria de *rationales*: justificaciones explícitas y basadas en evidencia —fuentes de información, supuestos explicitados, metodología de cálculo y registro de revisiones— y métricas significativas para cada decisión o valor asignado.

3.2.2. Métodos Cualitativos y sus Condiciones de Uso

La guía permite métodos cualitativos para tratamientos de riesgo bajo o medio —matrices de riesgo clásicas, análisis de escenarios y método Delphi—, bajo condiciones estrictas: el tratamiento no involucra datos sensibles a gran escala, no utiliza tecnologías emergentes, el impacto potencial sobre derechos fundamentales es limitado, y existe documentación exhaustiva de los *rationales* para cada estimación.

[24] SPDP, *Guía de Gestión de Riesgos*, p. 18.

[25] Asamblea Nacional, LOPDP, art. 40.

3.2.3. Métodos Cuantitativos para Alto Riesgo

Para tratamientos que entrañan riesgo alto —conforme al artículo 42 de la LOPDP—,[26] la guía ecuatoriana impulsa y en la práctica exige el uso de modelos matemáticos avanzados: el estándar FAIR (*Factor Analysis of Information Risk*) y las simulaciones de Monte Carlo.

FAIR descompone el riesgo en factores primitivos cuantificables: Riesgo = Probabilidad de Evento de Pérdida × Magnitud de Pérdida. Permite expresar el riesgo en términos financieros anualizados (por ejemplo, «Pérdida Anual Esperada de USD 500.000 con 90% de confianza»), adaptándolo a protección de datos como número de titulares afectados o gravedad de vulneración de derechos.[27] Monte Carlo, por su parte, simula miles de escenarios posibles ante la incertidumbre en las variables, generando distribuciones de probabilidad del resultado. La guía introduce además el Pd-VaR (*Personal Data Value at Risk*), análogo al VaR financiero, que expresa la máxima pérdida probable en términos de derechos afectados con un nivel de confianza dado.

La versión 2 (2026) de la guía profundiza esta exigencia, rechazando explícitamente las lógicas de cumplimiento normativo "en el papel" basadas en listas de verificación documentales (checklists), exigiendo en su lugar la construcción de modelos de riesgo alimentados por datos confiables y dictámenes de expertos calibrados[28]. Jurídicamente, esto significa que la justificación de una medida de seguridad deja de ser un argumento retórico-legal para convertirse en una demostración analítica obligatoria.

3.2.4. Comparación con el Marco Europeo: Flexibilidad vs. Prescripción

El contraste con el enfoque europeo es marcado. El sistema GESTIONA de la AEPD[29] guía al responsable a través de un análisis de riesgos fundamentalmente cualitativo: identificación de activos y amenazas, evaluación en escalas bajo/medio/alto, y mapeo automático a controles del Esquema Nacional de Seguridad o ISO 27001. No calcula pérdidas esperadas ni emplea simulaciones estocásticas. Su fortaleza es la accesibilidad para cualquier pequeña empresa; su debilidad, la limitada precisión para tratamientos complejos.

La metodología PIA de la CNIL[30] también es predominantemente cualitativa: proporciona escalas (negligible / limitada / importante / máxima) sin exigir cuantificación matemática. La apuesta ecuatoriana por la cuantificación obligatoria para alto riesgo representa un *trade-off* claro: Ecuador sacrifica accesibilidad y flexibilidad a cambio de objetividad y verificabilidad.

3.3. El Proceso de Gestión de Riesgos

La gestión de riesgos requiere un objeto de análisis definido. Por ello, la normativa ecuatoriana estructura el proceso partiendo de una base documental obligatoria: el Registro de Actividades de Tratamiento (RAT).

3.3.1. Registro de Actividades de Tratamiento (RAT)

El RAT —exigido por el artículo 38 del Reglamento General de la LOPDP, concordante con el artículo 30 del RGPD—[31] es el inventario vivo y dinámico de los flujos de datos de la organización. Sin un RAT actualizado que detalle qué datos personales se tratan, para qué finalidades, quién los trata, dónde se almacenan, cómo se protegen, cuándo se eliminan y hacia dónde fluyen, resulta imposible identificar riesgos con precisión.

Ecuador eleva el RAT de herramienta útil a requisito de validez técnica, cerrando la posibilidad de presentar evaluaciones de riesgos «de escritorio» desconectadas de la realidad operativa. La guía

[26] Asamblea Nacional, LOPDP, art. 42; SPDP, *Guía de Gestión de Riesgos*, p. 35.

[27] The FAIR Institute, "FAIR Standard for Privacy Risk", White Paper (2020), p. 5. Véase también Luis Enríquez, "FAIR Model Privacy Uncertainty Quantification GDPR", *The FAIR Institute* (blog), 3 de diciembre de 2024, accedido el 14 de enero de 2026, https://www.fairinstitute.org/blog/fair-model-privacy-uncertainty-quantification-gdpr

[28] SPDP, Guía de *Gestión de Riesgos*, p. 2.

[29] Agencia Española de Protección de Datos, "GESTIONA: Manual de Usuario" (2019). Herramienta disponible en: https://www.aepd.es/guias-y-herramientas/herramientas/facilita-rgpd. Accedido el 15 de diciembre de 2024.

[30] Commission Nationale de l'Informatique et des Libertés, *Privacy Impact Assessment (PIA): Methodology* (París: CNIL, 2018). Accedido el 29 de diciembre de 2025. https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf

[31] Presidencia de la República, *Reglamento General LOPDP*, art. 38.

establece explícitamente que cualquier evaluación realizada sin RAT previo y actualizado carece de validez técnica y no satisface las obligaciones de *accountability*. Esta posición es más estricta que la europea: ni la AEPD ni la CNIL invalidan automáticamente evaluaciones realizadas sin RAT formal, aunque ambas reconocen su utilidad instrumental.

3.3.2. El Ciclo de Gestión de Riesgos: Cinco Etapas

Sobre la base del RAT, la guía de la SPDP estandariza el flujo de trabajo en cinco fases secuenciales inspiradas en ISO 31000:2018 y NIST SP 800-39:

- Etapa 1 — Establecimiento del Contexto: Define los parámetros del análisis, incluyendo el contexto externo (marco legal, obligaciones contractuales, panorama de amenazas), el contexto interno (estructura organizacional, cultura de cumplimiento, recursos disponibles), el alcance (global, por proceso, por proyecto o por incidente) y los criterios de aceptación de riesgo (apetito de riesgo). Es responsabilidad de la alta dirección —no del CISO ni del DPO— definir formalmente qué nivel de riesgo está dispuesta a tolerar.

- Etapa 2 — Identificación de Riesgos: Construye un inventario exhaustivo de amenazas (humanas intencionales, humanas no intencionales, tecnológicas y ambientales) y vulnerabilidades, combinando talleres multidisciplinarios, listas de verificación basadas en estándares, revisión de incidentes históricos y análisis de brecha.

- Etapa 3 — Análisis de Riesgos: Estima la probabilidad de ocurrencia y la magnitud del impacto para cada escenario. Para métodos cuantitativos, el riesgo se calcula como Pérdida Anual Esperada (LAE) = Frecuencia Anual × Magnitud de Pérdida; con Monte Carlo se obtiene una distribución de probabilidad del resultado.

- Etapa 4 — Evaluación de Riesgos: Compara el nivel calculado contra los criterios de aceptación definidos en la Etapa 1, clasificando cada riesgo como aceptable, tolerable o inaceptable. Si el riesgo residual es alto para los derechos y libertades de los titulares, se activa la obligatoriedad de EIPD conforme al artículo 42 de la LOPDP.

- Etapa 5 — Tratamiento de Riesgos: Selección e implementación de controles mediante las estrategias de mitigación (controles técnicos, organizativos y físicos), transferencia (seguros, contratos con encargados), evitación (no iniciar el tratamiento) o aceptación consciente (con aprobación formal de la alta dirección y documentación del rationale). El ciclo PDCA integrado garantiza la revisión periódica y actualización del RAT.

3.4. La Evaluación de Impacto en la Protección de Datos (EIPD)

La EIPD —regulada en el artículo 42 de la LOPDP, análogo al artículo 35 del RGPD—[32] representa una iteración más profunda de la gestión de riesgos, enfocada exclusivamente en los riesgos para los derechos y libertades de las personas físicas.

3.4.1. Disparadores de Obligatoriedad

La EIPD es obligatoria cuando el tratamiento presenta al menos dos de los criterios de alto riesgo establecidos en las Directrices del Grupo de Trabajo del Artículo 29 (WP29), adoptadas por el EDPB:[33] (1) evaluación o puntuación/*scoring*; (2) decisiones automatizadas con efectos jurídicos significativos; (3) observación sistemática a gran escala; (4) datos sensibles a gran escala; (5) datos tratados a gran escala; (6) cruce de datos de diversas fuentes; (7) datos de personas vulnerables; (8) uso innovador o aplicación de nuevas tecnologías (IA, biometría, IoT); (9) transferencias internacionales fuera de países con decisión de adecuación; y (10) tratamientos que impidan ejercer derechos o acceder a servicios esenciales.

3.4.2. Contenido Mínimo y Consulta Previa

Una EIPD completa requiere: (1) descripción sistemática del tratamiento basada en el RAT; (2) evaluación de necesidad y proporcionalidad, incluyendo análisis de legitimación, minimización y

[32] Asamblea Nacional, LOPDP, art. 42.

[33] Grupo de Trabajo del Artículo 29, "Directrices sobre la evaluación de impacto relativa a la protección de datos (EIPD) y para determinar si el tratamiento 'entraña probablemente un alto riesgo' a efectos del Reglamento (UE) 2016/679", WP248 rev.01, adoptadas el 4 de octubre de 2017, accedido el 29 de diciembre de 2025, https://www.aepd.es/documento/wp248rev01-es.pdf

compatibilidad de finalidades; (3) gestión de riesgos detallada con documentación completa de rationales; (4) especificación de medidas de seguridad y garantías; (5) consulta formal al DPO y, cuando sea apropiado, a partes interesadas; y (6) conclusión y plan de acción con responsables, plazos y métricas de seguimiento.

Cuando, tras realizar la EIPD, el riesgo residual sigue siendo alto y el responsable no puede mitigarlo suficientemente, el artículo 43 de la LOPDP (equivalente al art. 36 del RGPD) exige consultar a la SPDP *antes de iniciar el tratamiento*. Este mecanismo de control preventivo (*ex ante*) permite a la autoridad revisar la EIPD, solicitar información adicional, recomendar medidas adicionales o, en casos extremos, prohibir el tratamiento. La EIPD no es un documento estático; debe revisarse y actualizarse ante cambios en las circunstancias del tratamiento, nuevas amenazas, brechas de seguridad, o periódicamente —al menos cada tres años para tratamientos de alto riesgo—.

4. Análisis Comparativo Estructural: Ecuador vs. RGPD

El análisis comparado entre el RGPD y la LOPDP revela una divergencia fundamental en la filosofía de implementación. Aunque ambos cuerpos normativos comparten un ADN principialista común y buscan la tutela del mismo derecho fundamental, las estrategias para operacionalizar este mandato difieren sustancialmente.

4.1. Divergencia en los Principios Rectores: Granularidad vs. Condensación

Es un error común asumir una equivalencia exacta entre los principios del RGPD (art. 5) y los de la LOPDP (art. 10). El RGPD condensa sus mandatos en seis principios materiales más el de accountability. La LOPDP despliega trece principios rectores explícitos, decisión que no es meramente retórica, sino que tiene consecuencias jurídicas operativas:

- Autonomía de la Transparencia y la Lealtad: Al separar la «Transparencia» (art. 10.c) de la «Lealtad» (art. 10.b), Ecuador impide que la mera disponibilidad de avisos de privacidad justifique prácticas desleales.
- Seguridad vs. Confidencialidad: La LOPDP distingue el principio de «Confidencialidad» (art. 10.g) del de «Seguridad» (art. 10.k), clarificando que la obligación de secreto es distinta de la obligación técnica de protección integral (CID).
- Aplicación Favorable al Titular: El literal (l) del artículo 10 introduce explícitamente el principio pro homine digital, que en Europa es construcción jurisprudencial del TJUE, mientras que en Ecuador es mandato de ley positiva.

4.2. Flexibilidad vs. Prescripción: El Costo de la Certeza y sus Límites

La diferencia más profunda radica en cómo cada sistema gestiona la incertidumbre del cumplimiento. El modelo europeo se basa en la neutralidad tecnológica y la madurez institucional: asume que las organizaciones poseen (o pueden contratar) el criterio necesario para determinar qué medidas son «apropiadas». Esta flexibilidad favorece la innovación, pero genera inseguridad jurídica para las PYME.

El modelo ecuatoriano, diagnosticando una falta de madurez en la cultura de cumplimiento regional, opta por la reducción de incertidumbre mediante prescripción. Al exigir metodologías cuantitativas como FAIR y Monte Carlo para riesgos altos, la SPDP elimina la ambigüedad del término «apropiado»: si el análisis matemático demuestra que la probabilidad de pérdida excede el apetito de riesgo, la medida es inapropiada; si lo reduce por debajo del umbral, es conforme.

Este enfoque ofrece objetividad y verificabilidad, pero impone barreras significativas: la implementación de modelos estocásticos y estándares ISO requiere recursos financieros y técnicos —actuarios, ingenieros de datos— fuera del alcance de la mayoría de las empresas locales. Ecuador, paradójicamente, construye un sistema de *compliance* de élite en una economía en desarrollo, apostando a que la exigencia técnica elevará el nivel del mercado, aunque a riesgo de asfixiar a los actores más pequeños que no puedan costear la «ingeniería de la privacidad». La evaluación definitiva de este modelo requerirá evidencia empírica sobre su efectividad en la protección real de derechos versus los costos de cumplimiento generados.

Un límite adicional que merece profunda reflexión crítica es el riesgo de que la sofisticación técnica desplace la centralidad de los derechos fundamentales y desborde la propia capacidad institucional

del Estado. Si una gran corporación tecnológica presenta una Evaluación de Impacto sustentada en un complejo modelo estocástico alimentado por inteligencia artificial que legitima un tratamiento intrusivo, cabe preguntarse si la SPDP cuenta con el músculo auditor, los actuarios y científicos de datos necesarios para refutar y deconstruir dicho modelo. Cuando el cumplimiento se convierte en un intrincado ejercicio matemático, la objetivación cuantitativa de la privacidad puede reproducir, a una escala inaudita, el cumplimiento cosmético y excluyente que buscaba erradicar.

5. Conclusiones

La evolución normativa de la protección de datos en Ecuador, cristalizada en la LOPDP y su desarrollo reglamentario y técnico entre 2021 al 2026, representa un caso de estudio singular en el derecho comparado. Lejos de una simple trasplantación del RGPD europeo, Ecuador ha reconfigurado el modelo de *accountability* para adaptarlo a una realidad institucional distinta, transitando desde un garantismo abstracto hacia una metarregulación.

En primer lugar, el análisis evidencia que Ecuador ha abandonado la neutralidad metodológica europea. La imposición de jerarquías analíticas que obligan al uso de métodos cuantitativos (FAIR, Monte Carlo) y rechazan las listas de verificación documentales para tratamientos de alto riesgo busca cerrar la brecha de implementación mediante su objetivación. Se asume que, ante la falta de una cultura de cumplimiento orgánica, la matemática actuarial ofrece un refugio de certeza jurídica superior a la discrecionalidad cualitativa.

En segundo lugar, la estructura de gobernanza de la LOPDP refuerza el control estatal sobre la autorregulación. Al convertir el RAT en un requisito de validez técnica y al subordinar los códigos de conducta a verificación estricta, el regulador ecuatoriano transforma el modelo de «cumplir y demostrar» en «documentar, cuantificar y demostrar».

En tercer lugar, la comparación estructural con el RGPD revela que la tecnificación normativa ecuatoriana es una espada de doble filo. Ofrece métricas claras y reduce la ambigüedad interpretativa, lo que constituye un avance deseable para la seguridad jurídica. Sin embargo, eleva sustancialmente los costos de transacción y las barreras de entrada al cumplimiento, con el riesgo adicional de que la sofisticación técnica desplace la efectividad sustantiva de la protección de derechos.

El éxito de este modelo dependerá de si la sofisticación técnica exigida logra ser internalizada por el tejido empresarial —o si, por el contrario, deriva en un cumplimiento formalista de élite que deja desprotegida a la gran masa de titulares cuyos datos son tratados por actores incapaces de costear la ingeniería de privacidad prescrita—, y de si la SPDP mantiene una supervisión sustantiva orientada a la efectividad real de los derechos, más allá de la verificación formal de los modelos matemáticos. Ecuador ha apostado por la ciencia de datos como garante del derecho; el tiempo dirá si la fórmula es sostenible.

Bibliografía

Agencia Española de Protección de Datos. "Directrices sobre la evaluación de impacto en la protección de datos (wp248 rev.01)." Madrid: AEPD. Accedido el 29 de diciembre de 2025. https://www.aepd.es/documento/wp248rev01-es.pdf.

Agencia Española de Protección de Datos. "Facilita RGPD." Herramienta digital. Madrid: AEPD. Accedido el 15 de diciembre de 2025. https://www.aepd.es/guias-y-herramientas/herramientas/facilita-rgpd.

Agencia Española de Protección de Datos. "La AEPD presenta su herramienta Gestiona como ayuda para realizar análisis de riesgos y evaluaciones de impacto en la protección de datos." Nota de prensa. Madrid: AEPD. Accedido el 29 de diciembre de 2025. https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/la-aepd-presenta-su-herramienta-gestiona-como-ayuda-para

Asamblea Constituyente de la República del Ecuador. *Constitución de la República del Ecuador*. Registro Oficial 449. Quito: 20 de octubre de 2008. Accedido el 15 de diciembre de 2025 https://www.lexis.com.ec/biblioteca/constitucion-republica-ecuador.

Asamblea Nacional del Ecuador. *Ley Orgánica de Protección de Datos Personales*. Registro Oficial Suplemento 459. Quito: 26 de mayo de 2021. Accedido el 15 de diciembre de 2025 https://www.lexis.com.ec/biblioteca/ley-organica-proteccion-datos-personales.

Commission Nationale de l'Informatique et des Libertés (CNIL). "PIA, methodology." París: CNIL. Accedido el 29 de diciembre de 2025. https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf.

Comité Europeo de Protección de Datos. "Guidelines 3/2019 on processing of personal data through video devices." Version 2.0. Bruselas: EDPB, 2020. Accedido el 29 de diciembre de 2025. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32019-processing-personal-data-through-video_en.

Cox, Louis Anthony Jr. "What's Wrong with Risk Matrices?" *Risk Analysis* 28, no. 2 (2008): 497-512. Accedido el 29 de diciembre de 2025. https://doi.org/10.1111/j.1539-6924.2008.01030.x.

Enríquez, Luis. "FAIR Model Privacy Uncertainty Quantification GDPR." *The FAIR Institute* (blog), 3 de diciembre de 2025. Accedido el 14 de enero de 2026. https://www.fairinstitute.org/blog/fair-model-privacy-uncertainty-quantification-gdpr.

Hoepman, Jaap-Henk. "Privacy Design Strategies." En *ICT Systems Security and Privacy Protection: 29th IFIP TC 11 International Conference*, editado por Nora Cuppens-Boulahia et al., 446-459. Cham: Springer, 2014. Accedido el 16 de enero de 2026. https://doi.org/10.1007/978-3-642-55415-5_38.

International Organization for Standardization. *ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements*. Ginebra: ISO, 2022. Accedido el 10 de enero de 2026.

International Organization for Standardization. *ISO/IEC 27701:2019 Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*. Ginebra: ISO, 2019. Accedido el 10 de enero de 2026.

International Organization for Standardization. *ISO 31000:2018 Risk management — Guidelines*. Ginebra: ISO, 2018. Accedido el 10 de enero de 2026.

National Institute of Standards and Technology. *Framework for Improving Critical Infrastructure Cybersecurity, Version 2.0*. Gaithersburg, MD: NIST, 2024. Accedido el 10 de enero de 2026. https://doi.org/10.6028/NIST.CSWP.29.

National Institute of Standards and Technology. *NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management, Version 1.0*. Gaithersburg, MD: NIST, 2020. Accedido el 10 de enero de 2026. https://doi.org/10.6028/NIST.CSWP.01162020.

National Institute of Standards and Technology. *Zero Trust Architecture*. NIST Special Publication 800-207. Gaithersburg, MD: NIST, 2020. Accedido el 11 de enero de 2026. https://doi.org/10.6028/NIST.SP.800-207.

National Institute of Standards and Technology. *Guide for Conducting Risk Assessments*. NIST Special Publication 800-30 Revision 1. Gaithersburg, MD: NIST, 2012. Accedido el 11 de enero de 2026. https://doi.org/10.6028/NIST.SP.800-30r1.

Parlamento Europeo y Consejo de la Unión Europea. "Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos." *Diario Oficial de la Unión Europea* L 119 (2016). Accedido el 15 de diciembre de 2025. https://www.boe.es/doue/2016/119/L00001-00088.pdf.

Presidencia de la República del Ecuador. *Reglamento General a la Ley Orgánica de Protección de Datos Personales*. Decreto Ejecutivo 904. Registro Oficial Tercer Suplemento 435. Quito: 13 de noviembre de 2023. Accedido el 15 de diciembre de 2025. https://www.gob.ec/sites/default/files/regulations/2025-01/02%20Reglamento%20General%20a%20la%20Ley%20Org%C3%A1nica%20de%20Protecci%C3%B3n%20de%20Datos%20Personales_0.pdf.

Superintendencia de Protección de Datos Personales del Ecuador. *Resolución No. SPDP-SPD-2025-0003-R: Guía de Gestión de Riesgos y Evaluación de Impacto del Tratamiento de Datos Personales*. Quito: SPDP, 2025. Accedido el 15 de diciembre de 2025. https://spdp.gob.ec/wp-content/uploads/2025/05/GUIA-DE-GESTION-DE-RIESGOS-E-IMPACTO-VERSION-1.pdf.

Superintendencia de Protección de Datos Personales del Ecuador. *Guía de Gestión de Riesgos y Evaluación de Impacto del Tratamiento de Datos Personales – Versión 2*. Quito: SPDP, 2026. Accedido el 20 de marzo de 2026. https://spdp.gob.ec/wp-content/uploads/2026/03/ggrispdp2026v2.pdf

Datos de carácter personal e Inteligencia Artificial en la Administración local. Cuestiones relevantes

Personal Data and Artificial Intelligence in Local Government: Relevant Issues

Patricio Monreal Vilanova
Responsable del Centro de Competencia de Privacidad, Govertis Part of Telefónica Tech

Resumen:
El dato es el principal activo de la Inteligencia Artificial (IA). En la Administración pública, hallamos una infinidad de conjuntos de datos estructurados, semiestructurados o desestructurados, los cuales podrán ser el nutriente de la IA. Ahora bien, entre dichos datos, los datos de carácter personal precisan de una especial atención, puesto que las operaciones de tratamiento de los mismos, asociadas a los sistemas de IA, entrañan un riesgo para los derechos y libertades de las personas físicas afectadas. Es por esto que, a través del presente, tratamos de ofrecer a la Administración local, responsable del tratamiento, un primer acercamiento, de carácter general y práctico, sobre cómo atender a una parte de las principales obligaciones y responsabilidades dimanantes de la legislación en protección de datos de carácter personal, en consonancia con el Reglamento Europeo de IA.

Palabras clave:
Protección de datos de carácter personal; Inteligencia Artificial; Administración pública; Actividad de tratamiento; Licitud del tratamiento; Responsable o encargado de tratamiento.

Sumario:
1. Introducción. **2.** Concepto y tipología de datos aplicados o utilizados por los sistemas de IA. **3.** Operaciones y actividades de tratamiento de datos personales en sistemas de IA. **4.** Bases jurídicas de licitud del tratamiento de datos personales en la IA. **5.** Transparencia e información relativa a protección de datos de carácter personal y el uso de sistema de IA. **6.** Responsables y encargados de tratamiento de datos personales en el uso de sistemas de IA **7.** Conclusiones.

Abstract:
The primary asset of Artificial Intelligence (AI) is data. In public administration, we find an endless variety of dataset structured, semi-structured, or unstructured which can serve as the fuel for AI systems. However, among these datasets, personal data require special attention, since the processing operations associated with AI systems may pose risks to the rights and freedoms of the individuals concerned. For this reason, the present document aims to provide local public administrations, acting as data controllers, with an initial, general, and practical overview of how to address some of the main obligations and responsibilities arising from personal data protection legislation, in line with the European AI Regulation.

Keywords:
Protection of Personal Data; Artificial Intelligence; public Administration; Processing activity; Lawfulness of processing; Controller or processor.

Summary:
1. *Introduction.* ***2.*** *Concept and typology of data applied or used by AI systems;* ***3.*** *Operations and activities involving the processing of personal data in AI systems;* ***4.*** *Legal bases for the lawful processing of personal data in AI;* ***5.*** *Transparency and information regarding personal data protection and the use of AI systems;* ***6.*** *Data controllers and data processors in the use of AI systems;* ***7.*** *Conclusions.*

1. Introducción

Desde la denominada Administración pública electrónica, pasando por la digital, estamos en los albores de la acuñada administración inteligente. En el discurrir de toda esta modernización o transformación del sector público[1], nos centraremos en un elemento que ha tenido mayor significación, a fin de lograr, entre otros, la mejora en la personalización[2], eficiencia, efectividad, equidad y transparencia de los servicios públicos[3]: el dato.

El dato o conjunto de datos son, de manera innegable, el activo fundamental de la inteligencia artificial (IA). Así, la IA se desarrolla mediante el uso de algoritmos y datos, si bien, haciendo alusión a la metáfora citada por Ponce Solé[4], en la cocina de la IA, los primeros serían las recetas y los segundos, los ingredientes. En el presente trabajo, nos ceñiremos únicamente a los datos, centrándonos, en primer lugar, en el análisis de la tipología de datos o conjuntos de datos, que pueden nutrir la IA. Después, nos centraremos en los datos de carácter personal obrantes en los sistemas de información de la Administración local, a los efectos de analizar, de manera pragmática, una serie de cuestiones de interés en la salvaguarda del derecho fundamental a la protección de datos personales frente a la IA.

En particular, se analizarán las cuestiones relativas a qué operaciones o actividades de tratamiento se desgajarían al utilizar herramientas basadas en la IA; qué legitimaría el tratamiento de los mismos a través de IA; la transparencia e información en el ámbito de la protección de datos y la IA ; y, finalmente, qué condición ostentan los personas jurídico-públicas y, en su caso, privadas en el tratamiento y cómo articular la relación entre éstas, a fin de desgranar las obligaciones y responsabilidades de cada una.

A nuestro entender, estas cuestiones resultan primordiales a los efectos de acometer un tratamiento de datos de carácter personal en relación con la IA, por parte de la Administración local. A partir de aquí o de manera simultánea, deberán atenderse el resto de las cuestiones, como la evaluación de impacto en protección de datos y, en su caso, en los derechos y libertades fundamentales, la privacidad desde el diseño o por defecto, así como el análisis de riesgos y, en consecuencia, la seguridad a aplicar al tratamiento.

De este modo, el presente trabajo pretende ofrecer una guía u orientación, una primera aproximación, a nivel general y con carácter práctico, sobre cómo atender las citadas obligaciones en protección de datos, en la medida en que el diseño, el desarrollo o el uso de la IA, implique un tratamiento de datos de carácter personal.

No serán objeto de análisis otras aplicaciones de la IA en el ámbito de la Administración local[5] ni tampoco el resto de las obligaciones, de uno y otro marco normativo.

Por último, hay que decir que, al referirnos a «sistemas de IA», de forma general, englobamos

[1] Martínez Martínez, Ricard, "De la limitación de la informática a la transformación digital", *Revista Agenda Pública. Análisis de políticas públicas* (2021) https://agendapublica.es/noticia/13475/limitacion-informatica-transformacion-digital: "La transformación digital comporta necesariamente la analítica de datos no personales, personales y anonimizados y la definición de entornos de tratamiento confiables. Para ello, se ha diseñado un ecosistema normativo que incluye, entre otros, el Reglamento General de Protección de Datos, el Reglamento sobre uso de datos no personales y la Directiva 'Open Data'. A ellas se sumarán la 'Data Governance' y el Reglamento sobre la IA".

[2] Cerrillo i Martínez, Agustí "La personalización de los servicios digitales", en *La administración digital*, 311-342 (Madrid: Dykinson, 2022).

[3] En relación al potencial transformador de la Inteligencia Artificial (IA) en el sector público, en cuanto a transparencia, objetividad, legalidad, eficacia, eficiencia y calidad en la prestación de servicios esenciales para el bienestar de la ciudadanía, léase a Almonacid Lamelas, Víctor, "Impacto transformador de la inteligencia artificial en la Administración pública", *Revista DerechoLocal.es* ed. Lefebvre (2024). Según el autor, la aplicación de la IA en áreas como la automatización de actos administrativos, toma de decisiones y personalización de servicios se presenta como una oportunidad para llevar a cabo avances muy tangibles e inéditos hasta la fecha en la calidad de los servicios públicos que paga la ciudadanía a través de los impuestos.

[4] Esta metáfora se recoge en un vídeo, "En la cocina de la IA, los primeros serían las recetas y los segundos los ingredientes" https://ed.ted.com/on/ktzb2Muu/think, que incluye Ponce Sole, Julio en "Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico". *Revista General de Derecho Administrativo*, nº. 50 (Iustel, 2019): 19-0.

[5] Por ejemplo, aplicación de IA en la clasificación y reciclaje de residuos. A través de sistemas de visión artificial y aprendizaje profundo, las plantas de tratamiento pueden identificar y separar automáticamente materiales reciclables, aumentando la eficiencia del proceso y reduciendo el volumen de desechos destinados a vertederos.

cualquiera de las variantes objeto de regulación[6].

2. Concepto y tipología de datos aplicados o utilizados por los sistemas de IA

La definición del concepto de «datos», así como la naturaleza, categorización o tipología de los mismos no es unívoca. Es por esto que, a los únicos efectos de escoger una, optamos por la definición del legislador europeo, al articular el "*Data Act*" y el "*Data Governance Act*"[7]. En estos reglamentos se define como «datos» cualquier representación digital de actos, hechos o información y cualquier recopilación o compilación de tales actos, hechos o información, incluso en forma de grabación sonora, visual o audiovisual. No obstante, ante la amplitud de la citada definición es preciso acotar. En este sentido, podemos distinguir dos grandes grupos de datos, personales y no personales.

<table>
<tr><th>DATOS PERSONALES</th><th>DATOS NO PERSONALES</th></tr>
<tr><td>«datos personales[8]» : toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona.</td><td>«datos no personales[9]» : aquellos que no sean datos personales.</td></tr>
<tr><td>Carácter identificativo (nombre y apellidos; DNI/Pasaporte/NIE; núm. SS.)</td><td rowspan="3">De personas jurídicas (datos relativos a la denominación o razón social, número de identificación fiscal, dirección, correo electrónico u otro tipo de información relativa a sociedades mercantiles, cooperativas, asociaciones, fundaciones o cualesquiera otras formas jurídicas)</td></tr>
<tr><td>Contacto (dirección postal o electrónica; teléfono fijo o móvil)</td></tr>
<tr><td>Características personales (ej. estado civil; edad; datos de familia; sexo; fecha de nacimiento; nacionalidad; lugar de nacimiento; lengua materna)</td></tr>
</table>

[6] A estos efectos, más que la distinción hecha por el Reglamento europeo de Inteligencia Artificial, es más comprensiva la definición hecha por el artículo 3 del Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial: Sistemas de IA de alto riesgo, *Sistemas de propósito general y modelos fundacionales*.

[7] El Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo, de 30 de mayo de 2022, de Gobernanza de Datos (en inglés, "Data Governance Act") y el Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo, de 13 de diciembre de 2023, sobre normas armonizadas para un acceso justo a los datos y su utilización ("Reglamento de Datos" o "Data Act") son dos regulaciones clave en el marco de la Estrategia Europea de Datos, a través de la cual la Unión Europea trata de establecer y asentar la libre circulación de datos y la creación de un mercado único europeo de datos. Ambas normativas recogen una definición similar del concepto de «dato», artículo 2.1 Reglamento de datos y artículo 2.1 del Reglamento de Gobernanza de Datos.

[8] Artículo 4.1 del RGPD; Artículo 3.50 del RIA

[9] Artículo 3.1 del Reglamento (UE) 2018/1807 relativo a un marco para la libre circulación de datos no personales en la Unión Europea; Artículo 3.51 del RIA; Artículo 2.4 del Reglamento (UE) 2023/2854 Reglamento de Datos; Artículo 2.4 del Reglamento (UE) 2022/868 (Reglamento de Gobernanza de Datos.

DATOS PERSONALES	DATOS NO PERSONALES
Circunstancias sociales (ej. características de alojamiento, vivienda; situación militar; propiedades, posesiones; aficiones y estilo de vida; pertinencia a clubes, asociaciones; licencias, permisos, autorizaciones)	**De personas jurídicas** (datos relativos a la denominación o razón social, número de identificación fiscal, dirección, correo electrónico u otro tipo de información relativa a sociedades mercantiles, cooperativas, asociaciones, fundaciones o cualesquiera otras formas jurídicas)
Académicos y profesionales (ej. formación; titulaciones; historial de estudiante; experiencia profesional; pertenencia a colegios o asociaciones profesionales)	
Detalle de empleo (ej. cuerpo/escala; categoría/grado; puestos de trabajo; datos no económicos de nómina; historial del trabajador)	Anonimizados, agregados o disociados
Información comercial (ej. actividades y negocios; creaciones artísticas, literarias, científicas o técnicas; licencias comerciales; suscripciones a publicaciones/medios de comunicación)	**Estadísticos** (de manera general, podríamos distinguir las siguientes categorías o subcategorías. cualitativos: nominales (ej. nacionalidad, color de piel o religión) ordinales (ej. nota de un examen o resultado de una baremación) y cuantitativos: discretos (ej. salario, el número de veces que una persona acude al médico al año o cuántos viajes hace una persona al mes en transporte público) o continuos (ej. la cantidad de agua que se puede acumular en los depósitos municipales)
Económicos, financieros y de seguros (ej. ingresos, rentas; inversiones, bienes patrimoniales; créditos, préstamos, avales; datos bancarios; planes de pensiones, jubilación; datos económicos de nómina; datos deducciones impositivas/impuestas; seguros; hipotecas; subsidios, beneficios; historial créditos; tarjetas crédito)	**Metadatos** (una descripción estructurada del contenido o de la utilización de los datos que facilita la búsqueda o la utilización de esos datos)
Transacciones (ej. bienes y servicios suministrados por el afectado; bienes y servicios recibidos por el afectado; transacciones financieras; compensaciones/ indemnizaciones)	**Sintéticos** (son información fabricada, artificialmente, que imita las características y distribuciones de los datos reales, sin contener información personal o sensible. Estos datos se generan mediante algoritmos y técnicas que preservan la estructura y las propiedades estadísticas de los datos originales. Los datos sintéticos pueden ser de audio, imagen, vídeo o texto)
Categoría especial (origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física)	**De sensores** (ej. ambientales o meteorológicos (lluvia, humedad, gas viento, presión atmosférica, calidad del aire); sobre el tráfico (medición del flujo del tráfico, detección y supervisión de los vehículos en movimiento y parados en las intersecciones señalizadas y recopilación de datos de tráfico en intersecciones o carreteras interurbanas) u otros (ej. gestión de residuos, medición del llenado de contenedores, o pulseras teleasistencia, para la protección de personas mayores con localizador GPS, botón SOS, caídas y llamadas).

DATOS PERSONALES	DATOS NO PERSONALES
Categoría especial (origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física)	**De sensores** (ej. ambientales o meteorológicos (lluvia, humedad, gas viento, presión atmosférica, calidad del aire); sobre el tráfico (medición del flujo del tráfico, detección y supervisión de los vehículos en movimiento y parados en las intersecciones señalizadas y recopilación de datos de tráfico en intersecciones o carreteras interurbanas) u otros (ej. gestión de residuos, medición del llenado de contenedores, o pulseras teleasistencia, para la protección de personas mayores con localizador GPS, botón SOS, caídas y llamadas).
Equiparables al régimen de categoría especial (ej. datos de víctimas de violencia de género o terrorismo; de personas con riesgos de exclusión o vulneración social)	
Operativos sensibles (los datos operativos relacionados con actividades de prevención, detección, investigación o enjuiciamiento de delitos cuya divulgación podría poner en peligro la integridad de las causas penales)	
De identificación de conexión Internet (dirección IP (un número que identifica de manera lógica y jerárquicamente a una interfaz de un dispositivo (habitualmente una computadora) dentro de una red que utilice el protocolo de Internet (*Internet Protocol*); dirección mac (número físico que es asignado a la tarjeta o dispositivo de red (viene impuesta por el fabricante y no varía en toda su vida útil). datos necesarios para determinar la fecha y hora de la comunicación).	
Seudonimizados (datos resultantes de un proceso de «seudonimización»[10], esto es, un tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable. ejemplo, sustituir el nombre y los apellidos por un número, alias, o seudónimo)	

Fuente: Tabla de elaboración propia

En nuestro análisis, dejaremos a un lado los datos o conjuntos de datos no personales, puesto que sólo nos interesa el tratamiento de datos de carácter personal, que, en su caso, se efectúe en las etapas del ciclo de vida de una solución de IA.

Ahora bien, a pesar de que los datos no personales aplicados o utilizados por la IA no tienen, *a priori*, un impacto en la privacidad o protección de datos de carácter personal, el hecho de que estén entremezclados con conjuntos de datos personales y/o, en su caso, no puedan desenmarañarse o escindirse de éstos, deberá atenderse también al régimen de la legislación en protección de datos. En especial, cabe subrayar que, si un dato inferido permite identificar o tomar decisiones sobre una persona, debe tratarse con las mismas garantías que los datos personales obtenidos directamente. Por tanto, ante el riesgo de identificación de personas físicas, al aplicar inferencia, los datos inferidos los incardinaremos en la categoría de personales, pero, realmente, están entre una y otra categoría.

[10] Artículo 4.5 del RGPD.

A los efectos de la IA, "se entiende que los sistemas de IA se desarrollan y utilizan de conformidad con normas en materia de protección de la intimidad y de los datos, al tiempo que tratan datos que cumplen normas estrictas en términos de calidad e integridad"[11].

Pero, esta presunción o suposición del legislador europeo, realmente es un deber de los operadores[12] de la IA, para que, en cualesquiera de las fases del ciclo de vida del sistema de IA en las que intervienen, en su caso, respeten del derecho fundamental a la protección de datos de carácter personal.

Cabe decir que, aparte de las prohibiciones o limitaciones en la aplicación o utilización de datos de carácter personal por los sistemas de IA, podemos encontrar que, en los ficheros o bases de datos de la Administración pública, se contienen determinadas categorías de datos, como datos comerciales confidenciales, las informaciones amparadas por el secreto estadístico y los datos protegidos por derechos de propiedad intelectual de terceros, incluidos los secretos comerciales, cuya aplicación en la IA también resultaría prohibida o limitada[13]. Sin embargo, tales restricciones no son objeto del presente estudio, centrado únicamente en el tratamiento de datos personales por los sistemas de IA de la Administración local.

Dicho esto, los datos personales pueden hallarse en la Administración local de manera estructurada, semiestructurada o desestructurada. Pero, con independencia de la forma en la que estén dispuestos deben aplicarse técnicas para proteger la privacidad de las personas físicas, tales como la anonimización, la privacidad diferencial, la generalización, la supresión y la aleatorización, el uso de datos sintéticos o de métodos similares[14]. Asimismo, el sometimiento a la privacidad desde el diseño o por defecto, las evaluaciones de impacto en protección de datos y otras salvaguardas, puede contribuir a una mayor seguridad en la utilización de los datos personales.

Más, estos conjuntos de datos deben ser de buena calidad, puesto que, en caso contrario, es posible que señalen a personas de manera discriminatoria, incorrecta o injusta[15]. Por tanto, la utilización de datos por los sistemas de IA debe ser de alta calidad[16]. De acuerdo con el Reglamento de Inteligencia Artificial[17], es preciso instaurar prácticas adecuadas de gestión y gobernanza de datos para lograr que los conjuntos de datos para el entrenamiento, la validación y la prueba sean de alta calidad. Los conjuntos de datos para el entrenamiento, la validación y la prueba, incluidas las etiquetas, deben ser pertinentes, lo suficientemente representativos y, en la mayor medida posible, estar libres de errores y ser completos en vista de la finalidad prevista del sistema[18]. Por tanto, componentes clave de la calidad de los datos en IA incluyen, entre otros, la exactitud, la consistencia, la completitud, la oportunidad y la relevancia[19].

[11] Considerando 27 del RIA «gestión de la privacidad y de los datos».

[12] Artículo 3.8 del RIA «operador»: un proveedor, fabricante del producto, responsable del despliegue, representante autorizado, importador o distribuidor.

[13] Considerando 6 del Reglamento (UE) 2022/868 de Gobernanza de datos.

[14] No se analizarán estas técnicas y salvaguardas en el presente estudio, dada limitación en la extensión o alcance del mismo.

[15] Considerando 59 del RIA.

[16] El Reglamento de Inteligencia Artificial no define qué se entiende por alta calidad. A nuestro entender, se trata de que los datos tengan atributos como, por ejemplo, la exactitud, actualización, relevancia, coherencia, accesibilidad, comparabilidad e integridad. También podría considerarse, como premisa de alta calidad, la fuente de la cual emanan los datos que alimenten la IA. Así, consideraríamos alta calidad, por ejemplo, los datos obtenidos de las bibliotecas nacionales y no de una extracción indiscriminada de Internet. Otra apreciación, en línea con la Directiva (UE) 2019/1024 de 20 de junio de 2019 de Datos Abiertos, sería que los datos fueran de alto valor, esto es, artículo 2.2.10) Directiva, «conjuntos de datos de alto valor»: documentos cuya reutilización está asociada a considerables beneficios para la sociedad, el medio ambiente y la economía, en particular debido a su idoneidad para la creación de servicios de valor añadido, aplicaciones y puestos de trabajo nuevos, dignos y de calidad, y del número de beneficiarios potenciales de los servicios de valor añadido y aplicaciones basados en tales conjuntos de datos; (Ej. los datos de observación meteorológica) Si bien, el legislador europeo, en cuanto a la IA, no ha incidido en este aspecto.

[17] Considerando 67 y artículo 10 del RIA.

[18] En el anexo IV del RIA, obliga a que en la documentación técnica del sistema de IA, cuando proceda, se incluyan los requisitos en materia de datos, en forma de fichas técnicas que describan las metodologías y técnicas de entrenamiento, así como los conjuntos de datos de entrenamiento utilizados, e incluyan una descripción general de dichos conjuntos de datos e información acerca de su procedencia, su alcance y sus características principales; la manera en que se obtuvieron y seleccionaron los datos; los procedimientos de etiquetado (p. ej., para el aprendizaje supervisado) y las metodologías de depuración de datos (p. ej., la detección de anomalías).

[19] Alonso Peña, Carlos, "Calidad del dato en sistemas de IA desde su adecuado gobierno y gestión", *El Consultor de los*

En particular, debe prestarse una atención especial a la mitigación de los posibles sesgos en los conjuntos de datos que puedan afectar a la salud y la seguridad de las personas físicas, que puedan tener repercusiones negativas en los derechos fundamentales o que puedan dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión. Además, los conjuntos de datos deben tener en cuenta, en la medida en que lo exija su finalidad prevista, los rasgos, características o elementos particulares del entorno geográfico, contextual, conductual o funcional específico en el que esté previsto que se utilice el sistema de IA[20].

A fin de obtener u operar con datos, deben utilizarse los espacios de datos[21] para los municipios y provincias, e incluso a nivel autonómico también se ha previsto[22]. Así, en los bancos o entornos controlados de pruebas, para el ensayo de sistemas de inteligencia artificial (en inglés, *sandbox*)[23] permite a las entidades locales el tratamiento de datos personales en un entorno aislado, sin poner en riesgo la privacidad.

La Administración local, en condición de proveedora de IA o usuaria participante de estos entornos controlados, debe atender también a la legislación europea y española en protección de datos[24].

Por último, también entendemos pertinente, para la organización pública, adoptar una gobernanza de los datos[25].

3. Operaciones y actividades de tratamiento de datos personales en sistemas de IA

En primer lugar, cabe decir que los sistemas de IA no son una actividad de tratamiento de datos de carácter personal por sí mismos[26]. Los sistemas de IA son medios para llevar a cabo operaciones[27] de una actividad de tratamiento de datos personales, así como, una actividad de tratamiento de datos personales podría ser implementada por diferentes sistemas de IA simultáneamente. De modo que, a los efectos del inventario o registro, los sistemas de IA no se incluyen como una actividad

[19] [cont.] *Ayuntamientos. Revista técnica especializada en Administración local y justicia municipal*, nº extraordinario III, edit. La Ley (2024).

[20] Considerando 67 del RIA.

[21] El Reglamento de Inteligencia Artificial prevé e impulsa la creación de entornos de experimentación y prueba controlados, en particular, para el desarrollo de sistemas de un interés público sustancial (en salud, medio ambiente, sostenibilidad energética, movilidad, calidad del servicio público y seguridad de infraestructuras críticas). A este respecto, interesa señalar, cuanto regula acerca del tratamiento de datos personales en estos entornos (vid. artículo 59 del RIA). Entre otras prescripciones, establece que los datos personales que se traten en el contexto del espacio controlado de pruebas se encuentren en un entorno de tratamiento de datos funcionalmente separado, aislado y protegido, bajo el control del proveedor potencial, y que únicamente las personas autorizadas tengan acceso a dichos datos.

[22] Ejemplo, artículo 7 del Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura.

[23] El objeto de este entorno controlado de pruebas, según se explica en la exposición de motivos del Real Decreto 817/2023, de 8 de noviembre, será doble: por una parte, establecer un entorno controlado de pruebas para ensayar el cumplimiento de ciertos requisitos por parte de algunos sistemas de inteligencia artificial que puedan suponer riesgos para la seguridad, la salud y los derechos fundamentales de las personas; por otra, regular el procedimiento de selección de los sistemas y entidades que participarán en el entorno controlado de pruebas. Más información acerca de estos entornos controlados de pruebas, entre otros, Fernández Hernández, Carlos Diario, "Estructura y contenido del Real Decreto 817/2023, de 8 de noviembre, por el que se establece un entorno controlado de pruebas (*sandbox*), para el ensayo de sistemas de inteligencia artificial", *Diario La Ley*, nº 18 (2023).

[24] Artículo 16 del Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial.

[25] Vid. "Guía 7. Datos y gobernanza del dato". Fecha versión: 10 de diciembre de 2025. Y, a nivel doctrinal, léase Loza Corera, María, "Datos y gobernanza de datos y conexiones con principios de protección de datos en el art. 10 del Reglamento", en *Tratado sobre el Reglamento de Inteligencia Artificial de la UE*, ed. Aranzadi (2024): 473-497.

[26] En el blog de la Agencia Española de Protección de Datos se publica "*Inteligencia Artificial: Sistema vs. tratamiento, medios vs. finalidad*", el 10 de abril de 2023, en el cual señala que "Los sistemas de IA no son tratamientos por sí mismos, son medios para implementar operaciones en una actividad de tratamiento. Un tratamiento de datos personales podría ser implementada por diferentes sistemas de IA simultáneamente, y también, estos sistemas de IA podrían implementarse en local o en la nube, podrían involucrar encargados del tratamiento, etc. Los sistemas IA formarán parte de la naturaleza del tratamiento cuando hayan sido incluidos en algunas de las operaciones necesarias para llevar a cabo el tratamiento con su finalidad explícita".

de tratamiento, sino como medios. Pero, como se expondrá más adelante, los tratamientos de datos personales efectuados en las etapas de creación o desarrollo (entrenamiento, validación, despliegue) de un sistema de IA, sí comportan una actividad de tratamiento, que debe ser inventariada o registrada.

En relación al registro de la actividad de tratamiento, a pesar de que no sea una información obligatoria del mismo, en nuestra opinión, debería especificarse que la actividad de tratamiento de datos personales se lleva a cabo mediante un sistema de IA. Así, dado que, la publicación por medios electrónicos del inventario de actividades de tratamiento es un ejercicio de publicidad activa o transparencia, ésta puede ligarse a la transparencia[28] en cuanto al uso de sistemas de IA. De manera que, se indique el sistema de IA utilizado entre la información que conste en el registro de actividades de tratamiento. Con esto, el interesado podrá saber que en la actividad de tratamiento de datos personales se ha utilizado IA, en una o varias de las operaciones de tratamiento. En todo caso, sugerimos que se trate de una mera indicación, de manera que, una información más completa o adicional acerca del sistema de IA, podría aportarse, por ejemplo, dirigiendo al interesado al registro de la Unión Europea, en caso de que sea de alto riesgo[29].

La legislación europea en IA sólo obliga a que en el registro de actividades de tratamiento se incluyan las razones por las que un tratamiento de categorías especiales de datos personales era estrictamente necesario para detectar y corregir sesgos, y por las que ese objetivo no podía alcanzarse mediante el tratamiento de otros datos[30]. Pero, según se ha indicado, entendemos que es interesante señalar en el registro las actividades de tratamiento, en qué actividades se ha utilizado IA, ya sea en una o varias operaciones de tratamiento asociadas a la actividad.

En segundo lugar, es necesario distinguir los ciclos de vida de un sistema de IA del ciclo de vida del dato[31].

Teniendo esto último claro, por un lado, se aportan una serie de ejemplos a los efectos de discernir qué es en sí la actividad de tratamiento de datos personales y qué operaciones dentro de la misma pueden ser realizadas mediante un sistema de IA.

ACTIVIDAD DE TRATAMIENTO	OPERACIONES DE TRATAMIENTO CON SISTEMAS DE IA
Asistencia básica a la ciudadanía[32]	↘ Chatbots y asistentes virtuales para respuestas inmediatas o consultas frecuentes.

[27] Esto es, cualquier operación o conjunto de operaciones técnicas realizadas sobre datos personales o conjuntos de datos personales, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción.

[28] De acuerdo con las directrices del Grupo independiente de expertos de alto nivel sobre IA "(...) Por «transparencia» se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos".

[29] Las autoridades públicas, instituciones, órganos u organismos de la Unión, o personas que actúen en su nombre, que desplieguen o implementen sistemas de IA de alto riesgo se registrarán, seleccionarán el sistema y registrarán su utilización en la base de datos de la UE (Artículos 49 y 71 del RIA).

[30] Artículo 10.5.f) del RIA.

[31] En la guía de la Agencia Española de Protección de Datos relativa a la "Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción", febrero 2020, señala que "No hay que confundir las etapas del ciclo de vida de un componente IA, de un componente de un tratamiento, con las fases en las que se puede dividir un tratamiento para su análisis. La Guía Práctica para Evaluaciones de Impacto en la Protección de Datos de la AEPD, expone una división de los tratamientos en fases desde el punto de los datos en: captura de datos, clasificación/almacenamiento, cesión de datos a terceros y destrucción de los datos. Esta división es genérica, una primera aproximación para analizar un tratamiento en el momento de la explotación del mismo, y cada tratamiento tiene que adaptar esta división a sus condiciones específicas".

[32] ALos asistentes virtuales es una de las tecnologías más extendidas en la Administración local. Entre los muchos supuestos, citaremos el Ayuntamiento de Alicante ha creado "ALI" un asistente virtual de uso en todo tipo de dispositivos encargado de atender e informar a la ciudadanía; el Ayuntamiento de Barcelona: en colaboración con la Fundación Mobile World Capital, ha creado "ARI" un asistente dedicado a atender a personas mayores de 65 que viven solas. ARI es un robot de aproximadamente 1 metro de altura y 12 kg de peso que se desplaza de forma autónoma por la casa interactuando con

ACTIVIDAD DE TRATAMIENTO	OPERACIONES DE TRATAMIENTO CON SISTEMAS DE IA
Asistencia básica a la ciudadanía[32]	↘ Chatbots y asistentes virtuales para respuestas inmediatas o consultas frecuentes.
Proceso selectivo de personal	↘ Operación de filtrado concurrencia de requisitos (admitidos y excluidos). ↘ Preparar una prueba de aptitud. ↘ Realizar la prueba de aptitud. ↘ Realizar la evaluación final.
Contratación pública[33]	↘ Predicción de corrupción. ↘ Planificación de decisiones de compra. ↘ Análisis de duplicidades e ineficiencias dentro de la propia organización. ↘ Generación automatizada de borradores de pliegos. ↘ Asignación automatizada del CPV. ↘ Control y fiscalización durante la vigencia de los contratos.

Fuente: Tabla de elaboración propia

Por otro lado, en las distintas etapas de creación de un sistema de IA pueden apreciarse tratamientos de datos de carácter personal. Así, por ejemplo, un sistema de IA basado en técnicas de *Machine Learning*[34] podrían utilizar datos personales en las etapas de desarrollo del mismo. Según aprecia la Agencia Española de Protección de Datos, pueden desgranarse una serie de operaciones de tratamiento de datos personales en las etapas de creación del sistema IA[35].

[32] [cont.] la persona usuaria a través de conversaciones sencillas; el Ayuntamiento de Madrid pone a disposición "Línea Madrid Chat Online- Asistente Virtual"; el Ayuntamiento de Valladolid dispone de ANA, un 'chatbot' pionero en su ámbito a nivel nacional, basado 100% en inteligencia artificial generativa, que asiste a la ciudadanía ante preguntas de diversa tipología vinculadas a las competencias municipales.

[33] En relación a la IA y contratación pública, resulta de interés la lectura de los artículos recogidos en la publicación *Contratación Administrativa Práctica: revista de la contratación administrativa y de los contratistas*, nº 182 (2022). Por otro lado, de interés la lectura de Montoro Montarroso, Andrés, "Propuesta de metodología para la implantación de sistemas de Inteligencia Artificial para la prevención del fraude en la contratación Pública" y Sánchez Bolívar, Francisco José "Contratación pública para la digitalización del Sector Público: optimizando la actividad administrativa a través de la inteligencia artificial", ambos artículos incluidos en la obra "*Inteligencia artificial y administraciones públicas: una triple visión en clave comparada*" (2024): 401-409 y 411-420, respectivamente. También se sugiere la lectura de Almonacid Lamelas, Víctor "Irrupción y repercusión de la inteligencia artificial (IA) en las funciones reservadas: una visión muy práctica" *El Consultor de los Ayuntamientos: Las entidades locales ante el reto de la inteligencia artificial* (2024).

[34] El aprendizaje automático (*Machine Learning*) es una rama específica de la IA, aplicada a la resolución de problemas específicos y limitados, como tareas de clasificación o predicción. A diferencia de otros tipos de IA que intentan emular la experiencia humana (por ejemplo, sistemas expertos); el comportamiento de los sistemas de aprendizaje automático no está definido por un conjunto predeterminado de instrucciones. En relación a esta técnica de la IA y la protección de datos, léase "*10 Malentendidos sobre el Machine Learning (Aprendizaje Automático)*" publicado por la Agencia Española de Protección de Datos y el Supervisor Europeo de Protección de Datos.

[35] En la guía de la Agencia Española de Protección de Datos relativa a la "*Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*", febrero 2020 (Págs. 12-14).

Con todo esto, a fin de inventariar las actividades de tratamiento de datos personales por parte de la Administración local, responsable del tratamiento, sugerimos la siguiente distinción:

- Actividad de tratamiento de datos personales asociada a la creación o desarrollo del sistema de IA, de manera que englobe todos los tratamientos realizados en las etapas de entrenamiento y validación. De este modo, los tratamientos de datos personales que se efectúen quedan registrados o inventariados, aunque finalmente no se pusiera en funcionamiento el sistema de IA. Cabe decir también que, en el supuesto de que se efectúe un proceso de eliminación de datos personales o anonimizado[36], esto supone un tratamiento subsumido en esta actividad de tratamiento.

- Actividad de tratamiento de datos personales llevada a cabo por sistema de IA. En este caso, la IA se entiende como medio para llevar a cabo la actividad de tratamiento de datos personales, con unos fines o usos. Las distintas operaciones de tratamiento que se efectúan en la explotación de la IA quedan subsumidas en la concepción de la actividad de tratamiento de datos. Por tanto, los sistemas IA formarán parte de la actividad del tratamiento cuando hayan sido incluidos en algunas de las operaciones necesarias para llevar a cabo el tratamiento con su finalidad explícita.

4. Bases jurídicas de licitud del tratamiento de datos personales en la IA

La Administración local, antes de efectuar un tratamiento de datos personales mediante IA, debe verificar la concurrencia de condición o base jurídica de licitud. Además, la base jurídica puede ser diferente para cada una de las etapas del ciclo de vida del sistema de IA, en la que se produzca el tratamiento de datos personales. Además, una base jurídica no habilita para el uso de los datos para cualquier propósito y en todo momento.

A su vez, téngase en cuenta que estamos considerando la IA como un medio, utilizado en operaciones de una actividad de tratamiento de datos personales. De modo que, en principio, no se trata de hallar fundamento, amparo o habilitación legal para el uso de la IA, sino para el tratamiento de los datos personales. No obstante, al utilizar IA en un tratamiento de datos personales, en una mayoría de los supuestos, deberá realizarse, de manera previa, una evaluación de impacto en protección de datos (EIPD). Y, en consecuencia, debe hacerse un análisis de la necesidad y proporcionalidad del tratamiento de los datos, pudiendo resultar que la IA sea un medio que no supere el juicio de «necesidad», y, por tanto, deba optarse por un medio alternativo equivalente igual de eficaz y menos intrusivo. No podemos ahondar más en esta última cuestión, dada la limitación de la publicación del estudio, ya que puede generar un amplio debate, como el acontecido en el tratamiento de datos biométricos en los controles de presencia.

Dicho esto, la selección de las bases jurídicas está estrechamente relacionada con la finalidad del tratamiento y la finalidad perseguida en cada una de las fases del ciclo de vida del sistema y, en el caso de las Administraciones públicas, con sus competencias atribuidas.

En el ámbito de la Administración pública, el «interés legítimo» como base de licitud del tratamiento de los datos personales (por ejemplo, para el entrenamiento de un sistema de IA), en principio, no puede aplicarse[37]. No obstante, esto no implica que, si un tercero ha desarrollado un sistema de IA, utilizando dicha base jurídica, el tratamiento construido sobre la IA no pueda ser empleado por las Administraciones públicas. Al respecto, la Agencia Española de Protección de Datos señala que no se podrá utilizar dicha base jurídica por parte de las Administraciones públicas para justificar tratamientos ulteriores.

En la mayoría de las ocasiones, el tratamiento de datos de carácter personal estará legitimado, con base a una misión de interés público o el ejercicio de los poderes públicos conferidos. Por tanto,

[36] Al respecto, no debe olvidarse la responsabilidad proactiva en el proceso de disociación o anonimización, puesto que hay que demostrar que estas operaciones han sido realmente efectivos y debe evaluarse cuál es el riesgo de reidentificación.

[37] De acuerdo con el artículo 6.f) *in fine* del RGPD el interés legítimo "no será de aplicación al tratamiento realizado por las autoridades públicas en el ejercicio de sus funciones". En nuestra opinión, la salvedad se aplica para el ejercicio de funciones o potestades jurídico-públicas de la Administración. En el caso de que sean funciones de carácter privado, podría aplicarse. Además, la propuesta de un nuevo reglamento denominado «ómnibus digital» de la Comisión europea, noviembre de 2025, prevé que el interés legítimo del responsable del tratamiento sea base jurídica para el tratamiento de datos personales en el contexto del desarrollo y la implantación de modelos o sistemas de IA. Aunque el legislador europeo, en esta propuesta, no alude a si será de aplicación también para entes públicos. Todo esto amerita un análisis más profundo, el cual no podemos acometer en la presente publicación.

será tarea del legislador que en nuestro ordenamiento jurídico hallemos normas que habiliten a los poderes públicos en el uso de la IA[38].

En particular, el interés público se erige como base jurídica de licitud para el tratamiento ulterior de datos personales en el desarrollo, entreno y testeo, de determinados sistemas de IA, en el contexto de un espacio controlado de pruebas para IA[39]. No obstante, tal y como ha expresado el legislador europeo debe ser un interés público esencial en uno o varios de los siguientes ámbitos: i) la seguridad y la salud públicas, incluidos la detección, el diagnóstico, la prevención, el control y el tratamiento de enfermedades y la mejora de los sistemas sanitarios, ii) un elevado nivel de protección y mejora de la calidad del medio ambiente, la protección de la biodiversidad, la protección contra la contaminación, las medidas de transición ecológica, la mitigación del cambio climático y las medidas de adaptación a este, iii) la sostenibilidad energética, iv) la seguridad y la resiliencia de los sistemas de transporte y la movilidad, las infraestructuras críticas y las redes, v) la eficiencia y la calidad de la Administración pública y de los servicios públicos. A mayor abundamiento, a fin de proteger los derechos de terceros frente a la discriminación que podría provocar el sesgo de los sistemas de IA, los proveedores deben ser capaces de tratar también categorías especiales de datos personales, como cuestión de interés público esencial[40], en el sentido del artículo 9.2.g) del RGPD y del artículo 10.2.g) del Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo, de 23 de octubre de 2018.

En cuanto al consentimiento, decir que se aplica, de manera residual, como condición de licitud en los tratamientos de datos personales en el ámbito de la Administración pública[41]. A pesar de que, en el ámbito de la IA, se admite el consentimiento informado en determinados supuestos, debe distinguirse de la manifestación de voluntad para tratar los datos personales[42]. De manera extrapolada, también es interesante señalar que, la Generalitat Catalana dispone de un registro digital o sistema equivalente[43], que recoge y centraliza los consentimientos de las personas, que desean la prestación de servicios públicos proactivos y la personalización de los mismos. No obstante, el uso de sistemas de la IA por parte de la Administración local no puede regirse por el consentimiento de los interesados, ya que, con el consentimiento se admite siempre su revocación, sin necesidad de condición alguna[44].

No puede estarse a expensas de los vecinos que manifiestan la voluntad de tratar los datos personales por medio de IA, frente a los que no. Es por esto que, la base jurídica del tratamiento, en nuestra opinión, ha de ser el interés público, la cual debe tener presente, en su caso, el derecho de oposición con expresión de la causa, y la consiguiente ponderación.

5. Transparencia e información relativa a protección de datos de carácter personal y el uso de sistema de IA

La Administración pública debe garantizar transparencia, en relación con los tratamientos de datos de carácter personal y en el uso de la inteligencia artificial. Sin embargo, la transparencia tiene un

[38] En este sentido, Valero Torrijos, Julián, "Las singularidades del tratamiento de datos de carácter personal en entornos de inteligencia artificial en el sector púbico", en *Inteligencia artificial y sector público: retos, límites y medios*, 335-396 (Valencia: Tirant Lo Blanch, 2023), 368.

[39] En concreto, en el espacio controlado de pruebas, los datos personales recabados lícitamente con otros fines podrán tratarse únicamente con el objetivo de desarrollar, entrenar y probar determinados sistemas de IA en el espacio controlado de pruebas cuando se cumplan todas las condiciones previstas en el artículo 59.1 del RIA.

[40] Considerando 70 del RIA.

[41] Informe Jurídico AEPD 175/ 2018 "(…) Con carácter general, la base jurídica del tratamiento en las relaciones con la Administración, en aquellos supuestos en que existe una relación en la que no puede razonablemente predicarse que exista una situación de equilibrio entre el responsable del tratamiento (la Administración), y el interesado (el administrado) no sería el consentimiento (art. 6.1.a) RGPD), sino, según os casos, el cumplimiento de una obligación legal (art. 6.1.c) RGPD) o el cumplimiento de una misión de interés público o en el ejercicio de poderes públicos (art. 6.1.e) RGPD)".

[42] Según el artículo 61.1 del RIA "A los efectos de las pruebas en condiciones reales, se obtendrá de los sujetos de las pruebas un consentimiento informado dado libremente antes de participar en dichas pruebas y después de haber recibido información concisa, clara, pertinente y comprensible en relación con (…)". Ahora bien, con atención al Considerando 141 del RIA, decir que "El consentimiento de los sujetos para participar en tales pruebas en virtud del presente Reglamento es distinto del consentimiento de los interesados para el tratamiento de sus datos personales con arreglo al Derecho pertinente en materia de protección de datos y se entiende sin perjuicio de este".

[43] Decreto 76/2020, de 4 de agosto, de Administración digital (artículo 31).

[44] Artículo 7.3 del RGPD.

significado diferente, en uno y otro ámbito, por lo que no debe confundirse. Así, la información a transmitir en cada supuesto es diferente, en cuanto a tipo, cantidad y destinatarios de la misma[45].

Dicho esto, en el caso de que las personas físicas interactúen con un sistema de IA, la Administración pública, responsable del despliegue, debe informar del uso de la IA[46]. Y, dado que, al interactuar con el sistema de IA, seguramente, se recaben datos personales, la Administración pública, en calidad de responsable del tratamiento, debe informar de dicho tratamiento.

Es por esto que, a nuestro entender, una y otra información, en este caso, pueden ir de manera conjunta. Además, en el caso de que el sistema de IA adopte decisiones automatizadas y/o realice elaboración de perfiles, se añadiría a la información anterior.

Ahondando un poco más, la Administración pública, en la toma de decisiones individuales automatizadas, a partir de datos personales obtenidos de una persona física, debe garantizar al interesado el derecho a obtener la intervención humana, expresar su punto de vista e impugnar la decisión[47]. Es por esto que, en los procedimientos administrativos para la adopción de decisiones administrativas automatizadas, en los que se empleen sistemas de IA, debe preverse el momento, modo y alcance de la intervención de personas físicas para garantizar los derechos y libertades.
De acuerdo con Mir Puigpelat[48] "podría exigirse expresamente que dicha información se facilite a los interesados en la motivación de la decisión. Como mínimo, debería indicarse que esta ha sido producida de forma automatizada y ofrecerles la posibilidad, si lo desean, de solicitar el acceso a los detalles del tratamiento automatizado".

En definitiva, a efectos prácticos, ligamos, en estos supuestos, la información obligatoria en protección de datos y la relativa al uso de sistema de IA.

6. Responsables y encargados de tratamiento de datos personales en el uso de sistemas de IA

La Administración local puede ostentar la condición de responsable[49] y/o encargada[50] de tratamientos de datos de carácter personal, con relación a la creación de un sistema de IA y/o explotación o implementación de dicho sistema. En cualquier caso, cuando un sistema de IA trata datos personales, debe prestarse especial atención a la coherencia de estos roles y responsabilidades, ya que ambas normas, RGPD y RIA, no son congruentes[51].

Nos podemos encontrar con una infinidad de supuestos, pero no podemos analizar toda la casuística en el presente trabajo. Así, para una mejor comprensión, señalaremos los supuestos más habituales.

[45] Vid. "Inteligencia artificial: Transparencia", 20 de septiembre de 2023. Agencia Española de Protección de Datos https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia.

[46] Considerando 132 del RIA.

[47] Artículo 22.3 del RGPD

[48] Mir Puigpelat, Oriol. "Capítulo XVI. Algoritmos, inteligencia artificial y procedimiento administrativo: principios comunes en el Derecho de la Unión Europea" en *Inteligencia artificial y sector público: retos, límites y medios*, 685-725 (Valencia: Tirant Lo Blanch, 2023), 699.

[49] Artículo 4.7 del RGPD «responsable del tratamiento» o «responsable»: la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros.

[50] Artículo 4.8 del RGPD «encargado del tratamiento» o «encargado»: la persona física o jurídica, autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento.

[51] "*El Comité Europeo de Protección de Datos y el Supervisor Europeo de Protección de Datos, acogen con satisfacción la participación en la regulación de todas las partes interesadas de la cadena de valor de la IA y la introducción de requisitos específicos para los proveedores de soluciones, ya que desempeñan un papel importante en los productos que utilizan sus sistemas. Sin embargo, las responsabilidades de las distintas partes (usuario, proveedor, importador o distribuidor de un sistema de IA) deben estar claramente definidas y asignadas. En particular, en el tratamiento de datos personales deberá prestarse especial atención a la coherencia de estas funciones y responsabilidades con los conceptos de responsable y encargado del tratamiento contenidos en el marco de protección de datos, ya que ambas normas no son congruentes*". CEPD-SEPD Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 2021.

Por una parte, la Administración local puede ostentar la condición de responsable de tratamiento en los siguientes supuestos:

- En calidad de proveedor[52] único – que desarrolla un sistema de IA o un modelo de IA de uso general y sea utilizado para sí, esto es, para fines o usos propios – es el responsable de tratamiento de los datos personales que se efectúe en las etapas de entrenamiento y validación o prueba del sistema. En caso de ser proveedor o desarrollador junto a otra Administración pública o privada de un sistema de IA, podría considerarse a ambas entidades corresponsables del tratamiento.

  En el caso de que la Administración local contratase a un tercero para el entrenamiento del sistema IA, éste será considerado encargado de tratamiento; tanto si la Administración es quien cede los datos como si los obtiene por sí mismo el tercero contratado. Cabe añadir, que, en tal caso la Administración es la que fija los términos que definen los fines del tratamiento y las características sustanciales de los datos. Y, en el supuesto que el tercero sea quien decida qué datos personales y cómo utilizarlos para entrenar el sistema de IA, asumirá la condición de responsable.

  En el caso de que la Administración local, más adelante, comunique o ceda a otra entidad los datos de salida o resultantes del uso del sistema de IA, será responsable de dicha transmisión[53].

- En calidad de Administración pública que efectúa el tratamiento de datos personales mediante un sistema IA. Así, en el caso de que sea responsable del despliegue[54] o implementación del sistema IA para llevar a cabo actividades de tratamiento, asume las obligaciones y responsabilidades en condición de responsable.

Por otra parte, la Administración local, aunque por el momento sea más improbable en la práctica, puede ostentar también la condición de encargado de tratamiento en los siguientes supuestos:

- En calidad de proveedor del sistema de IA para otras entidades, públicas o privadas, cuando la introduzca en el mercado o ponga en servicio con su propio nombre o marca, previo pago o gratuitamente, la IA que ha desarrollado.

- En calidad de distribuidor[55] del sistema de IA para otras entidades, públicas o privadas.

Además, también deben tenerse presente los supuestos de colaboración público-privada, a efectos de que la Administración local logre herramientas, tecnologías y servicios en torno al uso de la IA, así como la transferencia de conocimiento a la sociedad, de sus aplicaciones y usos. En esta relación colaborativa, el tratamiento de datos personales que, en su caso, subyace en el sistema de IA, podría atribuirse la responsabilidad del tratamiento a la Administración pública en cuanto titular de la competencia de los servicios públicos, en los que se haya utilizado la IA; siendo el ente privado encargado del tratamiento. O, en su caso, también podría articularse un acuerdo de corresponsabilidad[56] entre el ente público y privado. Se opte por una u otra forma de articular la relación, en cuanto concierne a la protección de datos, el contrato o concesión administrativa debería recoger, bien los extremos del encargo de tratamiento o, en su caso, los del acuerdo de corresponsabilidad. Además, ha de añadirse la exigencia de contemplar en los Pliegos de cláusulas administrativas particulares y/o en los Pliegos de prescripciones técnicas, la sujeción de los datos a las medias de seguridad del Esquema Nacional de Seguridad[57].

[52] Artículo 3.3 del RIA «proveedor»: una persona física o jurídica, autoridad pública, órgano u organismo que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema de IA o un modelo de IA de uso general y lo introduzca en el mercado o ponga en servicio el sistema de IA con su propio nombre o marca, previo pago o gratuitamente.

[53] El artículo 4.2 del RGPD define como tratamiento también la "…comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión…"

[54] Artículo 3.4 del RIA «responsable del despliegue»: una persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional.

[55] Artículo 3.7 del RIA «distribuidor»: una persona física o jurídica que forme parte de la cadena de suministro, distinta del proveedor o el importador, que comercialice un sistema de IA en el mercado de la Unión.

[56] Artículo 26 del RGPD.

[57] En cuanto a la incorporación a los Pliegos de cláusulas administrativas de los requisitos subjetivos y materiales exigidos por el ENS, léase Fondevila Antolin, Jorge, "Breves consideraciones sobre la ciberseguridad y su implementación en la Contratación Pública", *Boletín del Observatorio de Contratación Pública (2023)*: https://www.obcp.es/opiniones/breves-consideraciones-sobre-la-ciberseguridad-y-su-implementacion-en-la-contratacion

A mayor abundamiento, además de las formas directas e indirectas de gestión de los servicios públicos locales, de la asistencia técnica de las Administraciones supramunicipales[58] y del eventual uso de las técnicas de delegación intersubjetiva, los municipios también pueden utilizar técnicas o instrumentos jurídicos que, como la encomienda de gestión o los encargos a medios propios, implican la transferencia de determinadas funciones, aunque no la de la titularidad de la competencia o el servicio o, también, mediante formas asociativas de gestión compartida con otros entes locales o Administraciones, a través de fórmulas colaborativas que pueden dar lugar a la creación de nuevas personas jurídicas dotadas de personalidad jurídica propia y patrimonio independiente[59].

En el caso de que el sistema de IA se provea y/o despliegue mediante las anteriores fórmulas de la Administración local, debe contemplarse, en su caso, en las encomiendas, encargos, convenios o acuerdos, las obligaciones y responsabilidades que asumen cada una de éstas, por cuanto al tratamiento de datos personales, que pueda operarse con el uso de la IA en cualquiera de las etapas de su ciclo de vida.

Dicho esto, también es de interés analizar la forma de articular, de manera clara y completa, la relación entre la Administración local y el tercero, a quien se encargase la creación de un sistema de IA. Al respecto, aparte de articular el acto jurídico o contrato de encargo de tratamiento[60], como la declaración de ubicación de servidores y servicios asociados a los mismos[61], en nuestra opinión, resulta imprescindible atender también las siguientes diligencias, de forma enunciativa y no limitativa:

a) Diligencias *in eligendo*[62] específicas en cuanto a un sistema de IA:

- Seleccionar la solución tecnológica más adecuada, en particular cuando contrata su desarrollo o adquiere un sistema de IA. La Administración local debe "exigir y analizar las especificaciones de calidad de la solución; y determinar la extensión del tratamiento y la carga de hacer frente a las consecuencias de sus decisiones. El que toma la decisión de realizar el tratamiento es responsable, y no puede escudarse en la carencia de información o el desconocimiento técnico para evadir su responsabilidad a la hora de auditar y decidir la adecuación del sistema"[63].

- Solicitar la declaración de conformidad al proveedor de un sistema de IA, cuando es alto riesgo[64].

- Verificar que el sistema de IA, cuando es de alto riesgo, tiene el marcado CE o, cuando no sea posible, en su embalaje o en la documentación que lo acompañe, para indicar la conformidad con Reglamento de Inteligencia Artificial[65],

- En el caso de que el proveedor del sistema de IA sea una autoridad pública o institución, órgano u organismo de la Unión, podrá solicitarse que aporte documento acreditativo del registro del sistema en la base datos de la Unión Europea[66]. En el caso de que descubra que el sistema de IA de alto riesgo que tienen previsto utilizar no ha sido registrado en la base de datos de la UE, no utilizarán dicho sistema e informarán de ello al proveedor o al distribuidor.

[58] Diputaciones provinciales, Cabildos.

[59] En este caso, nos estamos refiriendo a las mancomunidades de municipios y los consorcios, desde los que se trata del servicio público, de manera conjunta y descentralizadamente.

[60] Artículo 28.3 del RGPD, artículo 33 de la LOPDGDD y Disposición adicional vigésima quinta de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público.

[61] Artículo 122.2 de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público.

[62] Artículo 28.1 del RGPD

[63] Guía de la Agencia Española de Protección de Datos relativa a la "*Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*", febrero 2020 (pág. 19). Si bien, en cuanto a la integración efectiva de la normativa de protección de datos de carácter personal y al principio de responsabilidad proactiva en el ámbito de la contratación administrativa pública léase a Guasp Martínez, Virginia, "Incidencia práctica de la normativa de protección de datos en la contratación administrativa", *Revista jurídica de Castilla y León*, nº. 51 (2020): 205-254.

[64] Artículo 16.g) del RIA.

[65] Artículo 16.h) del RIA.

[66] Considerado 131 y artículos 26.8 y 71 del RIA.

- Solicitar al proveedor de un sistema de IA de alto riego las fichas técnicas que describan las metodologías y técnicas de entrenamiento, así como los conjuntos de datos de entrenamiento utilizados, e incluyan una descripción general de dichos conjuntos de datos e información acerca de su procedencia, su alcance y sus características principales; la manera en que se obtuvieron y seleccionaron los datos; los procedimientos de etiquetado (p.ej., para el aprendizaje supervisado) y las metodologías de depuración de datos (p.ej., la detección de anomalías).

En la adquisición de sistema de IA, la Administración local podría optar por licitar un contrato de prestación de servicios consistente en el desarrollo de una solución basada en IA o, en su caso, se licite un contrato de suministro para la adquisición de licencias de uso de soluciones ya existentes. Pero, en todo caso, tal y como señala Berning Prieto[67] "para conseguir una verdadera actuación administrativa garantista se deberá apostar por el desarrollo de aplicaciones para uso exclusivo por las administraciones públicas y posterior compartición con el resto a través de las herramientas en materia de transferencia de tecnología reguladas en el artículo 158 LRJSP materializado en el Centro de Transferencia de Tecnología de la Administración General del Estado".

b) Diligencias in vigilando durante el uso del sistema de la IA:

- Los proveedores de sistemas de IA de alto riesgo que consideren o tengan motivos para considerar que un sistema de IA de alto riesgo que han introducido en el mercado o puesto en servicio no es conforme con el RIA adoptarán inmediatamente las medidas correctoras necesarias para que sea conforme, para retirarlo del mercado, desactivarlo o recuperarlo, según proceda. Es por esto que, se podría prever en el pliego o contrato administrativo la comunicación a la Administración que haya desplegado o implementado dicho sistema de IA[68].

- En el caso de que el sistema de IA de alto riesgo presente un riesgo y el proveedor tenga conocimiento de dicho riesgo, este investigará inmediatamente las causas, en colaboración con la Administración local (responsable del despliegue= que lo haya notificado, en su caso, e informará a las autoridades de vigilancia del mercado competentes respecto al sistema de IA de alto riesgo de que se trate y, cuando proceda, al organismo notificado que haya expedido un certificado para dicho sistema de conformidad 44, en particular sobre la naturaleza del incumplimiento y sobre cualquier medida correctora adoptada[69].

- Los proveedores de sistemas de IA de alto riesgo colaborarán con la Administración local, responsable del despliegue o implementación, aportando la documentación técnica del sistema, a fin de llevar a cabo una evaluación de impacto relativa a la protección de datos[70].

7. Conclusiones

La Administración local y, podemos decir, la Administración pública, en general, debe ser capaz de aplicar una gobernanza de los datos obrantes en sus sistemas de información, a fin de lograr datos de calidad y no sesgados para la IA.

Entre la ingente diversidad de datos, debe tenerse especial atención a los datos de carácter personal. Así, a fin de llevar a cabo tratamientos de datos personales, en cualesquiera etapas del ciclo de vida del sistema de IA, la Administración local, en calidad de responsable del tratamiento, ha de atender a las obligaciones de la legislación en protección de datos de carácter personal.

En una aproximación, a qué y cómo debe atender en primer lugar, en el diseño, el desarrollo o el uso de sistemas de IA que impliquen el tratamiento de datos personales, tal y como se ha expuesto, la Administración local deberá comenzar por solventar estas cuestiones de carácter relevante:

- Dilucidar la condición que ostenta, responsable o encargado de tratamiento, proveedor o responsable del despliegue, a fin de deslindar, de manera apropiada, las obligaciones y responsabilidades de todas las Partes, en aplicación del marco normativo de la protección de

[67] Berning Prieto, Antonio David "El uso de sistemas basados en inteligencia artificial por las Administraciones públicas: estado actual de la cuestión y algunas propuestas ad futurum para un uso responsable", *Revista de Estudios de la Administración local y Autonómica*. Núm. 20 (2023): 163-185 DOI: https://doi.org/10.24965/reala.11247

[68] Artículo 20.1 del RIA.

[69] Artículo 20.2 del RIA.

[70] Artículo 26.9 del RIA.

datos y la IA. Dado que gran parte de nuestra Administración local española está abocada a implementar las soluciones de IA de terceros, debe prever en la contratación, de la manera en la que se ha expuesto, que estos proveedores cumplan con las exigencias del Reglamento de Inteligencia Artificial europeo. A nuestro entender, del mismo modo que la Ley de Contratos de Sector público incluyó, explícitamente, la obligación del futuro contratista de someterse en todo caso a la normativa nacional y de la Unión Europea en materia de protección de dato[71], también debería incorporar disposiciones en cuanto a la sujeción del mismo a la normativa, europea y española, en IA.

- Saber qué operaciones o actividades de tratamiento de datos personales, en la aplicación de la IA, deben inventariarse o registrarse. A nuestro entender, en la publicidad activa de las actividades de tratamiento realizadas por la Administración local[72], debe constar el propio hecho de aplicar IA.

- Dilucidar la condición/es de licitud de los tratamientos de datos personales, que respondan, sin perjuicio de aplicar otras bases de legitimación, a un interés público esencial; sobre todo, cuando el uso de la IA está orientado a la satisfacción de servicio público. El invocar el interés público, en aplicación de la legislación en proteccion de datos, conduce a que haya Leyes o normas con rango de ley, que habiliten o amparen el tratamiento. Además, no basta, según la doctrina constitucional, con una mera remisión genérica a las normas generales de protección de datos[73]. Es por esto que, el legislador español, entendemos que, progresivamente, debe incluir disposiciones en el ordenamiento jurídico que sirvan de condición o licitud de estos tratamientos de datos personales. De lo contrario, nos podemos encontrar, como en otros supuestos, huérfanos de legitimidad.

- Aplicar transparencia, en la actividad de tratamiento de datos y en el uso de la IA. A pesar de que es diferente por cuanto, a la información a transmitir, sí encontramos que puede conjugarse, sobre todo, cuando de un tratamiento de datos personales, automatizado y con aplicación de IA, se tomen decisiones con efectos jurídicos; en el ámbito que nos ocupa, en las decisiones administrativas automatizadas.

Para toda esta labor, entendemos esencial el papel de la figura del delegado de protección de datos en la aplicación de la IA, a los efectos de garantizar la protección de los datos personales, así como el respeto a los derechos y libertades de los ciudadanos[74]. Del mismo modo que, abogamos también por el papel crucial que deben tener las diputaciones provinciales, cabildos, mancomunidades o las entidades públicas supramunicipales, en la implantación efectiva de la IA en el ámbito local español[75].

Bibliografía

Almonacid Lamelas, Víctor, "Impacto transformador de la inteligencia artificial en la Administración pública", *Revista DerechoLocal.es* ed. Lefebvre (2024).

"Irrupción y repercusión de la inteligencia artificial (IA) en las funciones reservadas: una visión muy práctica" *El Consultor de los Ayuntamientos: Las entidades locales ante el reto de la inteligencia artificial* (2024).

[71] Artículo 122.1 de la LCSP.

[72] Artículo 6bis de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.

[73] Esta doctrina del TC español es la llamada "teoría de las garantías adecuadas". Según la cual, la limitación legal de un derecho fundamental, como en este caso, el de la protección de datos, debe incorporar, dentro de la propia norma habilitante, "garantías adecuadas de tipo técnico, organizativo y procedimental" que sean específicas, proporcionadas al riesgo y suficientes para "procurar el respeto del contenido esencial del propio derecho fundamental" (STC 76/2019, fj. 6).

[74] Mendoza Balladares, Carmen Patricia, "El rol del delegado de protección de datos en el sector público y el uso de la IA", *Revista Canaria de Administración pública*, nº 4 (2025): 243-296.

[75] Según manifiesta la Federación Española de Municipios y Provincias (FEMP) en la "Guía Práctica y Políticas de uso de la IA en las entidades locales" ed. 2025, es crucial el papel de los entes supralocales, en particular, en aquellos municipios con menores recursos técnicos, económicos y humanos. Estos entes supralocales pueden facilitar economías de escala y posibilitar que municipios más pequeños puedan acceder a tecnologías avanzadas que, individualmente, estarían fuera de su alcance.

Alonso Peña, Carlos, "Calidad del dato en sistemas de IA desde su adecuado gobierno y gestión", *El Consultor de los Ayuntamientos. Revista técnica especializada en Administración local y justicia municipal*, nº extra III, edit. La Ley (2024).

Berning Prieto, Antonio David "El uso de sistemas basados en inteligencia artificial por las Administraciones públicas: estado actual de la cuestión y algunas propuestas ad futurum para un uso responsable", *Revista de Estudios de la Administración local y Autonómica*. Núm. 20 (2023): 163-185. DOI: https://doi.org/10.24965/reala.11247

Cerrillo i Martínez, Agustí, "La personalización de los servicios digitales", en "La administración digital" 311-342 editado por Dykinson (2022).

Fernández Hernández, Carlos,"Estructura y contenido del Real Decreto 817/2023, de 8 de noviembre, por el que se establece un entorno controlado de pruebas (*sandbox*), para el ensayo de sistemas de inteligencia artificial", *Diario La Ley*, nº 18 (2023).

Fondevila Antolin, Jorge "Breves consideraciones sobre la ciberseguridad y su implementación en la Contratación Pública", *Boletín del Observatorio de Contratación Pública* (2023): https://www.obcp.es/opiniones/breves-consideraciones-sobre-la-ciberseguridad-y-su-implementacion-en-la-contratacion

Guasp Martínez, Virginia, "Incidencia práctica de la normativa de protección de datos en la contratación administrativa", *Revista jurídica de Castilla y León*, nº. 51 (2020): 205-254.

Loza Corera, María, "Datos y gobernanza de datos y conexiones con principios de protección de datos en el art 10 del Reglamento". En *Tratado sobre el Reglamento de Inteligencia Artificial de la UE*, editado por Aranzadi, 473-497, 2024.

Martínez Martínez, Ricard "De la limitación de la informática a la transformación digital", *Revista Agenda Pública. Análisis de políticas públicas* (2021) https://agendapublica.es/noticia/13475/limitacion-informatica-transformacion-digital.

"Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo", *Revista catalana de dret públic*, nº. 58 (2019): 64-81.

Mendoza Balladares, Carmen Patricia, "El rol del delegado de protección de datos en el sector público y el uso de la IA", *Revista Canaria de Administración pública*, nº 4 (2025): 243-296. https://doi.org/10.36151/RCAP.4.9.

Mir Puigpelat, Oriol. "Capítulo XVI. Algoritmos, inteligencia artificial y procedimiento administrativo: principios comunes en el Derecho de la Unión Europea". En *Inteligencia artificial y sector público*: retos, límites y medios, editado por Tirant Lo Blanch, 685-725, 2023.

Montoro Montarroso, Andrés "Propuesta de metodología para la implantación de sistemas de Inteligencia Artificial para la prevención del fraude en la contratación Pública". En *Inteligencia artificial y administraciones públicas: una triple visión en clave comparada,* editado por Iustel, 401-409, 2024.

Ponce Sole, Julio "Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico". *Revista General de Derecho Administrativo*, nº. 50 (Iustel, 2019): 19-0.

Sánchez Bolívar, Francisco José "Contratación pública para la digitalización del Sector Público: optimizando la actividad administrativa a través de la inteligencia artificial". En *Inteligencia artificial y administraciones públicas: una triple visión en clave comparada*, editado por Iustel, 411-420, 2024.

Valero Torrijos, Julián, "Las singularidades del tratamiento de datos de carácter personal en entornos de inteligencia artificial en el sector púbico". En *Inteligencia artificial y sector público: retos, límites y medios*, editado por Tirant Lo Blanch, 335-396, 2023.

Recensiones

DemocracIA. Un análisis en clave constitucional, Dykinson, Madrid, 2025
Jorge Castellanos Claramunt

Sistemas de Inteligencia Artificial en la Agencia Estatal de Administración Tributaria: despliegue tecnológico y control jurídico
Atelier, Madrid, 2026
Olivares Olivares, Bernardo

DemocracIA. Un análisis en clave constitucional

Jorge Castellanos Claramunt
1.ª ed.
Dykinson, Madrid, 2025
178 pp.
ISBN 9791370060626
DOI 10.14679/3871
Depósito legal M-3840-2025

La obra de Jorge Castellanos Claramunt propone una lectura constitucional del giro tecnológico asociado a la inteligencia artificial, con un foco explícito en la calidad democrática, la ciudadanía y la inteligibilidad del marco normativo. Desde la introducción, el autor sitúa el problema en términos de agencia política: "no se puede dimitir de la condición de ciudadano" (p. 19), porque la fascinación tecnológica puede traducirse en pérdida de capacidad de organización en libertad y en aceptación de un "nuevo paradigma político" que ya no se sostiene sin fricciones sobre la idea de que "el poder descansa en el pueblo" (p. 19). Su tesis de fondo se articula en una advertencia doble: la democracia se enfrenta a riesgos estructurales (desinformación, microsegmentación, opacidad algorítmica, concentración de poder) y, simultáneamente, la regulación corre el riesgo de ser tan técnica que se vuelva políticamente inasumible, un "galimatías" que impide comprender "aquello que… nos influye" (p. 13).

El libro está organizado en cuatro capítulos más conclusiones y bibliografía. El capítulo I contextualiza la IA en términos tecnológicos y jurídicos, subrayando su aceleración histórica y su inserción en la vida cotidiana; el capítulo II analiza el *AI Act* (Reglamento de Inteligencia Artificial de la UE) y lo contrasta con el Convenio del Consejo de Europa; el capítulo III se centra en la erosión del Estado constitucional y en el impacto sobre procesos democráticos (incluyendo un caso contemporáneo de anulación electoral en Rumanía); el capítulo IV examina efectos sociales con atención a colectivos vulnerables, educación, idealización ("divinización"), burbuja tecnológica y uso retórico de la dignidad.

La tesis se despliega desde una tensión central: la IA como "utensilio" y la IA como vector de transformación política. El autor ancla la primera idea en una formulación explícita: "La inteligencia artificial es un utensilio humano" (p. 24). A partir de ahí, construye una advertencia: el problema democrático no es solo tecnológico, sino constitucional, porque afecta a condiciones de posibilidad del autogobierno (información pública, pluralismo, control del poder, comprensión normativa) y a la relación efectivo-simbólica entre ciudadanía y reglas.

En términos estructurales, el propio índice del libro (pp. 9–10) muestra una disposición progresiva: del contexto (cap. I) a la regulación (cap. II), de esta a la democracia (cap. III) y, finalmente, a impactos sociales (cap. IV), cerrando con conclusiones (p. 159) y bibliografía (p. 167). (pp. 9–10, 159, 167).

La introducción fija, además, un criterio de legitimidad: una regulación democrática de "materias sensibles" debe ser "asumible por la ciudadanía", evitando escapatorias que habiliten manipulación (p. 19) y denunciando la opacidad técnica como estrategia no inocente: "ese galimatías técnico…" (p. 13).

Introducción

El autor abre con una afirmación de fuerte carga doctrinal: el derecho es "piedra angular" de la organización política y el derecho constitucional sostiene el edificio democrático (p. 11). Ese encuadre habilita su crítica doble: (i) la opacidad tecnológica como dificultad epistemológica para la ciudadanía y (ii) la opacidad normativa como dificultad política.

La idea más rotunda, por su tono de exigencia cívica, aparece en un pasaje que condensa su noción de "DemocracIA" como mutación sin ruptura: si el ciudadano acepta pasivamente tecnologías orientadas —"muy presumiblemente"— a la manipulación, asume "un nuevo paradigma político" que no es "decididamente antidemocrático", pero tampoco compatible sin tensiones con la etimología

y la práctica del poder popular (p. 19). La frase programática "no se puede dimitir de la condición de ciudadano" (p. 19) sitúa al lector en una ética constitucional de la responsabilidad política, más cercana al ensayo crítico que a una exposición neutral.

Críticamente, esta estrategia tiene un riesgo: el salto desde "posible manipulación" a "presumible orientación manipuladora" requiere apoyos empíricos o delimitaciones conceptuales más finas (¿qué modelos de negocio?, ¿qué actores?, ¿qué técnicas?, ¿qué umbrales probatorios?). En el texto, el argumento se sostiene sobre plausibilidad y ejemplos posteriores, pero la delimitación inicial opera como advertencia general más que como tipología analítica.

Capítulo I: contexto tecnológico y jurídico de la IA

El capítulo I cumple una función pedagógica: ordenar el fenómeno y preparar el tránsito a la regulación. En su arranque, el autor insiste en que el análisis constitucional exige contextualizar la "realidad tecnológica y jurídica" y subraya la "dependencia" cotidiana: "No hemos empleado el término dependencia de manera casual" (p. 21). El mensaje es inequívoco: el peso social de la tecnología convierte a la IA en cuestión constitucional.

En la sección histórica, el texto recuerda que el término "inteligencia artificial" fue acuñado formalmente en 1959 por John McCarthy (p. 25) y que el fenómeno no es estrictamente contemporáneo, aunque su aceleración sí lo sea. El interés para la recensión no está en la erudición histórica —tratada con concisión—, sino en la función argumental: demostrar que la IA no es "mágica", sino técnica humana; por eso el autor insiste: "La inteligencia artificial es un utensilio humano…" (p. 24) y, por tanto, no debe "superponerse la realidad tecnológica a la humana" (p. 24).

La segunda sección aborda la aproximación regulatoria europea "desde una perspectiva crítica". Aquí se formula una caracterización muy cargada: bajo el discurso de "confianza" y "derechos fundamentales", la acción regulatoria adolecería de "lentitud, falta de coherencia, excesiva burocratización y una desconexión palpable" con un sector que evoluciona deprisa (p. 36). Es un diagnóstico reconocible en debates de política pública: el desfase regulatorio ante tecnología acelerada. La virtud del libro es conectar ese desfase con riesgos democráticos: si el derecho llega tarde o llega incomprensible, pierde capacidad de orientar conductas y de habilitar control público.

Capítulo II: el AI Act y el Convenio del Consejo de Europa

El capítulo II es el más "jurídico-institucional" del volumen. Arranca afirmando que el Reglamento de IA representa "el primer marco jurídico integral" para abordar riesgos de IA y pretende posicionar a Europa como líder (p. 41). Esta tesis general se alinea con la caracterización oficial del Reglamento (UE) 2024/1689 como marco armonizado para el mercado interior y para una IA "centrada en el ser humano y fiable".

La sección sobre "fases de implantación" es especialmente útil para el lector no especializado porque ofrece un calendario escalonado (p. 43). En cuanto al diseño regulatorio, el autor expone la clasificación de riesgos y conecta la definición de sistema de IA con la recomendación de la OCDE (p. 45), insistiendo en la inferencia como rasgo distintivo (p. 45). El interés constitucional de este análisis reside en dos planos: (i) cómo se traduce riesgo técnico en categorías jurídicas y (ii) qué tipos de prácticas se prohíben o restringen para proteger autonomía, no discriminación o seguridad.

El tramo de "análisis crítico" introduce observaciones que se apoyan en el problema clásico de la gobernanza multinivel: la eficacia del régimen depende de capacidades nacionales y de coordinación. El autor advierte, por ejemplo, que determinadas flexibilidades podrían "generar inseguridad jurídica" (p. 77) cuando no está claro cómo se garantizará la supervisión en pruebas en condiciones reales. La crítica es coherente con debates actuales sobre *sandboxes*, supervisión y enforcement: no basta con definir obligaciones; hay que asegurar cumplimiento institucional.

El capítulo culmina con una comparación con el Convenio del Consejo de Europa sobre IA. La formulación central es que el Convenio se distingue por su "enfoque transversal" y por pretender un "marco jurídico global que trascienda las fronteras europeas" (p. 84), articulado en principios como transparencia, responsabilidad y no discriminación (p. 84). Esta lectura coincide con el carácter paneuropeo (y abierto a no miembros) señalado por el Consejo de Europa al anunciar el tratado como "primer" instrumento internacional vinculante en la materia y con la participación de múltiples Estados y de la propia UE.

Capítulo III: influencia en el desarrollo democrático

El capítulo III desplaza el centro desde la regulación hacia la democracia como práctica constitucional. El subepígrafe sobre desgaste del Estado constitucional anuncia el movimiento conceptual: la IA no afecta solo a derechos concretos, sino a "pilares estructurales del Estado constitucional" (p. 99), generando dificultades para soberanía, equilibrio de poderes y principio de legalidad (p. 99). Esta formulación es relevante porque evita reducir la IA a un "tema sectorial" y la trata como fenómeno transversal con capacidad de reconfigurar condiciones de legitimidad.

En la sección de desinformación, el autor adopta un tono deliberadamente admonitorio: "Una democracia… puede debilitarse con sorprendente facilidad" (p. 103) si se descuidan valores o libertades, y la manipulación informativa se convierte en "baza" de desestabilización (p. 103). La pertinencia constitucional aquí es clara: pluralismo, deliberación pública y formación de la voluntad política son presupuestos de calidad democrática.

El libro refuerza este punto con un caso concreto: la anulación de la primera vuelta de las elecciones presidenciales en Rumanía por el tribunal constitucional del país, tras la desclasificación de informes de inteligencia que apuntaban a interferencias y "ataques híbridos" (pp. 110–111). El uso del ejemplo es metodológicamente significativo: el libro no se queda en una crítica abstracta, sino que conecta IA/ciberinjerencia con integridad electoral, aunque no desarrolla un test constitucional comparado (márgenes de apreciación, estándares probatorios, garantías jurisdiccionales), lo que abre un campo de trabajo para futuras investigaciones.

El subepígrafe sobre transparencia continúa una línea ya apuntada en la introducción: la opacidad algorítmica dificulta el escrutinio público y la rendición de cuentas (p. 116). En clave de política pública, esta preocupación dialoga con exigencias de transparencia del RGPD y con el derecho a no ser objeto de decisiones automatizadas (cuando concurren sus requisitos), tal como lo explican guías de autoridades competentes como la Agencia Española de Protección de Datos. En clave estrictamente interna del libro, la transparencia aparece como condición de posibilidad del control democrático más que como mero requisito documental (pp. 116–118).

El capítulo concluye con un subepígrafe de tono reflexivo: "No es nuestro cometido jugar a ser "pitoniso"" (p. 122), pero sí advertir que la velocidad de la evolución tecnológica exige un ajuste democrático consciente. En páginas posteriores, el autor llega a anticipar un escenario en el que, si se "eliminan progresivamente los principios democráticos", tenderán posiciones autoritarias que se sirvan de la IA para mantenerse en el poder, erosionando la rendición de cuentas y la transparencia (p. 124). Esta proyección es un argumento prospectivo: su fuerza depende de la plausibilidad política más que de una demostración empírica, pero funciona como cierre coherente con la tesis general de responsabilidad ciudadana.

Capítulo IV: impacto social (vulnerabilidad, educación, idealización, burbuja, dignidad)

El capítulo IV amplía el análisis hacia impactos sociales y culturales. El arranque enumera avances con "serias dudas éticas", incluyendo automatización de trabajos "incluyendo el de los jueces" y "tecnologías de cibervigilancia" (p. 127). La relevancia constitucional es evidente: debido proceso, independencia judicial, límites al poder de vigilancia, privacidad y libertad.

La sección sobre grupos vulnerables insiste en que la IA puede profundizar brechas (por edad, discapacidad, situación socioeconómica) y que el diseño inclusivo es condición para evitar discriminación estructural (pp. 128–131). Este foco es uno de los aportes más valiosos del libro, también porque conecta con debates sobre sesgos algorítmicos y con literatura ampliamente conocida (y citada en bibliografía) sobre discriminación, sistemas predictivos y desigualdad (pp. 167–168).

En educación, el texto aterriza en terreno constitucional español al recordar que la cuestión abarca la "vertiente constitucional relativa al artículo 27 CE" (p. 137).

El autor, además, reconoce el papel de la IA generativa (menciona herramientas como ChatGPT) y advierte que la tecnología debe mantener una posición de "mera herramienta", con preeminencia del enfoque ético para no subsumir la tarea educativa (p. 137). Aquí el libro ofrece un espacio fértil: el constitucionalismo democrático tiene en la educación un mecanismo de reproducción cívica; si la educación se tecnologiza sin criterios, se resiente la base cultural del autogobierno.

El subepígrafe sobre "divinización" es un punto alto en originalidad retórica. Se apoya en la idea de fe "ciega" en la infalibilidad de la IA (p. 140) y cita a Cathy O'Neil para calificar ciertos usos de datos

masivos como "armas de destrucción matemática" (p. 140). El movimiento es argumentativamente interesante: el autor transforma una crítica socio-técnica en una advertencia constitucional contra la entrega acrítica de capacidad decisoria. En contexto, esta crítica dialoga con literatura internacional sobre opacidad, escala y efectos regresivos de algoritmos (p. 140; y, como referencia externa, investigaciones como la de ProPublica sobre sesgo en *risk assessment*).

En "burbuja de la IA", el autor enfatiza que advertir problemas no equivale a tecnofobia y sostiene que "el ser humano no puede delegar todas las decisiones… en un ente tecnológico externo" (p. 150). La sección sugiere una economía política de la IA: promesas infladas, incentivos de mercado y narrativas legitimadoras que pueden reconfigurar prioridades públicas (pp. 150–153). Aquí la obra se acerca, indirectamente, a diagnósticos contemporáneos sobre tecnología, poder y progreso que también aparecen en su bibliografía, como el libro de Daron Acemoglu y Simon Johnson (p. 167), aunque el diálogo no se desarrolla de manera sistemática "capítulo a capítulo" con esa obra.

Finalmente, la sección sobre dignidad resume la sospecha del autor frente a ciertos discursos públicos: la dignidad "no deja de ser" usada como una "coartada" que suaviza la "noción negativa" del desarrollo tecnológico ocultando dificultades sociales (p. 157). Este pasaje conecta con el art. 10 CE (dignidad como fundamento del orden político y de la paz social) y con el riesgo de convertir una categoría constitucional en retórica legitimadora sin control material. La tesis es provocadora y útil para el debate público: obliga a preguntar cuándo la dignidad funciona como límite normativo y cuándo como eslogan.

Evaluación de metodología, argumentos y fuentes

Metodología. El libro mezcla tres registros: (i) contextualización tecnológica (cap. I), (ii) comentario institucional de marcos normativos europeos (cap. II) y (iii) ensayo constitucional sobre democracia (caps. III–IV). Esta hibridación es, al mismo tiempo, fortaleza y limitación. Es fortaleza porque permite construir un "hilo" que va de la técnica al núcleo democrático; es limitación porque algunos lectores echarán en falta una mayor separación entre descripción positiva (qué dice el Reglamento) y evaluación normativa (qué debería decir o producir). Por ejemplo, el capítulo II ofrece calendarios y categorías (pp. 43–45) que son útiles, pero la precisión técnico-jurídica debe cuidarse, como se ve en el matiz entre "publicación" y "entrada en vigor" del Reglamento (p. 43 vs. textos oficiales). [8]

Argumentación constitucional. El núcleo del razonamiento del autor puede resumirse así: la IA afecta derechos y también condiciones estructurales del Estado constitucional; por ello, la transparencia, la rendición de cuentas y la inteligibilidad normativa son exigencias democráticas, no solo técnicas (pp. 99, 116, 13). Esta tesis es consistente, y el libro la refuerza con ejemplos —como el caso rumano— y con conexiones entre educación y democracia (pp. 110–111, 137). [23]

Uso de fuentes. La bibliografía es amplia y heterogénea: incluye obras doctrinales españolas relevantes (p. ej., la monografía de Francisco Balaguer Callejón sobre el "algoritmo" y tratados sobre el AI Act) y también periodismo de investigación y agencias (p. 167). Esta mezcla es funcional para una obra que pretende captar riesgos democráticos: en desinformación y sesgo, el periodismo de investigación puede ser fuente empírica de primer orden, aunque siempre con cautela metodológica. (p. 167).

Fortalezas. Destacan cuatro: (1) planteamiento constitucional claro y "cívico" (p. 19); (2) crítica a la opacidad normativa accesible al lector no experto (p. 13); (3) atención a ámbitos usualmente periféricos en debates de IA (educación, vulnerabilidad) (pp. 128–137); (4) lectura integrada AI Act–Convenio del Consejo de Europa (p. 84), relevante para juristas que trabajen con múltiples niveles normativos.

Debilidades o zonas mejorables. Sin desmerecer el valor del conjunto, pueden señalarse: (a) tendencia a afirmaciones generales ("lentitud" y "desconexión" de la UE) que ganarían con indicadores comparativos (p. 36); (b) un enfoque más propositivo ayudaría a traducir la crítica en agendas de reforma (qué estándares exigibles, qué mecanismos institucionales, qué controles jurisdiccionales); (c) la parte estrictamente "constitucional española" podría dialogar más con el desarrollo jurisprudencial y con instrumentos internos como los derechos digitales en la LOPDGDD o la Carta de Derechos Digitales, que estructuran el ecosistema nacional.

Aportación de la obra

La contribución del libro es doble:

- 1) Integra IA y democracia como problema constitucional, no meramente regulatorio. Frente a un enfoque tecnocrático, exige reconstruir condiciones de ciudadanía, deliberación y control (pp. 13, 19, 99, 116).

- 2) Amplía el foco: no se queda en el AI Act, sino que conecta con educación, vulnerabilidad, narrativa de dignidad y economía política de la IA (pp. 128–157).

Este planteamiento dialoga con líneas doctrinales del constitucionalismo digital. Por ejemplo, trabajos en España han discutido si la Carta de Derechos Digitales y los reglamentos europeos (DSA, DMA, IA) apuntan hacia un "constitucionalismo digital" de nuevo cuño; la literatura insiste en el carácter no normativo de la Carta y en su función de marco de referencia. Asimismo, la monografía de Balaguer sobre "la constitución del algoritmo" explora el encaje difícil entre constitución analógica y mundo digital, ofreciendo un trasfondo doctrinal que refuerza el movimiento conceptual del libro reseñado.

Y, en materia de gobernanza multinivel, análisis españoles del Convenio del Consejo de Europa han destacado su complementariedad respecto al AI Act y su alcance no limitado a alto riesgo, reforzando la intuición que el libro formula al contraponer ambos instrumentos (p. 84). [32]

Conclusión

Conclusión evaluativa. *DemocraclA* es una monografía recomendable para lectores que quieran pensar la IA más allá del cumplimiento normativo y que necesiten un puente entre derecho constitucional y política tecnológica. Su principal virtud es convertir la IA en problema de ciudadanía: si el marco regulatorio es un "galimatías" incomprensible, la democracia pierde una capa de control; si el ciudadano "dimite" de su función, el cambio tecnológico puede traducirse en cambio político no deliberado (pp. 13, 19). El libro también gana valor al iluminar áreas menos transitadas en ciertos comentarios del AI Act (educación, vulnerabilidad, idealización tecnológica) (pp. 128–157) y al integrar, en un mismo mapa, AI Act y Convenio del Consejo de Europa (p. 84).

En términos críticos, su tono de advertencia —con formulaciones intensas— es útil para despertar preguntas constitucionales, pero a veces requiere un "segundo paso" metodológico: mayor precisión técnico-jurídica en cronologías y conceptos, y mayor traslación a propuestas institucionales o estándares de control. En cualquier caso, su aportación al debate sobre constitucionalismo digital es clara y su relevancia para políticas públicas de IA es alta: sirve para repensar la gobernanza de la IA como gobernanza de la democracia.

La valoración global es positiva: la monografía es oportuna, conceptualmente ambiciosa y útil para juristas que busquen conectar el debate regulatorio con categorías constitucionales (soberanía, legalidad, rendición de cuentas, derechos fundamentales). A la vez, su aproximación ensayística y normativa deja espacios para una crítica constructiva: hay pasajes donde el diagnóstico supera a la operativización (qué instrumentos concretos, qué estándares interpretativos, qué test de proporcionalidad o garantías procedimentales en cada contexto), y algunas formulaciones sobre cronología regulatoria requieren precisión técnica contrastada en textos oficiales. En conjunto, el libro contribuye al campo emergente del "constitucionalismo digital" y resulta especialmente relevante para el diálogo entre derecho público, gobernanza de IA y políticas democráticas en la Unión Europea.

Cristina Fernández González
Universidad Europea de Valencia

Sistemas de inteligencia artificial en la Agencia Estatal de Administración Tributaria: Despliegue tecnológico y control jurídico[1]

Bernardo Olivares Olivares
Atelier, Barcelona, 2026
563 pp.
ISBN 979-13-88096-79-2
Depósito legal B 6807-2026

El uso de la inteligencia artificial ha experimentado un crecimiento exponencial y sin precedentes en la última década. Este, asumido ya como habitual, impregna en sus distintas formas todos los ámbitos de la sociedad y genera nuevas realidades que van más allá del ámbito tecnológico –social, económico o jurídico, entre otros–. Esto se debe a la mayor disponibilidad de sistemas capaces de gestionar la ingente cantidad de datos que pueden obtenerse a partir de fuentes muy diversas.

El sector público no ha sido ajeno a esta transformación. En su recién publicado informe *Administración Tributaria 2025*, la OCDE da muestra de este fenómeno. Por un lado, el 87 % de las 58 administraciones tributarias analizadas utiliza macrodatos en el ejercicio de sus funciones; la mayoría de ellas para mejorar sus actuaciones en materia de cumplimiento. Por otro lado, casi todas declaran utilizar herramientas de ciencia de datos o analítica, y la mayoría ha introducido ya la inteligencia artificial, incluido el aprendizaje automático, en los procedimientos previstos para evaluar el riesgo, detectar los posibles incumplimientos y combatir el fraude fiscal.

Escenario de vanguardia tecnológica en la Administración Pública, la Agencia Estatal de Administración Tributaria (AEAT) es pionera en la implementación de estas herramientas. Sin embargo, esta *metamorfosis* no ha estado exenta de profundas controversias doctrinales, en las que no han faltado alusiones a realidades distópicas e, incluso, a clásicos de la literatura de ficción cuyas tramas parecen haber sido superadas por la realidad.

El eje vertebrador del debate no es otro que la existencia de una incuestionable tensión entre la búsqueda de la eficacia y eficiencia administrativas en la aplicación del sistema tributario, de un lado, y la incidencia de la denominada «informática tributaria» sobre la esfera jurídica de los obligados tributarios, de otro. Un debate que, si bien no es nuevo, sí que ha ocupado con mayor intensidad a la doctrina en la última década por el modo en que el uso de los nuevos sistemas exacerba las históricas preocupaciones sobre la protección de los derechos y garantías.

Pues bien, es precisamente en este contexto donde nace la obra *Sistemas de inteligencia artificial en la Agencia Estatal de Administración Tributaria: Despliegue tecnológico y control jurídico* que en estas líneas se reseña y en la que Bernardo D. Olivares, Profesor Titular de Derecho Financiero y Tributario de la Universidad Complutense de Madrid, demuestra que los riesgos que los distintos grupos de investigación de la universidad española han venido manifestando con respecto al uso de estos sistemas en la relación jurídico-tributaria no son meras hipótesis.

La monografía, presentada en acceso abierto, recoge, con la destreza y el rigor que caracterizan al autor, un exhaustivo análisis jurídico-documental del ecosistema tecnológico de la AEAT, lo que le permite llevar a cabo, posteriormente, una evaluación de impacto de estos sistemas en los derechos fundamentales de los obligados tributarios y proponer *in fine* vías de mejora de los mecanismos de control jurídico.

Para ello, el profesor Olivares teje una extraordinaria urdimbre que refuerza la cohesión de la investigación mediante cinco capítulos. Lejos de ser compartimentos estancos, cada uno de ellos actúa como el eslabón necesario para la consecución de un estudio crítico y exhaustivo que da respuesta a la demanda de análisis proactivos de tecnologías emergentes en las administraciones

[1] https://docta.ucm.es/rest/api/core/bitstreams/a7b62650-e7b2-4545-9a8c-42c3f224623a/content

públicas de la Agencia Española de Protección de Datos (AEPD) en su Estrategia 2025-2030 para el actual despliegue algorítmico.

Así las cosas, en el primer capítulo, el investigador se enfrenta al tradicional hermetismo de la Administración tributaria española y levanta lo que el propio autor denomina el «velo de la opacidad». En la actualidad, no cabe duda de la compleja y potente red de sistemas analíticos, fuentes de datos masivas y herramientas de integración desplegadas al amparo del artículo 96 de la Ley General Tributaria (LGT). Sin embargo, la sistematización de estos sistemas siempre ha sido compleja, debido a la dispersión y ambigüedad del contenido relativo a estas tecnologías en un considerable volumen de memorias, estrategias y directrices generales.

Es precisamente en este punto donde reside el innegable valor añadido del capítulo. El autor lleva a cabo una encomiable labor de revisión de la documentación de la AEAT relativa a 70 licitaciones de la Dirección General, 993 licitaciones de la Dirección del Servicio de Gestión Económica y 668 contratos menores, correspondientes a la última década, y describe, en términos totalmente aprehensibles para el lector, el funcionamiento de herramientas analíticas –como Zújar, MIDAS, HERMES, NIDEL, entre otras–, susceptibles de ser usadas para el control tributario.

Este mapa revela una certeza, a mi juicio, inobjetable: la «marcada asimetría» en la transparencia institucional. Hasta la publicación del Plan Anual de Control Tributario y Aduanero de 2025, la AEAT no reconoce el desarrollo de nuevas herramientas de inteligencia artificial en el marco de las actividades de comprobación e investigación; la relega a un papel auxiliar. Sin embargo, a pesar de no emplear «técnicas canónicas» de inteligencia artificial, muchos de estos sistemas pueden tener una «influencia sustancial» sobre las decisiones y la esfera jurídica de los obligados tributarios.

Partiendo de este escenario, en el capítulo segundo, se delimitan qué herramientas y aplicaciones empleadas por la AEAT encajan en la categoría jurídica de Sistema de Inteligencia Artificial (SIA), conforme al artículo 3.1 del Reglamento de Inteligencia Artificial de la Unión Europea (RIA). El autor lleva a cabo un test de subsunción tecnológica y supera, de este modo, la narrativa –a menudo simplificada– de la AEAT con respecto al desarrollo y uso de sistemas que deben encuadrarse en dicha categoría.

Ahora bien, como es ya por todos conocido, los SIA empleados por las autoridades tributarias y aduaneras han sido excluidos de la categoría de alto riesgo. Esta decisión del legislador europeo ha sido objeto de críticas por parte de la doctrina. En el capítulo tercero, el autor, lejos de dar por cerrada la cuestión, centra su estudio en las dos únicas vertientes del Anexo III que podrían incidir en la calificación de los SIA en la Administración tributaria: por un lado, la evaluación para el acceso y disfrute de servicios y prestaciones esenciales (apartado 5); y, por otro, los sistemas empleados para la garantía del cumplimiento del derecho (apartado 6).

Así pues, se advierte que, si bien la calificación exige un análisis caso por caso, las herramientas utilizadas en la obtención de datos y en la evaluación de riesgos operan en «zonas grises». Esta ambigüedad puede conducir a una mutación de un posible incumplimiento hacia una investigación por fraude fiscal con inusitada rapidez. No obstante, frente a los restrictivos umbrales de la normativa europea –analizados de forma pormenorizada en la obra–, resulta altamente improbable que estas herramientas alcancen la calificación de alto riesgo; y menos aún que puedan encuadrarse en el catálogo de prácticas prohibidas del artículo 5 del RIA.

Con respecto a este último, cabe destacar el modo en que el profesor Olivares conecta la investigación desarrollada con otra de sus líneas más consolidadas: el estudio de las ciencias conductuales en el ámbito tributario. La conjunción de los conocimientos sobre el comportamiento de los obligados tributarios y el uso de tecnologías avanzadas constituye un binomio especialmente interesante en la consecución de la eficacia y eficiencia administrativas, a la vez que inquietante para los derechos y garantías de los obligados tributarios.

La Administración tributaria española es una de las pioneras en la implementación de este binomio, como puede desprenderse del anexo 4.3 del informe *Administración Tributaria 2022* de la OCDE. La aplicación práctica de este sistema se materializó en su integración con la plataforma RentaWeb, con el fin de mostrar mensajes conductuales (*nudges*) a los contribuyentes que presentaran una alta probabilidad de cometer errores al modificar ciertas casillas.

Pues bien, es aquí donde el autor somete a escrutinio esta práctica y, huyendo de todo apriorismo subjetivo o sesgo doctrinal que tienda a condenar la actuación administrativa, apunta que la mayoría de estos acicates operan como una lícita «persuasión legal» orientada a fomentar el cumplimiento. No obstante, advierte que existen casos paradigmáticos de *sludge* conductual –como el bloqueo de opciones legales válidas en RentaWeb–, mediante los que la Administración puede imponer unilateralmente el criterio administrativo y que podrían socavar los principios de legalidad y buena administración.

Otra de las cuestiones fundamentales que se aborda en el capítulo es la necesaria creación de un organismo supervisor dotado de verdaderas facultades de control. En un contexto como el que acaba de exponerse, desde los grupos de investigación de la universidad española, se ha reclamado –en no pocas ocasiones– la puesta en marcha de auditorías –tanto internas como externas– de los sistemas utilizados por la AEAT.

En este sentido, el profesor Olivares evidencia que el diseño institucional de la recién creada Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), adscrita orgánicamente al poder ejecutivo, carece de total independencia e imparcialidad plenas, como exige la jurisprudencia y la normativa europeas. Esta arquitectura orgánica compromete su capacidad para actuar como contrapeso eficaz frente al «panóptico algorítmico» que ha construido la AEAT.

En el cuarto capítulo, se examina cómo el «arsenal tecnológico» de la AEAT incide sobre los derechos y garantías de los obligados tributarios, tensionando el histórico desequilibrio de posiciones en la relación jurídico-tributaria. Entre estos, destacan la dignidad humana, la igualdad, la no discriminación o el derecho de defensa, que debe entenderse vulnerado «desde el momento en que la actuación administrativa previa impide o dificulta gravemente al ciudadano articular su defensa y comprender plenamente la actuación seguida en su contra». Esta incapacidad para entender la lógica subyacente a un acto tributario no puede considerarse un mero problema técnico o un vicio formal del procedimiento.

No obstante, si hay una vertiente en la que el profesor Olivares destaca, es –como ya apunta el profesor Cotino Hueso en el prólogo– en su absoluto dominio sobre el derecho a la protección de datos de carácter personal. El autor advierte que, por un lado, la obtención y el cruce masivo de información –especialmente, el que proviene de fuentes abiertas o que no se ha proporcionado para un fin tributario–, y, por otro, la automatización masiva de actos administrativos y la elaboración de perfiles de riesgo, plantea serias dudas con respecto a los principios que rigen este derecho.
Lo cierto es que esta cuestión, si bien no es nueva, ha despertado un especial interés en la doctrina en los últimos años. Y es que la Administración ha consolidado su capacidad de obtener información de la mayor –y más actualizada– base de datos: las redes sociales. El uso de este ingente volumen de información para controlar el cumplimiento fiscal, con base en análisis inferenciales sobre los hábitos, las relaciones o la localización de los obligados tributarios, exacerba las clásicas controversias sobre la licitud, la transparencia y la proporcionalidad en el tratamiento de datos personales.

Un ejemplo –muy alarmante– de esta problemática lo constituye la previsión administrativa de emplear «avatares» o perfiles falsos para obtener información de plataformas de contenido restringido, como es el caso de *Patreon* o de *OnlyFans*. Como subraya el autor, «esta práctica no solo podría constituir una violación de los términos de servicio de las plataformas, sino que representa una intrusión deliberada y planificada en la esfera más reservada de la intimidad de los contribuyentes».

Finalmente, la monografía desemboca en su quinto capítulo, donde el autor estructura sus propuestas de mejora en tres líneas de actuación. La primera, la necesidad de una intervención legislativa decidida y garantista ante la constatación de una brecha entre las capacidades tecnológicas de la AEAT y la suficiencia del marco normativo actual. La segunda, la imperativa mejora de la gobernanza interna y las prácticas operativas de la AEAT en relación con el uso de SIA, con vistas a mitigar los riesgos identificados y fortalecer la protección de los derechos y garantías. La tercera, el crucial refuerzo del papel de los actores externos encargados de la supervisión, el control judicial y la asistencia al obligado tributario.

La obra, fruto de un notorio esfuerzo empírico, está llamada a convertirse en una referencia en la materia y debe recibirse, a mi juicio, con profundo sentimiento de satisfacción por quienes, en los últimos años, hemos navegado por los procelosos mares de la «informática tributaria». El profesor Olivares no solo aporta claridad a un fenómeno que ha ocupado de forma intensa a la doctrina, sino que, además, lo hace con una innegable generosidad intelectual y respeto hacia los trabajos de los distintos grupos de investigación que, en línea con los objetivos de la AEPD, se vuelcan en la consecución de una «innovación responsable» y en la construcción de un marco normativo robusto. Por todo lo expuesto, no cabe más que recomendar su lectura. Francamente, resulta arduo y complejo sintetizar en estas líneas un trabajo tan completo, riguroso y oportuno como el que se presenta. Espero haberlo conseguido; no obstante, pido disculpas al autor y al lector –a quien sugiero encarecidamente una lectura detenida– si no lo hubiera logrado por omitir partes que pudieran merecer una mayor atención.

Yeray Villegas Almagro
Profesor Lector de Derecho Financiero y Tributario
Universitat Oberta de Catalunya (UOC)