

Derecho de la neurología: Libro blanco para el desarrollo de la
Carta de Derechos Digitales en el ámbito de la neurotecnología
Beatriz Gregoraci, Jorge Castillo Abella y Francisco Velasco Caballero (dirs.)
ISBN: 979-13-87913-93-3
Madrid, 2026
pp. 729-769
DOI: 10.37417/derecho-de-la-neurotecnologia/18
Marcial Pons Ediciones Jurídicas y Sociales
Editado bajo licencia Creative Commons Attribution 4.0 International License

CAPÍTULO XVIII

INTELIGENCIA ARTIFICIAL Y NEUROTECNOLOGÍA

Sebastián LÓPEZ MAZA
Universidad Autónoma de Madrid
sebastian.lopez@uam.es

Gemma MINERO ALEJANDRE
Universidad Autónoma de Madrid
gemma.minero@uam.es

Carla BRAGADO HERRERO DE EGAÑA
Universidad Autónoma de Madrid
carla.bragado@uam.es

1. INTRODUCCIÓN

Hablar de IA y de neurotecnologías es hablar de dos tecnologías emergentes y disruptivas que postulan avances prometedores para la sociedad. A su vez, ambas son fuente de importantes riesgos para la persona, tanto en su dignidad, como en los derechos y libertades fundamentales que le son inherentes.

A grandes rasgos, la IA surge como un sistema de inteligencia paralelo a la inteligencia humana, susceptible de superarla en términos de eficiencia y agilidad. En particular, los sistemas de IA tienden a me-

jorar la capacidad de procesamiento y análisis de ingentes cantidades de datos o información hasta niveles que el humano difícilmente podría alcanzar, o bien, ello sería el resultado de un proceso largo y costoso. A través de las técnicas de aprendizaje automático, estos sistemas de IA están asimismo dotados de capacidad para tomar decisiones de forma autónoma, así como para inferir o generar resultados de diversa índole. En este marco, el individuo ha de lidiar e interrelacionarse con estas nuevas inteligencias artificiales que se despliegan en múltiples ámbitos: así, cuando formula una simple consulta en línea, cuando se sirve de ellas como herramienta de estudio o de trabajo, o cuando otros las emplean como herramientas de *marketing* y de contratación. Más allá de las capacidades de la IA para inferir datos sobre su persona, puede someterla también a manipulaciones o engaños de los que difícilmente es consciente.

Por su parte, y sin ánimo de proporcionar aquí una definición, diremos simplemente que las neurotecnologías remiten a un conjunto de herramientas tecnológicas susceptibles de descifrar la actividad cerebral humana y, eventualmente, manipularla o cambiarla. Las primeras aplicaciones en este campo de la innovación surgen con fines eminentemente terapéuticos: se trata de corregir deficiencias neurológicas de diverso tipo u ofrecer nuevos tratamientos para determinadas dolencias psiquiátricas (desde la depresión o enfermedades como el alzhéimer o el párkinson, hasta una parálisis cerebral que impide al afectado comunicarse con su entorno). Con todo, rápidamente estas nuevas tecnologías, capaces de leer la mente e influir en su actividad, se han trasladado también al ámbito del bienestar e incluso del mero entretenimiento (así, de manera particular, al terreno de los videojuegos). Tanto en su aplicación clínica o pseudoclínica, como en la eminentemente lúdica, los riesgos son evidentes: las neurotecnologías vienen a ser una suerte de «llave» para acceder al cerebro, el órgano central que pilota toda la actividad humana. Se comprenderá así que el potencial intrusivo en el individuo y en su mente es aún mayor que el de la IA. A partir de la captación y tratamiento de las señales cerebrales, se habla de poder traducir los pensamientos a palabras escritas. Asimismo, se habla de la posibilidad de modular o cambiar esa actividad cerebral con fines diversos: suprimir o editar recuerdos, aumentar las capacidades cognitivas, solucionar problemas relacionados con el aprendizaje, y cambiar estados de humor, sensaciones o patrones de comportamiento, entre muchos otros. De ahí que algunas voces se alcen pidiendo el reconocimiento de nuevos derechos fundamentales como son los derechos a la privacidad mental, al libre albedrío, a la identidad del individuo, a su protección contra los sesgos; más allá de la garantía al acceso equitativo a la llamada «neuroaumentación».

Es cierto que muchos de estos avances en el ámbito de la neurotecnología se encuentran aún en una fase incipiente o rudimentaria de su desarrollo, por lo que se desconoce aún su potencial real. Ello no obsta

para que el derecho, ante los usos proyectados de estas neurotecnologías, deba tratar de ofrecer respuestas óptimas.

Tampoco hay que olvidar que las dos tecnologías a las que aquí nos estamos refiriendo no son compartimentos estancos. Antes bien, la asociación de ambas, creando importantes sinergias, es ya una realidad, y la tendencia apunta a ir en aumento. En este sentido, la IA parece ser una potente herramienta de soporte de las neurotecnologías (así, por ejemplo, a efectos de analizar en tiempo real ingentes secuencias de imágenes sobre la actividad cerebral de un individuo que se encuentra sometido a un determinado estímulo).

En lo que a la IA se refiere, el legislador europeo hizo suyas estas preocupaciones éticas, renunciando a que la Unión Europea se convirtiera en un territorio líder en el desarrollo de dicha tecnología si la misma no era acorde con los valores presentes en su Carta de Derechos Fundamentales (en adelante, CDFUE)¹. Antes bien, el liderazgo debería venir de la mano del desarrollo de una IA centrada en el ser humano y fiel al respeto de sus derechos fundamentales. A estos efectos, la primera parte del presente informe ofrece un estudio detallado del Reglamento Europeo de Inteligencia Artificial (en adelante, RIA)². En particular, se incide en la clasificación que este instrumento normativo hace de los sistemas de IA en función de los riesgos que crean en materia de derechos fundamentales, poniendo el foco de atención en la relación directa o indirecta que algunos de estos sistemas guardan con determinados usos en el campo de la neurotecnología.

Descendiendo al terreno concreto del derecho civil, se presenta seguidamente un análisis riguroso sobre la manera en que tanto la IA como las neurotecnologías —separada o conjuntamente— pueden incidir en una noción fundamental en este campo jurídico, como es el consentimiento. En efecto, de entre los múltiples riesgos que entraña el uso de ambas tecnologías, están aquellos que inciden en la formación del consentimiento, poniéndose aquí el foco en la esfera contractual del individuo.

El informe finaliza con un apartado en el cual se exponen las principales conclusiones. En él se explora en qué medida el RIA constituye ya una fuente de tutela de los llamados «neuroderechos». Ello sin perjuicio de la conveniencia de dotar a las neurotecnologías de un marco jurídico propio y específico, que podría revelarse imprescindible en un futuro cercano.

¹ DOUE C núm. 202 de 7 de junio de 2016, pp. 389-405.

² 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) núm. 300/2008, (UE) núm. 167/2013, (UE) núm. 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. DOUE núm. 1689, de 12 de julio de 2024, pp. 1-144.

2. CLASIFICACIÓN DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN FUNCIÓN DE LOS RIESGOS QUE CREAN EN MATERIA DE DERECHOS FUNDAMENTALES

A la hora de estudiar las implicaciones que el derecho civil tiene en la IA y la neurotecnología, resulta imprescindible acudir al RIA. Esta norma, que es la primera regla clasificadora de los sistemas de IA y reguladora, de manera transversal, de las obligaciones de sus operadores, se caracteriza por su enfoque antropocéntrico: el ser humano, como titular de derechos fundamentales reconocidos en la CDFUE, es la vara de medir o de clasificar los sistemas de IA, permitiéndose su introducción y uso en la Unión Europea —o no— en función del grado de afectación del ejercicio de los citados derechos fundamentales³. En este sentido, el considerando 1.º del RIA es claro al afirmar que «[e]l objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme [...] a fin de promover la adopción de una IA centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la CDFUE, incluidos la democracia, el Estado de derecho y la protección del medio ambiente⁴. Claridad que se predica, asimismo, del considerando 6.º del RIA, que prevé que «la IA debe ser una tecnología centrada en el ser humano. Además, debe ser una herramienta para las personas y tener por objetivo último aumentar el bienestar humano». El nivel de riesgo de afectación de los derechos fundamentales asociado a usos específicos de sistemas de IA es el parámetro de medición de la clasificación de un sistema de IA como permitido o, por el contrario, como prohibido, así como de la categoría de sistema de IA permitido pero de alto riesgo o sistema de IA permitido de no alto riesgo. De acuerdo con el considerando 5.º del RIA, el menoscabo de los derechos fundamentales tenido en cuenta al elabo-

³ De acuerdo con el art. 3 del RIA, el término «operador» engloba la totalidad de eslabones de la cadena de creación y distribución del sistema de IA, lo que incluye el proveedor, el fabricante del producto, el responsable del despliegue, el representante autorizado, el importador o el distribuidor. Este art. 3 contiene una definición específica de cada tipo de operador enunciado.

⁴ De acuerdo con las Directrices éticas para una IA fiable del Grupo independiente de expertos de alto nivel sobre inteligencia artificial, creado por la Comisión Europea en junio de 2018, «La fiabilidad de la inteligencia artificial (IA) se apoya en tres componentes que deben satisfacerse a lo largo de todo el ciclo de vida del sistema: *a)* la IA debe ser lícita, es decir, cumplir todas las leyes y reglamentos aplicables; *b)* ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos; y *c)* debe ser robusta, tanto desde el punto de vista técnico como social, puesto que los sistemas de IA, incluso si las intenciones son buenas, pueden provocar daños accidentales» (p. 4). Entre las fuentes jurídicas cuyo cumplimiento es imprescindible para cumplir con el primer requisito figuran el derecho primario de la UE, compuesto por los Tratados de la Unión Europea y por su Carta de Derechos Fundamentales.

rar la clasificación de sistemas de IA es tanto tangible como intangible «e incluye los perjuicios físicos, psíquicos, sociales o económicos».

El considerando 48.º del RIA identifica los derechos fundamentales que, con mayor probabilidad, sufrirán las consecuencias adversas de un sistema de IA⁵. No solo están en juego el derecho a la intimidad, el derecho a la integridad, la libertad de expresión y el derecho a la protección de datos personales, que son, seguramente, los derechos o libertades fundamentales en los que se piensa cuando se oye hablar de neurotecnologías que puedan tener acceso al cerebro o a los estados mentales, sino que también pueden y suelen quedar afectados el derecho a la no discriminación, el derecho a la educación, la protección de los consumidores, los derechos de los trabajadores, el derecho a la tutela judicial efectiva y a un juez imparcial, los derechos de la defensa y la presunción de inocencia⁶.

De resultados de lo anterior se puede afirmar que la Unión Europea, implícitamente, ha renunciado a convertirse en un referente mundial en producción de sistemas de IA. Por el contrario, busca convertirse «en líder en la adopción de una IA fiable», aplicando el RIA «de conformidad con los valores de la Unión consagrados en la Carta» (considerando 2.º del RIA)⁷. Está por ver si el legislador europeo consigue convencer a otros legisladores nacionales o supranacionales de las virtudes de las que goza el RIA y conseguir un efecto expansivo similar al que se produjo con el Reglamento General de Protección de Datos Personales, que sirvió de inspiración de tantas normativas nacionales sobre la materia⁸.

⁵ En último término, se trata de proteger frente a la neurodiscriminación, como han expuesto MAGEE, P., IENCA, M. y FARAHANY, N. (2024), «Beyond neural data: Cognitive biometrics and mental privacy». *Neuron*, 112(18), pp.3017-3028; y MUHL, E. y ANDORNO, R. (2023), *Neurosurveillance in the workplace: do employers have the right to monitor employees' minds? Frontiers in Human Dynamics*.

⁶ Sobre la identificación de los principales derechos fundamentales afectados por las neurotecnologías y comúnmente reconocidos en el contexto europeo nos remitimos al informe de Roberto ANDORNO (2025), «Analysis of the existing European human rights framework concerning the human rights issues raised by neurotechnologies and their applications», comisionado por el Comité de Derechos Humanos del Consejo de Europa.

⁷ Así se indicó desde los primeros pasos del legislador europeo dados en esta materia. Ya en la propuesta del RIA se identificaba «el objetivo de que la Unión sea un líder mundial en el desarrollo de inteligencia artificial segura, digna de confianza y ética, [...] y garantiza la protección de los principios éticos, como solicitó específicamente el Parlamento Europeo» [apartado 1.1 de la Propuesta de Reglamento, presentada el 21 de abril de 2021, COM(2021) 206 final, 2021/0106(COD)]. Cuestiones que ya habían sido consensuadas en la reunión extraordinaria del Consejo Europeo (1 y 2 de octubre de 2020), tal y como figuran en sus Conclusiones EUCO 13/20, 2020, p. 6 y en la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, 2020/2012(INL).

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. *DOUE* L núm. 119 de 4 de mayo de 2016, pp. 1 a 88.

La clasificación de sistemas de IA no es inocua. Son cuantiosas las sanciones previstas por el uso de sistemas prohibidos, como no podía ser de otra manera si se busca el cumplimiento de esta norma —que habrá de asociarse, a su vez, con un sistema de vigilancia exigente—⁹. Además, las reglas sobre evaluación de riesgos, gestión de datos y transparencia impuestas a los operadores de sistemas de IA de alto riesgo son una importante carga. El legislador europeo buscó regular obligaciones previsibles, proporcionadas y claras a todos los operadores participantes en la cadena de valor, con el fin de garantizar la seguridad y el respeto de la legislación vigente protegiendo los derechos fundamentales durante todo el ciclo de vida de los sistemas de IA, desde la fase de diseño, realizando pruebas *ex ante* e implementando un sistema de gestión de riesgo, así como una obligación de vigilancia humana, con las que se busca reducir al mínimo el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas en esferas críticas como la educación, el empleo, servicios importantes o el poder judicial, pasando por obligaciones posteriores de transparencia y trazabilidad, unidas a controles *ex post* sólidos que permitan ofrecer a las personas afectadas una compensación efectiva. En el caso de determinados sistemas de IA, como son los sistemas de IA de uso general, sean o no sistemas de alto riesgo, se proponen obligaciones mínimas en materia de transparencia, en particular cuando se utilizan robots conversacionales o ultrafalsificaciones, con el objetivo de concienciar al usuario de que lo que hay detrás no es otro ser humano, sino un sistema de IA¹⁰.

La clasificación de sistemas de IA prevista en el RIA es tan compleja, con múltiples remisiones internas a sus diferentes anexos y excepciones recurrentes, así como excepciones a las referidas excepciones, que su comprensión requiere de una lectura sosegada, y una consulta paralela de las Directrices de la Comisión sobre las prácticas de inteligencia artificial prohibidas que se establecen en el RIA, publicadas

⁹ De acuerdo con el art. 99.3 RIA: «El no respeto de la prohibición de las prácticas de IA a que se refiere el art. 5 estará sujeto a multas administrativas de hasta 35 000 000 euros o, si el infractor es una empresa, de hasta el 7 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior».

¹⁰ De acuerdo con el art. 53 del RIA, los proveedores de estos sistemas de IA elaborarán y mantendrán actualizada la documentación técnica del modelo, incluida la información relativa al proceso de entrenamiento, con el fin de facilitarla, previa solicitud, a la Oficina de IA y a las autoridades nacionales competentes —en el caso español, la Agencia Española de Supervisión de la IA (AESIA), con sede en A Coruña. Por su parte, el art. 50 del RIA exige a los proveedores de sistemas de IA destinados a interactuar directamente con personas físicas que informen a los usuarios de que están relacionándose con un sistema de IA y no con otra persona física. A los proveedores de sistemas de IA que generen contenido sintético de audio, imagen, vídeo o texto, se les exige que el *output* permita detectar que han sido generados o manipulados de manera artificial. Asimismo, a los proveedores de sistemas de IA capaces de generar imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación se les exige que informen de que esos contenidos o imágenes han sido generados o manipulados de manera artificial.

el 29 de julio de 2025¹¹. En lo que a su entrada en vigor se refiere, las reglas referidas a sistemas de IA prohibidos ya están vigentes¹², mientras que el resto de normas contenidas en el RIA tienen una entrada en vigor progresiva, permitiendo a los operadores adaptarse al nuevo escenario normativo¹³.

Los sistemas de IA de riesgo inaceptable para el respeto y ejercicio de los derechos fundamentales son los previstos en el art.5 RIA y su uso en el seno de la UE está prohibido. Son un total de 8 tipos de sistemas de IA, todos ellos relacionados, directa o indirectamente, con la neurotecnología:

1) Sistemas de IA que permiten la manipulación, sirviéndose de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas. Por ejemplo, un *chatbot* de IA que promueva la autolesión entre los usuarios o incentive el suicidio o el perjuicio a otras personas o colectivos de personas mediante la promoción de contenidos terroristas o la incitación a la violencia contra minorías. Los perjuicios a los que se refiere esta norma pueden ser físicos, psicológicos, financieros, económicos o sociales¹⁴.

El considerando 29.º del RIA se refiere a «estímulos de audio, imagen o vídeo que las personas no pueden percibir —ya que dichos estímulos trascienden la percepción humana—, u otras técnicas manipulativas o engañosas que socavan o perjudican la autonomía, la toma

¹¹ C(2025) 5052 final. En esa misma fecha se publicaban las Directrices de la Comisión relativas a la definición de sistema de inteligencia artificial establecida en el RIA, C(2025) 5053 final. Si bien las Directrices ofrecen información valiosa sobre la interpretación de las prohibiciones por parte de la Comisión, no son vinculantes y las interpretaciones autorizadas de los conceptos previstos en el RIA quedan reservadas al TJUE.

¹² La entrada en vigor se produjo el 2 de febrero de 2025.

¹³ Se regulan tres fechas de entrada en vigor diferentes, en función de las reglas de que se trate: el 2 de agosto de 2025, el 2 de agosto de 2026 y el 2 de agosto de 2027.

¹⁴ Sobre el concepto de perjuicio psicológico, la Comisión Europea señala el siguiente ejemplo: «una aplicación de compañía de IA diseñada para emular patrones de habla, comportamientos y emociones humanos utiliza rasgos antropomórficos y estímulos emocionales para influir en los sentimientos, actitudes y opiniones de los usuarios; esto hace que dependan emocionalmente del servicio, lo que incentiva comportamientos similares a la adicción y puede provocar perjuicios considerables, como comportamientos suicidas y riesgos de perjudicar a otras personas» (p.40). La Comisión Europea se refiere a comportamientos adictivos, ansiedad y depresión, que pueden provocar más adelante perjuicios físicos, tales como insomnio, deterioro de la salud física o debilitamiento del sistema inmunitario.

de decisiones o la capacidad de elegir libremente de las personas de maneras de las que estas no son realmente conscientes o, cuando lo son, pueden seguir siendo engañadas o no pueden controlarlas u oponerles resistencia. Esto podría facilitarse, por ejemplo, mediante interfaces cerebro-máquina o realidad virtual, dado que permiten un mayor grado de control acerca de qué estímulos se presentan a las personas, en la medida en que pueden alterar sustancialmente su comportamiento de un modo que suponga un perjuicio considerable».

En sus Directrices sobre prácticas prohibidas, la Comisión Europea señala que «un juego puede utilizar neurotecnologías que posibilita la IA e interfaces cerebro-máquina que permiten a los usuarios controlar un juego o partes del mismo con un casco que detecte la actividad cerebral. La IA puede utilizarse para entrenar el cerebro del usuario de forma subrepticia y sin su conocimiento para revelar o extraer de los datos neuronales información que puede ser muy intrusiva o sensible (por ejemplo, información bancaria personal, información íntima, etc.) de un modo que pueda provocarle perjuicios considerables»¹⁵.

Sin embargo, el citado considerando 29.º prevé: «La prohibición de las prácticas de manipulación y explotación establecida en el presente Reglamento no debe afectar a prácticas lícitas en el contexto de un tratamiento médico, como el tratamiento psicológico de una enfermedad mental o la rehabilitación física, cuando dichas prácticas se lleven a cabo de conformidad con el derecho y las normas médicas aplicables, por ejemplo, con el consentimiento expreso de las personas o de sus representantes legales. Asimismo, no debe considerarse que las prácticas comerciales comunes y legítimas (por ejemplo, en el campo de la publicidad) que cumplan el derecho aplicable son, en sí mismas, prácticas de manipulación perjudiciales que posibilita la IA».

En relación con las técnicas de manipulación, la Comisión Europea hace hincapié en que, para estar prohibida la práctica, el perjuicio provocado ha de ser considerable¹⁶. En sus Directrices sobre prácticas prohibidas, señala: «[I]a prohibición establecida en el art. 5, apartado 1, letra a), del Reglamento de Inteligencia Artificial se refiere únicamente a los casos de manipulación subliminal considerablemente perjudicial; no se refiere a las aplicaciones de la interfaz cerebro-máquina en general cuando se diseñan de manera segura y protegida y de tal forma que respetan la privacidad y la autonomía individual. [...] Un ejemplo de técnica deliberadamente manipuladora es la manipulación sensorial, en la que el sistema de IA utiliza audio o imágenes de fondo que alteran

¹⁵ *Vid.*, p. 31.

¹⁶ «Debe considerarse que alcanzan dicho umbral los sistemas de IA que sugieren a una persona que cometa actos delictivos, como abuso y explotación sexuales, contenidos de violencia extrema o terroristas, o que incitan a las personas a cometer delitos, autolesiones o perjuicios a otras personas» (p. 42).

el estado de ánimo, por ejemplo, generando un aumento de la ansiedad y la angustia mental que influya en el comportamiento de los usuarios hasta el punto de provocar perjuicios considerables. Otro ejemplo es la manipulación personalizada, en la que un sistema de IA que crea y adapta mensajes muy persuasivos basados en los datos personales de una persona o explota otras vulnerabilidades individuales influye en su comportamiento o sus elecciones hasta el punto de provocar perjuicios considerables» (p. 32).

En relación con las técnicas engañosas, la Comisión Europea sostiene: «Una de las técnicas engañosas que puede utilizar la IA es, por ejemplo, un chatbot de IA que se haga pasar por un amigo o un pariente de una persona usando una voz sintética y que intente suplantar su identidad, dando lugar a estafas y provocando perjuicios considerables. Otro ejemplo es un sistema de IA que aprenda a detectar cuándo está siendo evaluado y detenga temporalmente todo comportamiento no deseado, para reanudarlo una vez finalizada la evaluación. Este comportamiento engañoso es especialmente peligroso, ya que evita cualquier supervisión humana externa del sistema y puede prohibirse si es razonablemente probable que provoque perjuicios considerables. En cambio, no puede considerarse que un sistema de IA generativa que presente incidentalmente alucinaciones o información falsa o engañosa se sirva de técnicas engañosas en el sentido del art. 5, apartado 1, letra a), del Reglamento de Inteligencia Artificial, teniendo en cuenta las limitaciones y el estado actual de la técnica de la IA generativa» (p. 34).

Téngase en cuenta que, para estar prohibido, el sistema de IA debe generar una «alteración sustancial» en la capacidad de tomar decisiones informadas y autónomas, lo que conlleva un grado de coacción, manipulación o engaño que trasciende la persuasión lícita¹⁷. Para que se aplique la prohibición, se necesita siempre una relación causal plausible o razonablemente probable entre la técnica subliminal, deliberadamente manipuladora o engañosa utilizada por el sistema de IA y sus efectos en el comportamiento. El hecho de que un sistema de IA sea capaz de generar comportamientos que mermen de manera apreciable la capacidad de las personas para tomar una decisión informada y socaven su capacidad de elegir libremente basta para cumplir esa condición. En sus Directrices, la Comisión Europea pone el siguiente ejemplo: «el proveedor concibe un chatbot de bienestar basado en la IA con el objetivo de ayudar y guiar a los usuarios para que lleven un estilo de vida saludable y proporcionar asesoramiento personalizado para la práctica

¹⁷ A la hora de interpretar cuándo puede producirse esta «alteración sustancial del comportamiento», la Comisión Europea, en sus Directrices sobre Prácticas Prohibidas, se remite a la Directiva 2005/29/CE (Directiva sobre las prácticas comerciales desleales) y a la jurisprudencia dictada por el TJUE en la que interpreta esta norma.

de ejercicios psicológicos y físicos. Sin embargo, si el chatbot explota las vulnerabilidades de una persona para que adopte hábitos poco saludables o practique actividades peligrosas (por ejemplo, hacer deporte en exceso sin descansar ni beber agua) y existen fundamentos para esperar que determinados usuarios sigan esas recomendaciones, que de otro modo no habrían seguido, y sufran perjuicios considerables (por ejemplo, un ataque al corazón u otros problemas de salud graves), dicho sistema de IA estaría sujeto a la prohibición del art. 5, apartado 1, letra a), del Reglamento de Inteligencia Artificial, aunque el proveedor no hubiera tenido la intención de que se produjera ese comportamiento ni de que las personas sufrieran consecuencias perjudiciales» (p. 39).

2) Sistemas de IA que exploten alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra.

En sus Directrices sobre prácticas prohibidas, la Comisión Europea pone diversos ejemplos, entre los que se pueden destacar los siguientes: «[u]n juego utiliza la IA para analizar el comportamiento y las preferencias individuales de los niños, a partir de los cuales crea recompensas personalizadas e impredecibles mediante programas de refuerzo adictivo y bucles con un efecto similar a la dopamina con el objetivo de fomentar el juego excesivo y el uso compulsivo» (p. 47); «[u]n chatbot terapéutico destinado a prestar apoyo en materia de salud mental y estrategias de adaptación a las personas con discapacidad intelectual puede explotar sus capacidades intelectuales limitadas para incitarlas a comprar productos médicos caros o a adoptar un comportamiento perjudicial para ellas u otras personas» (p. 48); y «un algoritmo predictivo de IA para dirigir publicidad de productos financieros predatorios a personas, seleccionadas por su código postal en localidades de bajos ingresos, que se encuentran en una situación financiera grave. De esta manera, explota su vulnerabilidad a este tipo de publicidad debido a la desesperación que pueden sentir, provocándoles perjuicios financieros considerables» (p. 49).

3) Sistemas de IA de puntuación ciudadana de personas físicas, esto es, sistemas para evaluar o clasificar a personas físicas o a colectivos de personas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas, de forma que la puntuación ciudadana resultante provoque un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas en contextos sociales que no guarden relación con los contextos donde se generaron

o recabaron los datos originalmente o que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este.

Debe prestarse especial atención al uso de sistemas de IA en la evaluación crediticia de las personas. En función de los casos, el sistema de IA se subsumirá en la prohibición del art. 5 del RIA o, por el contrario, en la categoría de sistema de IA permitido, pero de alto riesgo, del art. 6 del RIA. En sus citadas Directrices sobre prácticas prohibidas, la Comisión Europea señala: «en la sentencia SCHUFA I, el TJUE examinó un sistema de puntuación de solvencia utilizado en Alemania¹⁸. En ese asunto, el score [«puntuación»] generado por el programa informático era un «valor de probabilidad» relativo a la capacidad de una persona para cumplir sus compromisos de pago, que el TJUE calificó de «elaboración de perfiles»¹⁹. Más concretamente, el sistema estableció «un pronóstico sobre la probabilidad de un comportamiento futuro de una persona (score o calificación), como el reembolso de un préstamo, a partir de determinadas características de dicha persona [...]. La generación de valores de score (scoring) se basa en la hipótesis de que la clasificación de una persona en una categoría de personas con características comparables, que han revelado cierto comportamiento, permite prever un comportamiento similar». Según el TJUE, esta actividad respondía a la definición de «elaboración de perfiles» en el sentido del art. 4, punto 4, del RGPD²⁰. También puede considerarse que esta forma de elaboración de perfiles constituye una evaluación de las personas basada en sus características personales en el sentido del art. 5, apartado 1, letra c), del Reglamento de Inteligencia Artificial, que está prohibida si se lleva a cabo con sistemas de IA y siempre que se cumplan las demás condiciones para que se aplique dicha disposición» (pp. 67-68). Así, por ejemplo, un sistema de evaluación crediticia que decida la no concesión de un crédito o la aplicación de un interés más alto con base en datos tales como la calle en la que reside la persona, el origen étnico de su familia o haber recibido alguna multa de circulación.

Dos prácticas prohibidas que pueden destacarse, de entre los muchos ejemplos señalados por la Comisión Europea en sus Directrices, son las siguientes: «[u]na agencia de asistencia social utiliza un sistema de IA para estimar la probabilidad de fraude de los beneficiarios de prestaciones familiares sobre la base de características recabadas o inferidas a partir de contextos sociales sin relación o pertinencia aparente para evaluar el fraude, como tener un cónyuge de una determinada nacionalidad u origen étnico, disponer de conexión a internet,

¹⁸ Sentencia del Tribunal de Justicia de 7 de diciembre de 2023, SCHUFA Holding (Scoring), C-634/21, ECLI:EU:C:2023:957.

¹⁹ Apartado 14.

²⁰ Apartado 47.

el comportamiento en plataformas sociales, el rendimiento en el lugar de trabajo, etc.» (pp. 72-73); y «[u]n organismo público utiliza un sistema de IA para elaborar los perfiles de familias con vistas a la detección precoz de menores en situación de riesgo basándose en criterios como la salud mental de los progenitores y el desempleo, así como la información sobre el comportamiento social de los progenitores procedente de múltiples contextos. En función de la puntuación resultante, se selecciona a las familias para inspeccionarlas y se separa de ellas a los niños considerados “en situación de riesgo”, incluso en caso de transgresiones leves cometidas por los padres, como faltar en ocasiones a las citas médicas o recibir multas de tráfico» (pp. 73-74).

4) Sistemas de IA para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito basándose únicamente en la elaboración del perfil de una persona física o en la evaluación de los rasgos y características de su personalidad²¹. En el considerando 42 del RIA se justifica esta prohibición: las personas físicas deben ser juzgadas basándose en su comportamiento real y no en comportamientos predichos por una IA basados únicamente en la elaboración de sus perfiles o en los rasgos o características de su personalidad. En sus Directrices sobre prácticas prohibidas, la Comisión Europea plantea, entre otros, el siguiente ejemplo: «[u]n departamento de policía utiliza una herramienta de evaluación de riesgos basada en la IA para evaluar el riesgo de que niños pequeños y adolescentes estén implicados en “futuros delitos violentos y contra la propiedad”. El sistema evalúa a los menores sobre la base de sus relaciones con otras personas y sus supuestos niveles de riesgo, es decir, se puede considerar que los menores presentan un mayor riesgo de cometer un delito simplemente por estar vinculados a otra persona con una evaluación de alto riesgo, como un hermano o un amigo. Los niveles de riesgo de los padres también pueden afectar al nivel de riesgo de un menor. Tras estas evaluaciones de riesgos, la policía “registra” a estos menores en sus sistemas, los vigila por medio de inspecciones adicionales y los deriva a servicios de “atención” a la juventud»²².

²¹ La prohibición no se aplica a los sistemas de IA utilizados para apoyar la valoración humana de la implicación de una persona física en una actividad delictiva que ya se basa en hechos objetivos y verificables directamente relacionados con una actividad delictiva que han sido revisados por un ser humano.

²² *Vid.* p. 87. Las predicciones de delitos basadas en la ubicación o en los datos geoespaciales se centran en el lugar del delito o en la probabilidad de que se cometa un delito en esas zonas y quedan fuera del ámbito de aplicación de la prohibición porque este tipo de actividad policial no implica una evaluación de una persona concreta. Un ejemplo dado por la Comisión Europea en sus directrices es el siguiente: «[u]na autoridad aduanera utiliza herramientas de análisis de riesgos basadas en la IA para predecir la probabilidad de la localización de estupefacientes o mercancías ilícitas, basándose, por ejemplo, en las rutas de tráfico conocidas» (p. 90).

5) Sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión²³. De acuerdo con el considerando 43.º del RIA, esta práctica interfiere gravemente con los derechos de las personas a la intimidad y a la protección de datos, las priva del derecho a permanecer en el anonimato y crea el sentimiento de vigilancia masiva. Si una herramienta de extracción tiene instrucciones de recoger imágenes o vídeos que contengan únicamente los rostros humanos de determinadas personas o de un colectivo predefinido de personas, la extracción para, por ejemplo, encontrar a un delincuente concreto o para identificar a un grupo de víctimas, se consideraría selectiva y, por tanto, no estaría dentro de esta prohibición.

6) Sistemas de IA para inferir las emociones de una persona física a partir de sus datos biométricos en sus lugares de trabajo y en sus centros educativos. En el considerando 44.º del RIA se explica que la expresión de las emociones varía de forma considerable entre culturas y situaciones, e incluso en una misma persona. Además, se menciona que algunas de las deficiencias principales de estos sistemas son la fiabilidad limitada, la falta de especificidad y la limitada posibilidad de generalizar, lo que puede dar lugar a resultados discriminatorios e invasión de los derechos y las libertades de las personas afectadas. «Teniendo en cuenta el desequilibrio de poder en el contexto laboral o educativo, unido al carácter intrusivo de estos sistemas, dichos sistemas podrían dar lugar a un trato perjudicial o desfavorable de determinadas personas físicas o colectivos enteros»²⁴. Un sistema de IA que infiere emociones de las pulsaciones de la forma de teclear, expresiones faciales, posturas corporales o movimientos se basa en datos biométricos y entra en el ámbito de aplicación de la prohibición. Sin embargo, no entraría dentro de esta prohibición un sistema de IA que infiera emociones basándose en un texto escrito por un trabajador o un estudiante²⁵.

7) Sistemas de IA para la categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir determinadas características sensibles, como su raza, opiniones políticas, afiliación sindical, convicciones re-

²³ Una base de datos de reconocimiento facial es capaz de cotejar un rostro humano de una imagen digital o una imagen de vídeo con una base de datos de caras, comparar dicha imagen con las de la base de datos y determinar si existe una coincidencia probable entre ambas.

²⁴ En el considerando 18 del RIA se aclara que no forman parte de las emociones o intenciones «los estados físicos, como el dolor o el cansancio, como, por ejemplo, los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes».

²⁵ De acuerdo con la Comisión Europea, «No está prohibido utilizar un sistema de IA que permite detectar a los estudiantes que hablan entre ellos a través de sus teléfonos u otros canales durante las clases en línea de un centro educativo, ya que no deduce emociones. En cambio, el uso del sistema para detectar emociones, como la agitación emocional, la ansiedad y el interés, entraría en el ámbito de aplicación de la prohibición» (p. 107 de las Directrices sobre Prácticas Prohibidas).

ligiosas o filosóficas, vida sexual u orientación sexual²⁶. Estos sistemas de categorización biométrica menoscaban la dignidad humana y presentan un riesgo significativo para otros derechos fundamentales, como el derecho a la intimidad y la no discriminación. Entre los ejemplos dados por la Comisión Europea sobre sistemas de IA prohibidos pueden destacarse los siguientes: «[u]n sistema de IA categoriza a las personas activas en una plataforma de redes sociales en función de su supuesta orientación política, mediante el análisis de los datos biométricos de las fotografías que han subido a la plataforma, para enviarles mensajes políticos específicos» (p. 113 de las Directrices de prácticas prohibidas); «[u]n sistema de IA que categoriza a las personas activas en una plataforma de redes sociales en función de su supuesta orientación sexual mediante el análisis de los datos biométricos de las fotografías compartidas en dicha plataforma y, sobre esa base, les ofrece publicidad» (p. 114); y «un sistema de categorización biométrica que afirma ser capaz de deducir la orientación religiosa de una persona a partir de sus tatuajes o su rostro» (p. 115). Por su parte, entre los ejemplos que no se englobarían dentro de esta prohibición pueden destacarse los siguientes: «el etiquetado de datos biométricos para evitar los casos en los que un miembro de un colectivo étnico tiene menos posibilidades de ser invitado a una entrevista de trabajo porque el algoritmo ha sido “entrenado” con datos en los que ese colectivo concreto obtiene peores resultados que otros» (p. 115) y «la categorización de los pacientes usando imágenes en función del color de la piel o de los ojos puede ser importante para los diagnósticos médicos, como, por ejemplo, los diagnósticos de cáncer» (p. 116).

8) Sistemas de IA de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del derecho. De acuerdo con el considerando 32.º del RIA, el uso de estos sistemas de IA invade de forma especialmente grave los derechos y las libertades de las personas afectadas, «en la medida en que puede afectar a la vida privada de una gran parte de la población, provocar la sensación de estar bajo una vigilancia constante y disuadir indirectamente a los ciudadanos de ejercer su libertad de reunión y otros derechos fundamentales» [...] «la inmediatez de las consecuencias y las escasas oportunidades para realizar comprobaciones o correcciones adicionales en relación con el uso de sistemas que operan “en tiempo real” acrecientan el riesgo que estos conllevan para los derechos y las libertades de las personas afectadas»²⁷.

²⁶ Entre los datos biométricos que habitualmente se emplearán para realizar este tipo de categorizaciones se encuentran algunos sobre la base del aspecto físico, como la cara, la altura o el color de la piel, los ojos y el pelo (o una combinación de ambos), de los que deducir edad, género y etnia, así como otros rasgos, tales como cicatrices o tatuajes.

²⁷ De acuerdo con el considerando 19.º del RIA, ni las prisiones ni las zonas en que se realizan inspecciones fronterizas son espacios de acceso público. Tampoco los espacios en línea, tales como las salas para chatear, las redes sociales o las plataformas en línea.

Un sistema de identificación biométrica remota es «un sistema de IA destinado a identificar a las personas físicas sin su participación activa y generalmente a distancia comparando sus datos biométricos con los que figuran en una base de datos de referencia» (art. 3.41 del RIA). Las tecnologías de reconocimiento biométrico detectan, recogen y transforman características físicas mensurables (como la distancia entre los ojos y su tamaño, la longitud de la nariz, etc.) o características conductuales (como el modo de andar o la voz) en datos biométricos legibles por máquina. Estos datos están disponibles en diferentes formas: imágenes o plantillas, que constituyen una representación matemática de las características más destacadas de una persona, utilizadas a efectos de reconocimiento. Cuando una autoridad garante del cumplimiento del derecho toma de forma encubierta una fotografía de una persona con un dispositivo móvil y la carga en una base de datos para hacer una búsqueda inmediata, en función de las circunstancias, podrá aplicarse la prohibición establecida en el art. 5.1.h) del RIA.

Sin embargo, el propio art. 5 del RIA establece algunas excepciones a las prohibiciones anteriores, que también pueden relacionarse, en mayor o menor medida, con la neurotecnología. Como excepciones que son, habrán de interpretarse restrictivamente para garantizar un alto nivel de protección de los derechos fundamentales:

1) Se permiten los sistemas de IA para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito que se utilicen para apoyar la valoración humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva²⁸.

2) Se permiten los sistemas de IA que infieran emociones en el centro de trabajo o educativo por motivos médicos o de seguridad²⁹. Por ejemplo, para ayudar a empleados o a estudiantes con autismo y para mejorar la accesibilidad de las personas con discapacidad visual o auditiva. En cambio, el reconocimiento de emociones para evaluar el bienestar de los estudiantes o los empleados, los niveles de ansiedad, de aburrimiento, de motivación o satisfacción laboral o de aprendizaje estaría prohibido³⁰.

²⁸ De acuerdo con las Directrices de la Comisión Europea sobre prácticas prohibidas, a pesar de la exclusión del ámbito de aplicación del RIA de los sistemas de IA con fines de seguridad nacional, las actividades de Europol y de otras agencias de seguridad de la Unión, como Frontex, entran en el ámbito de aplicación del RIA (p. 16 de las citadas Directrices).

²⁹ «[U]n sistema de IA destinado a detectar el desgaste profesional o la depresión en el lugar de trabajo o en centros educativos no estaría contemplado en la excepción y seguiría estando prohibido» (p. 107 de las Directrices de la Comisión Europea sobre prácticas prohibidas).

³⁰ «[A] menos que el elevado nivel de estrés o la falta de concentración suponga un peligro específico como, por ejemplo, si se manejan máquinas peligrosas o se manipulan sustancias químicas peligrosas. En este último caso, el empleador no puede utilizar los datos para otros

Los sistemas de reconocimiento de emociones utilizados en un contexto comercial para dirigirse a los clientes no entran en el ámbito de aplicación de la prohibición del art. 5.1.f) del RIA tanto si se basan en datos biométricos como si no³¹. Tampoco entrarían dentro de la prohibición del art. 5 del RIA los sistemas de IA de reconocimiento de emociones empleados en el análisis del comportamiento de los clientes y la publicidad personalizada y el neuromarketing; en el sector del entretenimiento, por ejemplo, pueden servir para hacer recomendaciones personalizadas o predecir las reacciones a las películas; en la medicina y la asistencia sanitaria, para detectar la depresión o el autismo o prevenir el suicidio.

3) Se permiten los sistemas de IA para el etiquetado o filtrado de conjuntos de datos biométricos adquiridos lícitamente, como imágenes, basado en datos biométricos y la categorización de datos biométricos en el ámbito de la garantía del cumplimiento del derecho. Así, por ejemplo, la clasificación de imágenes en función del color del pelo o del color de ojos, que pueden utilizarse en el ámbito de la garantía del cumplimiento del derecho.

4) Se permiten los sistemas de IA de identificación biométrica remota en tiempo real en espacios de acceso público con fines de garantía del cumplimiento del derecho en la medida en que dicho uso sea estrictamente necesario para:

a) la búsqueda selectiva de víctimas concretas de secuestro, trata de seres humanos o explotación sexual de seres humanos, así como la búsqueda de personas desaparecidas,

b) la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas o de una amenaza real y actual o real y previsible de un atentado terrorista³²,

c) la localización o identificación de una persona sospechosa de haber cometido un delito castigado con una pena o una medida de seguridad privativas de libertad cuya duración máxima sea de al me-

fines, como para evaluar el rendimiento laboral del empleado» (p. 109 de las Directrices de la Comisión Europea sobre prácticas prohibidas).

³¹ En este sentido, en sus Directrices sobre prácticas prohibidas, la Comisión Europea afirma «la prohibición no se aplica a, por ejemplo, los sistemas de IA que permiten el reconocimiento de emociones a partir de la pulsación de tecla o de los mensajes de voz de los clientes (como mensajes de chat, el uso de asistentes de voz virtuales)» (p. 110).

³² A modo de ejemplo, la Comisión Europea señala, en sus Directrices sobre prácticas prohibidas, una perturbación grave y destrucción de infraestructuras críticas (por ejemplo, una central eléctrica, un sistema de suministro de agua o un hospital) pueden suponer una amenaza inminente para la vida o la seguridad física de las personas cuando existe un perjuicio grave derivado de la interrupción del suministro de productos básicos a la población (privación de electricidad o de agua potable durante un largo período de tiempo, en condiciones meteorológicas especialmente cálidas o frías, etc.) (p. 131).

nos cuatro años a fin de llevar a cabo una investigación o un enjuiciamiento penales o de ejecutar una sanción penal³³.

En estos casos, el sistema de IA únicamente puede emplearse para confirmar la identidad de la persona que constituya el objetivo específico³⁴ y su uso debe contar con autorización previa por parte de una autoridad judicial o una autoridad administrativa cuya decisión sea vinculante, que se expedirá previa solicitud motivada y de conformidad con las normas detalladas del derecho nacional. No obstante, en una situación de urgencia debidamente justificada, se podrá empezar a utilizar tal sistema sin autorización siempre que se solicite en un plazo máximo de 24 horas.

Adicionalmente, el art. 2 del RIA excluye expresamente de su ámbito de aplicación los sistemas de IA que «se introduzcan en el mercado, se pongan en servicio o se utilicen, con o sin modificaciones, exclusivamente con fines militares, de defensa o de seguridad nacional, independientemente del tipo de entidad que lleve a cabo estas actividades», lo que incluye el uso por operadores privados a los que los Estados miembros hayan encomendado las tareas relacionadas con esas competencias³⁵.

La segunda categoría de sistemas de IA a la que el legislador presta especial importancia es la referida a los sistemas de alto riesgo, regula-

³³ Son los delitos enunciados en el anexo II del RIA. A saber: terrorismo, trata de seres humanos, explotación sexual de menores y pornografía infantil, tráfico ilícito de estupefacientes o sustancias psicotrópicas, tráfico ilícito de armas, municiones y explosivos, homicidio voluntario, agresión con lesiones graves, tráfico ilícito de órganos o tejidos humanos, tráfico ilícito de materiales nucleares o radiactivos, secuestro, detención ilegal o toma de rehenes, delitos que son competencia de la Corte Penal Internacional, secuestro de aeronaves o buques, violación, delitos contra el medio ambiente, robo organizado o a mano armada, sabotaje y participación en una organización delictiva implicada en uno o varios de los delitos enumerados en esta lista. El considerando 33.º del RIA explica que «[f]ijar ese umbral para la pena o la medida de seguridad privativas de libertad con arreglo al derecho nacional contribuye a garantizar que la infracción sea lo suficientemente grave como para llegar a justificar el uso de sistemas de identificación biométrica remota “en tiempo real”».

³⁴ De conformidad con el art. 5.3 del RIA, no se adoptará ninguna decisión que produzca efectos jurídicos adversos para una persona exclusivamente sobre la base de los resultados de salida del sistema de identificación biométrica remota en tiempo real. Debe existir un agente humano en la verificación e interpretación de los resultados de salida que cuente con la correspondiente formación sobre el funcionamiento del sistema de IA.

³⁵ Así se explica en la p. 14 de las Directrices de la Comisión sobre las prácticas prohibidas, remitiéndose a la jurisprudencia del TJUE a la hora de interpretar el término «seguridad nacional» (STJUE de 6 de octubre de 2020, *La Quadrature du Net* y otros, C-511/18, C-512/18 y C-520/18, ECLI:EU:C:2020:791, apartado 135 y el resto de sentencias citadas). Si una agencia nacional de inteligencia encomienda a una agencia de seguridad nacional o a un operador privado que utilice sistemas de identificación biométrica remota en tiempo real con fines de seguridad nacional (como, por ejemplo, recopilar información), dicho uso quedaría excluido del ámbito de aplicación del Reglamento de Inteligencia Artificial. Los mayores problemas interpretativos se plantearán con los sistemas de IA de doble uso, militar y civil y de garantía del cumplimiento del derecho.

dos en el art. 6 del RIA. Entre ellos, se destacan los siguientes, que pueden estar directa o indirectamente relacionados con la neurotecnología:

1) Los sistemas de IA que funcionan como componentes de seguridad de productos o sistemas de IA que son productos que deban someterse a una evaluación de la conformidad para su introducción en el mercado o puesta en servicio con arreglo a la normativa europea³⁶. Por ejemplo, los robots cada vez más autónomos que se utilizan en las fábricas o con fines de asistencia y atención personal o los sistemas de IA empleados en el sector sanitario para el diagnóstico y apoyo de las decisiones humanas, que pueden tener repercusiones especialmente importantes en la vida y la salud.

2) Los sistemas de IA de biometría, en la medida en que su uso esté permitido por el derecho de la Unión o nacional aplicable, para:

- a) identificación biométrica remota (no en tiempo real)³⁷ o;
- b) reconocimiento de emociones (en ámbitos distintos del laboral y el educativo).

3) Los sistemas de IA destinados a ser utilizados como componentes de seguridad en la gestión y el funcionamiento de las infraestructuras digitales críticas, del tráfico rodado o del suministro de agua, gas, calefacción o electricidad, dado que un fallo o un defecto de funcionamiento puede poner en peligro la vida y la salud de las personas a gran escala y alterar de manera considerable el desarrollo habitual de las actividades sociales y económicas (considerando 55.º del RIA).

4) Sistemas de IA empleados en educación y formación profesional con el fin de:

- a) determinar la admisión de personas físicas;
- b) evaluar los resultados del aprendizaje;
- c) detectar comportamientos prohibidos en exámenes.

De acuerdo con el considerando 55.º del RIA, estos sistemas se consideran de alto riesgo porque pueden «decidir la trayectoria formativa

³⁶ Son los especificados en el anexo I del RIA: máquinas, juguetes, ascensores, equipos y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, equipos de embarcaciones de recreo, instalaciones de transporte por cable, aparatos que queman combustibles gaseosos, productos sanitarios, productos sanitarios para diagnóstico in vitro, automoción y aviación.

³⁷ Quedan excluidos de la clasificación de sistema de IA de alto riesgo los sistemas de IA destinados a ser utilizados con fines de verificación biométrica (como escaneo facial o dactilar) cuya única finalidad sea confirmar que una persona física concreta es la persona que afirma ser, muy habituales en controles de seguridad de aeropuertos y de mecanismos de control de acceso o entrada, así como en sistemas de desbloqueo de dispositivos, tal y como prevé el considerando 54.º del RIA.

y profesional de una persona y, en consecuencia, puede[n] afectar a su capacidad para asegurar su subsistencia. 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, por ejemplo, contra las mujeres, determinados grupos de edad, las personas con discapacidad o las personas de cierto origen racial o étnico o con una determinada orientación sexual».

5) Sistemas de IA utilizados en el empleo y la gestión de los trabajadores para:

a) la contratación o la selección de personas físicas, en particular para analizar y filtrar las solicitudes de empleo y evaluar a los candidatos;

b) la toma de decisiones sobre condiciones de las relaciones laborales, promoción o rescisión de relaciones contractuales o evaluación del rendimiento.

De acuerdo con el considerando 57.º del RIA, estos sistemas de IA «pueden afectar de un modo considerable a las futuras perspectivas laborales, a los medios de subsistencia de dichas personas y a los derechos de los trabajadores», además de perpetuar patrones históricos de discriminación.

6) Sistemas de IA para el acceso y disfrute de servicios públicos o privados esenciales y prestaciones públicas esenciales, tales como los utilizados para:

a) evaluar la admisibilidad de las personas físicas para beneficiarse de servicios y prestaciones esenciales de asistencia pública, incluidos los servicios de asistencia sanitaria;

b) evaluar la solvencia de personas físicas o establecer su calificación crediticia, salvo los sistemas de IA utilizados para detectar fraudes financieros;

c) evaluar riesgos de las personas físicas en el caso de los seguros de vida y de salud;

d) evaluar las llamadas de emergencia realizadas por personas físicas o establecer prioridades en el envío de servicios de primera intervención en situaciones de emergencia, por ejemplo, policía, bomberos y servicios de asistencia médica, y en sistemas de triaje de pacientes.

De acuerdo con el considerando 58.º del RIA, las personas que solicitan estas prestaciones o servicios «se encuentran en una posición de vulnerabilidad respecto de las autoridades responsables». Estos sistemas de IA «podría[n] tener un efecto considerable en los medios de subsistencia de las personas y vulnerar sus derechos fundamentales, como el derecho a la protección social, a la no discriminación, a la dignidad

humana o a la tutela judicial efectiva y, por tanto, deben clasificarse como de alto riesgo» [...] y «adoptan decisiones en situaciones sumamente críticas para la vida y la salud de las personas y de sus bienes».

7) Sistemas de garantía del cumplimiento del derecho, en la medida en que su uso esté permitido por el derecho de la Unión o nacional aplicable, y se realice por las autoridades garantes del cumplimiento del derecho para:

- a) evaluar el riesgo de que una persona física sea víctima de delitos;
- b) polígrafos o herramientas similares;
- c) evaluar la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de delitos;
- d) evaluar rasgos y características de la personalidad o comportamientos delictivos pasados de personas físicas o colectivos;
- e) elaborar perfiles de personas físicas, durante la investigación o el enjuiciamiento de delitos.

En relación con este grupo de sistemas de IA, el considerando 59.º del RIA contiene la siguiente declaración: «No debe ignorarse el impacto del uso de herramientas de IA en los derechos de defensa de los sospechosos, en particular la dificultad para obtener información significativa sobre el funcionamiento de dichos sistemas y la consiguiente dificultad para impugnar sus resultados ante los tribunales, en particular por parte de las personas físicas investigadas».

8) Sistemas de IA sobre migración, asilo y gestión del control fronterizo empleados por las autoridades públicas competentes, en la medida en que su uso esté permitido por el derecho de la Unión o nacional aplicable, para los siguientes fines:

- a) polígrafos o herramientas similares;
- b) evaluación de un riesgo para la seguridad, la salud o de migración irregular, de personas físicas que tengan la intención de entrar en el territorio nacional o hayan entrado;
- c) apoyo en el examen de las solicitudes de asilo, visado o permiso de residencia y las reclamaciones conexas;
- d) identificación de personas físicas³⁸.

La razón de ser de esta clasificación la encontramos en el considerando 60.º del RIA: «Los sistemas de IA empleados en la migración, el asilo y la gestión del control fronterizo afectan a personas que con frecuencia se encuentran en una situación especialmente vulnerable y que dependen del resultado de las actuaciones de las autoridades públicas competentes. Por este motivo, es sumamente importante

³⁸ Con excepción de los sistemas de IA para verificación de documentos de viaje, que no se consideran de alto riesgo.

que los sistemas de IA que se utilicen en estos contextos sean precisos, no discriminatorios y transparentes, a fin de garantizar que se respeten los derechos fundamentales de las personas afectadas»³⁹.

9) Sistemas de IA para la Administración de Justicia y procesos democráticos para:

a) ayudar en la investigación e interpretación de hechos y de la ley;

b) influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto.

Estos sistemas de IA se clasifican como de alto riesgo «dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial» (considerando 61.º del RIA).

Sin embargo, el art.6.3 excluye la subsunción en la categoría de sistemas de alto riesgo cuando cualquiera de los sistemas anteriores —de biometría, de infraestructuras críticas, de educación, de empleo, de migración, etc.— no plantee un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones, como sucede cuando realizan una tarea de procedimiento limitada⁴⁰; mejoran el resultado de una actividad humana previamente realizada⁴¹; detectan patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores sin sustituir la valoración humana previamente realizada ni influir en ella⁴², o cuando simplemente realizan una tarea preparatoria para una evaluación⁴³.

³⁹ Este considerando recuerda la necesidad de que el empleo de los sistemas de IA en la migración, el asilo y la gestión del control fronterizo no pueda ser utilizado como medio para eludir sus obligaciones internacionales en virtud de la Convención de las Naciones Unidas sobre el Estatuto de los Refugiados, ni para infringir el principio de no devolución, ni para negar unas vías jurídicas seguras y efectivas de acceso al territorio de la Unión, incluido el derecho a la protección internacional.

⁴⁰ El considerando 53.º del RIA pone los siguientes ejemplos: «un sistema de IA que clasifique en categorías los documentos recibidos o un sistema de IA que se utilice para detectar duplicados entre un gran número de aplicaciones».

⁴¹ «[P]or ejemplo, a los sistemas de IA destinados a mejorar el lenguaje utilizado en documentos ya redactados, por ejemplo, en lo referente al empleo de un tono profesional o de un registro lingüístico académico o a la adaptación del texto a una determinada comunicación de marca» (considerando 53.º del RIA).

⁴² El considerando 53.º del RIA pone el siguiente ejemplo «aquellos [sistemas de IA] que pueden utilizarse para comprobar a posteriori si un profesor puede haberse desviado de su patrón de calificación determinado, a fin de llamar la atención sobre posibles incoherencias o anomalías».

⁴³ De acuerdo con el considerando 53.º del RIA: «[e]sa condición comprende, entre otras cosas, soluciones inteligentes para la gestión de archivos, lo que incluye funciones diversas

Con el objetivo de mitigar los riesgos que presentan los sistemas de IA de alto riesgo que se introducen en el mercado o se ponen en servicio, y para garantizar un alto nivel de fiabilidad, deben aplicarse a los sistemas de IA de alto riesgo ciertos requisitos obligatorios que tengan en cuenta la finalidad prevista y el contexto del uso del sistema de IA y estén en consonancia con el sistema de gestión de riesgos que debe establecer el proveedor y que habrán de cumplirse durante todo el ciclo de vida del sistema de IA de alto riesgo. Son los requisitos previstos en los arts. 7 a 27 del RIA. A modo de ejemplo, desde las fases de diseño, se debe procurar que los conjuntos de datos empleados en el entrenamiento, en la mayor medida posible, sean completos y estén libres de errores. Asimismo, se exige un sistema de pruebas previas a la comercialización, un proceso continuo de gestión de calidad y una vigilancia poscomercialización de la trazabilidad de los sistemas de IA de alto riesgo, lo que supone disponer de información comprensible sobre el modo en que se han desarrollado y sobre su funcionamiento durante toda su vida útil. Especialmente relevante resulta la obligación de vigilancia humana del funcionamiento del sistema de IA, como factor corrector de las decisiones adoptadas por el sistema automatizado, de modo que se pueda decidir sobre la no utilización del sistema o el descarte o invalidación o reversión del output de dicho sistema, así como la interrupción del funcionamiento de este pulsando el botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura. A ello se añaden importantes obligaciones de documentación y registro.

3. IMPACTO DE LAS NEUROTECNOLOGÍAS Y DE LA IA EN LA FORMACIÓN DEL CONSENTIMIENTO CONTRACTUAL

3.1. El consentimiento en las relaciones contractuales

El consentimiento constituye uno de los elementos esenciales para que exista un contrato en el ámbito del derecho civil, junto con el objeto y la causa. Así se indica expresamente en el art. 1261 CC. El consentimiento es la declaración de voluntad que emite una persona. Para que sea válido, debe formarse de una manera libre, consciente y voluntaria. La IA y las neurotecnologías introducen una complejidad sin precedentes en la concepción y protección del consentimiento, y afectan a estas tres cualidades que debe tener la declaración de voluntad. La IA, con su capacidad para procesar ingentes volúmenes de datos, perso-

tales como la indexación, la búsqueda, el tratamiento de texto y del habla o la vinculación de datos a otras fuentes de datos, o bien los sistemas de IA utilizados para la traducción de los documentos iniciales».

nalizar interacciones y operar con un grado considerable de opacidad, puede influir en las decisiones humanas de manera sutil y masiva, incluso a nivel subconsciente⁴⁴. Las neurotecnologías, por su parte, al permitir la conexión directa entre el cerebro y dispositivos externos, abren un portal para interferir en las funciones mentales y cognitivas, planteando interrogantes fundamentales sobre la autonomía y la integridad mental del individuo.

El consentimiento, en toda relación contractual, debe prestarse libre de vicios. Según el art. 1265 CC, será nulo el consentimiento prestado por error, violencia, intimidación o dolo. Estos vicios afectan a la formación del consentimiento, comprometiendo la eficacia del contrato mismo. El error es la falsa representación de la realidad y, para que vicie el negocio jurídico, debe ser esencial (recaer sobre la sustancia de la cosa objeto del contrato) y excusable, es decir, no debe haberse podido evitar empleando una diligencia media o razonable por parte de quien lo padece (art. 1266 CC). El dolo se produce cuando uno de los contratantes emplea palabras o maquinaciones insidiosas para inducir al otro a celebrar el contrato, sin las cuales no se habría celebrado (art. 1269 CC). Se exige que el dolo sea grave y unilateral (art. 1270.I CC). La violencia tiene lugar cuando para arrancar el consentimiento se emplea una fuerza irresistible (art. 1267.I CC). Finalmente, la intimidación ocurre cuando se inspira en uno de los contratantes un temor racional y fundado de sufrir un mal inminente y grave en su persona o bienes, o en la persona o bienes de su cónyuge, descendientes o ascendientes (art. 1267.II CC).

3.2. Las neurotecnologías y su impacto en la voluntad

Podríamos entender las neurotecnologías como aquellas herramientas que establecen una conexión directa entre el cerebro humano y dispositivos externos, pudiendo interferir con las funciones mentales y cognitivas. La preocupación se agrava porque esta interferencia puede producirse sin el consentimiento o incluso sin el conocimiento de la persona⁴⁵. Al contrario que la IA, no tenemos aún una regulación de las neurotecnologías que evite los riesgos derivados de su utilización.

⁴⁴ Sobre la influencia de la IA en el comportamiento de una persona, *vid.* BELTRÁN DE HEREDIA RUIZ, I., *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*, Aranzadi, 2023.

⁴⁵ Es posible aplicar ciertas neurotecnologías de forma encubierta (no invasivas, fundamentalmente) para condicionar estados cognitivos y emocionales de una persona sin que sea plenamente consciente, lo que tendría graves repercusiones en la validez del consentimiento contractual. Así, podemos plantear los siguientes casos: *a)* estimulación subliminal mediante *neurofeedback* disfrazado de juego (ej.: piénsese en un consumidor que juega a un juego de atención en la web de una financiera, a la par que el sistema le induce un estado de mayor confianza y predisposición al riesgo, facilitando que acepte un préstamo con intereses altos); *b)* publicidad

Se han llevado a cabo ya experimentos y pruebas clínicas que muestran sus efectos en la conducta y en la toma de decisiones⁴⁶. De los mismos, se deriva que las neurotecnologías son capaces de alterar las emociones y la personalidad (DBS, tDCS), inducir cambios en preferencias y juicios (*neurofeedback*, estimulación cortical), generar errores en la comunicación de la voluntad (BCI en ELA) o implantar recuerdos falsos (optogenética en modelos animales). Así, cabe decir que las neurotecnologías pueden afectar negativa o positivamente a la declaración de voluntad. El impacto negativo en la formación del consentimiento ocurrirá cuando lleven a un contratante a celebrar un contrato realmente no querido, pues, a través de ellas, se pueden provocar todo tipo de vicios en el consentimiento, aunque mayormente estaremos ante casos de dolo o error. Téngase en cuenta, no obstante, que el hecho de usar una neurotecnología no significa que se conozcan todos los riesgos y efectos que puede tener sobre una persona, dado que estamos ante una fase muy primaria aún en su aplicación y desarrollo. Por lo que respecta al error, podrían plantearse los siguientes casos subsumibles en él: 1) una aplicación de *neurofeedback* explica contratos con información sesgada o simplificada en exceso, haciendo que el contratante no comprenda los riesgos esenciales⁴⁷; 2) un banco ofrece a sus clientes un sistema de contratación mediante dispositivos BCI (interfaz cerebro-ordenador), y un usuario, agotado tras una larga jornada de trabajo, acepta mediante un gesto mental un préstamo con unas condiciones un tanto perjudiciales. Este consentimiento estaría prestado en condiciones de fatiga mental o sin la suficiente información, por lo que el contrato sería anulable por error esencial.

En cuanto al dolo, este podría tener lugar cuando se usan técnicas de *neuromarketing* subliminal en contratos de consumo. Por ejemplo, una empresa de telefonía organiza una feria promocional donde expone a los asistentes a estímulos auditivos y visuales combinados con una ligera estimulación cerebral no invasiva (tDCS), y los consumidores, tras la experiencia, contratan masivamente planes con cláusulas de perma-

inmersiva en realidad virtual con estímulos neuronales (ej.: en una feria tecnológica, un cliente prueba un visor de realidad virtual que promociona hipotecas a través de una experiencia inmersiva; durante la demostración, estímulos neuronales le generan una sensación de seguridad y confianza artificial, favoreciendo la firma inmediata); c) interfaces de consumo que captan neurodatos sin informar (ej.: una aplicación de meditación con sensor EEG detecta que el usuario está cansado y es más influenciable, lo que es aprovechado para lanzarle publicidad contractual agresiva); d) implantes o dispositivos médicos usados con una finalidad distinta a la sanitaria (ej.: un paciente con un implante de estimulación cerebral profunda para tratar el párkinson se ve expuesto a ajustes sin su conocimiento, alterándole la impulsividad y la toma de decisiones).

⁴⁶ Se pueden consultar algunos de ellos en: <https://lableni.webs.upv.es/mariano-alcaniz-sa2/>.

⁴⁷ El *neurofeedback* es una técnica de entrenamiento cerebral que usa sensores para medir la actividad eléctrica del cerebro y la traduce en información visual o auditiva en tiempo real. El objetivo es que la persona aprenda a autorregular el funcionamiento de su propio cerebro a través de recuerdos (ej.: premios, recompensas), buscando mejorar aspectos cognitivos.

nencia de veinticuatro meses. Esto se asemeja al dolo, pues la voluntad se ve viciada por una manipulación indirecta. El consentimiento no se ha formado libremente, sino que ha sido formado bajo una alteración intencionada de la percepción. Tales técnicas entrarían dentro del concepto de «maquinaciones insidiosas» empleado en el art. 1269 CC a la hora de definir el dolo.

Lo mismo ocurriría si se tratara de engañar a una persona utilizando *neurocookies*. Las aplicaciones de neurotecnología pueden contener *neurocookies*, que permiten identificar las preferencias de compra de una persona y, de manera insidiosa, implantar otras nuevas. Un contrato de compraventa o de suscripción de un servicio que se celebrara debido a una preferencia que no es del consumidor, sino que ha sido implantada por esta tecnología sin su consentimiento, sería un caso de dolo, ya que se le estaría engañando para que adquiriera algo que no deseaba genuinamente.

Y, en tercer lugar, cabría plantearse la concurrencia de violencia o de intimidación, tal como sucedería con una neuroestimulación para provocar ansiedad o malestar hasta que la persona acepte las condiciones contractuales. Podría pensarse en una empresa de comercio electrónico que distribuye a clientes premium cascos EEG gratuitos que registran reacciones cerebrales ante publicidad. Con estos datos, les envía ofertas diseñadas para activar su sistema de recompensa neuronal y potencia la compra compulsiva de productos de lujo. Se genera aquí una especie de coacción interna, asimilable a la violencia psicológica. Por tanto, el contrato celebrado bajo técnicas de inducción emocional mediante neurodatos podría considerarse viciado por una forma de violencia psíquica.

A pesar de todo lo señalado, las neurotecnologías también pueden tener un efecto positivo en el consentimiento contractual, en la medida en que pueden coadyuvar a su formación. En este sentido, podrían utilizarse: para llevar a cabo la evaluación de la capacidad cognitiva en mayores o en personas con discapacidad, dado que permiten comprobar que la persona comprende el objeto y las consecuencias del contrato, reforzando así la validez del consentimiento; para verificar la comprensión contractual —a través de determinados sistemas, se puede confirmar si un consumidor realmente comprende cláusulas clave (ej.: intereses, vencimientos, riesgos), lo que evitaría el riesgo de error—; para detectar el estrés o la coacción en la firma de contratos —ciertos dispositivos permiten medir parámetros fisiológicos, como el estrés o la ansiedad, y podrían indicar si la persona está sometida a coacción o presión indebida, sirviendo como garantía preventiva contra vicios como la violencia o la intimidación—.

3.3. La IA y su influencia en la formación del consentimiento

Igualmente, la IA también puede afectar a la formación del consentimiento. Se han llevado a cabo múltiples experimentos con IA aplicados a la toma de decisiones humanas en el campo de la psicología, que demuestran cómo las personas aceptan, consienten o se adhieren a un contrato bajo la influencia de esta tecnología⁴⁸. La IA puede ser más persuasiva que los humanos, llegando a manipular decisiones, pues tiene la capacidad de acceder al «yo inconsciente» y subliminalmente condicionar el comportamiento⁴⁹. Además, la IA puede reconocer emociones, modificarlas o provocarlas durante las interacciones, lo que incrementa la vulnerabilidad de las personas.

Como se ha señalado en el apartado anterior de este estudio, determinadas técnicas de IA están prohibidas por el RIA por constituir «prácticas de riesgo inaceptable». Así, según el art. 5.1.a) del Reglamento, queda prohibida:

la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas debidamente manipuladoras o engañosas con el objetivo o con el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas.

Y, conforme al artículo 5.1.b) del Reglamento, está prohibida:

la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que explote alguna de las vulnerabilidades de una persona física o de un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra.

⁴⁸ Sobre algunos de ellos, *vid.*: <https://issues.org/ai-ethics-human-experimentation-taylor/>.

⁴⁹ Llega a señalar Pertíñez Vilchez que cuanto mayor sea la intervención del sistema de IA empleado en la formación del contrato, menor será la incidencia de la voluntad humana y más en entredicho estará la consideración misma del contrato como producto del acuerdo de dos voluntades. *Vid.* PERTÍÑEZ VILCHEZ, F., «La contratación entre sistemas de inteligencia artificial», en ÁLVAREZ LATA, N. (coord.), *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*, Aranzadi, 2024, p. 103.

Conforme a estas disposiciones, estarían prohibidas las siguientes prácticas. En primer lugar, el uso de técnicas subliminales para condicionar la aceptación de una cláusula (ej.: una empresa financiera diseña un asistente de IA en una aplicación que, cuando el usuario va a contratar un préstamo en un momento de fatiga o distracción, porque lleva mucho tiempo de uso del dispositivo, introduce estímulos subliminales auditivos o visuales que inducen confianza, urgencia o disminuyen la percepción de riesgo). En segundo lugar, la explotación de vulnerabilidad por edad o discapacidad (ej.: una aseguradora de vida ofrece pólizas a personas mayores a través de un sistema de IA que adapta el discurso, interfaz y presentación para una persona mayor con cognición disminuida, como letra más grande o voz calmada, pero al mismo tiempo omite o minimiza riesgos o costes). Y, en tercer lugar, el empleo de reconocimiento emocional domiciliario para inducir decisiones (ej.: un asistente doméstico inteligente equipado con cámara y micrófono recoge expresiones de frustración, tristeza o preocupación del usuario, analiza esas señales emocionales con IA y, tras detectar vulnerabilidad emocional, envía ofertas de servicios financieros o productos complejos, como seguros o créditos justo en ese momento). En todos estos casos, además de contravenir las prohibiciones establecidas en el art. 5.1 RIA, se genera un consentimiento que no es plenamente libre ni plenamente informado.

Sin embargo, el RIA deja un espacio gris para muchas otras técnicas de IA que no estarían prohibidas, pero que pueden afectar negativamente a la formación del consentimiento contractual. Se trataría de técnicas legales bajo el prisma del Reglamento, pero problemáticas desde el punto de vista del derecho civil y de la protección del individuo, al derivar en vicios del consentimiento. En el marco de la IA, son el error y el dolo los que más se plantean en la práctica cuando se explotan vulnerabilidades de un sujeto, existen sesgos en el algoritmo o se omite información relevante⁵⁰.

Por lo que se refiere al error, y como se ha indicado *supra*, este debe ser esencial y excusable (art. 1266 CC). La diligencia razonable que se espera de un usuario se ve comprometida por la inherente opacidad y la creciente sofisticación de los sistemas de IA. Esta tecnología puede generar un error de manera sistemática, no por una acción humana, sino por la propia arquitectura del algoritmo y los datos con los que ha sido entrenada. Los sesgos algorítmicos son resultados distorsionados que surgen de sesgos humanos presentes en el momento del entrenamiento o en el momento del diseño del algoritmo mismo. A ello se une que, en los entornos mediados por IA, la omisión de in-

⁵⁰ Así lo considera también NAVAS NAVARRO, S., «Deshumanizando al contrato. Una mirada desde la inteligencia artificial», en ÁLVAREZ LATA, N. (coord.), *Derecho de contratos...*, op. cit., p. 67.

formación es frecuente y relevante. De ahí que quepa concluir que la intervención de la IA puede volver problemático el requisito de la excusabilidad en el error. La diligencia razonable de un usuario no puede prever ni corregir errores que son inherentes al diseño de un algoritmo.

Cabe plantear los siguientes casos de error: *a)* un *chatbot* opaco que oculta comisiones o riesgos en un contrato de seguro, generando error sobre sus condiciones esenciales; *b)* los asistentes virtuales contractuales sesgados (ej.: un *chatbot* de IA explica a un cliente las condiciones de un contrato, pero lo hace con un sesgo intencionado, resaltando beneficios y minimizando riesgos, sin llegar a ocultarlos, lo que le hace creer que tiene un conocimiento correcto de las condiciones, pero su percepción está distorsionada); *c)* el empleo de técnicas de gamificación en la contratación (ej.: utilizar IA para convertir el proceso contractual en una experiencia lúdica, haciendo que el consumidor se vea empujado a aceptar rápidamente, sin leer con detenimiento las condiciones).

En cuanto al dolo, la maquinación a la que se refiere el art. 1269 CC puede ser perpetrada por algoritmos diseñados específicamente para influir en el comportamiento del usuario de manera engañosa. Entre los supuestos que podrían ampararse en este vicio del consentimiento se encuentran los *deepfakes* que inducen la contratación. Se trata de vídeos generados por IA en los que una celebridad recomienda contratar un producto financiero con unas condiciones un tanto peculiares y que el consumidor contrata convencido por la supuesta recomendación. El art. 50.2 RIA obliga a que esos vídeos estén marcados en un formato legible para las máquinas y que sea posible detectar que han sido generados o manipulados de manera artificial, pero no indica de qué manera deben marcarlos, por lo que dicha indicación puede aparecer de tal forma que el consumidor no sea siquiera consciente de que está. Otro ejemplo, en este caso de dolo por omisión, es el de un *robot-advisor* (asesor financiero basado en IA), que recomienda a un cliente un producto de inversión; el algoritmo, tras analizar el perfil del cliente, sabe que el producto es inadecuado y que tiene una alta probabilidad de generar pérdidas, pero la IA está programada para priorizar la venta de productos de su propia entidad y omite deliberadamente esta información crucial, presentando únicamente las ventajas genéricas. Todas estas prácticas pueden ser calificadas como dolo, ya que buscan intencionadamente inducir un contrato que el usuario no habría celebrado bajo condiciones de información y libertad plenas. Si bien la IA carece de ánimo de engaño en el sentido humano, esta tecnología es el resultado de un diseño humano con una finalidad clara: optimizar el comportamiento del usuario para obtener un beneficio comercial. Esto significa que el ánimo de engaño debe atribuirse no al algoritmo en sí, sino a los programadores, diseñadores o empresas que lo despliegan.

Especial referencia merecen los denominados *dark patterns* (patrones oscuros). Constituyen «trampas digitales» que condicionan nuestras decisiones diarias, incitándonos a tomar decisiones de compra que pueden no ser las más convenientes. Ejemplos de estas técnicas son: los temporizadores de cuenta atrás falsos (crean una sensación artificial de urgencia para presionar al consumidor a comprar, aunque la oferta siga siendo válida una vez que el tiempo expira); las jerarquías falsas (el diseño de la interfaz guía sutilmente al consumidor hacia opciones específicas, como suscripciones o alternativas de entrega más costosas) o la ocultación de información esencial (datos cruciales sobre un producto o servicio se presentan con fuentes muy pequeñas, colores de bajo contraste o en ubicaciones poco visibles, impidiendo que el consumidor tome una decisión informada). El art. 25.1 del Reglamento 2022/2065, sobre servicios digitales, prohíbe los patrones oscuros, al señalar que «los prestadores de plataformas en línea no diseñarán, organizarán ni gestionarán sus interfaces en línea de manera que engañen o manipulen a los destinatarios del servicio o de manera que distorsionen u obstaculicen sustancialmente de otro modo la capacidad de los destinatarios de su servicio de tomar decisiones libres e informadas». Quedan incluidos dentro de esta prohibición, independientemente de que se utilice tecnología basada en IA o no: a) los diseños que inducen decisiones que el usuario no habría tomado libremente (ej.: botones que resaltan una opción más perjudicial para el usuario, flujos de compra confusos o estructuras que dificultan rechazar una opción, como *cookies*); b) las interfaces que generan presión indebida o manipulación emocional (ej.: temporizadores falsos, mensajes de urgencia no reales, colores o sonidos diseñados para provocar ansiedad o aceptar más rápido); c) la organización de la interfaz de tal manera que oculte información relevante (ej.: colocar la opción más protectora en zonas escondidas del menú); d) la repetición coercitiva o *nagging* (ej.: la insistencia repetida para aceptar una suscripción)⁵¹; e) la obstrucción por complejidad (ej.: un *clic* para suscribirse, pero doce *clicks* para cancelar la suscripción). Esta prohibición del art. 25.1 del Reglamento 2022/2065 se aplica a plataformas en línea, como redes sociales, *marketplaces* o aplicaciones de comercio electrónico.

Finalmente, en cuanto a la violencia, no parece que pueda hablarse en el ámbito de la IA de una fuerza irresistible en el sentido del art. 1267 CC. Más común puede ser la intimidación. La capacidad de la IA para generar una coacción moral o psicológica es considerable. Tácticas de manipulación emocional, como infundir miedo, *gaslighting* (distorsionar la realidad para minar la autoconfianza) o generar culpa

⁵¹ El *nagging* es un término inglés que significa reclamar de manera muy persistente y constante a otra persona.

pueden ser automatizadas y ejecutadas por sistemas de IA, lo que debería llevar a su prohibición conforme al art. 5.1.b) RIA. Si estas técnicas son lo suficientemente graves y persistentes, podrían constituir una forma de intimidación al coartar la libertad de decisión del individuo. Se hablaría, en estos casos, de intimidación cognitiva y emocional. Piénsese, por ejemplo, en un sistema de IA que gestiona el cobro de deudas de una empresa; el algoritmo accede a una gran cantidad de datos personales del deudor, incluyendo su red de contactos y su actividad profesional; al detectar un impago, el sistema envía automáticamente un mensaje al deudor amenazando con comunicar su situación de morosidad a su empleador o publicar su deuda en un grupo de una red social donde están sus amigos, a menos que acepte de inmediato un acuerdo de refinanciación. El problema del precepto señalado es que está repleto de conceptos jurídicos indeterminados que pueden poner en duda su eficacia (ej.: «alteración sustancial del comportamiento de una persona», «provocación de perjuicios considerables a la persona»).

En definitiva, la IA puede afectar negativamente al consentimiento contractual de varias formas: induciendo error (falsas percepciones, ocultación de riesgos) o creando dolo (engaños intencionales mediante *deepfakes* o *chatbots* opacos), propiciando asimetrías de poder e información que rompen la igualdad contractual. No obstante, la IA también puede servir como garantía en su formación, reforzando su validez y reduciendo los litigios, en términos muy similares a los que se han señalado con ocasión de las neurotecnologías. Así, la IA puede utilizarse para verificar la comprensión de contratos financieros (ej.: un banco incorpora un sistema de IA que, antes de firmar productos complejos, analiza las respuestas del cliente a preguntas clave y evalúa si ha comprendido riesgos y costes, bloqueando la celebración del contrato si el cliente no demuestra una comprensión mínima). También puede usarse como apoyo a personas con discapacidad (ej.: un hombre con discapacidad cognitiva leve desea arrendar una vivienda y la IA le explica las cláusulas en un lenguaje sencillo, verifica si las entiende y genera un informe que se adjunta al contrato). En el ámbito notarial, puede utilizarse para detectar posibles coacciones (ej.: un notario emplea un sistema de IA que analiza expresiones faciales, el tono de la voz y patrones de comportamiento de la persona que comparece para otorgar una donación, alertando de un posible caso de intimidación por parte de los familiares presentes, suspendiendo el notario la firma). Igualmente, puede resultar útil para simplificar contratos complejos (ej.: una aseguradora utiliza un sistema de IA que genera resúmenes personalizados de pólizas, traduciendo el lenguaje técnico a términos comprensibles para el cliente, y verifica que este ha leído y comprendido estos resúmenes antes de la aceptación), para registrar el consentimiento en la *blockchain* (ej.: en una plataforma de contratación electrónica, la IA verifica la identidad biométrica y registra cada fase

de la manifestación de la voluntad) o para asegurar el cumplimiento normativo (ej.: un modelo de IA que identifica patrones abusivos en contratos y bloquea su inclusión).

3.4. Neurotecnologías e IA en el ámbito del derecho de consumo

La normativa sobre derecho de consumo constituye un buen refuerzo de la posición del consumidor destinatario de ambos tipos de tecnologías. Dado que la inmensa mayoría de los supuestos que se pueden plantear se enmarcan en una relación entre un empresario y un consumidor, resultan aplicables todas las garantías previstas en este ámbito. El derecho de consumo protege también la autonomía de la voluntad del consumidor, lo que significa que su consentimiento debe ser libre, informado, no manipulado ni distorsionado cognitivamente por técnicas agresivas o desleales.

A este respecto, no se puede perder de vista la normativa sobre competencia desleal. Así, el art. 8 de la Directiva 2005/29/CE, relativa a las prácticas comerciales desleales de las empresas en sus relaciones con consumidores, considera práctica agresiva toda práctica comercial que merme o pueda mermar de forma importante, mediante acoso, coacción o influencia indebida, la libertad de elección o conducta del consumidor medio respecto al producto y le haga o pueda hacerle tomar una decisión sobre una transacción que, de otra manera, no hubiera tomado. En el mismo sentido se recoge en el art. 8.1 Ley 3/1991, de competencia desleal. Estos preceptos prohíben, por tanto, cualquier práctica que distorsione el comportamiento económico del consumidor mediante presiones indebidas, manipulación emocional o técnicas que distorsionen su capacidad de decisión. Aunque no menciona tecnologías concretas, su redacción abierta permitiría incluir: *a)* las interfaces neuronales que induzcan estados de ansiedad o urgencia antes de contratar; *b)* la IA persuasiva adaptativa que personalice mensajes justo en momentos de vulnerabilidad detectada; *c)* el uso de datos biométricos o neurométricos para prever y explotar sesgos individuales.

Por tanto, cualquier técnica neurotecnológica o algorítmica que altere de forma relevante el estado emocional o cognitivo del consumidor puede considerarse que vicia el consentimiento no solamente desde la perspectiva civil clásica, sino también desde la deslealtad comercial o desde la prohibición regulatoria. Se podrían plantear los siguientes supuestos:

- 1) *Chatbots* persuasivos en la contratación digital. Los asistentes conversacionales que imitan el estilo humano generan mayor disposición a aceptar ofertas que normalmente los usuarios habrían cuestionado.

2) La técnica del *dynamic pricing* (fijación dinámica de precios por IA). En determinadas plataformas, la IA ajusta precios en tiempo real según la disposición del consumidor a pagar, de manera que los consumidores aceptan precios distintos para el mismo producto, sin saberlo, influidos por sus hábitos de compra y el perfil digital.

3) IA emocional en publicidad personalizada. Determinados sistemas de reconocimiento facial por IA pueden detectar expresiones emocionales (ej.: tristeza, ansiedad), adaptando los anuncios en tiempo real, de tal manera que quienes presenten estados emocionales vulnerables serán más proclives a aceptar ofertas de forma inmediata, incluso desventajosas.

4) Recomendadores en plataformas de *streaming* y compras, capaces de generar tal confianza extrema en el sistema y, en ocasiones, tal sensación de inevitabilidad, que llevan a los usuarios a aceptar automáticamente términos de suscripción para acceder a contenido recomendado, aceptando obligaciones de pago o cláusulas adicionales poco beneficiosas para el sujeto.

5) Una aseguradora que usa un algoritmo que detecta estrés o nerviosismo a través del teléfono móvil (detección de voz, micropausas, cambios sutiles de tono) y muestra opciones más caras y mensajes del tipo «protege ya a tu familia». No se trata de IA prohibida, pues no necesariamente es subliminal, pero sí puede ser entendida como una práctica agresiva, pues aprovecha una vulnerabilidad emocional momentánea.

6) Una empresa de *e-commerce* que coloca un sistema de análisis de pupila, parpadeo y microexpresiones en probadores virtuales de gafas. La IA detecta cuándo el usuario está más fatigado o menos atento y lanza ofertas *flash* con temporizadores agresivos en ese momento.

En definitiva, el empleo de IA o neurotecnologías en consumidores es lícito cuando: 1) no afecte de manera significativa a la libertad de elección, por no utilizar técnicas de diseño, color, tamaño o disposición engañosa; 2) no manipule estados emocionales, cognitivos o conductuales; 3) no explote vulnerabilidades derivadas de la edad o de discapacidad; 4) sea transparente y el consumidor comprenda cómo funciona; 5) el consentimiento se obtenga de manera libre, específica e informada; 6) la finalidad sea legítima y proporcional. Aunque no exista una regulación específica de las neurotecnologías, el derecho de consumo ya ofrece un marco suficiente para prohibir o cuestionar usos que alteren cognitivamente la decisión de contratar. La información previa y la transparencia son esenciales. Cualquier mecanismo (tecnológico o no) que manipule la voluntad del consumidor causa problemas de validez contractual y puede ser considerado ilícito, incluso aunque la tecnología en sí no esté prohibida.

3.5. Replanteamiento del consentimiento

A la vista de las consideraciones anteriores, convendría proceder a la reconceptualización del consentimiento informado, que se torna fundamental en estos ámbitos (IA y neurotecnologías). La capacidad de estas tecnologías para acceder a datos cerebrales o influir en el subconsciente plantea una pregunta sin precedentes: ¿cómo puede ser informado un consentimiento si el individuo no es consciente de los procesos mentales que están siendo monitorizados o manipulados? Esto exige una nueva categoría de consentimiento, quizá un consentimiento neurológicamente informado, que no solo abarque la información explícita, sino también salvaguardas para los procesos cognitivos subconscientes.

Así, cabría hablar de dos tipos de consciencia: la formal (el sujeto sabe que está celebrando un contrato porque percibe el acto externo) y la material (el sujeto comprende las implicaciones jurídicas y económicas de esa decisión, y que esta se adopta en condiciones libres, no manipuladas). En los casos tratados, lo que se ve afectado no es la consciencia formal, sino la material, pues la persona ignora que su estado interno ha sido condicionado. En los usos positivos de la IA y de las neurotecnologías, la manipulación existe, pero se hace con transparencia y para facilitar una decisión más libre e informada (ej.: *neurofeedback* para reducir la ansiedad o mejorar la atención antes de leer un contrato). El sujeto consiente expresa y deliberadamente el uso de la tecnología con esa finalidad («quiero que me ayuden a concentrarme para comprender mejor el contrato»). Se considera un refuerzo de la autonomía. En los usos negativos (ej.: IA que explota estados emocionales, *neurofeedback* que induce euforia para aceptar inversiones), la manipulación es oculta y desleal (la persona no sabe que está siendo predispuesta para celebrar el contrato en un sentido concreto). Se altera su estado mental de forma incompatible con la libertad contractual porque el sujeto, aunque sí tiene consciencia de que está celebrando un contrato, cree que decide por sí mismo cuando en realidad se le dirige, no es plenamente consciente del mecanismo técnico que actúa sobre su mente. En estos casos, el consentimiento está manipulado (el sujeto sabe que contrata, pero no percibe que su estado mental ha sido alterado para decidir de cierta manera). La diferencia entre ambos tipos de usos está en la finalidad y la transparencia: si es informado y en beneficio de la persona, podemos aceptarlo como refuerzo de la autonomía; si es oculto y en su perjuicio, puede generar un vicio del consentimiento y no tanto la ausencia total de este elemento esencial de los contratos.

Por otro lado, los vicios del consentimiento tradicionales operan en el momento de la celebración del contrato. Sin embargo, los neu-

roderechos actúan como una salvaguarda previa a la manifestación de la voluntad contractual. Si la privacidad mental es infringida o la identidad personal es alterada por neurotecnologías o por IA antes de que se forme el consentimiento, entonces cualquier consentimiento subsiguiente deberá considerarse inherentemente defectuoso. Esto implica que la infracción de un neuroderecho debería ser una causa suficiente para presumir el vicio del consentimiento en un contrato, incluso si no se cumplieran todos los requisitos específicos de ese vicio. Esta perspectiva eleva la protección a un nivel fundamental, donde la integridad mental es un prerequisite para la validez de cualquier acto jurídico.

Asimismo, surge la necesidad de revisar los vicios del consentimiento. Hay vicios que ya consideran la vulnerabilidad a la hora de determinar su concurrencia (ej.: la edad en la intimidación). Sin embargo, ahora se introduce la vulnerabilidad sistémica cuando existen sesgos en el algoritmo o se omite información relevante. Los sistemas pueden identificar y explotar debilidades cognitivas o emocionales de manera masiva y sutil. Debemos plantearnos si la explotación de la vulnerabilidad puede tratarse como un vicio del consentimiento *per se* o como un factor agravante que facilite la demostración del error o el dolo.

En lo que respecta a la carga de la prueba en los vicios del consentimiento, esta recae tradicionalmente en la parte que los alega. Sin embargo, esta cuestión se vuelve ineficaz ante la opacidad de la IA y la dificultad de las neurotecnologías. La imposibilidad de acceder al funcionamiento interno de estos sistemas hace que la carga de la prueba tradicional sea prácticamente insuperable para el afectado. Dada la asimetría de información y el efecto de caja negra de la IA e incluso de las neurotecnologías, exigir al afectado demostrar el vicio del consentimiento es, en la práctica, negarle el acceso a la justicia. Esto lleva a la conclusión de que, en casos de consentimiento viciado por IA o neurotecnologías, la carga de la prueba debería invertirse.

Se plantea igualmente un tema no menor, como es la protección de la confianza del tercero que contrata con una persona cuyo consentimiento no ha sido prestado libre de vicios porque se ha utilizado una neurotecnología o IA. Cuando se manipula cognitivamente a una persona para que preste un consentimiento contractual, ella misma cree que consiente libremente y el tercero cree de buena fe que el consentimiento fue válido. A diferencia de la embriaguez, la manipulación con IA o neurotecnologías no es visible externamente. Además, el tercero no tiene forma de detectar anomalías cognitivas. Se produce, así, un conflicto entre la protección de la víctima, que tiene derecho a no quedar vinculada por un consentimiento manipulado, y la protección del tercero de buena fe (seguridad jurídica y tráfico económico). A la hora de resolver estos casos, debemos distinguir según el tipo de vicio.

Cuando estemos ante una IA o neurotecnología empleada para generar coacción en un sujeto para que celebre el contrato con un tercero, este dispondrá de una acción indemnizatoria contra el manipulador. El art. 1268 CC contempla la posibilidad de que la intimidación provenga de un tercero. En este caso se resarce el interés contractual negativo, es decir, los daños derivados de la invalidez del contrato, debiéndose colocar al tercero de buena fe en la misma posición que si el contrato no se hubiera celebrado⁵². Sin embargo, cuando el vicio ante el que nos encontremos sea dolo o error, primará la seguridad del tráfico y la protección de la confianza generada en el tercero. Si este no ha intervenido en la manipulación de la víctima, se le deberá proteger, manteniendo la validez del contrato y obligando al manipulador a indemnizar a la víctima por haber afectado su autonomía. Así será siempre y cuando el tercero y el manipulador no se hayan confabulado contra el contratante víctima de las tecnologías.

Asimismo, cabe recordar que, en nuestro ordenamiento, se prevén actos jurídicos de carácter personalísimo respecto de los cuales no cabe acudir a la representación o asistencia de terceros [ej.: el consentimiento matrimonial (art. 56 CC) o el otorgamiento de testamento (arts. 663 a 665 CC)]. Cuando se trate de este tipo de actos, no es válido el consentimiento que se haya «fabricado» artificialmente. El CC exige una voluntad humana propia, entendida como un querer auténtico. La IA o las neurotecnologías reconstruyen intenciones, interpretan la actividad cerebral y regulan estados cognitivos. En definitiva, estarían sustituyendo la capacidad de una persona, no apoyándola, invadiendo de esta forma el núcleo personalísimo del acto. Si el acto no es personalísimo, no debería haber problema para que el apoyo que necesita esa persona lo desarrolle la IA o una neurotecnología. La clave está en que el apoyo no sustituya la voluntad, sino que facilite la comprensión del contrato, incremente la capacidad de deliberación, mejore la comunicación, provea de información adaptada y reduzca las barreras cognitivas (ej.: asistentes de IA que permiten adaptar la información jurídica a lenguaje fácil, interfaces que ayudan a recordar opciones). Ambas tecnologías deberían permitirse siempre que la decisión final sea efectivamente de la persona, respetándose la voluntad, los deseos y las preferencias de la persona que tenga discapacidad, que es lo que pretende proteger el sistema de apoyos de los arts. 249 y siguientes CC.

Finalmente, conviene plantearse cuáles serán las consecuencias derivadas del uso de la IA o de neurotecnologías para formar el consentimiento. Dependiendo de cómo se haya hecho uso de estas tecnologías, el contrato podrá ser nulo o anulable. Será nulo, por aplicación

⁵² MARÍN LÓPEZ, M. J., «Requisitos esenciales del contrato. Elementos accidentales del contrato», en BERCOVITZ RODRÍGUEZ-CANO, R. (coord.), *Tratado de contratos*, Tirant lo Blanch, Valencia, 2024, p. 699.

del art. 6.3 CC, en tres casos: 1) cuando el uso prohibido sea un elemento esencial del contrato (ej.: se contrata a alguien para crear un sistema de IA prohibido para perfilar a consumidores como parte de la operación contractual); 2) cuando sea parte de la causa o del objeto del contrato (ej.: se contrata un servicio cuyo funcionamiento es en sí mismo ilícito porque emplea *dark patterns* expresamente prohibidos por el Reglamento de servicios digitales); 3) cuando el contrato se celebre gracias a un mecanismo prohibido que sustituye o suplanta enteramente la voluntad de las partes (ej.: el consentimiento se obtiene mediante manipulación conductual significativa, estimulación neurocognitiva que altera emociones o la conducta, o se induce a la otra parte a un estado mental artificial). En el segundo caso, el contrato es nulo *per se* porque vulnera normas prohibitivas. En el tercer caso, porque el consentimiento es jurídicamente inexistente. Así podría ocurrir, por ejemplo, en una plataforma de compraventa que activara automáticamente contratos inteligentes cuando detectara mediante sensores EEG que el usuario experimenta entusiasmo frente a un producto; el usuario no confirma conscientemente la compra, pero el contrato se ejecuta (el consentimiento se reduce a una mera reacción neurofisiológica). Si la tecnología elimina por completo la voluntad del sujeto, cabría concluir que no concurrirá el consentimiento y, por consiguiente, el contrato celebrado será radicalmente nulo, por faltar uno de los elementos esenciales del contrato (art. 1261 CC). En cambio, cabría predicar la anulabilidad cuando la tecnología prohibida vicia el consentimiento de una de las partes, sin que el propio contrato celebrado por esta sea, en sí mismo, ilícito. La consecuencia del art. 6.3 CC se aplica a los contratos que tengan un contenido ilícito o que persigan una finalidad ilícita, cosa que no ocurre en estos casos. Por tanto, cuando las neurotecnologías y la IA se utilicen con la finalidad de presionar o engañar a la otra parte (ej.: un empresario que utiliza IA prohibida para predecir vulnerabilidades de un consumidor y explotarlas a su favor; uso de *dark patterns* para empujar a un consumidor a contratar una suscripción), se producirá entonces un vicio en el consentimiento.

4. CONCLUSIONES

El auge de la IA planteó de inmediato el problema de la compatibilidad entre los usos de esta tecnología y el respeto de los derechos fundamentales de las personas. En esa preocupación ética, la Unión Europea se decantó rápidamente por garantizar, dentro de su ámbito territorial de competencias, una IA centrada en el ser humano. Lograr este postulado únicamente parecía posible a través de un instrumento normativo. El instrumento elegido fue el Reglamento, norma de alcance general, obligatoria en todos sus elementos y directamente aplicable

en cada Estado miembro⁵³. Las neurotecnologías, por el contrario, carecen a día de hoy de una regulación propia.

Por otra parte, ha de tenerse en cuenta que la IA y la neurotecnología no son compartimentos estancos. Antes bien, las sinergias entre ambas existen y van en aumento. La IA es el soporte de numerosas neurotecnologías, aunque no constituya un elemento imprescindible o inherente a las mismas. En consecuencia, parece legítimo plantearse si existe ya una tutela —ciertamente incipiente y parcial— de los neuroderechos a través del RIA. En este sentido, la premisa fundamental para que proceda la aplicación del RIA es la existencia de un sistema de IA que responda a la definición presente en su art. 3.1). El mismo alude a:

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.

La definición pone el acento en la capacidad de «inferencia», siendo este el parámetro que distingue el sistema de IA de los sistemas de *software* o de los planteamientos de programación tradicionales y más sencillos. Esta capacidad de inferencia se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones. Dice al respecto el considerando 12 que:

las técnicas que permiten la inferencia al construir un sistema de IA incluyen estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse.

El mismo considerando añade que: «los sistemas de IA pueden utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado)».

Numerosas son, en este sentido, las neurotecnologías que caben dentro de esta definición. La detección automatizada de patologías neurológicas a través de los datos inferidos de imágenes del cerebro, la interpretación en tiempo real de datos neuronales obtenidos a través de interfaces cerebro-ordenador o las tecnologías de neuromodulación adaptativa se basan en métodos de aprendizaje automático y, por ello, pueden entrar en el ámbito de aplicación del RIA.

⁵³ Art. 288 TFUE.

A partir de esta constatación, se observa que el RIA ofrece efectivamente una tutela de la actividad cerebral del ser humano prohibiendo radicalmente algunas prácticas de IA o bien considerándolas de alto riesgo, según los casos. En concreto, el art. 5 RIA enumera una serie de prácticas prohibidas, entre las cuales cabe destacar las «técnicas subliminales que trasciendan la conciencia de una persona» así como las «técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado». Hasta qué punto este art. 5 RIA está relacionado con las neurotecnologías es una cuestión abierta⁵⁴, si bien el propio considerando 29 RIA hace una referencia explícita a un dispositivo propio del campo de las neurotecnologías como son las interfaces cerebro-máquina. Por su parte, las Directrices publicadas por la Comisión parecen incluir la estimulación cerebral en la categoría de técnicas subliminales y, por ende, como práctica de IA prohibida⁵⁵. En este sentido, resulta llamativa la mención expresa de las neurotecnologías en el contexto de las técnicas subliminales, presente en el párrafo 66 de las mencionadas Directrices («El rápido desarrollo de la IA y las tecnologías relacionadas, como el análisis de macrodatos, las neurotecnologías, las interfaces cerebro-máquina y la realidad virtual, aumenta el riesgo de manipulación subliminal sofisticada y su capacidad para influir eficazmente en el comportamiento humano de forma subconsciente. La IA también puede extenderse a las interfaces cerebro-máquina emergentes y a técnicas avanzadas como la manipulación del sueño y el espionaje neuronal»).

En un plano distinto, el apartado f) del mismo art. 5 RIA se refiere al uso de sistemas de IA para inferir las «emociones» de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad. Fuera de estos contextos, el art. 6 RIA cataloga los sistemas de reconocimiento de emociones como sistemas de IA de alto riesgo. El considerando 18 define el «sistema de reconocimiento de emociones» como:

un sistema de IA destinado a distinguir o deducir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos. El concepto se refiere a emociones o intenciones como la felicidad,

⁵⁴ BUBLITZ, J. C. (2025), «Brain Stimulation May Be a Subliminal Technique Under the European Union's Artificial Intelligence Act», *European Journal of Neuroscience*, 2025; 61:e70115; BUBLITZ, J. C., MOLNÁR-GÁBOR, F. y SOEKADAR, S. R. (2024), «Implications of the novel EU AI Act for neurotechnologies», *Neuron*, núm. 112, Septiembre 25, 2024, pp. 3013-3016, en p. 3013.

⁵⁵ BUBLITZ, J. C. (2025), «Brain Stimulation May Be a Subliminal Technique Under the European Union's Artificial Intelligence Act», *op. cit.*, p. 3014.

la tristeza, la indignación, la sorpresa, el asco, el apuro, el entusiasmo, la vergüenza, el desprecio, la satisfacción y la diversión. No incluye los estados físicos, como el dolor o el cansancio, como, por ejemplo, los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes. Tampoco incluye la mera detección de expresiones, gestos o movimientos que resulten obvios, salvo que se utilicen para distinguir o deducir emociones.

Continúa el considerando 44 expresando la existencia de una gran preocupación respecto a los sistemas de IA que procuran detectar o deducir las emociones. Dice así 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 de las personas afectadas», insistiendo además en el «carácter intrusivo de estos sistemas.

En cualquier caso, deja fuera de la prohibición a los «sistemas de IA introducidos en el mercado estrictamente con fines médicos o de seguridad, como los sistemas destinados a un uso terapéutico».

Ciertamente, el RIA habla de «emociones» e «intenciones». No así de «pensamientos», que es lo que muchos usos de la neurotecnología (y la IA a ella asociada) podrían llegar a controlar (así se desprende de algunos experimentos que se han realizado, en los que se ha podido leer la mente de personas con determinadas deficiencias, como puede ser una parálisis cerebral⁵⁶). Ambas categorías (emociones e intenciones) remiten a información relacionada con la actividad cerebral. La emoción parece derivar del orden de lo irracional (esto es, aquello que en principio la persona no puede controlar, sino que experimenta en función de su sensibilidad). Por el contrario, los pensamientos se situarían en el orden de lo racional. Pero, si se observa, el RIA también menciona las «intenciones», parámetro que implica ya, en nuestra opinión, una actividad volitiva, más cercana al ámbito del pensamiento (ej.: la intención de votar a un partido político o la intención de decantarse por la compra de un producto de una marca determinada, y no el de un competidor). A través de su caracterización como sistemas de IA de alto riesgo, o de prácticas de IA prohibidas (en este caso, cuando tienen lugar en lugares de trabajo o en centros educativos), se estaría garantizando ya una suerte de derecho a la privacidad mental.

Cabría, en todo caso, preguntarse si el hecho de limitar la prohibición a las prácticas que tienen lugar en esos lugares de trabajo o en los centros educativos seguiría estando justificada cuando, más allá de las emociones e intenciones, la aplicación neurotecnológica per-

⁵⁶ Vid. <https://www.cell.com/cell/fulltext/S0092-8674%2825%2900681-6> [04/12/2024].

mitiera descifrar los meros pensamientos (experimentos de este tipo ya se han realizado con cierto éxito en personas afectadas por parálisis cerebral⁵⁷, o sin ella⁵⁸). Cuestión distinta es que, en aquellos procesos que implican la lectura de la mente en tiempo real, la prohibición pudiera quedar levantada en los casos en que las neurotecnologías se empleen con fines estrictamente médicos o de seguridad, siempre acompañándolos de las garantías adecuadas.

En definitiva, y a pesar de que el Reglamento de IA no ignora por completo las neurotecnologías, se insiste en la necesidad de dotar a estas herramientas tecnológicas de un marco normativo propio, general, transversal y de alcance como mínimo europeo. Sobre este marco, en un plano de derecho privado, cabría avanzar en cuestiones más concretas como es la protección del consentimiento —noción central de la disciplina— frente al uso de las neurotecnologías y de la IA. También en materia de consentimiento, un terreno a explorar estaría relacionado con las medidas de apoyo a las personas con discapacidad en el ejercicio de su capacidad jurídica ante el uso de la IA y de las neurotecnologías, cuestión particularmente delicada⁵⁹.

5. BIBLIOGRAFÍA

- ANDORNO, R. (2025), «Analysis of the existing European human rights framework concerning the human rights issues raised by neurotechnologies and their applications», Derechos Humanos del Consejo de Europa, accesible en <https://rm.coe.int/steering-committee-for-human-rights-in-the-fields-of-biomedicine-and-h/4880290460> [4/12/2025].
- BELTRÁN DE HEREDIA RUIZ, I. (2023), *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*. Pamplona: Aranzadi.
- BUBLITZ, J. C. (2025), «Brain Stimulation May Be a Subliminal Technique Under the European Union's Artificial Intelligence Act», *European Journal of Neuroscience*, 2025; 61:e70115.
- BUBLITZ, J. C., MOLNÁR-GÁBOR, F. y SOEKADAR, S.R (2024), «Implications of the novel EU AI Act for neurotechnologies», *Neuron*, núm. 112, Septiembre 25, 2024, pp. 3013-3016, en p. 3013.
- Directrices de la Comisión relativas a la definición de sistema de inteligencia artificial establecida en el RIA, C(2025) 5053 final. 29 de julio de 2025. Disponibles en: <https://digital-strategy.ec.europa.eu/es/library/>

⁵⁷ KUNZ, E.M. *et al.*, «Inner speech in motor cortex and implications for speech neuroprostheses», *Cell* 188, 4658-4673, August 21, 2025 (disponible en: <https://www.cell.com/cell/fulltext/S0092-8674%2825%2900681-6>) [04/12/2024].

⁵⁸ <https://www.uts.edu.au/news/2023/12/portable-non-invasive-mind-reading-ai-turns-thoughts-text>.

⁵⁹ *Vid. la Convención sobre los derechos de las personas con discapacidad*, Nueva York, 13 de diciembre de 2006.

- commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application* [4/12/2025].
- Directrices de la Comisión sobre las prácticas de inteligencia artificial prohibidas que se establecen en el Reglamento (UE) 2024/1689 (Reglamento de Inteligencia Artificial). C(2025) 5052 final. 29 de julio de 2025. Disponible en: <https://digital-strategy.ec.europa.eu/es/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act> [4/12/2025].
- KUNZ, E. M. *et al.*, «Inner speech in motor cortex and implications for speech neuroprostheses», *Cell* 188, 4658-4673, August 21, 2025 (disponible en: <https://www.cell.com/cell/fulltext/S0092-8674%2825%2900681-6>) [04/12/2024].
- MAGEE, P., IENCA, M. y FARAHANY, N. (2024), «Beyond neural data: Cognitive biometrics and mental privacy». *Neuron*, 112(18).
- MUHL, E. y ANDORNO, R. (2023), «Neurosurveillance in the workplace: do employers have the right to monitor employees' minds?». *Frontiers in Human Dynamics*. Disponible en: <https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2023.1245619/full> [4/12/2025].
- MARÍN LÓPEZ, M. J. (2024), «Requisitos esenciales del contrato. Elementos accidentales del contrato». En BERCOVITZ RODRÍGUEZ-CANO (coord.). *Tratado de contratos*. Valencia: Tirant lo Blanch.
- MUHL, E. y ANDORNO, R. (2023), «Neurosurveillance in the workplace: do employers have the right to monitor employees' minds?». *Frontiers in Human Dynamics*. Disponible en: <https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2023.1245619/full> [4/12/2025].
- NAVAS NAVARRO, S. (2024), «Deshumanizando al contrato. Una mirada desde la inteligencia artificial». En ÁLVAREZ LATA, N. (coord.). *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*. Pamplona: Aranzadi.
- PERTÍÑEZ VÍLCHEZ, F. (2024), «La contratación entre sistemas de inteligencia artificial». En ÁLVAREZ LATA, N. (coord.). *Derecho de contratos, responsabilidad extracontractual e inteligencia artificial*. Pamplona: Aranzadi.

