

Algoritmos y condicionamiento por debajo del nivel consciente: un análisis crítico de la propuesta de Ley de Inteligencia Artificial de la Unión Europea

Algorithms and Subconscious Conditioning: A Critical Analysis of the European Union's Artificial Intelligence Act

Ignasi BELTRAN DE HEREDIA RUIZ*

RESUMEN: La Ley de Inteligencia Artificial que está elaborando la Unión Europea contiene un precepto que prohíbe el uso de esta tecnología para condicionar el comportamiento por debajo del nivel consciente. Este estudio, a partir de una aproximación neurocientífica, critica las propuestas que se están barajando porque, al habilitar importantes espacios de intromisión, describen una profunda amenaza a la libertad y la identidad de las personas. Por este motivo, propone su prohibición absoluta, sin condicionamientos.

PALABRAS CLAVE: algoritmo; derechos del yo inconsciente de las personas; inteligencia artificial; condicionamiento por debajo del nivel consciente; aproximación neurocientífica.

ABSTRACT: The Artificial Intelligence Act Bill being developed

* Profesor Agregado, acreditado a Catedrático, de Derecho del Trabajo y de la Seguridad Social. Universitat Oberta de Catalunya (UOC), España. Autor del libro: *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*. Aranzadi, 2023. Contacto: <ibeltran_de_heredia@uoc.edu>. Más información en: <www.ignasibeltran.com>. Fecha de recepción: 19/06/2023. Fecha de aprobación: 07/07/2023.

by the European Union includes a provision that prohibits the use of this technology for subconscious conditioning. This study, based on a neuroscientific approach, criticizes the proposals that are being considered because, by enabling significant intrusions, they pose a profound threat to citizen's freedom and identity. Therefore, it proposes an absolute prohibition without any conditions.

KEYWORDS: Algorithm; Artificial Intelligence; Rights of the unconscious mind; Subconscious Conditioning; neuroscientific approach.

I. INTRODUCCIÓN: EL PENSAMIENTO INTERNO HA DEJADO DE SER UN REDUCTO INEXPUGNABLE

El cerebro humano describe una de las mayores fronteras del conocimiento científico. No obstante, el avance tecnológico está llegando a cotas, hace unas décadas, inimaginables. La decodificación de los procesos mentales más profundos parece que es un objetivo, en cierta medida, al alcance. Aunque todavía quedan importantes incógnitas por desvelar (y, quizás, dada su insondabilidad, muchas permanecerán tras el velo del conocimiento), lo cierto es que el avance prodigioso de interfaces mente-ordenador (o neurotecnologías) ya permiten acceder (con diversos niveles de intrusión y, por el momento, con fines esencialmente terapéuticos) a estratos neuronales profundos¹.

La iniciativa de los conocidos como *neuroderechos*, que lidera y promueve la *Neurorights Foundation*², aspira, precisamente, al reconocimiento de una nueva batería de mecanismos de protección dirigida a combatir los desafíos anudados a la proliferación de estos avances técnicos³. A la espera del resultado del análisis

¹ Un par de ejemplos, pueden ayudar a visualizar el estado de la técnica (y, probablemente, su potencial). *BrainPort* es un dispositivo – no mayor que un sello de correos – que, colocado en la lengua, permite *ver* a personas con discapacidad visual a través de los impulsos eléctricos que emite una pequeña cámara adosada a unas gafas de sol. O, el chaleco vibratorio VEST que, al convertir los sonidos en vibraciones, permite a las personas sordas oír. Y, recientemente, ha aparecido una nueva versión en forma de pulsera.

² Una síntesis del recorrido doctrinal de esta iniciativa en BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 146-156.

³ En apretada síntesis, estos serían: (1) el derecho a la identidad mental o personal, o al ‘sentido de sí mismo’; (2) el derecho al ‘libre albedrío’ (o derecho a la agencia, o la libertad de pensamiento); (3) el derecho a la privacidad mental; (4) el derecho al acceso equitativo a un aumento de la capacidad mental; y (5) el derecho a la protección contra sesgos y discriminaciones. Por

que, al cierre de este trabajo, está llevando por el Comité Asesor a instancias del Consejo de Derechos Humanos de la ONU sobre la suficiencia (o no) del marco regulatorio internacional vigente⁴, no

otra parte, es interesante reparar que, en marzo de 2023, el Comité Jurídico Interamericano, siguiendo la senda de declaraciones anteriores, ha aprobado la *Declaración de Principios interamericanos en materia de Neurociencias, Neurotecnologías y Derechos Humanos*. Consultado en: <https://www.oas.org/es/sla/cji/docs/CJI-RES_281_CII-O-23_corr1_ESP.pdf>, recogiendo 10 principios (que sólo enumero): Identidad, autonomía y privacidad de la actividad neuronal; Protección de los Derechos Humanos desde el diseño de las neurotecnologías; Los datos neuronales como datos personales sensibles; Consentimiento expreso e informado de los datos neuronales; Igualdad, No Discriminación y Acceso equitativo a las neurotecnologías; Aplicación terapéutica exclusiva respecto al aumento de las capacidades cognitivas; Integridad neurocognitiva; Gobernanza transparente de las neurotecnologías; Supervisión y fiscalización de las neurotecnologías; y Acceso a la tutela efectiva y acceso a remedios asociados al desarrollo y uso de las neurotecnologías. Otras iniciativas precedentes dignas de consideración, como recoge RECHE TELLO, Nuria, “Nuevos derechos frente a la neurotecnología: la experiencia Chilena”, *Revista de Derecho Político*, núm. 112, 2021, pp. 420 y 421, serían las siguientes: la *Recommendation on Responsible Innovation in Neurotechnology*, 11 de diciembre 2019, de la Organización para la Cooperación y el Desarrollo Económicos. OCDE. Consultado en: <<https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm>>, que “invita tanto a los poderes públicos como a empresas y desarrolladores a anticiparse y afrontar los déficits éticos, jurídicos y sociales que puedan surgir”. Por otra parte, debe mencionarse el *Report of the International Bioethics Committee of UNESCO (IBC) on the ethical issues of neurotechnology*, del 15 de diciembre 2021, del Comité de Bioética de la UNESCO. Consultado en: <<https://unesdoc.unesco.org/ark:/48223/pf0000378724>>.

⁴ La Resolución aprobada por el Consejo de Derechos Humanos el 6 de octubre de 2022 A/HRC/RES/51/3: Consultado en: <<https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/525/04/PDF/G2252504.pdf?OpenElement>>, explícitamente afirma “Teniendo presente que la neurotecnología permite conectar el cerebro humano directamente a las redes digitales mediante disposi-

conviene olvidar que este extraordinario y acelerado proceso corre paralelo a otra revolución tecnológica en ciernes.

El ser humano está en el umbral de un proceso de delegación generalizada en la toma de decisiones a las máquinas. Es lo que se ha venido a denominar *la Gran Desconexión*⁵. Aunque el ámbito empresarial ha estado históricamente a la vanguardia (por ejemplo, la dirección y el control de los trabajadores ha sido asumida desde hace mucho tiempo por *asistentes informacionales*), todo indica que estamos a las puertas de su expansión acelerada a todos los ámbitos de la sociedad. Esto nos aproxima a un punto de inflexión por dos motivos: en primer lugar, porque de forma paulatina, “la inteligencia se está desconectando de la conciencia”; y, derivadamente, los “humanos corren el peligro de perder su valor”⁶.

Esta delegación creciente está permitiendo la expansión de algoritmos extractivos. Es decir, *fórmulas* que proceden a una recopilación implacable de múltiples y heterogéneas dimensiones de la vida profesional y cotidiana de las personas. El efecto combinado de este *big data* y la expansión sin precedentes de esta estadística computacional (todavía lejos de la “inteligencia” en el sentido humano del término) está posibilitando una cartografía humana

tivos y procedimientos que pueden utilizarse, entre otras cosas, para acceder al sistema nervioso humano, monitorizarlo y manipularlo. Reconociendo que la neurotecnología podría ser prometedora para la salud humana y la innovación, pero que, al mismo tiempo, el desarrollo continuado de algunas de sus aplicaciones podría plantear una serie de cuestiones éticas, jurídicas y sociales que deben ser abordadas, también en términos de derechos humanos. Teniendo presente que no se comprenden plenamente los efectos que tiene, las oportunidades que crea y los retos que plantea la neurotecnología en lo relativo a la promoción, la protección y el disfrute de los derechos humanos, y que es necesario seguir analizándolos de forma coherente, integral, inclusiva y exhaustiva a fin de aprovechar al máximo todo el potencial que tiene la neurotecnología para apoyar el progreso humano y el desarrollo para todas las personas”.

⁵ HARARI, Yuval-N., *Homo Deus*, Debate, 2017, p. 341.

⁶ HARARI, *op. cit.*, p. 341.

exhaustiva. Nuestras intenciones, emociones y estados de ánimo, como si estuvieran tras una pared de cristal, ya pueden ser *leídos*. Este acceso al *patio trasero* neuronal describe un nuevo desafío porque estas máquinas están acumulando capacidad para agujinear la mente, acceder al yo inconsciente y condicionar subliminalmente el comportamiento. En definitiva, el pensamiento ha dejado de ser un reducto inexpugnable; y es obvio que, la más que previsible expansión de la realidad virtual inmersiva (o metaverso), podría agudizar (superlativamente) esta amenaza.

El escenario recién descrito plantea numerosos conflictos y dilemas. En términos estrictamente jurídicos, al menos, obliga a plantearse si el concepto de *acto propio y voluntario* queda en entredicho. Y, de forma derivada, si también impacta en el de *culpa y responsabilidad*. Si, realmente, estamos hablando de herramientas *efectivas* que trascienden la conciencia es posible que necesitemos un marco jurídico que (paradójicamente) nos dé amparo *frente a nosotros mismos* y, obviamente también, frente a quienes quieran aprovecharse de nuestros actos por debajo de dicho umbral.

La Unión Europea (UE), en el marco de su estrategia digital, está tramitando una Ley de Inteligencia Artificial (IA) que, entre otros propósitos, aspira a contener los efectos de esta amenaza⁷. Esta norma pretende establecer obligaciones para proveedores y usuarios en función del nivel de riesgo de la IA. A la espera de la redacción definitiva (que se espera que quede aprobada a finales de 2023), el propósito de este breve ensayo es diseccionar el borrador de un aspecto muy específico de esta propuesta: esto es, la prohibición de comercialización de una IA dirigida a condicionar por

⁷ Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión. Un rápido repaso sobre la evolución y documentación previa aquí (hasta junio 2023): Consultado en: <<https://www.europarl.europa.eu/news/es/headlines/society/20230601S-T093804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial>>.

debajo del nivel consciente el comportamiento de las personas y que se recoge en el apartado a) de su art. 5.1.

El análisis crítico de esta regla (y, en particular, de sus limitaciones en términos de protección de los derechos y libertades de los ciudadanos) requiere que, con carácter propedéutico, se expongan (de forma separada) dos aspectos importantes: en primer lugar, algunas nociones básicas sobre el cerebro y, en particular, sobre la distinción entre la mente consciente y la mente inconsciente; y, en segundo lugar, cómo los algoritmos extractivos están contribuyendo a identificar patrones por debajo del nivel consciente.

II. SOBRE EL CEREBRO Y LA PREDOMINANCIA DE LA MENTE INCONSCIENTE

El cerebro es un órgano esquivo y misterioso. A diferencia del corazón, con sus latidos, o el estómago, cuando se mueve o gruñe, no emite señal sensorial alguna de su existencia. Es una caja oscura y silenciosa. Sólo tiene “conocimiento” del mundo exterior de forma indirecta, a través de los impulsos que recibe de los sentidos, debidamente “formateados” a un “lenguaje común”. La corteza cerebral o neocórtex es la principal causante de la inteligencia del *homo sapiens*⁸. Aunque no se sabe con exactitud, se estima que la corteza podría tener hasta 30.000 millones de neuronas. Estas células nerviosas conforman la identidad como ser de cada persona. Además,

⁸ HAWKINS, Jeff y BLASKESLEE, Sandra, *Sobre la inteligencia*, Espasa, 2005, pp. 57 y 58. Como apuntan estos autores, su grosor es el equivalente aproximado a unas 6 tarjetas de visita colocadas una encima de la otra (unos 2 milímetros) y, si pudiéramos “plancharla”, con unas dimensiones similares a una servilleta grande. Está colocada “encima” del cerebro “viejo” y su aparición es, en términos evolutivos, relativamente “reciente”. Ocupa tres cuartas partes del volumen cerebral y sus pliegos característicos serían la resultante del escaso espacio del cráneo, como si tratáramos de poner una servilleta en una copa de vino grande.

el neocórtex es el responsable de la inteligencia humana y también el causante de la conciencia⁹.

El comportamiento humano es la resultante de procesos deliberados y controlados, pero también de otros que están por debajo del nivel consciente. Desde hace más de un siglo, sabemos que las personas poseen una vida inconsciente rica y activa que convive en paralelo a nuestros pensamientos y sentimientos conscientes, ejerciendo “un poderoso efecto a través de unos medios que solo ahora comenzamos a poder medir con cierto grado de precisión”¹⁰.

La relación entre mente consciente e inconsciente es difícil de descifrar¹¹. No obstante, los estudios neurocientíficos han evidenciado que el *yo consciente* “es el fragmento más pequeño de lo que ocurre en el cerebro”¹². No se halla en el centro de la acción, “sino más bien en un borde lejano, y no oye más que susurros de la actividad”¹³. El cerebro actúa (literalmente) por su cuenta. El procesamiento automático e inconsciente es tan importante que, para que se hagan una idea, “casi todas sus operaciones quedan fuera de la acreditación de seguridad de la mente consciente. El *yo* simplemente no tiene derecho de entrada”. En términos cuantitativos (aproximados) el dato es más elocuente si cabe: se estima que el 95% de nuestra actividad cerebral rige por debajo del nivel consciente¹⁴. El inconsciente de la mente (incluyendo las habilidades innatas como las que hemos aprendido y automatizado u otras,

⁹ Es el responsable de crear patrones que provienen de las fibras nerviosas sensoriales y de la capacidad de predicción que los mismos habilitan. Todos los mamíferos tienen corteza cerebral, aunque, en el caso de los humanos, es significativamente más grande en comparación al tamaño de nuestro cuerpo. HAWKINS y BLASKESLEE, *op. cit.*, pp. 76-79 y 230.

¹⁰ MLODINOW, Leonard, *Subliminal*, Crítica, 2013, p. 5.

¹¹ Extensamente al respecto, BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 47-52.

¹² EAGLEMAN, David, *Incógnito*, 3ª ed., Anagrama, 2022, pp. 12 y 13.

¹³ EAGLEMAN, *op. cit.*, p. 18.

¹⁴ MLODINOW, *op. cit.*, p. 46.

fruto de la práctica reiterada) es como un iceberg: la mayor parte de su masa queda oculta. Operamos conforme a subrutinas “escritas en un lenguaje de programación no descifrado de proteínas y sustancias neuroquímicas”, que merodean imperceptiblemente en nuestro cerebro para activarse cada vez que se las reclama¹⁵. El yo consciente es sólo la parte más pequeña de la actividad.

El problema es que no nos percatamos de este extraordinario torrente de actividad neuronal porque la interacción entre la mente consciente y el comportamiento inconsciente es sumamente compleja. Las fuerzas que guían nuestras vidas están muy lejos del alcance de nuestra conciencia o de nuestro control. Por este motivo, somos “ignorantes de las raíces de muchas de nuestras decisiones”¹⁶. No podemos observar cómo decidimos algo y, cuando lo hacemos, “nuestra decisión ya ha aflorado a nuestra conciencia, venida no sabemos de dónde. No hemos presenciado su *formación*; hemos presenciado su *llegada*”¹⁷. Apenas podemos captar una ínfima fracción del bullicio de los procesos mentales que dictan nuestras decisiones. De ahí que tengamos la extraña sensación de que el *cuartel de operaciones* no está donde nosotros estamos, sino en alguna parte más profunda e inaccesible. En cambio, “nada supera a la mente consciente a la hora de convencerse de que está al mando de la situación”¹⁸. Como apunta Daniel Kahneman, nuestra mente consciente (lo que denomina “Sistema 2”) es el personaje secundario que se cree protagonista de la película¹⁹.

A la luz de este dato, repárese que el Derecho ha estado predominantemente pendiente, *grosso modo*, del 5% consciente. El avance tecnológico y, en particular, la proliferación de algoritmos extractivos y el *big data*, sugieren que, quizás, tendríamos que aspi-

¹⁵ EAGLEMAN, *op. cit.*, p. 27.

¹⁶ JOHNSON-LAIRD, Philip N., *El ordenador y la mente*, Paidós, 1990, p. 343.

¹⁷ DENNETT, Daniel C., *La libertad de acción*, Gedisa, 2018, pp. 95-96.

¹⁸ EAGLEMAN, David, *El cerebro*, Anagrama, 2017, p. 92.

¹⁹ *Pensar rápido, pensar despacio*, Debolsillo, 2012, p. 26.

rar a dar protección al 95% restante. Especialmente porque, si estas *fórmulas* (como se expondrá a continuación) están contribuyendo a decodificarlo aceleradamente, dada su predominancia, la esencia de la esencia del ser humano podría quedar muy expuesta y amenazada.

III. ALGORITMOS EXTRACTIVOS UBICUOS Y PSICOMETRÍA HUMANA

La *psicometría*, también conocida como *psicografía*, es –como apunta Byung-Chul Han– un “procedimiento basado en datos para obtener un perfil de personalidad”. El *smartphone* (aunque no sólo) juega, hoy en día, un papel absolutamente medular en este proceso porque, como *informante* muy eficiente que vigila permanentemente a su usuario, lo convierte en “un dispositivo de registro psicométrico que alimentamos con datos día a día, incluso cada hora”. Por este motivo (y, especialmente, desde el instante que se ha convertido en nuestra nueva *extremidad artificial*), “puede utilizarse para calcular con precisión la personalidad del usuario”²⁰

El rastreo intensivo a través de las redes sociales y aplicaciones para móviles y ordenadores tiene, prioritariamente, un propósito publicitario y comercial. Mas, colonizado el *mundo virtual* (con el acopio de un volumen de datos descomunal), el *mundo real*, a través del internet de las cosas y de la *inteligencia ubicua*, se ha convertido en el nuevo espacio de minería humana por explotar. Aunque algunos datos, a través de la eufemística *personalización*, contribuyen a mejorar el servicio prestado (y así nuestra experiencia), el verdadero propósito es la apropiación despiadada de los datos *en la sombra* (los metadatos o *migas digitales*) que revelan aspectos íntimos e imperceptibles de nuestra conducta²¹. Esto explica, al

²⁰ HAN, Byung-Chul, *Infocracia*, Taurus, 2022, pp. 35-39.

²¹ ZUBOFF Shoshana, *La era del capitalismo de la vigilancia*, Paidós, 2020, pp. 96, 256 y ss.

albur del denominado *solucionismo tecnológico*²², la expansión de las *SMART cities* y la creciente inversión en productos domésticos *SMART* (TV, neveras, aspiradoras, lavadoras, bombillas, lámparas, altavoces, alarmas, vigilabebés, termostatos, teteras, asistentes domésticos activados por voz, etc.), *wereables* (pendientes, relojes, pulseras, juguetes, ropa, etc.), vehículos autónomos hiperconectados, drones o la realidad virtual aumentada.

El acopio de estos datos, la experiencia humana, es la materia prima de un nuevo orden económico: el *capitalismo de la vigilancia*. El propósito es aprovechar esta información para predecir y modificar el comportamiento. El concepto, concebido por Shoshana Zuboff,²³ es probablemente el que mejor describe el impulso empresarial predominante, pero también la tendencia que podría marcar el futuro sociopolítico. Hoy empleamos la IA, la ciencia y las matemáticas para extraer las más puras formas a partir del desordenado mundo material. Como en la metafísica platónica, la tecnología consigue extraer una forma (el modelo) del mundo de las apariencias (los datos)²⁴.

La mente inconsciente también está siendo asediada. Como apunta Oliver Sacks, aunque es “imposible predecir el comportamiento individual de cada elemento dentro de un sistema dinámico complejo (por ejemplo, las neuronas individuales o los grupos neuronales de la corteza visual primaria), se pueden distinguir patrones a un nivel superior utilizando modelos matemáticos y análisis por ordenador. Existen ‘comportamientos universales’ que representan la manera en que se autoorganizan los sistemas dinámicos no lineales. Estos suelen adquirir la forma de patrones reiterativos complejos en el espacio y el tiempo”²⁵.

²² MOROZOV, Evgeny, *La locura del solucionismo tecnológico*, Katz, 2015, p. 21.

²³ ZUBOFF, Shoshana, *op. cit.*, p. 6.

²⁴ COECKELBERGH, Mark, *Ética de la inteligencia artificial*, Cátedra, 2021, p. 32.

²⁵ SACKS, Oliver, *El río de la conciencia*, Anagrama, 2019, pp. 179-180.

Así pues, los algoritmos extractivos estarían en condiciones de levantar el velo del conocimiento y acceder (de forma externa e indirecta) a ciertos procesos neuronales profundos. En efecto, la cartografía humana no sólo se circunscribe al entorno exterior, nuestro universo interior (personalidad, emociones complejas, etc.) también está siendo escrutado. Para ilustrar, a través de lo que se conoce como *análisis de sentimientos* (y que se emplea en marketing y publicidad, lingüística, ciencia política, sociología y muchos otros ámbitos), es posible introducir una cadena de textos (por ejemplo, los tweets de un usuario o de miles) en un ordenador y un software especializado (como el programa *Vader*²⁶) puede evaluar si el contenido emocional es positivo o negativo, así como la intensidad de los sentimientos²⁷. Las expresiones faciales y la voz también están siendo objeto de parametrización. En el caso de las primeras, se espera medir las emociones²⁸. En el caso de la voz, el botón, no es tanto lo que decimos (que también), sino cómo lo hacemos: la estructura de nuestra habla e ideas, el vocabulario, la pronunciación, la entonación, la cadencia, la inflexión, el dialecto²⁹. Todo ello, para tratar de tener acceso a nuestras emociones, personalidad, estado de ánimo y/o nivel socio-cultural.

En todo caso, en lo que al reconocimiento de las emociones se refiere, conviene advertir que, en virtud de la denominada “teoría de la emoción construida”, el margen de error (de momento) podría ser extraordinario. En efecto, este tipo de tecnología (a través del tratamiento de datos biométricos) parte de un principio neuronal que podría ser totalmente erróneo (y, por consiguiente, estar

²⁶ *Valence Aware Dictionary and sEntiment Reasoner*.

²⁷ MŁODINOW, Leonard, *Emocional*, Crítica, 2023, pp. 236 y 237. El software *TextBlob*, por otra parte, permite obtener una métrica de polaridad y subjetividad.

²⁸ Una recopilación de los diversos instrumentos empleados para ello en MUÑOZ RUIZ, Ana Belén, *Biometría y sistemas automatizados de reconocimiento de emociones*, Tirant Lo Blanch, 2023, pp. 113-118; y 158 y ss.

²⁹ ZUBOFF, *op. cit.*, p. 355.

haciendo mediciones sumamente desviadas). Según la visión clásica de la emoción, como expone Lisa Feldman Barrett³⁰, tenemos muchos circuitos emocionales en el cerebro y cada uno da lugar a un conjunto característico de cambios, es decir, a una “huella dactilar”. De modo que, según esta concepción, estímulos externos a las personas son los responsables de activar las “neuronas de la ira” y que nuestra “presión sanguínea aumente, frunzamos el ceño, gritemos y sintamos un arranque de ira”. Desde esta visión clásica, las emociones son productos de la evolución y, dado que fueron positivos para la supervivencia, ahora son un componente fijo de nuestra naturaleza biológica. Por este motivo, durante mucho tiempo se ha entendido que son universales: “personas de todas las edades y culturas, y de cualquier parte del mundo, deberían experimentar la tristeza más o menos como nosotros, igual que hicieron nuestros antepasados homínidos que deambulaban por la sabana africana hace un millón de años”³¹.

Diversos estudios neurocientíficos, no obstante, han evidenciado el cerebro no tiene una *huella dactilar* física, única y universal para cada emoción, como si fuera un fichero estático *cargado de serie* en nuestro *hardware*, que se activa (*reacciona*) a consecuencia de una percepción externa a nosotros. Son muchas las evidencias que muestran que este patrón universal no existe. Por consiguiente, no son los estímulos externos los que precipitan las emociones de las personas. Las emociones “no son reacciones *al* mundo, son nuestras construcciones *del* mundo”³². De modo que, como evidencian numerosos experimentos con grupos sociales y etnias de todo el mundo, no es posible identificar un patrón emocional a partir de una expresión facial. Como apunta Feldman Barrett³³, “si reunimos todas las pruebas científicas no podemos afirmar, con

³⁰ FELDMAN BARRETT, Lisa, *La vida secreta del cerebro*, Paidós, 2018, p. 13.

³¹ FELDMAN BARRETT, Lisa, *op. cit.*, p. 141.

³² *Ibidem*, p. 13.

³³ FELDMAN BARRETT, *op. cit.*, p. 31.

una certeza razonable, que cada emoción tiene una expresión facial diagnosticada”.

Esto comporta que, si nuestras emociones no son meramente reactivas, conforme a un patrón neuronal predeterminado, sino que son construidas, podemos colegir que, los conceptos emocionales son un producto cultural, fruto del consenso humano y, por esta razón, variables entre grupos sociales y mutables en el tiempo. Por lo tanto, su parametrización a través de datos biométricos, con el objeto de identificar patrones universales, es mucho más compleja. Lo que implica que estos instrumentos de medición no son, hoy por hoy, fiables. O, dicho de otro modo, en estos momentos, el valor de esta parametrización es escaso o nulo. Están midiendo cosas con una desviación muy superior a la que sería deseable³⁴. Esto

³⁴ Así lo recoge también la enmienda del Parlamento al considerando 26 quater: “Existe una gran preocupación respecto a la base científica de los sistemas de IA que procuran detectar las emociones, los rasgos físicos o psicológicos, como las expresiones faciales, los movimientos, la frecuencia cardíaca o la voz. Las emociones o sus formas de expresión y su percepción varían de forma considerable entre culturas y situaciones, e incluso en una misma persona. Algunas de las deficiencias principales de estas tecnologías son la fiabilidad limitada (las categorías de emociones no se expresan de forma coherente a través de un conjunto común de movimientos físicos o psicológicos ni se asocian de forma inequívoca a estos), la falta de especificidad (las expresiones físicas o psicológicas no se corresponden totalmente con las categorías de emociones) y la limitada posibilidad de generalizar (los efectos del contexto y la cultura no se tienen debidamente en cuenta). Los problemas de fiabilidad y, por consiguiente, los principales riesgos de abuso, pueden surgir especialmente cuando se implanta el sistema en situaciones de la vida real relacionadas con la aplicación de la ley, la gestión de fronteras, el lugar de trabajo y las instituciones educativas. Por tanto, debe prohibirse la introducción en el mercado, la puesta en servicio y el uso de sistemas de IA diseñados para utilizarse en dichos contextos a fin de detectar el estado emocional de las personas físicas”. Esta valoración es la que ha llevado a que, en la propuesta del Parlamento (ver considerando 33 bis), estos sistemas sean, o bien, prohibidos, o bien, calificados

no significa que las emociones no puedan ser parametrizadas, sino que, probablemente, exigen una sofisticación y granularidad muy superior a la que estos métodos atesoran en la actualidad.

Al margen de estas cautelas (absolutamente necesarias, a la luz del estado de la técnica actual), lo paradójico de todo este proceso es que, aparentemente, la intromisión de estos “ladrones de sombras”³⁵ es consentida. Como si se tratara de un *Gran Hermano* amable, ha conseguido la complicidad extrema de sus súbditos, basada en el goce, el placer y la certeza. De tal modo que “los miembros de este nuevo panóptico participan activamente y gustosos en su propia vigilancia”; y lo hacen porque, debidamente estimulados, así colman una necesidad interna³⁶ (inconscientemente estimulada). Esta despreocupación ilustra una devaluación de la privacidad, quedando relevada a un nivel inferior al que le es propio. Por este motivo, apenas se exige su amparo. Con esta cándida pasividad se subestiman sus efectos, porque no se tiene en cuenta que esta dinámica puede liberar inmensas fuerzas de cambio en contra de los intereses de los ciudadanos que nadie imaginaba ni deseaba.

Téngase en cuenta que esta antropometría, a través de la datificación ubicua y los algoritmos extractivos, está posibilitando la *desencriptación indirecta* de los estratos más recónditos de la mente. Estas *fórmulas*, a base miles de iteraciones y procesos de ensayo-error derivados de la interacción con las personas, están contribuyendo a revelar ciertos patrones profundos de nuestro

de “alto riesgo” y sometidos a las condiciones de aprobación de art. 52.2 (en cambio, en la propuesta inicial de la Comisión estaban sometidos a obligaciones de transparencia del Título IV). Por este motivo, en el listado de sistemas de IA prohibidos, se propone un nuevo apartado iv, al art. 5.1.d: “La introducción el mercado, la puesta en servicio o la utilización de sistemas de IA para inferir las emociones de una persona física en los ámbitos de la aplicación de la ley y la gestión de fronteras, en lugares de trabajo y en centros educativos”.

³⁵ En una versión digital sofisticada del personaje que protagoniza la novela de CELA i OLLÉ, Jaume, *El ladrón de sombras*, Catamarán, 1990.

³⁶ MÈLICH, Joan-Carles, *La fragilidad del mundo*, Tusquets, 2021, p. 142.

comportamiento. En definitiva, están reconociendo “rasgos de la programación humana, de nuestro código fuente, que todavía no hemos sido capaces de articular con palabras (...). Los programas informáticos han detectado rasgos que guían nuestras preferencias, y que podemos intuir, pero no articular”³⁷. De hecho, hay poderosas señales que indican que estamos en la antesala de una gran transformación³⁸. Especialmente porque, aunque las perso-

³⁷ DU SAUTOY, Marcus, *Programados para crear*, Acantilado, 2020, pp. 110 y 111.

³⁸ Un botón de muestra del estado de la ciencia puede contribuir a ilustrar este posible escenario futuro. Por otra parte, el auricular EEG de la empresa *InteraXon*, que se emplea para mitigar migrañas, ayuda a los consumidores a meditar de manera más efectiva a través de la retroalimentación neuroauditiva, como el canto de los pájaros en respuesta a la actividad de las ondas cerebrales. Otras corporaciones como *Myontec*, *Athos*, *Delsys* y *Noraxon* están explorando otros ámbitos. Éstas ofrecen a los atletas y terapeutas deportivos información generada por electromiografía (EMG) sobre lo que está sucediendo en sus músculos durante el entrenamiento y las competiciones, como la velocidad de desarrollo de la fuerza (una medida de la fuerza explosiva), mejoras en la coordinación a través del entrenamiento y la simetría y asimetría en la activación muscular. El ámbito de la salud es obviamente otro sector en expansión. *Control Bionics* vende *NeuroNode*, un dispositivo EMG portátil para pacientes con trastornos neurológicos degenerativos como Esclerosis lateral amiotrófica (ELA) y otras enfermedades de la neurona motora (ENM). Les permite controlar un dispositivo digital o motor a través de las señales bioeléctricas que se envían a los músculos para activar los movimientos. *Kernel* ofrece en el mercado *Flow*, un dispositivo de espectroscopia funcional de infrarrojo cercano (fNIRS) y que mide los cambios en los niveles de oxigenación de la sangre en el cerebro para comprender y mejorar su funcionamiento. FARAHANY, Nita A., “The extent of employee surveillance is greater than you know”, *Duke Magazine*, 24 de Agosto 2021, Consultado en: <<https://alumni.duke.edu/magazine/articles/extent-employee-surveillance-greater-you-know>>; y FARAHANY, Nita A., *The battle for your brain*, St. Martin's Press, 2023, pp. 23 y 24. Otros testimonios en BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 162-164.

nas pueden ser conscientes de muchos aspectos de su entorno, no necesariamente tienen que serlo respecto de cómo estos aspectos ambientales ejercen influencia sobre sus pensamientos y comportamiento: “estas influencias no son perceptiblemente subliminales, pero siguen siendo subliminales en su influencia”³⁹.

Aunque la neurociencia todavía sigue sin poder dar luz sobre numerosos aspectos, este calibrage individual permite la identificación de correlaciones lo suficientemente fuertes como para, sin saber *todavía* el *por qué* (desde el punto de vista bioquímico y/o fisiológico), los datos *hablen por sí solos* y saquen a la luz ciertos patrones de la mente inconsciente. Con esta tecnología, no es necesario saber *cómo está* o *como es* el cerebro para determinar *qué es lo*

³⁹ FRANKLIN, Matija; ASHTON, Hal; GORMAN, Rebecca; ARMSTRONG, Stuart, “Missing Mechanisms of Manipulation in the EU AI Act”, *The International FLAIRS Conference Proceedings*, n. 35, p. 2. Consultado en: <<https://doi.org/10.32473/flairs.v35i.130723>>. Autores que, además, añaden: “La evidencia sugiere que la historia de comportamiento de una persona influye en sus preferencias (...). Las preferencias influyen en el comportamiento, pero a menudo el comportamiento precede y conduce a la aparición de nuevas preferencias. Los sistemas de recomendación a menudo utilizan el Aprendizaje Automático (AA) para aprender las preferencias de los usuarios y optimizar la entrega de algún servicio. Cuando se aplica un enfoque iterativo de AA a los sistemas de recomendación, se vuelve cada vez más difícil identificar si el sistema está aprendiendo las preferencias de sus usuarios o si el sistema de recomendación ha influenciado a los usuarios para que se comporten de cierta manera con el fin de maximizar su función objetivo (como métricas de atención del usuario o clics) (...). Por lo general, los elementos más populares se recomiendan con más frecuencia, lo que los hace aún más populares (...). Los algoritmos del sistema de recomendación utilizan estos datos de comportamiento para el entrenamiento. Dado que los datos de comportamiento provienen de comportamientos influenciados por el sistema de recomendación, entrenar con esos datos crea un bucle de retroalimentación (...). Con el tiempo, estos sistemas pueden cambiar las preferencias a un rango cada vez más estrecho de contenido”.

que hacemos⁴⁰. Al tener las llaves de la *puerta trasera* del código impreso en el circuito neuronal, se está en condiciones de sumergirse bajo la superficie de la percepción consciente y tratar de manipular sutilmente la conducta⁴¹.

⁴⁰ Aunque, de momento, los procesos cerebrales que se pueden *hacer visibles* mediante técnicas de imagen de la actividad cerebral no pueden *hacer visible el pensamiento*, se están dando algunos pasos que apuntan que, de algún modo, esto podría ser posible. Por ejemplo, recientemente, se ha creado un decodificador de lenguaje que, utilizando un transformador de inteligencia artificial (IA), puede traducir los pensamientos de una persona, captados con una máquina de resonancia magnética funcional (fMRI) en texto. TANG, Jerry; LEBEL, Amanda; JAIN, Shailee; y HUTH, Alexander G., “Semantic reconstruction of continuous language from non-invasive brain recordings”, *Nature Neuroscience*, vol. 26, Mayo 2023, pp. 858–866. Consultado en: <<https://doi.org/10.1038/s41593-023-01304-9>>. Estos avances cuestionarían las objeciones de la corriente anti-naturalistas de la actual filosofía de la mente (extensamente, GABRIEL, Markus, *Yo no soy mi cerebro*, Pasado & Presente, 8ª Ed., 2022).

⁴¹ Por ejemplo, algunas compañías están trabajando en campañas publicitarias a través de la “incubación de sueños dirigida” (Xataka, 5 de julio 2021: “El futuro de la publicidad es inducir la en tus sueños. Y los expertos creen que es jugar sucio”. Consultado en: <<https://www.xataka.com/magnet/futuro-publicidad-inducirla-tus-suenos-expertos-creen-que-jugar-sucio>>. O, como expone FARAHANY, *op. cit.*, p. 26: “Los hackers podrían incluso instalar programas espía cerebrales en las aplicaciones y dispositivos que estás utilizando. Un equipo de investigación liderado por la profesora de informática Dawn SONG de la Universidad de California en Berkeley probó esto en jugadores que estaban utilizando una interfaz neural para controlar un videojuego. Mientras jugaban, los investigadores insertaron imágenes subliminales en el juego y exploraron los cerebros inconscientes de los jugadores en busca de reacciones a estímulos, como direcciones postales, detalles bancarios o rostros humanos. Sin que los jugadores lo supieran, los investigadores pudieron robar información de sus cerebros al medir sus respuestas cerebrales inconscientes que indicaban reconocimiento de los estímulos, incluido un código PIN de la tarjeta de crédito de un jugador y su dirección de casa. Los datos neurales también

A la luz de todo lo expuesto, el *yo inconsciente* emergería como una *entidad* susceptible de protección jurídica. Y, admitiendo la oportunidad de la propuesta de neuroderechos (brevemente descrita al inicio de este trabajo), repárese que no es necesario esperar a la expansión de las neurotecnologías para que la mente inconsciente pueda ser asaltada. Por este motivo (y sin negar que puedan identificarse puntos de contacto con la propuesta de la *Neurorights Foundation*), quizás, sea oportuno hablar de los *derechos del yo inconsciente de la persona* (o, en una posible traducción al inglés, *Rights of the unconscious mind – RUM*), como un ámbito susceptible de protección autónomo y diferenciado (y que no es incompatible con la citada propuesta)⁴².

La singularidad de esta necesidad es manifiesta desde el instante que la propuesta de Reglamento de IA de la UE ha previsto un artículo específico dirigido a proteger esta esfera preconscious (aspecto que, cerrando este ensayo, se abordará en el epígrafe que sigue).

IV. SOBRE LA PROPUESTA DEL ARTÍCULO 5 DE LA LEY DE IA DE LA UE: UNA VALORACIÓN CRÍTICA

El condicionamiento por debajo del nivel consciente a través de la IA es una amenaza con entidad suficiente como para justificar la existencia de una regla prohibitiva. Precisamente, este es el propósito del art. 5.1 de la propuesta de Ley de la IA de la UE.

podrían ser interceptados mientras se envían a un teléfono celular emparejado si no están bien protegidos”.

⁴² De forma monográfica al respecto, BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 146-166.

En el marco de la negociación tripartita de la UE (Comisión, Consejo y Parlamento)⁴³, la propuesta original de la Comisión⁴⁴ ofrece un esperanzador (aunque, como se apuntará, limitado) instrumento para tratar de contener (al menos formalmente) el riesgo de la intromisión por debajo del nivel consciente. Su literalidad es la siguiente:

“Las siguientes prácticas de inteligencia artificial estarán prohibidas:

(a) la comercialización, puesta en servicio o uso de un sistema de IA que emplee técnicas subliminales más allá de la conciencia de una persona con el fin de distorsionar materialmente el compor-

⁴³ Algunas nociones sobre el procedimiento pueden contribuir a contextualizar esta cuestión. Como expone CARBONELL, Carlos, “Negociaciones tripartitas y transparencia en la Unión Europea”, *Agenda Pública*, 24 de enero 2017, consultado en: <<https://agendapublica.elpais.com/noticia/15539/negociaciones-tripartitas-transparencia-union-europea>> el proceso habitual y reconocido por los tratados (coloquialmente llamado codecisión, puede constar de hasta tres fases: “Una primera lectura, en la que las dos instituciones que ejercen como legislador, -el Consejo y el Parlamento-, analizan la propuesta de la Comisión separadamente e intentan negociar para llegar a un acuerdo: el Consejo acepta o rechaza la posición y/o enmiendas del Parlamento respecto al texto original de la Comisión. Si no lo consiguen, una segunda lectura tiene lugar. En ella, la propuesta vuelve al Parlamento para que éste se posicione de nuevo, ahora en base a la posición (y/o enmiendas) del Consejo, con el objetivo de que éste acepte la nueva propuesta del Consejo. Y una tercera, en la que si las dos fases anteriores fracasan, se crea un Comité de Conciliación (compuesto por una delegación similar a la de los trilogues más uno de los Vice-Presidente del Parlamento Europeo) que tiene por objetivo facilitar el acuerdo (no se suele llegar a ésta fase)”.

⁴⁴ Disponible en: <<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2021:206:FIN>>

tamiento de una persona de manera que cause o sea probable que cause daño físico o psicológico a esa persona o a otra persona”⁴⁵.

La posición común (“orientación general”) del Consejo (acordada en diciembre de 2022), no ha afectado al art. 5⁴⁶. No obstante, el Parlamento Europeo, en fecha 14 de junio de 2023, ha aprobado una batería de enmiendas sobre la propuesta inicial de la Ley⁴⁷, afectando, entre otros, al precitado precepto⁴⁸. Con esto se ha es-

⁴⁵ Para ayudar a contextualizar esta disposición, repárese que el considerando 16 de la propuesta original establece que “Debe prohibirse la introducción en el mercado, la puesta en servicio o el uso de determinados sistemas de IA destinados a alterar la conducta humana que es probable que provoquen perjuicios físicos o psicológicos. Dichos sistemas de IA despliegan componentes subliminales imperceptibles para el ser humano o se aprovechan de las vulnerabilidades de los menores y las personas por razones de edad o incapacidad física o mental. Lo hacen con la intención de alterar sustancialmente el comportamiento de una persona y de un modo que perjudique o es probable que perjudique a esa misma persona o a otra. Dicha intención no puede darse por supuesta si la alteración del comportamiento humano es el resultado de factores externos al sistema de IA que escapan al control del proveedor o el usuario”.

⁴⁶ Disponible en: <<https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf>>.

⁴⁷ Disponible en: <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.pdf>.

⁴⁸ De hecho, el Parlamento también ha aprobado una enmienda sobre el considerando 16, relativo al art. 5, incluyendo (entre otras novedades) una referencia explícita a las neurotecnologías: “Debe prohibirse la introducción en el mercado, la puesta en servicio o el uso de determinados sistemas de IA que tengan por finalidad o efecto provocar alteraciones sustanciales de la conducta humana que es bastante probable que provoquen perjuicios físicos o psicológicos. Debe entenderse que esta limitación comprende las neurotecnologías asistidas mediante sistemas de IA que se utilizan para supervisar, utilizar o influir en los datos neuronales recopilados a través de interfaces cerebro-ordenador, en la medida en que alteren sustancialmente el comportamiento de una perso-

tablecido la posición negociadora del Parlamento, tras lo cual, a través de las negociaciones interinstitucionales, comenzarán las conversaciones con los países de la UE en el Consejo sobre la forma final de la ley (que, como se ha apuntado, se espera aprobar a finales de 2023). La propuesta del Parlamento sería la siguiente (en negrita las novedades):

Las siguientes prácticas de inteligencia artificial estarán prohibidas: (a) la comercialización, puesta en servicio o uso de un sistema de IA que emplee técnicas subliminales más allá de la conciencia de una persona *o técnicas deliberadamente manipuladoras o engañosas, con el objetivo o el efecto* de distorsionar materialmente el comportamiento de una persona o de un grupo de personas, *al afectar considerablemente la capacidad de la persona para tomar una decisión informada, causando así que la persona tome una decisión que de otro modo no habría tomado de una manera* que cause o sea probable que

na física de una manera que cause o pueda probablemente causar un perjuicio significativo a esa misma persona o a otra. Dichos sistemas de IA despliegan componentes subliminales imperceptibles para el ser humano o se aprovechan de las vulnerabilidades de las personas y de grupos específicos de personas por razones de sus rasgos conocidos o previstos de personalidad, edad, incapacidad física o mental, situación social o económica. Actúan así con la finalidad o el efecto de alterar sustancialmente el comportamiento de una persona y de un modo que perjudique o es probable que perjudique de forma significativa a nivel físico o psicológico a esa misma persona o a otra, o a grupos de personas, incluyendo perjuicios que pueden acumularse con el tiempo. La intención de alterar el comportamiento no puede darse por supuesta si la alteración es el resultado de factores externos al sistema de IA que escapan al control del proveedor o el usuario, factores que el proveedor o el implementador del sistema de IA no puedan prever ni mitigar razonablemente. En cualquier caso, no es necesario que el proveedor o el implementador tengan la intención de causar perjuicios significativos, siempre y cuando el perjuicio se derive de las prácticas de manipulación o explotación que permite la inteligencia artificial (...).”

cause daño *significativo* a esa persona, a otra persona o a un grupo de personas.

La prohibición de un sistema de IA que emplea técnicas subliminales mencionada en el primer párrafo no se aplicará a los sistemas de IA destinados a ser utilizados con fines terapéuticos aprobados, siempre y cuando se obtenga un consentimiento informado específico de las personas expuestas a ellos o, en su caso, de su tutor legal.

En una primera lectura de las enmiendas introducidas podría concluirse que la nueva redacción describe un entorno más protector que en su redacción original. En este sentido, es positivo que, además de las técnicas subliminales también se incluyan técnicas deliberadamente manipuladoras o engañosas. Este incremento del espectro abarcable es, sin duda, positivo y necesario.

No obstante, una lectura más atenta, permite concluir que la resultante de los cambios es profundamente regresiva. En efecto, esta propuesta del Parlamento plantea serias objeciones porque los elementos configuradores del supuesto de hecho de la norma, lejos de restringir, amplían la posibilidad del condicionamiento por debajo del nivel consciente de forma sustancial. En concreto, estos factores habilitadores serían tres:

a) En primer lugar, las técnicas empleadas (subliminales, manipuladoras y/o engañosas) se prohíben si afectan a la “capacidad” de la persona para tomar una decisión (de modo que, si no se ve afectada, deben permitirse). Y, además, se exige que esta incidencia en la capacidad sea “considerable”.

Repárese que, formulada en estos términos, se refiere a una limitación próxima a la “no capacidad”. Lo que significaría que, para que se active la regla prohibitiva, se estaría exigiendo que el condicionamiento heterónimo prácticamente anule la voluntad humana (bien porque se niega la propia intención y/o bien, la autoconciencia de lo que se hace). Es decir, si debe afectar a la capacidad y con esta entidad, apela a una intromisión que reduzca el elemento volitivo a algo similar a la condición de *zombi*.

Con esta configuración, de forma indirecta, se delimita un espacio de intromisión en nuestra mente inconsciente absolutamente abrumador: se permite cualquier afectación que permanezca por debajo de un determinado umbral (o, incluso, por encima, si no se dan los otros dos factores que se describirán a continuación). Además, la delimitación de este umbral tolerado plantea serios inconvenientes. En particular, porque la “considerabilidad” de una afectación resulta difícil de medir. Como se ha expuesto, con la ciencia actual es muy complejo medir y/o evaluar cómo la mente inconsciente afecta a las decisiones del yo consciente. Además, en este entorno neuronal, es ilusorio pensar que seremos capaces de establecer en términos jurídicos y de acreditación probatoria un parámetro de comparación que permita saber si algo es más o menos “considerable”.

b) Segundo, la regla prohibitiva del art. 5 sólo se desplegará en el caso de que la intromisión heterónoma lleve a una persona a tomar una decisión que, de otro modo, no habría tomado. Es decir, si, antes del condicionamiento, ya tenía “pensado” ese curso de acción con plena determinación, no se activa la prohibición.

Es obvio que este requisito plantea un problema jurídico de extrema complejidad porque exige una comparación de dos escenarios (ya de por sí, difícilmente determinables):

Por un lado, exige valorar un estadio previo a la decisión tomada muy etéreo (¿qué *entidad* debe tener este pensamiento previo para que se entienda que, antes del condicionamiento imperceptible, era una decisión *ya tomada*?). Repárese que es muy posible que el propio afectado no sea siquiera capaz de apreciarlo; o, si puede hacerlo, no pueda medir la determinación con la que quería tomar esa decisión (pudiendo ser una de las muchas alternativas al alcance). Como puede imaginarse, acreditar este extremo raya lo imposible.

Por otro lado, exige constatar un contrafactual. En efecto, debe proyectarse un acontecimiento hipotético que, aunque pudo ser posible si las decisiones se hubieran tomado de otro modo, no se ha materializado en el mundo real. En estas condiciones (teniendo

siempre presente que ha mediado un “susurro” imperceptible), se está exigiendo la acreditación de un hecho ajeno a la realidad (porque apela a un escenario que no se ha materializado). En definitiva, es muy difícil saber qué hubiera podido pasar si no se hubiera producido la intromisión por debajo del nivel consciente. Como apuntan Judea Pearl y Dana Mackenzie, el razonamiento contrafactual, que se ocupa de dar respuesta a la pregunta “¿y si...?” (como, por ejemplo, “¿y si no hubiera ido a la playa, qué habría pasado?”), se caracteriza porque la observación empírica no puede ni confirmar ni rebatir las respuestas a tal clase de preguntas⁴⁹.

La resultante de esta comparación (sobre un estadio previo difícilmente medible y un contrafactual que ni es confirmable ni rebatible) arroja un problema en términos de identificación y acreditación procesal insuperables.

– Tercero, sólo se prohíben los condicionamientos que causen un daño “significativo” (con todos los problemas que la evaluación de este tipo de daño puede plantear, especialmente, en términos de relación de causa y efecto; y, obviamente - y de nuevo-, de su acreditación).

A la luz de todo lo expuesto, es difícil imaginar que la aplicación de esta regla prohibitiva sea pacífica. Además, adviértase que, en los términos propuestos por el Parlamento, sólo se activa si concurren simultáneamente los tres factores recién descritos (que la afectación en la capacidad de decisión sea considerable; que la decisión finalmente tomada haya sido provocada por el condicionamiento; y que se cause un daño significativo). Al margen de los problemas ya apuntados, si el texto definitivo se aprueba con esta configuración acumulativa, es obvio que el patio trasero neuronal de las personas quedará muy expuesto. Y las intromisiones en el inconsciente que no superen estos umbrales quedarán amparadas legalmente.

⁴⁹ PEARL, Judea, y MACKENZIE, Dana, *El libro del porqué, Pasado y Presente*, 2020, p. 20.

Esto describe una amenaza muy inquietante, porque abre la puerta a la devaluación severa de dos esferas determinantes del ser humano: la libertad y la identidad. De hecho, esta intromisión también es apreciable en la propuesta original de la Comisión, al condicionar la aplicación de la regla prohibitiva a la concurrencia de un “daño”. Aunque en este caso, no se exige un umbral y, por consiguiente, la propuesta es más garantista (de hecho, tampoco exige que la intromisión heterónoma afecte considerablemente a la capacidad, ni tampoco que la decisión tomada sea distinta a la que se hubiera tomado sin el condicionamiento imperceptible), téngase en cuenta que, si no se produce un daño (o éste no es acreditable), se admitirá el condicionamiento por debajo del nivel consciente. En tal caso, la libertad y la identidad, como se tratará de exponer brevemente a continuación, también serían objeto de una violación intolerable.

En relación a la libertad, debe tenerse en cuenta que el condicionamiento inconsciente pretende “susurrarnos” anticipadamente qué es lo que nos gusta, qué es lo que necesitamos, qué es lo que deseamos. El propósito es convertir a la sociedad en algo parecido a una gran Caja de Skinner a cambio de socavar el sentido de la libertad. O, como ha apuntado Alain Supiot, aspira a metamorfosear a los ciudadanos en perros de Pávlov⁵⁰.

Sin embargo, sólo podemos elegir con libertad “si no estamos indebidamente presionados (...) y sí razonablemente bien informados sobre las demás posibilidades”⁵¹. Los estímulos subyacentes en estas tecnologías operan como si fueran incentivos voluntariamente aceptados (consentidos). Conviene tener muy presente que los incentivos son una manifestación de ingeniería social y conductual (dirigidas por expertos) que ofrecen personas e instituciones con poder a quienes no lo tienen. En estas condiciones (asumiendo que son instrumentos de poder⁵²), es obvio que pueden

⁵⁰ SUPIOT, Alain, *El trabajo ya no es lo que fue*, Siglo XXI, 2023, p. 99.

⁵¹ SANDEL, Michael J., *Justicia*, Debolsillo, 2016, p. 113.

⁵² GRANT, Ruth W., *Los hilos que nos mueven*, Avarigani, 2021, pp. 70 y 57.

ser objeto de abuso. Especialmente porque, como apunta Ruth W. Grant, intentan de algún modo “evitar la necesidad de convencer a la gente, dándoles razones extrínsecas para escoger lo que las personas o instituciones que proponen los incentivos quieren que escojan”. Este diseño experto puede laminar la democracia porque en él “no entra en juego el tipo de debate público y el consentimiento que idealmente caracteriza los procesos democráticos de toma de decisiones”⁵³.

Las emociones también podrían verse afectadas (y esto también podría afectar a nuestro pensamiento racional). Como se ha expuesto anteriormente, la idea del cerebro basado en estímulos-respuestas es un mito. Lejos de este enfoque, la actividad cerebral es predicción y corrección y, esta construcción de experiencias emocionales se lleva a cabo de forma invisible, fuera de la conciencia. Recuerden que, aunque no seamos conscientes de ello, somos los arquitectos de nuestras emociones. No son *provocadas*. A partir de los patrones emocionales contruidos, el cerebro “crea una avalancha de predicciones, simula sus consecuencias como si fueran reales, y comprueba y corrige esas predicciones comparándolas con el input sensorial real”⁵⁴. Por consiguiente, es predictivo y no reactivo (“nuestros sistemas centrales están tratando de adivinar constantemente lo que pasará a continuación para que podamos sobrevivir”). De modo que existe una extraordinaria actividad previa preconsciente. Si el art. 5.1.a) de la Ley de IA, en los términos propuesto por la posición común del Parlamento, admite cierto tipo de condicionamiento por debajo del nivel consciente, estaremos dejando la puerta del “patio trasero neuronal” mucho más abierta de lo que sería deseable. Lo que implicaría que el estadio preconsciente que alimenta nuestras emociones también podría ser deliberada y subliminalmente alterado.

Para comprender la trascendencia de esta amenaza es importante advertir un aspecto que tiende a pasarse por alto. Como

⁵³ *Ibidem*, p. 48.

⁵⁴ FELDMAN BARRETT, *op. cit.*, p. 198.

apunta Antonio Damasio⁵⁵, en el pensamiento común, es frecuente pensar que las decisiones sensatas son “fruto de una mente serena, que emociones y razón no se mezclan mejor que el agua y el aceite” (decimos: “*¡no te dejes llevar por las emociones!*”; o “*¡Piensa con la cabeza y no con el corazón!*”). Es decir, tiende a pensarse que los mecanismos racionales existen “en una provincia mental separada”, a la que no deben tener acceso las emociones. No obstante, lejos de la existencia de sistemas neurales “distintos para emoción y razón”, emociones y sentimientos “no son para nada intrusos en el bastión racional”. Es cierto que “en ciertas circunstancias emociones y sentimientos puedan causar estragos en los procesos de razonamiento”. Este influjo potencialmente dañino de los sesgos emocionales convive con otra realidad sumamente importante: la ausencia de emoción y sentimiento puede “ser igualmente perjudicial, pueda comprometer la racionalidad que nos hace distintivamente humano”. O, como apunta Leonard Mlodinow⁵⁶, “en los humanos, la flexibilidad que brinda la emoción también permite contribuciones a nuestra mente racional que nos lleva a tomar mejores decisiones y reaccionar de manera más sofisticada”.

Si nuestra autonomía y capacidad de actuar y elegir libremente está comprometida (porque algunas opciones han sido deliberadamente descartadas o subliminal y deliberadamente sugeridas), ya no podremos decir (en términos de la filosofía kantiana) que actuamos conforme a una ley que nos damos a nosotros mismos (la validez del consentimiento libre consciente quedaría en entredicho). En tal caso, lo haremos sometidos a unos fines dados desde fuera de nosotros (porque han sido dictados por terceros)⁵⁷.

Esta “arquitectura de ejecución” tampoco es inocua en términos de la identidad. Especialmente, si todo nuestro concepto del *yo* presupone que existe algo que podemos denominar elección⁵⁸. La

⁵⁵ DAMASIO, Antonio, *El error de Descartes*, 2ª ed., Andrés Bello, pp. 11-13.

⁵⁶ MLODINOW, Leonard, *Emocional*, Crítica, 2023, p. 49.

⁵⁷ SANDEL, *op. cit.*, pp. 126-129.

⁵⁸ PEARL y MACKENZIE, *op. cit.*, p. 371.

matematización que alimenta el “gobierno de los números” (por emplear un concepto sugerido por Supiot) desvela una de las principales amenazas a nuestra individualidad, pues “reducen al hombre al nivel de un animal de conducta condicionada”⁵⁹.

En definitiva, libertad e identidad describen esferas del ser humano que deberían ser absolutamente inexpugnables. En tanto que apela a la “esencia de la esencia”, es difícil imaginar algún tipo de renuncia legítima⁶⁰.

De hecho, cualquier decisión procesada a partir del descifrado (por el medio que sea – algorítmicamente o no) de nuestros procesos mentales profundos y con el propósito de susurrarnos por debajo del nivel consciente, debe ser prohibida y declarada nula y sin efecto. La nulidad debería predicarse, aunque el yo consciente sea partícipe (en algún grado) en el curso de acción finalmente exteriorizado. La naturaleza del bien jurídico protegido, la entidad de la amenaza y el hecho de que sea materialmente imposible distinguir entre la aportación volitiva y la inconsciente así lo sugieren.

Por otra parte, en el eventual caso que el resultado final de la conducta condicionada (con participación volitiva o no) arroje un efecto positivo en el agente, tampoco bastaría para validarla (de otro modo, implícitamente, se estaría admitiendo que cierto tipo de condicionamiento subliminal es jurídicamente admisible)⁶¹.

⁵⁹ ARENDT, Hannah, *La condición humana*, Austral, 1993, p. 55.

⁶⁰ En el marco de los *derechos de yo inconsciente de la persona* (anteriormente apuntados), esta esfera de protección podría materializarse en el reconocimiento del *Derecho a la preservación de la mente inconsciente*. Extensamente al respecto, BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 157-177.

⁶¹ En el marco de los *derechos de yo inconsciente de la persona* (anteriormente apuntados), esta esfera de protección podría materializarse en el reconocimiento del *Derecho a la no intervención subliminal*. Extensamente al respecto, BELTRAN DE HEREDIA RUIZ, *op. cit.*, pp. 157-177.

V. VALORACIÓN FINAL

Todo condicionamiento por debajo del nivel consciente debe ser prohibido. El margen de tolerancia debe ser cero.

Es muy conveniente que las enmiendas del Parlamento que restringen la aplicación de la regla prohibitiva no se incorporen al texto definitivo de la Ley de IA. De hecho, tampoco es admisible, como se recoge en la propuesta inicial de la Comisión, que se exija la concurrencia de un daño para que se active la prohibición. No obstante, es obvio que hay muchos modelos de negocio que dependen de que esto no suceda o que incluso se relaje todavía más. Y aunarán todos los esfuerzos para conseguirlo.

Si, finalmente, el texto definitivo mantiene alguno de los factores que restringen la posibilidad de prohibir el condicionamiento por debajo del nivel consciente, como ciudadanos tendremos que asumir que nuestra individualidad y autonomía podrá ser anulada o limada. Como en el cuento del flautista de Hamelin, bailaremos sin saber por qué. Es decir, un *estar sin ser*.

Este escenario cobra dimensiones más preocupantes cuando se trata de las generaciones más jóvenes y su (irresistible) apego a la tecnología. Como apunta Supiot, la adicción a las redes sociales “prepara bien a los niños ‘adictos’ a internet para una vida de trabajo programada por las plataformas, encadenada a las pantallas, manejada por algoritmos y receptiva a los ‘nudges’”⁶² (esto es, al condicionamiento subliminal, a través de los famosos “empujoncitos”).

Blindar a nuestro yo inconsciente de todas las intromisiones externas es el mejor legado que podríamos dejar a las generaciones futuras⁶³.

⁶² SUPIOT, Alain, *op. cit.*, p. 118.

⁶³ Fecha de consulta de todos los enlaces web citados: 16 de julio 2023.