

Algoritmos, psicometría y derechos del yo inconsciente de la persona (o «neuroderechos») en el ámbito socio-laboral

Ignasi Beltran de Heredia Ruiz
Profesor Agregado y TU Acreditado. Universitat Oberta de Catalunya (UOC)

Working paper (v2)

Ponencia Presentada al Congreso «*La tecnología y la digitalización en las relaciones laborales: personas y competitividad*» (Valencia 17 y 18 de noviembre 2022)

Sumario

1. Introducción.....	2
2. <i>Homo sapiens</i> y limitaciones «de serie»	3
3. Corrigiendo el error humano: paternalismo libertario y automatización severa	5
4. El <i>homo sapiens</i> no es infalible: la gestión algorítmica (oportunamente utilizada por la inteligencia humana) puede ayudarnos.....	11
5. Psicometría y derechos del «yo inconsciente de las personas» (YIP).....	18
5.1. Psicometría y patrones conductuales emergentes	19
5.2. La amenaza del pastoreo social (¿los derechos del yo inconsciente de la persona como antídoto?)	25
6. Valoración final	36
7. Bibliografía citada	39

1. Introducción

El avance tecnológico está acelerándose de tal modo que está alcanzado cotas quiméricas. Entre otras innovaciones, la senda hacia la decodificación de los procesos mentales más profundos del ser humano parece que es un objetivo, en cierta medida, al alcance. Aunque todavía estamos lejos de observar lo que está detrás del *velo de la evolución*, el uso de *big data* y la estadística computacional están abriendo las puertas hacia estratos de nuestro cerebro hace unas décadas inimaginables.

Aunque debe advertirse que los escenarios transhumanistas de uso generalizado de lo que se conoce como *interfaces cerebro-ordenador (Brain-Computer Interfaces – BCI)* todavía quedan muy lejos. No obstante, por si no los conocen (y siguiendo a YUSTE y GOERING, 159 y 160), deben saber que estos instrumentos (quizás, de forma todavía rudimentaria) ya permiten la comunicación directa y bidireccional entre el cerebro y un dispositivo y también tenemos a nuestro alcance objetos portátiles (*wearables*), de todo tipo, con capacidad para leer la actividad del sistema nervioso periférico al tocar la cabeza o el cuerpo de una persona. Aunque el uso de estos dispositivos es, hoy por hoy, testimonial (y, principalmente, terapéutico), el impulso de la ciencia y la industria es de tal entidad que se espera una expansión sin precedentes (en este sentido, lo que se conoce como *metaverso* o realidad virtual inmersiva podría ser una de estas manifestaciones privilegiadas). Pero al margen de este proceso (o en paralelo), lo cierto es que la estructura computacional ubicua y la recopilación y tratamiento de lo que se conoce como *excedente conductual* (incluidos los metadatos o las *migajas digitales*) ya permite el acceso a las intenciones, emociones y decisiones de las personas. Este fenómeno, que tampoco se ha desplegado en toda su potencialidad, muestra signos amenazantes, en tanto que atesora capacidad suficiente para agujonear en el *yo inconsciente* de las personas e, incluso, perfilar su comportamiento.

Esta intromisión, tan inaudita como invasiva, sugiere la necesidad de articular mecanismos jurídicos para proteger a las personas. Y, aunque todo apunta a que la onda expansiva de este fenómeno afectará a la sociedad en su globalidad (es difícil pensar que algún ámbito quede libre de su impacto), el entorno profesional y del trabajo, dada la confluencia de fuerzas e intereses concurrentes, podría acabar siendo la *zona cero*.

El propósito de este trabajo es abordar esta problemática, distinguiendo, en su iter expositivo, dos estadios o etapas diferenciadas. En efecto, en primer lugar, se propondrá una descripción de la secuencia de factores que podrían explicar cómo se ha llegado al momento actual. Y, por este motivo, se partirá de algunas nociones básicas sobre el cableado mental que capitanea nuestro comportamiento. Especialmente porque la exteriorización de este universo introspectivo que nuestras acciones reflejan, paradójicamente, muestran una disociación nada desdeñable con el control consciente de la conducta. En efecto, los sesgos (y los consiguientes errores que precipitan) son patrones *típicamente* humanos. Y, como se apuntará, el acopio de numerosas evidencias sobre esta *racionalidad limitada* del *homo sapiens* (es decir, su carácter sistemático) ha alimentado una corriente de pensamiento (que se ha desarrollado en etapas y formas diferenciadas y) que sostiene que las personas necesitan ayuda heterónoma para alcanzar lo que *verdaderamente* desean (y que no son capaces de conseguir por sus propios medios). Y, todo parece indicar que la asistencia automatizada (pero no sólo) es,

hoy por hoy, *la elegida* para complementarnos en todo aquello que no somos capaces de alcanzar. De ahí la creciente suplantación del juicio humano en la toma de ciertas decisiones por parte de las máquinas. La particularidad del momento es que, en este proceso (*a priori* pensado para mejorar nuestra calidad de vida), nuestro *yo inconsciente* podría ya estar siendo víctima de un *pastoreo* intolerable¹.

Esta exposición será el punto de partida de la segunda parte de este estudio anunciada. En concreto (y como también ya se ha apuntado), se valorará la necesidad de articular un escudo jurídico que (a modo de cortafuegos) sea capaz de repeler esta invasión. No obstante, la sutileza e imperceptibilidad de estos ataques y el poder que su control generaría describen una amenaza sin precedentes y, por este motivo, surge la necesidad de sondear si los instrumentos a nuestro alcance podrán contener con efectividad estos embistes, o bien, debemos pensar en otros. Y aunque se trata de una cuestión que se encuentra en pleno debate internacional, en este contexto podrían cobrar protagonismo lo que podrían calificarse como derechos del *yo inconsciente de las personas* (o, en su denominación más conocida – aunque a mi entender menos precisa –, *neuroderechos*). El estado de la ciencia al respecto, la exposición de las primeras propuestas planteadas y el *tanteo* prospectivo de su posible impacto en el ámbito socio-laboral culminará el análisis que este trabajo persigue.

A la luz de lo expuesto, es obvio que el objeto de este estudio se caracteriza por una provisionalidad e incertidumbre extremas. Muchos problemas hoy sólo podemos intuirlos y/o especular sobre ellos. De hecho, hasta que no se corporicen, seremos víctimas de una aguda ceguera que invariablemente se irá disipando a medida que el futuro devenga presente. Como ha sucedido con todas las revoluciones tecnológicas precedentes, el desconocimiento del curso que acabará tomando la técnica y los riesgos que llevará aparejados, nos enfrenta al reto de imaginar *proactivamente* y *en el presente* el marco normativo del futuro incierto. Desde este punto de vista, este trabajo tiene un propósito meramente exploratorio y, por ende, muy modesto, pues, a modo de *globo sonda*, sólo aspira a palpar los límites *imaginables* a nuestro alcance. En definitiva, les invito a un paseo tanteando la oscuridad.

¿Quieren acompañarme?

2. *Homo sapiens* y limitaciones «de serie»

El comportamiento humano es la resultante de procesos deliberados y controlados, pero también de otros que están por debajo del nivel consciente. Aunque, propiamente (de acuerdo con los últimos avances de la neurociencia), no puede afirmarse que tengamos módulos cerebrales diferenciados², se han propuesto algunas metáforas que pueden ayudarnos a comprender esta *mecánica* cerebral.

¹ Lo que podría dar pie a un proceso paralelo de creciente obsolescencia humana (particularmente preocupante - ver extensamente al respecto en BELTRAN DE HEREDIA RUIZ, 113 – 124).

² FELDMAN BARRETT (218): «Cuando miramos cómo funciona el cerebro no vemos módulos mentales. Vemos sistemas centrales que interactúan continuamente de maneras complejas para crear muchas clases de mentes dependiendo de la cultura. El cerebro humano mismo es un artefacto cultural porque está cableado por la experiencia. Tenemos genes que el entorno enciende y apaga, y otros genes que regulan nuestra sensibilidad a ese entorno».

Una de las más célebres³ es la propuesta por Daniel KAHNEMAN, a resultas de sus investigaciones junto con Amos TVERSKY, a partir de la descripción metafórica de la vida mental mediante los agentes: el Sistema 1 y el Sistema 2. El papel de cada uno de ellos es muy ilustrativo para comprender nuestro comportamiento. Siguiendo con la exposición de KAHNEMAN (26), el primero produce pensamiento rápido e intuitivo y el segundo pensamiento lento y deliberativo, de modo que vendrían a ser «rasgos y disposiciones de dos caracteres existentes en nuestras mentes».

Ahondando un poco más en las particularidades de cada uno de ellos (35), el Sistema 1 opera de manera «automática, con poco o ningún esfuerzo y sin sensación de control voluntario»; y, de acuerdo con este patrón, «genera impresiones y sentimientos que son las fuentes principales de las creencias explícitas y las elecciones deliberadas del Sistema 2. Las operaciones automáticas del Sistema 1 generan patrones de ideas sorprendentemente complejos, pero solo el lento Sistema 2 puede construir pensamientos en una serie ordenada de pasos».

No obstante, debe advertirse que el Sistema 2 también puede *tomar las riendas* y anular «los irresponsables impulsos y asociaciones del Sistema 1». Así pues, «el Sistema 2 centra la atención en las actividades mentales esforzadas que lo demandan, incluidos los cálculos complejos». De hecho, sus operaciones «están a menudo asociadas a la experiencia subjetiva de actuar, elegir y concentrarse»; de modo que, «cuando pensamos en nosotros mismos, nos identificamos con el Sistema 2, con el yo consciente, racional, que tiene creencias, hace elecciones y decide qué pensar y qué hacer».

Sin embargo, el Sistema 2 (39) «se halla normalmente en un confortable modo de mínimo esfuerzo en el que solo una fracción de su capacidad está ocupada». En definitiva, es perezoso (se caracteriza [48] por «una renuencia a invertir más esfuerzo del estrictamente necesario»). Y, (39) «si cuentan con la aprobación del Sistema 2, las impresiones e intuiciones se tornan creencias y los impulsos acciones voluntarias». En definitiva, «Si todo se desarrolla sin complicaciones, como ocurre la mayor parte del tiempo, el Sistema 2 acepta las sugerencias del Sistema 1 con escasa o ninguna modificación».

En definitiva, como exponen SUNSTEIN y THALER (35 a 39), el Sistema 1, es rápido, instintivo y visceral (no implica lo que normalmente asociamos con la palabra *pensar*); y el Sistema 2 es más premeditado, autoconsciente y afecta al pensamiento consciente. Para ilustrar, si pudiéramos identificarlos con un personaje de ficción (espero que no se ofendan), el primero estaría presente predominante en el comportamiento de *Homer Simpson* y el segundo en el del *Señor Spock* de *Star Trek*.

Llegados a este estadio, es preciso reparar en tres aspectos de especial relevancia (para el propósito de este ensayo):

Primero: conviene reiterar que esta distinción es metafórica. Esto es, el Sistema 1 y el Sistema 2 son *personajes ficticios*. No son (KAHNEMAN, 46) «sistemas en el sentido corriente de entidades con aspectos o partes que interaccionan»; y, por este motivo,

³ Una exposición de otras posibles explicaciones en FELDMAN (217 y 218).

tampoco puede afirmarse que haya «una sola zona del cerebro que cada uno de los sistemas pueda llamar su hogar».

Segundo: aunque nos incomode, debemos aceptar que el Sistema 1 (el más *intuitivo*), es (26) «más influyente de los que nuestra experiencia nos dice, y es el secreto autor de muchas elecciones y juicios que hacemos». Y, como ilustrativamente expone KAHNEMAN, si se hiciera una película de ambos, «el Sistema 2 sería un personaje secundario que se cree protagonista»⁴.

Lo que, en definitiva, cuestiona de forma severa algunas concepciones de racionalidad (al menos, tal y como la manejan, por ejemplo, en la economía). Especialmente, al reparar que (36) «las capacidades del Sistema 1 incluyen destrezas innatas que compartimos con otros animales» (aunque también hay otras automáticas que se adquieren con la práctica prolongada - como recordar la capital de un país cercano).

Y, tercero, y quizás más relevante: el protagonismo del Sistema 1 precipita (inevitablemente) el uso de heurísticas (esto es, atajos en nuestras deliberaciones y juicios) y, con ellas, emergen sesgos (es decir, errores sistemáticos) de muy heterogénea naturaleza (y que se expondrán con un poco más de detalle posteriormente)⁵.

En relación a esto último, conviene hacer dos acotaciones: en primer lugar, aclarar que esta particularidad *tan humana* no debería identificarse con la idea de estupidez ni de irracionalidad (ni tampoco debería confundirse con las intuiciones acertadas que se explican mejor por los efectos de la práctica prolongada).

Y, en segundo lugar, advertir que la existencia de heurísticas y sesgos sí ha alimentado la idea de que el ser humano necesita ser asistido para suplir a estas carencias con las que venimos al mundo *equipados de serie*. Las propuestas dirigidas a complementarnos han sido de muy diversa naturaleza y en todas ellas subyacen profundas implicaciones. Extremo que se desarrollará en el epígrafe que sigue.

3. Corrigiendo el error humano: paternalismo libertario y automatización severa

La identificación de errores sistemáticos en el juicio humano ha impulsado la idea de que el *homo sapiens* necesita un complemento que colme estas carencias (y nos ayude a

⁴ O, como recoge WILLIAMS (49), «una parte considerable de nuestra experiencia cotidiana se compone de (...) procesos automáticos e inconscientes: nuestras vidas transcurren contra el telón de fondo de una ‘insoponible automaticidad del ser’».

⁵ En concreto, el término *heurística* (KAHNEMAN, 133) se refiere a un «procedimiento sencillo que nos ayuda a encontrar respuestas adecuadas, aunque a menudo imperfectas, a preguntas difíciles. La palabra tiene la misma raíz que eureka». Este concepto surge a partir de las investigaciones de KAHNEMAN y TVERSKY sobre cómo la gente puede hacer «juicios de probabilidad sin conocer con precisión lo que es la probabilidad», de modo que tendemos a «simplificar de algún modo esta tarea imposible», concluyendo que, cuando se pide «juzgar probabilidades, la gente realmente juzga algo y cree que ha juzgado sobre probabilidades». O, como apunta SUNSTEIN (2009, 57): «La heurística en general funciona mediante un proceso de ‘sustitución de atributos’, según la cual las personas responden una pregunta difícil que sustituye a una más fácil». En definitiva (THALER, 52 y 53), asumiendo que tenemos «un tiempo y una capacidad cerebral limitados (...) por ello nos servimos de sencillas reglas generales – las heurísticas – para facilitar la toma de decisiones». Y esto es lo que lleva a las personas a cometer *errores predecibles* (al respecto también KAHNEMAN, 183 y ss.).

hacer juicios más acertados y tomar mejores decisiones). Lo que ha derivado en dos reacciones diferenciadas (y, quizás, complementarias) que no son, en absoluto, inocuas:

La primera reacción ha propuesto que determinadas medidas políticas y las instituciones pueden proporcionar esta ayuda (*complementando* nuestras carencias). Y, en esta línea, la economía conductual (o *behavioral economics*) ha apostado por un tipo de *neointervencionismo* que ha sido denominado como *paternalismo libertario* (y, obviamente, ha recibido la crítica severa por los sectores más ortodoxos de la economía, como la *Escuela de Chicago*⁶). El efecto combinado de estos dos conceptos (SUNSTEIN y THALER, 19) aspira a legitimar a los *arquitectos de las decisiones* (del sector público y privado) para influir en la conducta de la gente (a través de los conocidos *nudges* o empujoncitos). Esto es, estos *planificadores* deben «esforzarse de forma consciente por orientar las decisiones de las personas en direcciones que mejoren [a juicio de ellas mismas] sus vidas», pero, manteniendo la libertad de elegir. Así (GRANT, 45), se pretende cambiar el comportamiento de la gente estructurando sus elecciones de cierta forma (como, por ejemplo, diseñado una distribución de la localización de la fruta en un comedor escolar – para fomentar más su consumo; o bien, estableciendo una inscripción a un plan de pensiones por defecto para fomentar su suscripción – dando la oportunidad de no hacerlo si no se desea).

Es importante advertir que esta corriente de pensamiento es una respuesta a la teoría de la elección racional (y la ilusoria hiperracionalidad) que se destila del *homo oeconomicus*. Y este contexto es fundamental para comprender el sentido crítico que atesora la psicología de la conducta (y también para observar cómo la automatización cognitiva severa la ha *tergiversado* al magnificar nuestra necesidad de asistencia).

Como expone ARIELY (258), «la economía estándar presupone que somos racionales; que conocemos toda la información pertinente relacionada con nuestras decisiones, que podemos calcular el valor de las distintas opciones que afrontamos, y que cognitivamente nada nos impide sopesar las ramificaciones de cualquier potencial decisión. El resultado es que se da por supuesto que tomamos decisiones lógicas y sensatas. Y aunque de vez en cuando tomemos una decisión equivocada, la perspectiva de la economía estándar sugiere que de inmediato aprenderemos de nuestros errores, bien por nosotros mismos, bien con la ayuda de las ‘fuerzas del mercado’»⁷.

Es obvio que lo anterior está muy alejado de la realidad. Siguiendo la exposición de THALER (29 a 32), el problema de la teoría económica reside en que emplea un modelo en el que se «sustituye al *Homo sapiens* por una criatura ficticia llamada *Homo oeconomicus*» y esto provoca que con los modelos económicos se alcancen predicciones erróneas. Añadiendo que «irónicamente, la existencia de modelos formales basados en esta idea equivocada del comportamiento humano es precisamente lo que concede a la ciencia económica su reputación de ser la más poderosa de las ciencias sociales, poderosa en dos sentidos claramente diferenciados. El primero es indiscutible: de entre

⁶ Una divertida exposición de esta controversia en THALER (233 y ss.).

⁷ Siguiendo con ARIELY (259 y 260), «según los supuestos de la economía estándar, todas las decisiones humanas son racionales e informadas, motivadas por un concepto preciso de la valía de todos los bienes y servicios, y de la cantidad de felicidad (utilidad) que probablemente producirá cada decisión. Bajo esta serie de supuestos, todos los participantes en el mercado tratan maximizar su beneficio y se esfuerzan en optimizar sus experiencias».

todos los conflictos sociales, los economistas son los que más influencia tienen en las políticas públicas (...). El otro es que la economía también está considerada como la más poderosa de las ciencias sociales en el ámbito intelectual, poder que se deriva del hecho de que posee un núcleo teórico unificado del que parte casi todo». Hasta el extremo de que «los economistas suelen comparar su campo con el de la física, ya que al igual que ésta, la economía se basa en unas pocas premisas iniciales»⁸.

El problema (siguiendo con el mismo autor) es que «premisas en las que se basa la teoría económica son imperfectas», porque, primero, «los problemas de optimización a los que se enfrenta la gente normal a menudo son demasiado difíciles como para que los puedan resolver, o incluso acercarse a su resolución»; segundo, «las creencias a partir de las cuales la gente toma sus decisiones no son imparciales. Puede que el exceso de confianza no aparezca en el diccionario de los economistas, pero es un rasgo muy arraigado en la naturaleza humana, y existen innumerables otros sesgos bien documentados por los psicólogos»⁹; y, tercero, «existen numerosos factores que el modelo de optimización no incluye». De hecho, añade que, lejos de ser propiedades humanas ilimitadas, en realidad operan las siguientes restricciones (362): «racionalidad limitada, fuerza de voluntad limitada y egoísmo limitado».

De modo que (ARIELY, 252, 262 y 263), «no somos nobles en nuestra razón ni infinitos en capacidad, y nuestra comprensión resulta ser más bien pobre». Especialmente porque hay una serie de fuerzas (emociones, relatividad, normas sociales, etc.) que influyen en nuestro comportamiento y (con independencia de si somos expertos o novatos) las subestimamos, provocando errores sistemáticos y previsibles. Somos víctimas de las «ilusiones decisorias» de nuestra mente (de un modo similar a lo que nos sucede con las ilusiones ópticas), pues, nuestras decisiones parten de la representación de la realidad que nuestro cerebro ha construido a partir de nuestros limitados sentidos que la naturaleza nos ha proporcionado¹⁰. No sólo no somos racionales todo el tiempo, sino que, incluso, podemos acabar tomando decisiones que nos perjudican o ser manifiestamente inconsistentes a lo largo del tiempo.

En definitiva, a partir de los descubrimientos pioneros de KAHNEMAN y TVERSKY, en la medida que nuestro sistema cognitivo y el juicio que de él se deriva está sesgado, hay elementos suficientes para poner en duda el fundamento medular de la teoría económica más ortodoxa, esto es, la teoría de la elección racional y, por ende, de la maximización de la utilidad esperada.

⁸ En otro momento del mismo trabajo (93) afirma que «No existe ninguna forma lógica de llegar a la conclusión de que los mercados transforman a las personas en agentes racionales».

⁹ Muchas de las actividades económicas, apuntan AKERLOF y SHILLER (2009, 10), lejos de tener motivaciones racionales, están gobernadas por *espíritus animales* (*animal spirits*), ya que los «estímulos que mueven a las personas no siempre son económicos ni su comportamiento es racional cuando persiguen este tipo de intereses». Y la *confianza* (32 y ss.), que implica que «una conducta sobrepasa un acercamiento racional en la toma de decisiones», y su impacto en las fases recesivas de la economía (especialmente cuando se disipa) es particularmente ilustrativa de esta aproximación.

¹⁰ Al respecto también, SUNSTEIN (2017, 10 y ss.). Y, especialmente, cuando afirma (12): «los descubrimientos sobre el comportamiento humano crean problemas para el argumento epistémico porque muestran que la gente comete muchos errores, algunos de los cuales pueden resultar extremadamente dañinos». Si bien es cierto que (13 y 14) «la mayor parte del tiempo, los mercados libres son la mejor salvaguarda contra los errores cognitivos», no puede olvidarse que «los mercados libres también pueden recompensar a los vendedores que explotan los errores humanos».

Ahora bien, (SUNSTEIN, 2012, 54) la existencia de estas limitaciones no «significa que el comportamiento de la gente es impredecible, sistemáticamente irracional, aleatorio, libre de reglas, o ajeno a los científicos sociales». Sino que responden a un determinado patrón que los sistematiza y los hace, en cierto sentido, previsibles: los sesgos. Y, como exponen SUNSTEIN y THALER (34 y 35) el conocimiento de nuestro sistema cognitivo ha posibilitado averiguar cómo inciden en nuestra forma de pensar.

En definitiva, esta primera reacción ha evidenciado que la economía tiene (severas) dificultades para seguir calificándose como la *física social*. Y, en paralelo, también ha alimentado al citado *paternalismo libertario* (y los *nudges* citados). Esta corriente de pensamiento, en definitiva, justifica una ingeniería social que, mediante estímulos externos, aspira a alterar el ambiente y, de este modo, orientar el comportamiento hacia resultados socialmente deseables (en este caso, *beneficiosos* para los propios sujetos – según sus promotores). En efecto, SUNSTEIN y THALER (20) entienden que el paternalismo libertario es un tipo de «paternalismo relativamente débil y blando, y que no supone una intromisión porque las opciones no se bloquean ni se eliminan, ni se gravan de forma significativa (...). El enfoque que recomendamos se considera paternalista porque los arquitectos de las decisiones públicos y privados no se limitan a intentar identificar o poner en práctica las decisiones que esperan de las personas. Más bien, intentan orientarlas en direcciones que mejorarán sus vidas. Dan nudges»¹¹.

La segunda reacción anunciada apuesta por acudir a la automatización para colmar las limitaciones humanas. En efecto, la estadística computacional (que no, todavía, *inteligencia* artificial – al menos, en el sentido humano del término) ha adquirido un protagonismo creciente en la toma de decisiones y aspira a relegar a un segundo plano al juicio humano. Y lo ha hecho, en especial, gracias a dos factores principalmente:

- por un lado, como si de un caballo de Troya se tratara, gracias a la concepción liberadora del esfuerzo que (como se ha descrito en los primeros capítulos) el avance de la técnica, a lo largo de los siglos, ha facilitado al ser humano; y,
- por otro lado (ZUBOFF, 397), gracias a los avances de la psicología y las evidencias sobre el carácter frágil e imperfecto del pensamiento humano (su racionalidad limitada) también recién expuestos.

No obstante, en relación a esto último, aunque es cierto que acude a una visión mecanicista del comportamiento similar a la que subyace en el paternalismo libertario y se nutre conceptual y materialmente de gran parte de sus avances científicos, lo cierto es que (como se desarrollará a lo largo de este trabajo) puede afirmarse que esta segunda reacción ha tomado su propia senda y proyecta algunos escenarios potenciales particularmente amenazantes.

¹¹ Llegados a este punto, conviene tener en cuenta que la arquitectura de la elección que impulsa esta corriente de la economía también ha sido objeto de crítica. Por ejemplo, entre otras objeciones, para GRANT (107), debe tenerse en cuenta que «cambia las condiciones que enmarcan nuestras elecciones pero, al contrario de lo que sucede con los incentivos, no incrementa nuestras opciones. De forma similar a los incentivos, pero de manera diferente a lo que sucede con los intercambios de mercado, la arquitectura de la elección intenta de forma deliberada influir en las elecciones de la gente, sin la transparencia asociada a los incentivos».

Estos dos factores están alimentando el desplazamiento de nuestra capacidad de juicio y decisión hacia procesos automatizados basados en la datificación masiva (o minería de la realidad)¹², la correlación estadística¹³ y el aprendizaje automático. Ante la convicción (probablemente exagerada) de que el ser humano tiene una tendencia irreprimible a tomar decisiones inconsistentes y/o perjudiciales para sí mismo, gracias a este *neopaternalismo* (o *solucionismo tecnológico* en términos de MOROZOV), se está consolidando la idea de que la tecnología es el nuestro mejor protector frente a uno mismo.

Además, se da la particularidad (como apunta ZUBOFF, 397) de que, a diferencia de la arquitectura de las decisiones propuesta de la economía conductual, en el contexto de lo que se conoce como *capitalismo de la vigilancia* (al que me referiré posteriormente), los nuevos *empujoncitos digitales* «buscan fomentar elecciones o decisiones que favorezcan al arquitecto, no al individuo que elige o decide».

En paralelo, son muchas las voces que han denunciado los riesgos asociados al empleo generalizado de la gestión algorítmica de nuestros juicios y decisiones, describiendo escenarios inquietantemente sombríos. Estos riesgos pueden sintetizarse como sigue¹⁴:

- Primero («*no free lunch*» e «*ignorancia objetiva*»): Es difícil que podamos codificar todos los datos y/o bien que todos los problemas sean computables y/o bien que las máquinas sepan interpretarlos adecuadamente (por ejemplo, tienen muchos problemas para comprender la ambigüedad y el contexto del lenguaje natural). En efecto, dado que no podemos recopilar todos los datos relevantes, debemos tener muy claro que, con una mera porción de los mismos, no se puede predecir el 100% de los resultados posibles (es lo que se conoce como el teorema «*no free lunch*» – o, en una posible traducción: «nadie de duros a cuatro pesetas»)¹⁵. Y esta limitación es particularmente acusada en lo que a la predicción del comportamiento humano se refiere, pues, se la da la circunstancia de que (KAHNEMAN, SIBONY, SUNSTEIN, 161), en este ámbito hay una gran cantidad de «ignorancia objetiva» (esto es, la imposibilidad objetiva de

¹² La era de los datos masivos (siguiendo con MAYER/CUKIER, 41) permite procesar ingentes cantidades de datos y, por este motivo, ya no tiene sentido recurrir a una muestra. De hecho, de lo que se trata es de recopilar todos los datos posibles y, cuando sea factible, absolutamente todos (incluso los erróneos). Esto es: «N = todo».

¹³ MAYER/CUKIER (72 y 73): Una correlación «cuantifica la relación estadística entre dos valores de datos. Una correlación fuerte significa que, cuando cambia uno de los valores, es altamente probable que cambie también el otro (...). Una correlación débil significa que, cuando un valor de datos cambia, apenas le ocurre nada al otro». Y, lo que es más importante: «al permitirnos identificar una aproximación realmente buena de un fenómeno, las correlaciones nos ayudan a capturar el presente y predecir el futuro: si A tiene lugar a menudo junto con B, tendremos que estar pendientes de B para predecir que va a ocurrir con A. Usar a B de aproximación nos ayuda a interpretar lo que probablemente esté ocurriendo con A, aunque no podamos medir A ni observarlo directamente (...). Las correlaciones no pueden vaticinar el futuro, solo predecirlo con cierta probabilidad. Pero esa cualidad resulta extremadamente valiosa».

¹⁴ En este sentido, sintetizo parcialmente una parte de lo expuesto en BELTRAN DE HEREDIA RUIZ (2022).

¹⁵ Entender lo contrario supondría caer en una falacia estadística, pues (MLODINOW, 43), «ningún número finito de pruebas parciales sumará nunca una certeza porque para componer probabilidades no las sumas: las multiplicas».

acceder a datos importantes limita la exactitud alcanzable de cualquier predicción)¹⁶. Sin olvidar que (MLODINOW, 12) el perfil de nuestras vidas (y nuestro destino) se va construyendo alimentado por hechos aleatorios que nos afectan y nuestras sucesivas reacciones a los mismos. En paralelo, en la medida que los algoritmos se alimentan de datos del pasado, pueden tener problemas para detectar irregularidades o hechos que están *fuera de muestra*. Y tampoco puede obviarse que (SILVER, 24 y 25) el aumento de datos provoca sobrecarga de información y, con ella, se incrementa el ruido, las hipótesis a evaluar y el número de correlaciones débiles (o espurias), que pueden llevar a pronósticos desviados.

- Segundo («*garbage in, garbage out*»): Hay numerosas pruebas (DU SAUTOY, 2020, 119) de que los algoritmos cobijan *sesgos ocultos*. Una preocupante derivada del fenómeno que se sintetiza con la expresión: *entra basura, sale basura* (*garbage in, garbage out* – «*GIGO*»). Como apunta BRIDLE (160), asumiendo que estas tecnologías dependen por completo de la información histórica que utiliza como datos de entrenamiento (con todos sus sesgos y discriminaciones), supone «codificar esa barbarie en nuestro futuro».

– Tercero (*cajas negras tecnológicas*): Las lógicas internas de algunos algoritmos (como los de aprendizaje profundo) se caracterizan por su opacidad. En estas *cajas negras* (O’NEIL), el rastreo y monitorización de las decisiones adoptadas es, en el mejor de los casos, *difícil* (¡incluso para sus propios creadores!)¹⁷. De ahí que para este tipo de tecnología se clame por la implantación de instrumentos de transparencia que permitan su *traducción* al lenguaje humano. Sin esta transparencia y comprensibilidad, especialmente exigible en los algoritmos de aprendizaje profundo (DU SAUTOY – 2020, 120 – habla de *metalenguaje*), difícilmente, seremos capaces de *entender* sus decisiones (y acabar siendo víctimas de las mismas – O’NEIL o FRY).

– Y, cuarto («*Minority Report*»): La dependencia de las correlaciones (MAYER/CUKIER, 188, 216 y 217) puede llevarnos a articular sanciones (o castigos) a partir de meras propensiones alejadas de la causalidad. Esto es, en vez de ser «responsables por nuestras acciones pasadas», a partir de predicciones estadísticas de unas acciones futuras, podemos «acabar juzgando a las personas no por lo que han hecho, sino por lo que hemos predicho que harían». Y es obvio que esto supondría un

¹⁶ Además, como expone PINKER (272, citando a DENETT y CLEG), «la conducta humana es intrínsecamente impredecible no solo a causa del ruido neuronal aleatorio en el cerebro, sino también como una adaptación que torna más difícil que nuestros rivales se anticipen a nosotros».

¹⁷ Y, llegados a este punto, creo que es importante detenerse en este aspecto. Siguiendo con PEARL/MACKENZIE (367), en el aprendizaje profundo, se utilizan «redes neuronales convolucionales. Estas redes no se ajustan a las reglas de probabilidad; no lidian con la incertidumbre de una forma rigurosa y transparente. Menos aún incorporan una representación explícita del entorno en el que operan. En lugar de esto, la arquitectura de la red tiene libertad para evolucionar por sí misma. Cuando se acaba de instruir una nueva red, quien la programa no tiene ni idea de qué computaciones está realizando o por qué funcionan. Si la red falla, tampoco tenemos ni idea de por qué». O, siguiendo con PINKER (135): «a diferencia de la inferencia lógica, el funcionamiento interno de la red neuronal es inescrutable. La mayor parte de los millones de unidades ocultas no representan ningún concepto coherente que podamos comprender, y los científicos computacionales que las entrenan no son capaces de explicar cómo llegan a una respuesta particular. Por eso muchos críticos de la tecnología temen que, a medida que se encomiendan a los sistemas de inteligencia artificial decisiones relativas al destino de las personas, estos puedan perpetuar sesgos que nadie pueda identificar y extirpar».

replanteamiento profundo sobre la responsabilidad individual (y, condicionaría de forma determinante nuestra libertad).

El efecto combinado de estos cuatro factores (y su particular capacidad destructiva) debería condicionar y/o modelar la expansión de esta tecnología. No obstante, también debemos ser muy conscientes de que, probablemente, dado su potencial, ya no estamos en condiciones de renunciar a su uso (del mismo modo que tampoco podemos hacerlo con respecto a otros métodos que nos ayudan a mejorar nuestros juicios: como reglas matemáticas y/o estadísticas sencillas y, obviamente, también las más complejas). En efecto, como se expondrá a continuación, existen poderosas razones para pensar que, lejos de reprimirla, debemos potenciar el empleo de esta automatización (siempre que seamos capaces de hacerlo *con inteligencia*).

4. El *homo sapiens* no es infalible: la gestión algorítmica (oportunamente utilizada por la inteligencia humana) puede ayudarnos

El contenido del epígrafe anterior describiría un escenario particularmente sombrío si, víctimas de una ceguera profunda y/o de una ingenuidad aguda, dejáramos que esta tecnología se expandiera sin límite alguno. A la espera de la definición del terreno de juego que la regulación (al menos) a nivel comunitario nos depare sobre la inteligencia artificial, y sin tratar de minusvalorar la gravedad de los riesgos que nos acechan, creo que también conviene tener en cuenta algunos factores sobre la condición humana que podrían justificar el uso (responsable) de esta tecnología:

- Primera (*cajas negras biológicas*): Tendemos a valorar la capacidad (limitada) de los instrumentos computacionales (se les ha denominado como *sabio-idiotas*) partiendo del *supremacismo humano* y, en particular, nuestra infalibilidad.

No obstante, aunque el ser humano es maravilloso y, como especie, hemos sido capaces de alcanzar cotas inimaginables, tampoco podemos olvidar que no somos inmunes al error (ni tampoco a otros comportamientos particularmente aberrantes y lesivos para nosotros mismos y nuestros semejantes). Además, aunque los avances de la neurociencia son deslumbrantes, apenas hemos sido capaces de rascar en la superficie y, por ello, nuestro cerebro sigue siendo una caja negra. De ahí que muchos de nuestros juicios y comportamientos (incluso, los adoptados a nivel *consciente*), son inescrutables para terceros (e, incluso, pueden serlo para nosotros mismos).

Este *oscurantismo*, como se acaba de exponer, también está presente en algunos algoritmos (en especial, los llamados de aprendizaje profundo) y es claro que debemos aunar esfuerzos para exigir su transparencia y así posibilitar la comprensibilidad humana y, por consiguiente, su exhaustiva monitorización (en este sentido, toda cautela es poca).

Para el resto (es decir, aquellos cuya *secuencia decisoria* sea rastreable), la delegación de ciertas funciones humanas a las máquinas podría contribuir a dar luz sobre juicios cuya explicabilidad se ve imposibilitada por la insondabilidad de nuestro cerebro. Así pues, si dejamos que nos complementen, podremos aspirar a una mayor transparencia en la toma de decisiones y, de este modo, combatir, por ejemplo, la arbitrariedad, la

discriminación y/o la disparidad de criterios sin un motivo justificado (y que puede calificarse como *ruido*¹⁸).

En todo caso, a pesar de lo recién apuntado, soy plenamente consciente que queda un largo camino por recorrer y es obvio que son muchos los obstáculos que todavía deben ser superados para llegar a este (hipotético y/o idílico) estadio (quedan muchas cosas por *pulir*).

- Segunda (*bolas de cristal defectuosas o chimpancés lanzando dardos*): tendemos a desconfiar de la capacidad predictiva de los algoritmos. Como ya se ha avanzado, existe un consenso generalizado alrededor de esta limitación, especialmente porque basan sus proyecciones a partir de datos del pasado. Y, obviamente, esto dificulta que puedan anticipar fenómenos novedosos o que quedan *fuera de muestra* (como, por ejemplo, los llamados *cisnes negros*). Y, por todo ello, no son pocas las voces que claman por recelar de su uso.

No obstante, también olvidamos (con demasiada frecuencia) que, el cerebro humano funciona de un modo similar y, por consiguiente, puede ser víctima de idénticos (e, incluso, peores) errores. Así, en respuesta a estos posicionamientos más escépticos respecto de la fiabilidad de esta tecnología, conviene detenerse brevemente sobre los problemas para confiar plenamente en la capacidad predictiva humana.

Por ejemplo (TETLOCK, 74 y 75), «nos empeñamos en buscar orden en las secuencias aleatorias» por nuestra renuencia a reconocer la imprevisibilidad. Esto ilustra (MLODINOW, 205) el conflicto fundamental entre «nuestra necesidad de sentir que tenemos el control y nuestra capacidad de reconocer la aleatoriedad», hasta el punto de que «interfiere la percepción correcta de sucesos aleatorios». Esto se debe a que los seres humanos interpretamos el mundo en términos mecanicistas, es decir (GAZZANIGA, 198), asumiendo que todos los acontecimientos tienen sus antecedentes causales. Y no podemos evitar buscar la causalidad de forma automática e inconsciente. Como expone KAHNEMAN (104), tenemos la necesidad de encontrar coherencia a lo que sucede a nuestro alrededor. Posteriormente, nuestro *yo consciente* recibirá y aceptará esta interpretación causal, incluso, justificándola, actuando (HAIDT, 79) como un portavoz a pesar de que no sepamos exactamente por qué nos hemos comportado de un determinado modo¹⁹. Así pues, (GAZZANIGA, 260) en el seno de la organización medular de nuestro sistema cognitivo, tenemos una capacidad especial que interpretar constantemente teorías sobre las «relaciones causales entre los acontecimientos elementales que ocurren en el interior y en el exterior de nuestras cabezas». Diversas investigaciones han desvelado que el hemisferio izquierdo es el que posibilita la

¹⁸ Este concepto empleado aquí se refiere (KAHNEMAN, SIBONY y SUNSTEIN, 33) a una variabilidad no esperada de juicios sobre el mismo problema (esto es, una dispersión aleatoria indeseada). Lo que, por ejemplo, en el ámbito de las resoluciones dictadas por los órganos judiciales o el personal médico en sus diagnósticos, resulta particularmente preocupante.

¹⁹ Es en este momento cuando entra en juego nuestra tendencia a buscar argumentos que preferencialmente ratifiquen esta idea (sesgo de confirmación) – en vez, de buscar argumentos que traten de falsearla. En estas situaciones (SILVER, 118), disponer de más información puede que no implique una mejora en la predicción, pues, puede ser utilizada para reforzar una visión preconcebida. El sesgo de confirmación, lejos de aproximarnos a una mejora predictiva, alimenta la desviación.

capacidad de hacer inferencias²⁰. Este cableado de serie podría explicar por qué somos tan proclives a caer en la *falacia narrativa* (TALEB). En efecto, no tenemos ningún reparo en construir relaciones causales concluyentes a pesar de sufrir una escasez de información abrumadora²¹. Así pues, reparen que esta intuición humana de la causalidad (MAYER/CUKIER, 85) no contribuye a acrecentar nuestra comprensión del mundo: «en muchos casos, es poco más que un atajo cognitivo que nos depara una ilusión de percepción, cuando en realidad nos deja en la inopia respecto al mundo que nos rodea».

En el fondo, y en el marco de los Sistemas cognitivos 1 y 2 anteriormente descritos, este efecto se explica (KAHNEMAN, 117) por la «notable asimetría existente entre las maneras en que nuestras mentes [tratan] la información disponible en el momento y la información que no tenemos». En concreto (117 y 118), «el Sistema 1 se distingue por construir la mejor historia posible que incorpore ideas activadas en el momento, pero no tiene en cuenta (no puede tener en cuenta) la información que no posee. La medida del éxito del Sistema 1 es la coherencia de la historia que se ocupa en crear. La cantidad y la cualidad de los datos en los que la historia se basa son en gran parte irrelevantes. Cuando la información es escasa, cosa que comúnmente ocurre, el Sistema 1 opera como una máquina de saltar a las conclusiones. La combinación de un Sistema 1 que busca la coherencia y un Sistema 2 perezoso supone que el Sistema 2 aprobará muchas

²⁰ GAZZANIGA (118): «el hemisferio dominante, normalmente el izquierdo, está comprometido en la tarea de interpretar nuestros comportamientos externos, así como las respuestas emocionales más encubiertas producidas por los módulos mentales independientes. El cerebro izquierdo construye teorías para explicar por qué han ocurrido estos comportamientos, y lo hace así debido a la necesidad que tiene el sistema cerebral de hacer coherente la totalidad de nuestros comportamientos. Se trata de una empresa característicamente humana, y sobre la misma descansa no sólo el mecanismo que genera nuestro sentido de la realidad subjetiva, sino también la capacidad mental que nos libera de los controles que nos atan a las contingencias externas». En paralelo, desde el punto de vista fisiológico (GILBERT, 34 y 35), diversas evidencias muestran que el lóbulo frontal es la maquinaria cerebral fundamental que permite a los seres humanos proyectarse a sí mismos en el futuro. Sin el lóbulo frontal (como le sucedió al famoso caso de Phineas CAGE), estaríamos atrapados en el presente. Esta reclusión en el espacio y el tiempo inmediatos, nos impediría imaginar el mañana y, por ello, permaneceríamos despreocupados por lo que nos pudiera deparar el más allá del momento presente. Dicho de otro modo, el lóbulo frontal «capacita a los adultos humanos sanos para pensar en la existencia a largo plazo», es una especie de «máquina del tiempo que nos permite abandonar el presente y experimentar el futuro antes de que ocurra». De hecho, probablemente la cualidad que más nos distingue del resto de seres vivos es nuestra capacidad de imaginar «objetos y acontecimientos que no existen en el reino de la realidad, y esta habilidad es la que nos permite pensar en el futuro. Como dijo un filósofo, el cerebro humano es una ‘máquina de anticipación’ y ‘crear futuro’ es lo más importante que hace». Y, pudiendo pensar en el pasado o el presente, de forma predominante, nos proyectamos hacia el futuro (de hecho, invertimos materialmente mucho más tiempo en ello). Y lo hacemos fundamentalmente por dos motivos: primero, porque nos provee placer; y, segundo, porque nos permite protegernos a través de la prevención. Y centrándome en este segundo, pensar en el futuro nos permite adoptar medidas preventivas para evitar el dolor (GILBERT, 40 a 43). Y lo hacemos, no sólo porque también nos produce placer, sino porque, al simular el futuro, el cerebro trata de controlar las experiencias que va a tener y, por consiguiente, podemos hacer predicciones. De modo que, al anticiparnos, podemos tratar de *controlar* el futuro, identificando y proyectando las posibles alternativas y escogiendo los mejores destinos y evitando los peores.

²¹ El concepto de «falacia narrativa» propuesto por TALEB se refiere a la tendencia humana a resumir y simplificar y, en consecuencia, a reducir la dimensión de las cosas. En concreto (116): «Tal falacia se asocia con nuestra vulnerabilidad a la interpretación exagerada y nuestra predilección por las historias compactas sobre las verdades desnudas, lo cual distorsiona gravemente nuestra representación mental del mundo; y es particularmente grave cuando se trata del suceso raro». O, como apunta KAHNEMAN (262), esta noción trata de describir cómo «historias dudosas del pasado conforman nuestras opiniones sobre el mundo y lo que esperamos del futuro. Las falacias narrativas surgen inevitablemente de nuestro continuo intento de dar sentido al mundo».

creencias intuitivas que reflejan directamente las impresiones generadas por el Sistema 1 (...). Y este salto a las conclusiones sobre la base de una evidencia limitada es determinante para entender el pensamiento intuitivo y KANHEMAN lo sintetiza con la frase «*what you see is all there is*» (*WYSIATI* – abreviatura de las iniciales): «lo que vemos es todo lo que hay». En definitiva, «el Sistema 1 es radicalmente insensible a la cualidad y a la cantidad de información que da lugar a las impresiones y las intuiciones».

Y la regla *WYSIATI* también podría explicar por qué olvidamos tan frecuentemente que «no sabemos que no sabemos». En efecto, como apunta KAHNEMAN (263 y 264): «La limitada información de que disponemos no puede bastarnos, porque en ella no está todo. Construimos la mejor historia posible partiendo de la información disponible, y si la historia es buena, la creemos. Paradójicamente, es más fácil construir una historia coherente cuando nuestro conocimiento es escaso, cuando las piezas del rompecabezas no pasan de unas pocas. Nuestra consoladora convicción de que el mundo tiene sentido descansa sobre un fundamento seguro: nuestra capacidad casi ilimitada para ignorar nuestra ignorancia».

Es muy frecuente que, a toro pasado, *sepamos* que las cosas iban a suceder (por ejemplo, una crisis financiera, el estallido de un conflicto bélico, los resultados electorales o el divorcio de unos amigos), pero quienes pensaban que sucederían estas cosas, no podían probarlo de forma concluyente, de modo que, en muchas ocasiones, se trata de acontecimientos que no son predecibles. Así, se alcanza una apreciación especialmente relevante (264): «Lo perverso del uso del verbo saber en este contexto no es que algunas personas creyeran en una presciencia que no poseen, sino que el lenguaje supone que el mundo es más cognoscible de lo que realmente es. Ello contribuye a perpetuar una perniciosa ilusión. El núcleo de la ilusión es que creemos entender el pasado, lo cual supone que también el futuro puede conocerse, pero la verdad es que entendemos el pasado menos de lo que creemos».

Lo descrito es la base de lo que se conoce como el sesgo de la retrospectión: tendemos a revisar la historia de nuestras creencias pasadas a la luz de acontecimientos reales (ya acaecidos). Transformamos la información escasa y parcial que somos capaces de captar de lo que ya ha pasado y la transformamos en *señales* clarividentes que nos permiten *identificar* la secuencia de causas que explican *irrefutablemente* nuestro presente. En definitiva (MAUBOUSSIN, 64), creamos una narrativa (que incluye el cómo y el por qué) de lo que ha sucedido. Y, a resultas del efecto combinado de estos factores, tenemos la poderosa ilusión de que nuestra capacidad de predicción está por encima de nuestra aptitud real. De modo que (KAHNEMAN, 268) «la maquinaria del Sistema 1, que a todo da sentido, nos hace ver el mundo más ordenado, predecible y coherente de lo que realmente es. La ilusión de que uno ha entendido el pasado alimenta la ilusión de que puede predecir y controlar el futuro. Estas ilusiones son reconfortantes. Reducen la ansiedad que experimentaríamos si reconociéramos francamente las incertidumbres de la existencia. Todos tenemos necesidad del mensaje tranquilizador de que las acciones tienen consecuencias previsibles y de que el éxito recompensará la prudencia y el valor. Muchos estatutos empresariales están hechos para satisfacer esta necesidad».

Lo que, a fin de cuentas, no deja de ser una derivada de lo que en el ámbito de la filosofía de la ciencia se conoce como el problema de la inducción (o problema del

conocimiento inductivo)²². El hecho de que nuestro conocimiento se induzca a partir de datos pasados hace que seamos incapaces de predecir los grandes fenómenos disruptivos personales o, incluso, fracturas históricas.

No cabe duda que lo anterior cuestiona el aprendizaje que nos puede ofrecer *la Historia* respecto del devenir futuro (KAHNEMAN, 285 y 286): «nuestra tendencia a construir - y creémoslas - narraciones del pasado hace que nos resulte difícil aceptar los límites de nuestra capacidad predictiva. Todo cobra sentido en una retrospectiva, un hecho que los entendidos en finanzas explotan cada mañana cuando dan cuenta convincente de los acontecimientos del día. Y no puede sorprendernos la poderosa intuición de que, en una retrospectiva, lo que hoy confiere sentido era ayer predecible. La ilusión de que entendemos el pasado fomenta el exceso de confianza en nuestra capacidad para predecir el futuro». Como apunta TETLOCK (72), «puede que la historia se repita, pero no lo hace mecánicamente. Sin embargo, gran parte del razonamiento analógico basado en la historia es mecánico».

Y, el papel de los expertos en la construcción de estos razonamientos y de las consiguientes predicciones es controvertido (por su discutible provecho), especialmente cuando se tratan de proyectar escenarios a largo plazo. En particular, porque crean apreciaciones inadecuadas de la incertidumbre y, consiguientemente, acaban favoreciendo el exceso de confianza; y, con él, precipitando cegueras colectivas sobre el riesgo. Las conclusiones de TETLOCK (80) al respecto son tan contundentes como clarividentes: «A medida que crece el saber experto, la seguridad de los analistas aumenta más rápido que la precisión de sus pronósticos». Dos factores contribuirían decisivamente a este efecto:

- El primero es que, como apunta KAHNEMAN (343), «el exceso de confianza es otra manifestación del *WYSIATI*: cuando estimamos una cantidad, nos basamos en información que recordamos y construimos una historia coherente en la que la estimación tiene sentido. Tener en cuenta la información que no recordamos - o que quizá nunca hemos conocido - es imposible». En todo caso, en descargo de los expertos (289), «los errores de predicción son inevitables porque el mundo es impredecible». No obstante, de lo recién expuesto tampoco deberíamos alcanzar conclusiones precipitadas. Como apuntan KAHNEMAN, SIBONY y SUNSTEIN (162), «es cierto que afirmar que el futuro es imprevisible no es un avance conceptual. Sin embargo, la obviedad de este hecho solo se ve igualada por la regularidad con la que se ignora, como demuestran los continuos hallazgos de exceso de confianza en la predicción».

- El segundo es el conocido sesgo de representatividad (y la consiguiente tendencia a olvidar las tasas base²³). En virtud del mismo (KAHNEMAN, 199), tendemos a juzgar y

²² Siguiendo a TALEB (87) este problema se refiere a la dificultad de extraer conclusiones generales a partir de la observación de casos específicos. O, «¿cómo podemos entender las propiedades de lo desconocido (infinito) basándonos en lo conocido (finito)». De lo que se extrae que, como expone McINTYRE (60), sintetizando el planteamiento de POPPER, «ni siquiera cuando una teoría tiene éxito, puede ser aceptada como verdadera - o como aproximadamente verdadera -, sino que ha de ubicarse siempre en el purgatorio de haber sobrevivido 'hasta ahora'. Por poderosas que sean las pruebas científicas, lo único que nos queda al final es un número potencialmente infinito de hipótesis que *podrían* encajar con los datos y un acopio infinito de evidencia empírica posible que *podría* derribar cualquier teoría».

²³ Un ejemplo sencillo puede ayudar a ilustrar este importante concepto: imaginen que tienen que sacar una bola de una urna y les preguntan si será roja o verde. Pues bien, para poder responder con algo de

a hacer predicciones a partir de juicios de similitud y de estereotipos que tengamos disponibles en la mente. De modo que (MLODINOW, 34) si las informaciones a nuestro alcance encajan con la imagen mental de algo, un incremento de las mismas nos induce a aumentar su certidumbre, aunque pueda disminuir su probabilidad (y, por ende, la predicción subsiguiente se desvíe significativamente)²⁴. De ahí que TETLOCK (92) afirme que «la clave decisiva de los juicios humanos de probabilidad no sería con qué frecuencia ocurren los acontecimientos de este tipo; sino más bien ‘¿con qué facilidad nos vienen a la cabeza escenarios causales convincentes?’ ». Con la particularidad de que los expertos al saber mucho sobre una materia son víctimas de una forma de arrogancia peculiar (103): «tienen tanto conocimiento de casos específicos a su alcance, son tan hábiles a la hora de ordenar ese conocimiento para construir situaciones de causa-efecto convincentes, que terminan convenciéndose a sí mismos de que hay que asignar probabilidades extremas que se alejan de las probabilidades objetivas medias»²⁵.

A la luz de lo anterior, no debería sorprender que las predicciones de los analistas más insignes no salgan bien paradas (TETLOCK, 47, 91, 93 y 94): su capacidad predictiva es, a duras penas, superior a la de unos chimpancés lanzando dardos... (es decir, a la de generadores aleatorios de predicciones)²⁶. Y, además, en la medida que estos expertos quedan lejos de los resultados que ofrecen ciertos algoritmos más complejos y sofisticados, su habilidad predictora está más cerca del «chimpancé que de los modelos estadísticos formales»²⁷.

Esta conclusión, por otra parte, tampoco debería sorprendernos. En especial porque, precisamente, el ser humano acudió a la estadística para gestionar el riesgo y la incertidumbre y, de este modo, tratar de superar nuestras carencias anticipatorias *de serie*. Y, podemos acudir a *reglas numéricas* más o menos complejas para superar estas limitaciones. Especialmente porque (MLODINOW, 164 y 165), aunque la vida de las personas es impredecible y varían imprevisiblemente, «cuando se recogen datos de grupos y se analizan en masa, surgen patrones regulares». De hecho (como se ha apuntado), los algoritmos, por el momento, no dejan de ser una expresión exuberante de

precisión es necesario saber cuántas bolas hay de cada color. De modo que la «proporción de bolas de un tipo particular recibe el nombre de tasa base». Si no tenemos más información, acudir a la tasa base (aunque no la conozcamos con exactitud) es el mejor recurso para dar respuesta a una pregunta o hacer una predicción. Por ejemplo, si nos preguntaran qué es más probable que una persona joven de 19 años estudie ciencias de la educación o matemáticas, la tasa base es la que nos permitiría afirmar que lo más probable es que estudie lo primero porque hay más estudiantes de esta especialidad que de matemáticas

²⁴ La predicción por representatividad (KAHNEMAN, 200 a 205) no es estadísticamente óptima. Específicamente porque se atribuye una probabilidad a un acontecimiento mucho más alta de lo que es. Y esto lo hacemos de forma extremadamente rápida (por la intervención del Sistema 1). No obstante, en muchos casos «hay cierto grado de verdad en los estereotipos que guían los juicios de representatividad, y las predicciones que siguen esta heurística pueden ser exactas. En otras situaciones, los estereotipos son falsos, y la heurística de la representatividad inducirá a error, especialmente si hace que las personas ignoren la información de la tasa base, que apunta en otra dirección».

²⁵ De hecho, del análisis llevado a cabo por TETLOCK (115) también se desprende que los analistas más conocidos (a partir de los resultados de búsqueda en Google) y, por consiguiente, los que tienen más papeletas para ser ensalzados por la prensa (y ser consultados por las instituciones), «ajustaron peor sus resultados que sus colegas de perfil más bajo».

²⁶ Aunque los expertos humanos aventajan a estos generadores aleatorios (91) «al asignar mayores probabilidades a los acontecimientos que ocurren que a los que no lo hacen».

²⁷ En concreto, el estudio de TETLOCK (93) sugiere que así es cuando se compara el juicio humano con los resultados extraídos de «algoritmos de extrapolación basados en casos específicos».

estadística computacional. Por este motivo (como hemos estado haciendo en el pasado reciente con versiones menos potentes de estas herramientas), pueden sernos de gran ayuda para identificar regularidades y hallar cierto *orden* frente a escenarios inciertos; eso sí, siempre que no se empleen de forma maximalista (el teorema *no free lunch* así lo ha evidenciado) y, además (MLODINOW, 192), no se olvide que los acontecimientos también pueden ser producto del azar. En definitiva, aunque pueda suscitar alguna resistencia admitirlo (KAHNEMAN, SIBONY y SUNSTEIN, 371), «aunque es poco probable que un algoritmo de predicción sea perfecto en un mundo incierto, puede ser mucho menos imperfecto que el ruidoso y, a menudo, sesgado juicio humano».

Así pues, a pesar de estos *hándicap*, no creo que sea *inteligente* que descartemos su uso (o lo vilipendamos). *Simplemente* se trata de no olvidar estas limitaciones y, como se ha apuntado (y tratando de acercarse a un hipotético punto de equilibrio), complementar esta tecnología con *inteligencia* humana.

Aunque los algoritmos son calificados como *sabio-idiotas* y, probablemente, lo sean, para ser más justos, esta valoración debería matizarse, pues, desde el instante que olvidamos sus limitaciones (y les concedemos todo el poder decisorio – o de forma excesiva), la estupidez única y exclusivamente debería atribuirse al ser humano.

- Tercera (*familiarizándonos con ellos - y los algoritmos del amor*): En la actualidad, no sufrimos aversión algorítmica (al menos, en un primer momento y mientras no nos fallen). Lejos de desconfiar de ellos, consciente o inconscientemente, estamos muy familiarizados con ellos. En efecto, no sé si han reparado que hemos normalizado su uso en muchas y heterogéneas facetas de nuestra vida.

Permítanme que me detenga en una de ellas: encontrar pareja sentimental (¿con relativo éxito?). Aunque todavía me pregunto cómo hemos llegado a este extremo (y si no estamos siendo víctimas de una candidez preocupante), creo que podemos inferir que, si confiamos en la fiabilidad de estos *algoritmos del amor* (por llamarlos de algún modo) para complementar una dimensión de nuestra vida personal tan sumamente compleja y con implicaciones tan humanamente profundas, ¿por qué no podemos pensar que pueden complementarnos en otras dimensiones (más prosaicas)?

- Cuarta (*un futuro más predecible*): lejos de la *utopía de certeza* a la que se refiere ZUBOFF y que preconizan algunos devotos de la gestión algorítmica, lo cierto es que esta automatización puede ayudarnos a articular modelos del futuro mejores de los que somos capaces de crear por nosotros mismos. Como apuntan KAHNEMAN, SIBONY y SUNSTEIN (160), «una tarea sólo puede considerarse imposible después de que muchos actores creíbles la hayan intentado y hayan fracasado». Partiendo de la base de que la suma mecánica de información es, en muchas ocasiones, superior al juicio humano (por ejemplo, así se ha evidenciado en algunos juicios clínicos), los algoritmos pueden proporcionar «una prueba mejor de lo intrínsecamente predecibles o impredecibles que pueden ser los resultados» de cualquier fenómeno que aspiremos a anticipar. Aunque están lejos de la predicción perfecta, hay motivos suficientes para afirmar que los juicios predictivos de los algoritmos pueden ser superiores a los que nunca seremos capaces de alcanzar. En cualquier caso, conviene dejar claro que esto último no significa que los algoritmos sean, de forma abrumadora, mejores que los juicios predictivos humanos. Aunque estos modelos son sistemáticamente mejores, no

lo son mucho más: «no hay en lo fundamental pruebas de situaciones en las que, con idéntica información, las personas lo hagan muy mal y los modelos lo hagan muy bien».

En definitiva, no parece que podamos imaginar un futuro prescindiendo por completo de estas innovaciones (de hecho, posiblemente no sería sensato). Al contrario, lo más probable es que cada vez esté más profundamente imbricada en nuestras vidas. Y, precisamente, por este motivo, debemos exigir un uso responsable y conforme a los parámetros de un Estado de Derecho. La cuestión (que será objeto de análisis en el epígrafe que sigue) es si esta potente tecnología, a partir de estos juicios predictivos, está desvelando dimensiones de nuestra conducta hasta la fecha inescrutables y, por consiguiente, debemos articular instrumentos de protección jurídica, paradójicamente, frente nosotros mismos (y, obviamente, frente a quienes aspiran a aprovecharse de este conocimiento).

5. Psicometría y derechos del «yo inconsciente de las personas» (YIP)

La profesora Lisa FELDMAN BARRETT en su fantástico libro *La vida secreta del cerebro* afirma (306) «entre neurociencia y el sistema jurídico hay una gran falta de sincronización en cuestiones fundamentales sobre la naturaleza humana. Estas discrepancias se deben abordar si queremos que el sistema jurídico siga siendo uno de los logros más importantes de la realidad social, y si queremos seguir protegiendo los derechos inalienables de las personas a la vida, la libertad y la búsqueda de la felicidad».

La necesidad de esta simbiosis podría acrecentarse con la irrupción de la estadística computacional y los algoritmos. Especialmente porque lo cierto es que, *de algún modo*, estas herramientas están contribuyendo a desvelar ciertos patrones de nuestro comportamiento. En efecto, (DU SAUTOY, 110 y 111) «los algoritmos de aprendizaje profundo reconocen 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».

Precisamente, la psicometría del ser humano (a la que me referiré en el epígrafe que sigue), y que está siendo posible gracias a la datificación (o minería de la realidad), es la herramienta que podría estar posibilitando esta particular *desencriptación* de nuestro comportamiento más profundo. De hecho, la comprensión del cerebro humano se ha erigido en uno de los últimos bastiones a conquistar por parte de la economía de mercado y, en particular, de lo que se ha venido a llamar (ZUBOFF), el *capitalismo de la vigilancia*. En una nueva versión *mejorada* del conductismo de Burrhus F. SKINNER (en este caso, calificable como *severo y digital*), se aspira a conducir el comportamiento y controlar los deseos mediante *estimaciones* externas sofisticadas.

Para contextualizar esta aproximación, debe tenerse en cuenta que John B. WATSON, de quien SKINNER fue su discípulo más aventajado, escribió en 1913 (siguiendo el relato de GRANT, 74) un artículo en el que abogaba por una nueva ciencia de la psicología, concentrada «únicamente en los aspectos observables del comportamiento humano y no en entidades que no podían observarse, como la conciencia (...)». De este modo el «comportamiento se concebía como un proceso en el que organismo se

adaptaba al medio. Y debido a que ese proceso (...) no estaba gobernado por las fuerzas internas del instinto, la investigación conductual debía centrarse en cuestiones de motivación y aprendizaje». El conductismo, en definitiva, aboga por la negación de la mente y propone (JOHNSON-LAIRD, 23) que la psicología debe convertirse en una ciencia objetiva, que tiene que controlar y predecir la conducta y, por este motivo, debe olvidar todo lo concerniente a la conciencia y a la mente, pues, son temas impropios de la investigación científica.

Así, abandonado el estudio de los procesos mentales y de las técnicas introspectivas, esta corriente de la psicología opta por la observación controlada de respuestas en el laboratorio. Y, como expone GRANT (231), en esencia, parte de dos ideas fundamentales: que la naturaleza de las personas es pasiva y moldeable; y que son las motivaciones extrínsecas las que exclusivamente disciplinan su conducta (y, por ende, ignorando completamente la influencia que puedan tener las motivaciones intrínsecas²⁸). De modo que el efecto combinado de ambos permite que el control de los estímulos presentes en el medio posibilite el del organismo en cuestión.

De hecho, el mismo WATSON (tal y como lo recoge GRANT, 77) llegó a afirmar (con carácter profético a la luz de la aparición y uso generalizado de algunas aplicaciones existentes en el mercado) que «llegará el día en que contemos con una ética conductista, basada en experimentos y capaz de recomendarnos, atendiendo a las adaptaciones pasadas y futuras de los individuos, cuándo resulta recomendable casarse con una o con varias mujeres, o cuándo resulta recomendable aplicar la pena de muerte por un delito o un castigo de otro tipo... estas recomendaciones dependerán del mayor o menor ajuste entre el individuo y el medio».

Así pues, como se expondrá a continuación, hay quien piensa que la tecnología nos estaría brindando la oportunidad para que este sueño (finalmente) se corporice.

5.1. Psicometría y patrones conductuales emergentes

La psicometría (HAN, 35 a 39), también conocida como *psicografía*, es un «procedimiento basado en datos para obtener un perfil de personalidad». El potencial de estas herramientas radica en el hecho de que estos perfiles «psicométricos permiten predecir el comportamiento de una persona mejor de lo que podría hacerlo un amigo o un compañero». De hecho, «con suficientes datos, es posible incluso generar información más allá de lo que creemos saber de nosotros mismos».

Y, en este contexto, el smartphone (aunque no sólo) juega un papel absolutamente medular, pues, como *informante* muy eficiente que vigila permanentemente a su usuario, lo convierte en «un dispositivo de registro psicométrico que alimentamos con

²⁸ La distinción entre las motivaciones intrínsecas de las extrínsecas puede ser pertinente en este estadio (GRANT 230 y 231): «una persona está motivada de forma intrínseca cuando realiza una actividad por el beneficio que le reporta la actividad misma, mientras que su motivación es extrínseca cuando realiza una actividad por un beneficio que no está conectado con la actividad misma. Los motivadores intrínsecos son enteramente internos – la curiosidad es posiblemente el mejor ejemplo -, así aquellas personas con un alto grado de motivación intrínseca están fuertemente automotivadas. Los ‘motivadores extrínsecos’ son externos, como los halagos, los premios» y los incentivos.

datos día a día, incluso cada hora»²⁹. 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».

Siguiendo con HAN (22 y 23), a partir de la idea de Walter BENJAMIN, que atribuía a la cámara cinematográfica la capacidad de acceder de forma especial al inconsciente de las personas (denominándola el *inconsciente óptico*), pues, los primeros planos y la cámara lenta permitían visibilizar «micromovimientos y las microacciones que escapan al ojo humano», en la actualidad, el big data y la estadística computacional operan como una «lupa digital que descubre el inconsciente oculto del agente tras el espacio consciente de la acción».

En efecto, los productos y servicios que proveen lo que se conoce como *capitalistas de la vigilancia* (como expone ZUBOFF a lo largo de su obra) son el *gancho* para extraer de nosotros los datos de nuestro comportamiento (el *excedente conductual*). Estos datos son utilizados para *fabricar* productos predictivos que son vendidos en mercados de futuros conductuales a las empresas clientes.

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 desvelan aspectos íntimos e imperceptibles de nuestra conducta (ZUBOFF, 256 y ss.). No obstante, en la actualidad, la antropometría digital a la que estamos subyugados, a través del flujo continuo de datos conductuales, nos somete a una intromisión profunda y sin precedentes, principalmente, con el siguiente propósito: una *publicidad dirigida* (96).

Siguiendo con ZUBOFF (125, 182 y 184), el éxito de la estrategia predictiva con fines publicitarios de Google, primero, y Facebook, más tarde, ha atraído a infinidad de competidores a «un ciclo de vida cada vez más implacable de secuestro de experiencia humana». Esta creciente competitividad entre los capitalistas de la vigilancia ha desembocado en una carrera sin cuartel por extraer más y más profundos excedentes conductuales (son - MAYER/CUKIER, 142 - el nuevo *polvo de oro* para fabricar lingotes).

Los incentivos que esta dinámica genera son de una extrema perversidad, especialmente porque retroalimentan una espiral de saqueo continuo. El botón no sólo es tentador (los beneficios de los *pioneros* han sido astronómicos), sino que abstenerse es prácticamente una condena a la inanición empresarial. Exprimiendo y rebañando esta información, se aspira a mejorar las técnicas predictivas y, así, satisfacer los intereses de terceros: los usuarios nos hemos convertido «en fines para los intereses de otros»³⁰. Y cuantos más

²⁹ Según los datos que recoge WILLIAMS (94), «el usuario medio consulta su teléfono móvil 150 veces al día (y lo toca más de 2.600 veces al día)».

³⁰ En todo caso, como apunta ZUBOFF (505), la inteligencia de máquinas «caza furtivamente nuestro comportamiento en busca de un excedente y deja tras de sí todo el significado alojado en nuestros cuerpos, cerebros y latidos, de un modo no muy diferente a como los furtivos dejan tras de sí los cadáveres de los elefantes tras vaciarlos de marfil. Olvídense del tópico de que si algo es gratis ustedes son el producto. Ustedes no son el producto; son el cadáver abandonado. El 'producto' es lo que se fabrica con el excedente que han arrancado de sus vidas».

datos tengan de nosotros mejor para ellos. Pues (ZUBOFF, 472, 273 y 275), cuanto más precisas sean las predicciones, más elevado será su valor.

La hipercompetitividad entre las empresas para captar nuestra atención y así extraer un excedente conductual mayor es tan despiada que (como expone WILLIAMS, 53) los diseñadores de estas estructuras «no han tenido más remedio que apelar a lo más bajo de nuestra naturaleza – dando aún mayor prioridad a nuestros impulsos sobre nuestras intenciones – y explotar el catálogo entero de sesgos cognitivos para la toma de decisiones que la psicología y la economía conductual llevan compilando diligentemente durante las últimas décadas». La caza de la atención humana es *la presa* y la industrialización digital de mecanismos de persuasión por debajo del nivel consciente cada vez más eficientes es el camino para alcanzarlo³¹. Y la configuración de arquitecturas extractivas que precipiten hábitos de uso compulsivo difíciles de neutralizar por el usuario se ha convertido en el modelo de negocio dominante y la lógica del diseño de la red. Pero ya no sólo.

En efecto, el rastreo intensivo a través de las redes sociales y aplicaciones para móviles y ordenadores tiene, prioritariamente, un propósito publicitario y comercial. No obstante, 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. Esto explica 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 aumentada³².

No obstante, esta expansión no sólo se circunscribe al entorno exterior, sino que nuestro universo interior (personalidad, emociones complejas, etc.) también está siendo

³¹ WILLIAMS (54): «existe una floreciente industria de escritores y consultores dedicados a ayudar a diseñadores el último grito de la psicología conductual y apretar así los resortes adecuados de nuestros cerebros con la máxima eficacia y fiabilidad. Uno de los objetivos primordiales del diseño persuasivo es asegurarse de que el usuario vuelve repetidamente a un producto o, en otras palabras, inducir hábitos».

³² Como apunta PEIRANO (117), muchos datos son generados involuntariamente por personas desprevenidas cuyas acciones y movimientos son registrados minuciosamente por cámaras, micrófonos y sensores sin que se den cuenta. Rastreo que incluye (ZUBOFF, 115) datos que los usuarios han optado intencionadamente por no compartir. Y la ocultación de los datos en la sombra y el eufemismo empleado para mantener este secretismo es el *modus operandi* de estas corporaciones. A diferencia del capitalismo industrial (102 y 623), dependiente de la explotación y del control de la naturaleza, en el de la vigilancia, los seres humanos somos la materia prima del proceso. Lejos de servirnos han pasado a vigilarnos (121). De hecho (132), «atender a las necesidades auténticas de las personas es menos lucrativo y, por lo tanto, menos importante que vender las predicciones de [nuestras] conductas». Y, en esta particular «cadena de producción» captar nuestra atención es fundamental (aunque ya no será suficiente). Las notificaciones y la cuantificación son elementos del diseño que juegan con la ansiedad del usuario, creando pequeños *chutes dopamínicos* para provocar nuestra adicción. ¿El motivo? Incrementar el tiempo ante la pantalla, desvelar más datos en la sombra y, así, aumentar la inteligencia del aprendizaje de máquinas para *afinar* su predicción. En todo caso, como apunta ZUBOFF (505), la inteligencia de máquinas «caza furtivamente nuestro comportamiento en busca de un excedente y deja tras de sí todo el significado alojado en nuestros cuerpos, cerebros y latidos, de un modo no muy diferente a como los furtivos dejan tras de sí los cadáveres de los elefantes tras vaciarlos de marfil. Olvídense del tópico de que si algo es gratis ustedes son el producto. Ustedes no son el producto; son el cadáver abandonado. El ‘producto’ es lo que se fabrica con el excedente que han arrancado de sus vidas».

escrutado. Por ejemplo, la captación de nuestra voz es, hoy en día, uno de los frentes de batalla más cruentos («la caza mundial de terabytes de habla humana» – ZUBOFF, 355). El botín, no es tanto lo que decimos, sino cómo lo hacemos: la estructura de nuestra habla e ideas, el vocabulario, la pronunciación, la entonación, la cadencia, la inflexión, el dialecto. Y, todo ello, para tener acceso a nuestras emociones y personalidad.

Esta diversificación de las arquitecturas extractivas aspira a garantizar unos nichos de producto, así como unas rutas y flujo constante de suministro (nuevas fuentes de excedente). Y, con este propósito, cualquier *aparatito* (un vigilabebés, por ejemplo) puede ser la puerta de acceso a un espacio de privacidad virgen susceptible de tabulación y procesamiento. Esto es lo que está impulsando la inundación en los espacios públicos y privados de sensores computacionales automatizados.

En el fondo, con la idea de *entender lo que queremos* y proporcionárnoslo al instante (ZUBOFF, 179 y 241), los capitalistas de la vigilancia exigen que no haya restricciones *territoriales* que limiten la caza de nuestros datos en la sombra. En síntesis (como si viviéramos en una casa con paredes de cristal), las condiciones de este nuevo paradigma es que ningún ámbito debe quedar eximido de este saqueo. Y, además, debe mantenerse la gratuidad de los suministros de materia prima, así como habilitar espacios inmunes (o cortafuegos) a la acción de la Ley.

La imperceptibilidad de esta actuación también es un propósito deliberado. En efecto, el capitalismo de la vigilancia persigue imbricar la omniscencia digital en la vida diaria de forma tan íntima que, como si de un personaje invisible se tratara, consiga pasar totalmente desapercibida.

En su conjunto estos factores persiguen un claro propósito (ZUBOFF, 137, 276, 313, 400 y 401): alcanzar productos predictivos de alta calidad para acercarse a la certeza absoluta. De este modo, se reducen los riesgos de los clientes a quienes se vende nuestras *conductas futuras*. Así pueden dirigirse a ellos ofreciendo *resultados conductuales garantizados*.

La persecución de esta meta lleva inevitablemente a perseguir otro propósito: influir en la conducta real y conseguir que actuemos en consecuencia. Especialmente porque es la mejor garantía de que cualquier predicción se cumple. De ahí que anticiparse a nuestros deseos, influir y provocar la modificación real de nuestro comportamiento (en secreto y con ánimo de lucro) sean (originariamente) indisociables del capitalismo de la vigilancia. A la luz de lo anterior (y sin exagerar), este fenómeno pueda ser calificado como (WILLIAMS, 112) un proyecto de manipulación de voluntades a gran escala e inaudito.

En definitiva, las imperfecciones (o la *racionalidad limitada*) exploradas por la psicología y la economía conductuales son el pretexto para (ZUBOFF, 490 y ss.) desvelar las leyes de la acción humana y mediante el condicionamiento severo anular todo desvío del comportamiento gregario.

En cualquier caso, en la medida que el propósito es inferir en la conducta real, ciertamente sería un error pensar que estamos *simplemente* ante un estadio evolucionado de las formas ocultas de propaganda que, iniciando una nueva era de la publicidad,

exponía hace décadas Vance PACKARD. En realidad, se trata de un salto que describe un nuevo orden de magnitud por lo siguiente:

- En primer lugar, porque la psicografía que se obtiene de esta información permite lo nunca antes alcanzado: un *microtargeting* con una granularidad tan fina que la personalización, literalmente, *a la carta* es ya una realidad. Conocer el psicograma de *cada una* de las personas se ha convertido en un propósito que, no sólo es posible, sino que está siendo perseguido crecientemente por empresas e instituciones³³.

En efecto, en otros ámbitos se están empleando estas herramientas con un impacto más profundo: hay investigadores que (a partir de una concepción errónea de cómo se generan las emociones FELDMAN, 45) emplean algoritmos para analizar escáneres cerebrales y creen que son capaces de detectar las huellas dactilares neurales de las emociones (como la ira o el miedo) y, de este modo, alcanzar lo que se conoce como *adivinación neural del pensamiento*.

- En segundo lugar, porque la proliferación de la tecnología, relegando la privacidad a un reducto marginal, aspira a reducir nuestra capacidad reflexiva, presentándose a sí misma como una fuerza de la naturaleza inevitable (esto es, como *tecnologías de la perfección*). Y esta arquitectura de la ejecución, dirigida a moldear nuestro comportamiento a demanda, puede articularse a través de la gamificación, la empatía, la imitación y/o la sugestión; y,

- En tercer lugar, porque es obvio que con estas estrategias se cercena nuestra autonomía, la posibilidad de decidir nuestro rumbo³⁴. En definitiva, se nos coarta la incertidumbre y, por consiguiente, la posibilidad de disponer de alternativas entre las que elegir. Se aspira a que, a través del *inconsciente digital*, pueda influirse sobre el comportamiento humano por debajo del umbral de la conciencia.

Por ejemplo, reparen en lo que se conoce como la *ilusión del libre albedrío* (FELDMAN, 88): el cerebro hace uso de la predicción para iniciar movimientos corporales (como alargar el brazo para coger una manzana o huir de una serpiente). Estas predicciones se dan antes de que seamos conscientes de la intención de mover el cuerpo. Es decir, «el cerebro emite predicciones motrices para mover el cuerpo mucho antes de que seamos conscientes de la intención de moverlo (...). Si el cerebro fuera sólo reactivo sería demasiado ineficiente». De hecho (DU SAUTOY, 2018, 422 a 424), estos estudios sobre la actividad cerebral (a través de un escáner fMRI³⁵), apuntan a que la decisión cerebral de llevar a cabo una acción (por ejemplo, apretar un botón) se

³³ Como apunta (ZUBOFF, 660), «Recordemos cómo alardeaba Mark Zuckerberg de que Facebook llegaría a conocer todos los libros, las películas y las canciones que una persona hubiera consumido en su vida, y que sus modelos predictivos nos dirían a qué bar ir cuando llegáramos a una ciudad extraña para nosotros, un bar en el que el camarero nos tendría ya preparada nuestra bebida favorita. Como el jefe del Departamento de Ciencia de Datos de Facebook reflexionaba en una ocasión, ‘esta es la primera vez que el mundo ha visto semejante escala y cantidad de datos sobre la comunicación humana. [...] Por primera vez, disponemos de un microscopio que [...] nos deja examinar la conducta social a un nivel muy preciso que nunca antes habíamos sido capaces de ver’. Un ingeniero de alto nivel de la empresa lo expresaba también en estos sucintos términos: ‘Estamos tratando de mapear el grafo de todo lo que hay en el mundo y sus interrelaciones’».

³⁴ Ver al respecto HARARI (373 y ss.); o MOROZOV (351 y ss.).

³⁵ «Imagen por resonancia magnética funcional».

produce hasta 6 segundos antes de que esa orden se envíe al cuerpo para que actúe (en operaciones que requieren más atención, el lapso es de 4 segundos). En definitiva, en ciertos contextos, antes de que nuestra conciencia asuma una decisión que creemos que estamos tomando por libre elección, la actividad cerebral preconsciente ya ha decidido³⁶. Lo que (como hipótesis) lleva a pensar que, en determinados ámbitos, «nuestra conciencia puede ser una función muy secundaria del cerebro», es decir, «una reflexión química tardía sin influencia alguna en nuestros actos»³⁷. Si esto es así, obviamente, describe un espacio (material y temporal – ¡6 segundos!) propicio para que terceros traten de interferir en estos estados *pre-intencionales* y, a pesar de ello, actuemos pensando que lo hacemos con plena autonomía y libertad.

En estas condiciones, y dado el impulso de estas fuerzas del mercado, el fin último es aguijonear en (HAN, 23) estas «capas prerreflexivas, instintivas y emotivas del comportamiento que van por delante de las acciones conscientes». Este apoderamiento ya no sólo persigue un propósito comercial, sino que, por ejemplo, ya ha dado el paso hacia la *psicopolítica* (el escándalo de *Cambridge Analytica* podría ser una buena muestra). Develados ciertos patrones profundos del comportamiento humano, (ZUBOFF, 398 a 400) las posibilidades para, de forma subliminal, moldear sutilmente el flujo de comportamiento o directamente condicionarlo en beneficio de terceros, son opciones sobre las que ya se está trabajando a través de empujoncitos digitales que impulsen a las personas hacia los comportamientos concretos deseados. La meta de esta *economía de la acción* (y en función del objetivo específico perseguido), oscila entre conseguir cierta pérdida de autocontrol de los sujetos hasta articular complejas matrices de refuerzos con el propósito de dirigir su conducta.

Lo cierto es que el esfuerzo intelectual y económico para calibrarnos es astronómico y no tiene parangón. Una vez hemos sido perfilados gracias a una exuberante versión de lo que SKINNER calificaba como *tecnología de la conducta*, hemos entrado en un nuevo estadio en el que (siguiendo a HAN, 48 – que cita a Eli PARISER), «máquinas pronosticadoras crean y refinan continuamente una teoría sobre su personalidad y predicen lo siguiente que usted querrá hacer. Juntas, estas máquinas crean un universo único de información para cada uno de nosotros y cambian fundamentalmente el modo en que accedemos a las ideas y a la información»³⁸. Es la *utopía de certeza* (a la que me he referido con anterioridad).

³⁶ No obstante, como advierte FELDMAN (88), «la palabra ‘ilusión’ no acaba de ser exacta; el cerebro no actúa a nuestras espaldas. ‘Somos’ nuestro cerebro, y toda la cascada de sucesos está causada por el poder de predicción del cerebro. Se llama ilusión porque el movimiento *parece* un proceso de dos pasos – decidir, después mover -, cuando la realidad es que el cerebro emite predicciones motrices para mover el cuerpo mucho antes de que seamos conscientes de la intención de moverlo». A fin de cuentas, no puede aceptarse que nuestra conciencia sea una especie de homúnculo dentro de nuestro cerebro, sentado detrás de nuestros ojos. Si se admitiera esta teoría dualista (como apunta SWAAB, 219), exigiría averiguar si, a su vez, tiene otro hombrecillo sentado detrás de los suyos (y así infinitamente...). Siguiendo con DIÉGUEZ (108 - en el marco de la crítica a las tesis funcionales del transhumanismo), «la mente es el cerebro en funcionamiento, no un *software* que el cerebro esté ejecutando».

³⁷ DU SAUTOY (2018, 424 y 425), sugiere que, las condiciones en las que se desarrollan estos experimentos podrían alterar estas conclusiones preliminares y que, ante determinadas decisiones más trascendentes, la actividad cerebral preconsciente sea «únicamente un factor añadido en la decisión que se toma en el momento de la deliberación consciente».

³⁸ En definitiva, la intensificación continua de los medios de modificación conductual, supone una violación de lo que ZUBOFF (443 y ss.) denomina el «derecho a un tiempo futuro». En cambio, en el

Y, para que vean que no se trata de meras especulaciones (o futuribles remotos), reparen que el art. 5.1.a) de la Propuesta de Reglamento en materia de Inteligencia Artificial ya describe el riesgo en estos términos: «un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona para alterar de manera sustancial su comportamiento»³⁹.

Quizás, a la luz de esta descripción y como se expondrá a continuación, deberíamos empezar a pensar seriamente en velar por nuestro *yo inconsciente*.

5.2. La amenaza del pastoreo social (¿los derechos del yo inconsciente de la persona como antídoto?)

Una de las paradojas de la sociedad de la información es que ha provocado una sobreabundancia informativa que, al impedir su procesamiento, acaba teniendo efectos desinformantes contraproducentes⁴⁰. En efecto, como expone WILLIAMS (29 - recogiendo la reflexión de Herbert SIMON), el superávit informativo provoca la escasez de aquello que la información consume, esto es, la atención de los receptores (una carestía atencional). Y este efecto se amplifica exponencialmente en el contexto de la red y la creación por parte de la industria de la persuasión digital de sutiles y muy sofisticados *clickbait*s (o *ciber anzuelos*) para captar nuestra atención a través del constante bombardeo de estímulos.

Una breve exposición de la *mecánica* implícita en algunos de ellos (notificaciones, gratificaciones, seguidores y el carrusel infinito) pueden ayudar a comprender su capacidad en términos atencionales:

En las *notificaciones*, el color y la localización de las notificaciones (rojas y en la esquina superior derecha del campo de visión) han sido deliberadamente escogidos para captar mejor la atención y potenciar el efecto persuasivo. Como apunta WILLIAMS (72) es un efecto basado en «la respuesta sensorial al color rojo y el instinto común de orden e higiene, por eso es a veces tan difícil resistirse a clicar en ellas». Y, añade (73 y 68), la exposición a constantes notificaciones «puede crear hábitos mentales que conducen al usuario a interrumpirse solo, incluso en ausencia de las propias tecnologías». Y, debido al carácter momentáneo de su influencia, es habitual pasar por alto el perjuicio que nos causa esta distracción en nuestras actividades hacia una determinada tarea.

En la exaltación de la acumulación de *seguidores* intervienen diversas palancas. En efecto, por un lado (ZUBOFF, 613) se observa un proceso de inflación del perfil

capitalismo de la vigilancia, nuestra voluntad (esto es, en tanto que conducta desviada de la previsión) es una amenaza para el flujo de ingresos.

³⁹ Y, como se expondrá posteriormente (pero conviene apuntar ya, por su importancia), téngase en cuenta que está previsto que estas *prácticas* de la inteligencia artificial (si se acaba aprobando en estos términos), estará prohibida.

⁴⁰ ZUBOFF (257): «Los científicos han advertido que la capacidad mundial de producción de información ha sobrepasado ya con creces la capacidad mundial de procesamiento y almacenamiento de información».

personal; y, por otro lado (HAN, 49) el culto al yo (que lleva a los individuos a representarse y producirse a sí mismos). Estos dos factores, consiguen que (WILLIAMS, 80) los usuarios se conviertan en emisores activos, transformando estos datos *reputacionales* en *bienes en sí mismos*. Y, por consiguiente, al atribuirles un valor intrínseco, se erigen en metas y objetivos a alcanzar y susceptibles de maximización (e, incluso, en algunos casos, se emplea la monetización para alimentar esta *puja* entre los individuos). La resultante es (ZUBOFF, 593), «una especie de cinta o teletipo continuo de cotizaciones de nuestro valor actualizado en el mercado social». La sociedad de la información, en definitiva (HAN 15 y 17), promueve una idea de la transparencia individual, convirtiéndose en un *imperativo sistémico*, en el que «todo debe presentarse como información» y, por consiguiente, debe circular con libertad (para evitar la ocultación). Aunque, paradójicamente, *al otro lado*, la gestión de esta transparencia no es recíproca, dado que opera sobre la base de una total opacidad.

Las *gratificaciones* (a través de reconocimientos sociales efímeros – como, entre otros, *likes* o *corazones*), tratan de explotar la vanidad, el egocentrismo y el ensimismamiento del usuario a través de lo que se ha denominado (WU, 453) la «celebración de la vida cotidiana»⁴¹. Y, con ella se persigue hacer realidad el anhelo de fama y reconocimiento social de personas anónimas⁴². En definitiva, las opciones para disfrutar de los famosísimos 15 minutos de Andy WARHOL se han incrementado sustancialmente. Las recompensas en este caso (ZUBOFF, 596, 597 y 608), especialmente «amoldadas a la estructura psicológica de la adolescencia y la adultez emergente»⁴³, persiguen «el reconocimiento, la aceptación, la pertenencia y la inclusión en el grupo». Lo que ha convertido a los *no-likes* en dolorosas condenas sociales; y, en paralelo (SADIN, 2020, 224), una vez interiorizado el principio de un impiadoso calibrado universal, retroalimenta un infatigable esfuerzo para hacer, con mayor o menor éxito, *un buen papel* en el nuevo foro digital.

Finalmente, el *carrusel infinito* (WILLIAMS, 55) consiste en la actualización aleatoria de contenido de webs o apps a medida que uno se desplaza por la página (y sólo si se hace así). Este efecto está inspirado en la barra de las máquinas tragaperras (y lo que se conoce como estado *zona máquina* que experimentan los adictos al juego⁴⁴) y hábilmente combina con la expectativa de encontrarse con una recompensa variable en

⁴¹ De hecho (SADIN, 2022, 110), el smartphone es un instrumento pensado para halagar al individuo contemporáneo; y éste, en correspondencia, invierte (o gasta) gran cantidad de energía con la única finalidad de experimentar el éxtasis de la importancia de sí mismo. Tendencia que tiene un efecto reflejo, pues, (HAN, 49), «La creciente atomización y narcisificación de la sociedad nos hace sordos a la voz del otro. También conduce a la pérdida de la empatía. Hoy todo el mundo se entrega al culto del yo. Todos los individuos se representan y se producen a sí mismos».

⁴² Como expone HAN (15), «en el régimen de la información, las personas se esfuerzan por alcanzar la visibilidad por sí mismas», a diferencia de los regímenes de la disciplinar en los que se las obliga a hacerlo. Así pues, lejos de evitarlo, «las personas se colocan de manera voluntaria en el foco, incluso desean hacerlo».

⁴³ Y que comprende (ZUBOFF, 601) la etapa que va de los 18 a los 30 años

⁴⁴ Recogiendo el estudio de SCHÜLL y SHAFFER, ZUBOFF (598 y 599) expone que «los jugadores adictos no buscan entretenimiento ni sacar el mítico premio gordo de la máquina. Andan detrás, más bien, de lo que [se ha denominado] (...) zona máquina: un estado de olvido de uno mismo en el que la persona se siente transportada por un irresistible ímpetu que hace que se sienta como si estuviera siendo ‘jugada por la máquina’. La zona máquina posibilita una sensación de inmersión completa (..) y que se asocia a una pérdida de la autoconciencia, a un comportamiento automático y a una absorción rítmica total arrastrada por una ola de compulsión».

forma de sorpresa (un contenido no esperado) y el conocimiento científico de que eso incrementa la probabilidad de que el usuario repita dicha acción. En efecto (e, intrínsecamente vinculado a la escasez atencional descrita), en este contexto (HAN 33), «la información tiene un intervalo de actualidad muy reducido. Carece de *estabilidad temporal*, porque vive del ‘atractivo de la sorpresa’. Debido a su inestabilidad temporal, fragmenta la percepción. Arrastra la realidad a un ‘permanente torbellino de actualidad’. Es imposible *detenerse* en la información. Esto deja al sistema cognitivo en un estado de inquietud. La necesidad de aceleración inherente a la información reprime las prácticas cognitivas que consumen tiempo, como *el saber, la experiencia y el conocimiento*».

Estos instrumentos están diseñados deliberadamente para explotar nuestras debilidades cognitivas y someter a nuestra atención finita a una sobreestimulación cegadora. Con la particularidad de que los recursos y talento destinados a la consecución de este objetivo persuasivo son tan descomunales que (WILLIAMS, 128) resistirse a ellos exige una capacidad de contención prácticamente inalcanzable. La resultante de este proceso es precipitar la pérdida de autocontrol y, de este modo, alimentar la conexión compulsiva con fines extractivos⁴⁵. Esta estrategia, que encajaría perfectamente en lo que se ha denominado (AKERLOF y SHILLER, 2015) *economía de la manipulación*⁴⁶, evidencia que la tecnología no está siendo diseñada para mejorar nuestro bienestar, sino con el propósito de agudizar los sumideros de atención existentes.

Esta sobrecarga cognitiva y la intensa distracción que puede acarrear tiene un efecto más profundo, pues, cortocircuita el uso de la voluntad. Esta automatización del comportamiento (o heteronomía conductual) es ilustrativa de una *desposesión de uno mismo* y se proyecta más allá de nuestra interacción con la red. El *clickivismo barato* (que denuncia KRASTEY, 71), el populismo, la desinformación deliberada (a través de las *fake news*) y la creciente desafección podrían ser manifestaciones, entre otras, de este proceso. En cualquier caso, repárese que el devenir tecnológico futuro podría empujar a este fenómeno a otra escala de magnitud, pues, la creciente datificación del mundo real (con la consiguiente gamificación, gratificaciones, reconocimientos, etc.), así como el intento en ciernes de virtualización de todo nuestro entorno a través del metaverso podrían tener un efecto multiplicador.

A la espera de nuevos avances de la neurociencia sobre la condición humana y nuestra capacidad de control real sobre nuestro comportamiento (desconociendo hasta qué profundidad podremos llegar), quizás, deberíamos empezar a exigir la protección frente

⁴⁵ Como expone ZUBOFF (599), recogiendo el comentario de Sean Parker (cofundador de Napster y, durante un tiempo, presidente de Facebook), «Facebook estaba diseñado para consumir la máxima cantidad posible del tiempo y la consciencia de los usuarios. La idea era enviar ‘un pequeño chute de dopamina cada cierto tiempo’ a los usuarios, o lo que es lo mismo, un ‘refuerzo variable en forma de clics en el botón de ‘me gusta’ y de comentarios. Se trataba de mantener a los usuarios pegados a la colmena, ansiosos de conseguir nuevas dosis y dejando, mientras tanto, un rastro de materias primas tras de sí».

⁴⁶ El término «economía de la manipulación» se refiere a la idea de que los mercados no sólo producen (17) «abundancia de cosas que la gente desea», sino que también «crean un equilibrio económico que es altamente adecuado para entidades económicas que manipulan y distorsionan nuestro juicio, utilizando prácticas empresariales que son análogas a cánceres biológicos que se instalan en el equilibrio normal del cuerpo humano». En definitiva, «en la medida que tengamos debilidades en saber lo que realmente deseamos, y también en la medida que esas debilidades puedan ser rentablemente generadas y alimentadas, los mercados aprovecharán la oportunidad de incidir en ellas. Se acercarán y se aprovecharán de nosotros. Nos harán caer como incautos».

a las violaciones deliberadas de nuestro *yo inconsciente*. Y no esperar a actuaciones reactivas una vez se constate la intromisión ilegítima (si es que somos capaces de detectarlas).

Sin perjuicio de otras derivadas (y salvo error y/o mejor doctrina), en el marco de este avance técnico, el derecho a la libertad, probablemente, es que esté más comprometido (o el que está más expuesto). Como punto de partida, deben tenerse en cuenta dos aspectos importantes:

- En primer lugar, sólo podemos elegir con libertad (SANDEL, 113) si no estamos indebidamente presionados (...) y sí razonablemente bien informados sobre las demás posibilidades». Los estímulos subyacentes en estas tecnologías operan como si fueran incentivos voluntariamente aceptados. No obstante (GRANT, 70 y 57), 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. Y, en estas condiciones (esto es, en la medida que son instrumentos de poder), es obvio que pueden ser objeto de abuso y cualquier demócrata debería estar preocupado por la frecuencia de su uso. Especialmente porque intentan de algún modo (48) «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». Y este diseño experto de las instituciones 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».
- En segundo lugar, siguiendo de nuevo con SANDEL (126 a 129 - a partir de la filosofía kantiana), en la medida que nuestra autonomía y capacidad de actuar y elegir libremente pueda verse comprometida, ya no podremos decir que estamos actuando conforme a una *ley que nos damos a nosotros mismos*. En tal caso, lo haremos sometidos a unos fines dados desde fuera de nosotros (porque han sido dictados por terceros). Si la capacidad de actuar autónomamente es lo que confiere a la vida humana su especial dignidad, una intervención heterónoma la laminaría.

A la luz de lo anterior, es fácil advertir que el imperativo predictivo y la arquitectura de ejecución descritos en el apartado anterior muestran una de las principales amenazas a nuestra libertad (y, derivadamente – como ya se ha avanzado, pero se detallará posteriormente -, a la democracia). Y debería preocuparnos.

La «administración automatizada de las conductas» implantada en China⁴⁷ es probablemente una primera versión de esta ingeniería social dirigida a *conducir*

⁴⁷ Como expone SADIN (2020, 219): «Este sistema consiste en establecer una ‘evaluación científica’ de los comportamientos. Primero concierne a los ciudadanos, con excepción de aquellos que poseen una causa judicial, que están dotados de un volumen inicial de mil puntos que van a evolucionar negativamente o positivamente según sus acciones. Por ejemplo, se contabilizan los actos reprobables en transportes públicos, el pago de facturas con atraso, no respetar una prohibición de fumar, el hecho de no ir a visitar a los propios padres, todo esto implica la supresión de una parte de ese crédito, impidiendo entonces a esos individuos beneficiarse de ciertas ventajas al mismo tiempo que se menoscaba su reputación. Para recuperar los puntos, hay que dar pruebas de ‘sentido cívico’, por ejemplo, dar sangre, destacarse como ‘trabajador modelo’ o realizar algunas ‘buenas acciones’. Son gestos que, entre otros tantos, y si se realizan no en vistas de ‘rescatarse’ sino movidos por un ‘impulso voluntario’, también

masivamente el comportamiento social. En todo caso, y al margen de esta manifestación *vanguardista* e hipertrofiada (y que debería describir un límite infranqueable que conviene mantener lo más alejado posible), en nuestro contexto más cercano e inmediato debe advertirse una circunstancia singularmente inquietante: en la medida que lo impredecible y/o divergente (esto es, lo que queda al margen del *curso programado*) resta certeza al producto predictivo, el capitalismo de la vigilancia (ZUBOFF, 213) aspira a neutralizarlo y evitarlo a toda costa, especialmente porque equivale a ingresos perdidos. Tendemos (SUPIOT, 244) hacia una gobernanza que aspira a la auto-regulación y que se caracteriza por el carácter unívoco del lenguaje lógico-matemático y que repulsa el discurso reflexivo (y que, en la medida que los decisores públicos son personas con sus propios sesgos y emociones, convierte en superfluo cualquier referencia a las leyes, salvo si reaccionan al dictado de datos cifrados). Lo *humano* (en tanto que fuente de errores e de impugnación de decisiones o resistencia) se percibe como una rémora que debe ser reducida a su mínima expresión. En efecto (siguiendo con ZUBOFF, en una extensa cita, pero particularmente ilustrativa, 414 y 415): «la consciencia humana misma es una amenaza para esos ingresos de la vigilancia, pues pone en peligro el proyecto de la modificación conductual en general. Los filósofos identifican la autorregulación, la autodeterminación personal y la autonomía con el ‘libre albedrío’. La palabra autonomía deriva del griego y significa literalmente ‘regularse por sí mismo’. Se opone a la heteronomía, que significa ‘regularse por otros’. La necesidad competitiva de crear economías de acción implica que los capitalistas de la vigilancia se sienten obligados a usar todos los medios disponibles para suplir la acción autónoma por la acción heterónoma».

Así pues, los *datos oscuros*, esto es, aquello no rastreado (y, especialmente, lo preconsciente), debe salir a la luz y ser cognoscible para *ellos*. En aras al «imperativo sistémico de la transparencia» anteriormente expuesto, debemos sacrificar nuestra libertad para posibilitar la acción y el conocimiento colectivo impuesto *por otros*. De modo que toda *fricción* (especialmente, la de origen individual) debe ser superada. En el fondo, el individualismo (entendido como la capacidad de una persona libre que actúa de forma autónoma) es visto como un fenómeno a erradicar (ZUBOFF, 581 y ss.; y HAN, 66 a 70). Y los encargados de pastorear a los rebaños humanos a partir de la nueva *matemática social* son la nueva *casta sacerdotal* tecnológica y sus máquinas. A diferencia del totalitarismo, que obró valiéndose de los medios de la violencia (ZUBOFF, 482, 483 y 472), la instrumentalización de la conducta a efectos de su modificación, predicción, monetización y control describe un nuevo poder: el instrumentarismo. Y el colectivo humano a él sometido es, en términos de ZUBOFF (571 y ss), la *sociedad instrumentaria*, en la que como principios generales (defendidos por PENTLAND, su *física social* y los *sociómetros*), pueden destacarse los siguientes: la conducta individual está sometida a un interés superior; la política es sustituida por planes predictivos basados en la verdad computacional; y la presión social debe converger hacia la armonía mediante la empatía.

En definitiva (ZUBOFF, 587), «una colmena de máquinas en la que nuestra libertad se sacrifica en un interés del conocimiento perfecto, administrado en provecho de otros». Aunque estos principios acarrear la muerte de la individualidad y la vida en la colmena, no se trata de un *totalitarismo digital*, porque persigue objetivos y medios claramente

umentan el ‘crédito’. Un buen puntaje puede dar derecho a un acceso prioritario a la salud, a un alojamiento social y a ciertos empleos públicos».

diferenciados del totalitarismo: es indiferente a la conversión de nuestra alma y tampoco aspira a inculcarnos ningún principio superior. Le basta con saber que lo que hacemos (lo que sea) es accesible a los efectos de configurar su producto predecible garantizado.

Sin embargo (483), es «tan alarmante, incomprensible y nuevo para el relato de la historia humana como lo fue el totalitarismo para sus testigos y sus víctimas en su tiempo (...). El instrumentalismo es un proyecto de mercado que, en convergencia con lo digital, alcanza a su propia y singular forma de dominación social». En opinión de ZUBOFF (679 y 683 y ss.) «el capitalismo de la vigilancia debe ser entendido como una fuerza social profundamente antidemocrática (...). Es una forma de tiranía que se nutre de las personas, pero no es de las personas»⁴⁸.

Y dado que es un fenómeno sin precedentes (pues, posibilita unas concentraciones exclusivas e ingentes de conocimiento y poder), no puede esperarse que los instrumentos legales a nuestro alcance sean capaces de protegernos de forma integral frente a lo inédito.

5.3. ¿Los derechos del yo inconsciente de la persona como antídoto?

Repáren que, si la estadística computacional es capaz de detectar patrones conductuales emergentes que inciden en nuestra capacidad decisoria y, por ende, nuestro comportamiento, deberíamos, como mínimo, conocerlos. Es obvio que quien tenga esta información (y, además, obtenga una ventaja competitiva) no estará dispuesto a compartirla – o será muy reacio a hacerlo (y, no nos engañemos, también es difícil que las personas acaben ejerciendo el derecho que les reconoce el art. 22 del RGPD). De ahí que la intensificación de la protección legal sea determinante.

Adviértase que, desde el punto de vista jurídico, el propio concepto de *acto propio y voluntario* puede quedar en entredicho. En efecto, si la conducta es perfilada y/o inducida deliberadamente por terceros, ¿en qué medida puede afirmarse que se ha dado un consentimiento *consciente* y, por ende, vinculante? O bien, en este contexto de comportamiento heterónomo, quizás, tendríamos que plantearnos si deben rediseñarse las fronteras de la *culpa* y de la responsabilidad derivada de nuestros propios actos.

Si, realmente, estamos hablando de herramientas *efectivas* que operan por debajo de nuestro nivel consciente, 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. No se trata sólo de saber que hemos sido *perfilados* (automatizadamente o no), sino de cortocircuitar jurídicamente a quienes eventualmente traten de pastorearnos, con la idea de subyugarnos a sus intereses y/o sacar un provecho a nuestra costa. Desde este punto de vista, la prohibición que contiene el art. 5.1.a) de la propuesta de Reglamento de IA (a la que se ha hecho referencia anteriormente) es muy interesante y oportuna.

⁴⁸ ZUBOFF (263): «Cuanto más se sabe de una persona, más fácil resulta controlarla. Garantizar la libertad de la que se nutre la democracia requiere de una correcta estructuración del uso social de la información e, incluso, de que se permitan ciertas dosis de ocultación de la información».

Especialmente porque su configuración (una prohibición expresa) sobrepasa lo que en la literatura especializada se conoce como *neurotecnología* (y a la que, sin emplear esta denominación se ha hecho una breve referencia al inicio de este trabajo). Este término está referido a instrumentos distintos a los citados hasta este momento. A grandes rasgos, de acuerdo con esta literatura, este concepto está referido a dos tipologías de instrumentos (GENSER, HERRMANN y YUSTE, 10 y11): invasivos y no invasivos. Los primeros, directamente, tienen contacto neuronal y potencialmente permiten la comunicación bidireccional entre el cerebro y el mundo exterior, ya sea exportando datos cerebrales o alterando la actividad cerebral. Los *no invasivos*, normalmente, se consideran dispositivos electrónicos portátiles (*wearables*), de todo tipo (cascos portátiles, gafas, diademas, gorras, muñequeras o cintas para la cabeza, entre otros), que pueden leer la actividad del sistema nervioso periférico al tocar la cabeza o el cuerpo de una persona (en lugar de tocar directamente el cerebro). Y, los avances en esta materia son extraordinarios y acelerados.

Sin menospreciar la capacidad de intrusión potencial de esta tecnología, especialmente, cuando se generalice su uso, lo cierto es que, lo que en este trabajo se quiere poner de relieve es que el *actual* rastreo y procesamiento de nuestros datos en la sombra, ya describen un escenario lo suficientemente amenazador como para adoptar *hoy* un posicionamiento (extremadamente) cauteloso.

En este contexto, en algunos foros se empieza a hablar de la necesidad de articular un conjunto de derechos de nueva generación y se aglutinan genéricamente bajo la denominación de *neuroderechos*. Aunque el término *neuro* tiene mucho *enganche* (y, probablemente, acabará predominando), quizás, sería más oportuno definirlos como los «derechos para proteger el *yo inconsciente* de las personas» (y, si quieren un acrónimo – quizás, un poco *kitsch* -, *derechos YIP*).

Al margen de esta discusión sobre la denominación y del debate sobre la oportunidad de esta regulación específica (pues se alega que supondría *inflacionar* o, incluso, devaluar la regulación existente – BUBLITZ⁴⁹), bajo el impulso de la *NeuroRights Foundation*, se han propuesto los siguientes derechos (siguiendo la exposición que de los mismos hace BUBLITZ, 3):

(1) *el derecho a la identidad mental, o al «sentido de sí mismo»*: «Deben desarrollarse límites para prohibir que la tecnología interrumpa el sentido de uno mismo. Cuando la neurotecnología conecta a las personas con redes digitales, podría desdibujar la línea entre la conciencia de una persona y los ‘inputs’ tecnológicos externos»

(2) *el derecho al «libre albedrío»*: «Las personas deben tener el control final sobre su propia toma de decisiones, sin manipulación desconocida de neurotecnologías externas»,

(3) *el derecho a la privacidad mental*: «Todos los datos obtenidos de la medición de la actividad neuronal (‘NeuroData’) deben mantenerse privados.

⁴⁹ En cambio, para GENSER, HERRMANN y YUSTE, en su informe, existen argumentos suficientes para entender que la configuración del marco normativo a nivel internacional vigente es, en términos generales, suficiente para atender a las demandas que esta nueva realidad tecnológica.

Además, la venta, la transferencia comercial y el uso de datos neuronales deben estar estrictamente regulados». De hecho, como sigue apuntando BUBLITZ, se ha sugerido que una nueva regulación para los ‘neurodatos’ debería «ser análoga a la legislación que prohíbe la venta de órganos humanos».

(4) *el derecho al acceso equitativo a un aumento de la capacidad mental*; «Deberían establecerse directrices tanto a nivel internacional como nacional que regulen el desarrollo y las aplicaciones de las neurotecnologías de mejora mental. Estas directrices deben basarse en el principio de justicia y garantizar la igualdad de acceso a todos los ciudadanos»; y

(5) *el derecho a la protección contra sesgos y discriminaciones*: «Las contramedidas para combatir el sesgo deberían ser la norma por defecto para el aprendizaje automático. El diseño de los algoritmos debe incluir aportaciones de grupos de usuarios para abordar los sesgos desde su creación»⁵⁰.

Y, de hecho, en octubre de 2022, Consejo de Derechos Humanos de la ONU⁵¹ solicitó al Comité Asesor que preparara un estudio sobre el impacto, las oportunidades y los desafíos de la neurotecnología con respecto a la promoción y protección de todos los derechos humanos.

A nivel interno (sin perjuicio de otras iniciativas a nivel internacional⁵²), debe destacarse la *Carta de Derechos Digitales*⁵³. Descartado su contenido normativo y tras afirmar que «no trata de crear nuevos derechos fundamentales sino de perfilar los más relevantes en el entorno y los espacios digitales o describir derechos instrumentales o

⁵⁰ Como apunta WILLIAMS (143), «la persuasión informática agresiva debería contar siempre con el consentimiento expreso del destinatario, en lugar de ser la opción por defecto que es preciso desactivar. En definitiva, la configuración predeterminada debería primar el control del usuario sobre su propia atención».

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

⁵² Véase al respecto MUÑOZ y MARINARO. No obstante, debe destacarse la propuesta de redacción del del art. 19.1º del proyecto de Constitución Chilena de 2022 (finalmente no refrendada) y que recogía el siguiente literal (aunque, finalmente, no fue incorporado [al texto final](#)): «El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella».

⁵³ https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

auxiliares de los primeros», en el apartado XXVI, bajo la rúbrica de *Derechos digitales en el empleo de las neurotecnologías*, establece lo siguiente:

«1. Las condiciones, límites y garantías de implantación y empleo en las personas de las neurotecnologías podrán ser reguladas por la ley con la finalidad de:

- a) Garantizar el control de cada persona sobre su propia identidad.
- b) Garantizar la autodeterminación individual, soberanía y libertad en la toma de decisiones.
- c) Asegurar la confidencialidad y seguridad de los datos obtenidos o relativos a sus procesos cerebrales y el pleno dominio y disposición sobre los mismos.
- d) Regular el uso de interfaces persona-máquina susceptibles de afectar a la integridad física o psíquica.
- e) Asegurar que las decisiones y procesos basados en neurotecnologías no sean condicionadas por el suministro de datos, programas o informaciones incompletos, no deseados, desconocidos o sesgados.

2. Para garantizar la dignidad de la persona, la igualdad y la no discriminación, y de acuerdo en su caso con los tratados y convenios internacionales, la ley podrá regular aquellos supuestos y condiciones de empleo de las neurotecnologías que, más allá de su aplicación terapéutica, pretendan el aumento cognitivo o la estimulación o potenciación de las capacidades de las personas».

La extensión de estas medidas de protección es ambiciosa y, *a priori*, describe un esperanzador punto de partida (siempre que, obviamente, se acabe incorporando en una norma). No obstante, convendría que el supuesto de hecho no quedara circunscrito a un tipo específico de avances tecnológicos (las *neurotecnologías*). Especialmente porque, como se ha expuesto, la computación estadística ya ha accedido a un cierto umbral del cableado mental (y seguirá perforando sin descanso). En efecto, del listado del citado *Apartado XXVI* puede colegirse que, como mínimo, la autodeterminación individual, soberanía y libertad en la toma de decisiones no está garantizada por completo. Tampoco puede afirmarse que tengamos el pleno dominio de los datos obtenidos o relativos a los *procesos mentales* (como tampoco, obviamente, tenemos garantizada su confidencialidad y seguridad). A su vez, hay decisiones que ya están siendo condicionadas por el suministro de datos, programas o informaciones incompletos, no deseados, desconocidos o sesgados.

En definitiva, una intervención legislativa es perentoria.

Al margen de estas iniciativas y de su futura (e incierta) progresión, la verdad es que, salvo que una intervención legislativa lo impida expresamente, *a priori*, no parece que puedan identificarse *nichos extractivos* ajenos a esta amenaza, pues, (autoconvencidos del determinismo tecnológico) nos hemos resignado a la estructura computacional ubicua. En cualquier caso, cuanto mayor sea el volumen de datos y la inteligencia de máquinas para procesarlo, mayor será la cartografía humana y la consiguiente optimización de la *tecnología de la conducta*.

A la luz de lo anterior, es claro que el ámbito socio-laboral no será un espacio ajeno a esta psicometría. Como expone SUPIOT (167), la organización científica del trabajo que perseguían por igual el capitalismo y el comunismo, simplemente ha cambiado de forma, pues, se manifiesta a través de algoritmos. La organización del trabajo se concibe como un sistema programable que «provoca unidades capaces de reaccionar a las señales que reciben en función de comunicarse con esta programación». Al albur de este *gobierno de los números*, en todos los niveles de la organización del trabajo «se pide a los hombres menos que obedezcan prescripciones a que alcancen objetivos cuyo logro se evalúa a partir de indicadores numéricos».

La particularidad del momento es que estamos a las puertas de la generalización exponencial del *data driven manufacture* (la empresa dirigida por los datos) que gravita alrededor del *individuo programado*, que debe objetivarse a sí mismo a través de la alineación de su rendimiento al patrón dictado por las máquinas. Como ya puede observarse en algunas actividades, el registro del desempeño profesional en tiempo real a través de, por ejemplo, el registro de conversaciones (teleoperadores), geolocalización (repartidores) o el uso de *ropa inteligente* (entre otros, deportistas⁵⁴) es ilustrativo del ímpetu extractivo en ciernes⁵⁵; y que se irá expandiendo a medida que surjan nuevos utensilios⁵⁶ y/o necesidades y, en paralelo, se corporice (MERCADER UGUINA, 138) la *gran delegación empresarial*. Estos instrumentos describen otro orden magnitud en la aspiración empresarial de monitorización del rendimiento y de la productividad y, probablemente, estén más cerca del moldeo conductual por debajo del nivel consciente para alcanzarlos.

La hipertrofia del poder empresarial (entendido en sentido amplio) que emergerá de la convergencia de estas fuerzas, obviamente, somete a los derechos de los trabajadores

⁵⁴ Con esta *ropa inteligente* (MUÑOZ RUIZ, en relación al «*Proyecto Bionic H2020*»), los «algoritmos predictivos pueden realizar valoraciones ergonómicas del riesgo de estrés físico del trabajador/usuario de la herramienta, mediante el desarrollo de sensores en su ropa que recogen información en tiempo real sobre la salud de la persona, a partir de un análisis de los parámetros examinados: postura, cargas, tiempos y estado fisiológico (frecuencia cardíaca, temperatura corporal, etc.) ». O también los chalecos que llevan algunos deportistas (por ejemplo, futbolistas) en los entrenamientos y partidos miden diversas variables (entre otras, frecuencia cardíaca, velocidad, distancia). De hecho, el mundo del deporte lleva algunos años experimentando una verdadera revolución a partir del uso de datos y la estadística (como expone LEWIS) y es claro que la monitorización exhaustiva de la conducta (como si de caballos de carreras se tratara) y la búsqueda de una mayor certeza en los pronósticos (de todo tipo) es el objetivo a alcanzar.

⁵⁵ Tal y como se reportan GENSER, HERRMANN y YUSTE (12 – traduzco del inglés), «en una fábrica de Hangzhou, China, los trabajadores de la línea de producción supuestamente están siendo equipados con sombreros y cascos que leen señales cerebrales para decodificar las emociones de los trabajadores, y luego estos datos se envían a algoritmos de inteligencia artificial para detectar cambios en la emoción que afecten a los niveles de productividad, aunque la precisión de esta tecnología es cuestionada».

⁵⁶ En algunos medios de comunicación digital (*La Vanguardia* - 01/07/19: <https://www.lavanguardia.com/economia/20190701/463176490145/empresa-control-trabajadores-blackrock-anillo-inteligente.html>) se recogía la iniciativa de una empresa de ofrecer un «anillo inteligente a sus empleados para controlar sus hábitos de sueño, su frecuencia cardíaca y otros indicadores fisiológicos». O bien, el utensilio de Amazon para captar los movimientos y la respiración mientras se duerme (*Wired* - 28/09/22: <https://www.wired.com/story/amazon-wants-to-cocoon-you-with-ambient-intelligence/>). O, bien, la implantación de pantallas y cámaras en las lentillas (*El País* - 26/10/22: <https://elpais.com/tecnologia/2022-10-26/pantallas-y-cameras-en-los-ojos-y-otras-promesas-de-las-lentillas-del-futuro.html>) es otro de los posibles avances del futuro.

(dignidad, privacidad, estabilidad en el empleo, etc.) a una tensión y a una potencial vulneración inaudita hasta la fecha (y, probablemente, de forma paralela, implique una redefinición del papel de la estadística en el entorno jurisdiccional y de los profesionales). Es obvio que el *texto en la sombra* de las decisiones basadas en algoritmos describe un ámbito de absoluta indefensión (y el papel del nuevo art. 64.4.d ET y el consiguiente rol de la representación legal de los trabajadores será determinante⁵⁷).

No obstante, en paralelo a este proceso (que requerirá una sintonía muy fina por parte de los tribunales), téngase en cuenta que, aunque los propósitos perseguidos por estas herramientas sean ciertamente heterogéneos (e, incluso, como se ha apuntado, alguno de ellos pueda acabar reportando un beneficio para las personas trabajadoras), lo que no cabe duda es que, sin excepción, ya acarrearán una intensificación de la minería de la realidad y, con ella, se habilita su explotación algorítmica. En este contexto (como hipótesis), es factible pensar que el mayor conocimiento sobre la conducta humana que esta información acabará facilitando (superior, incluso, a la que uno pueda tener de sí mismo), se acabe traduciendo en un acercamiento a un severo *moldeo conductual* (mediante incentivos, *empujoncitos digitales*, etc.), que, trascendiendo la conciencia, acaben teledirigiendo la conducta en aras a alcanzar los objetivos que se persigan.

En una adaptación del cuento de Hamelin, los trabajadores, al son de una música inaudible o imperceptible para sus oídos y compuesta por un flautista algorítmico, bailarían sin saber por qué

Desde este punto de vista, es posible tantear algunos escenarios posibles en el ámbito socio-laboral (y, quizás, podrían ser extrapolables a otros). Aunque, vaya por delante, este sondeo es especialmente complejo porque es difícil combatir un enemigo invisible y/o que todavía no se ha corporizado en toda su potencialidad. En cualquier caso, con el conocimiento actual (salvo error y/o mejor doctrina), podría afirmarse lo siguiente:

- Primero, la autonomía y el libre albedrío son espacios absolutamente inexpugnables, porque, fundamentan la validez del consentimiento libre consciente y, por consiguiente, son la base absolutamente irrenunciable de nuestra individualidad.

- Segundo, partiendo de lo anterior, cualquier decisión empresarial, 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, describe una múltiple violación de derechos fundamentales y debe ser prohibida y declarada sin efecto. De ahí que la prohibición que proclama el art. 5.1 del Proyecto de Reglamento de IA sea, como se ha reiterado, tan oportuna (recuérdese: «*un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona para alterar de manera sustancial su comportamiento*»).

Esto afectaría, a medidas empresariales que se exteriorizaran deliberadamente, por ejemplo, mediante incentivos y/o condicionantes extrínsecos de cualquier tipo y con un propósito *condicionante*.

⁵⁷ Ver al respecto, BELTRAN DE HEREDIA RUIZ (2022).

Ahondando en el tipo de violación (ciñéndome exclusivamente al marco constitucional) y partiendo de la premisa de que es posible que el *yo inconsciente* de las personas en cuanto tal no tenga un encaje preciso y/o directo, es difícil pensar que esta conducta pluriofensiva no impactara de algún modo (total o parcialmente) en el derecho al libre desarrollo de la personalidad (art. 10.1 CE), el derecho a la integridad (art. 15.1), el derecho a la libertad (art. 17.1 CE), el derecho a la intimidad (art. 18.1 CE) y/o el derecho a expresar y difundir libremente los pensamientos, ideas y opiniones (art. 20.1.a CE).

- Tercero, aunque es posible que no pueda percibirse esta intromisión en el *yo inconsciente* mientras *se está produciendo*, la legitimidad de un derecho de oposición o resistencia frente a la acción subliminal que la provoca quedaría fuera de toda duda.

- Cuarto, la persona trabajadora cuyo *yo inconsciente* ha sido condicionado no puede ser declarada responsable por sus actos.

6. Valoración final

La capacidad de autorreflexión del ser humano no puede penetrar en los estratos más profundos de la mente. Y, por este motivo (siguiendo con JOHNSON-LAIRD, 343), somos «ignorantes de las raíces de muchas de nuestras decisiones». A pesar de esta limitación, puede afirmarse que somos libres en la medida que «nuestros modelos de nosotros mismos nos permiten elegir la manera de elegir». De ahí que cualquier arquitectura de la decisión dirigida a restringir alguna de las opciones a nuestro alcance, *susurrándonos* por debajo del nivel consciente lo que más nos conviene, estará minando nuestra libertad como individuos y, si se aplica de forma generalizada, de la sociedad en su conjunto. Ante esta amenaza, sería absurdo que no permaneciéramos recelosos ante la evolución de una tecnología con esta capacidad de penetración y seríamos profundamente cándidos si no tratáramos de contenerla. Aunque todavía no sabemos si el armamento jurídico existente será lo suficientemente poderoso como para combatir una intrusión de esta magnitud, creo que sería muy oportuno que no se demorara el desarrollo legislativo de los *Derechos digitales en el empleo de las neurotecnologías* (apartado XXVI de la *Carta de Derechos Digitales*). En este sentido, la definición de estas innovaciones (y, por ende, el ámbito de aplicación objetivo de la norma) debe ser extensible a cualquier dispositivo con capacidad para afectar directa o indirectamente y de forma deliberada por debajo del nivel consciente y/o que *asista* a las personas a hacerlo (y no sólo a los conocidos como *interfaces cerebro-ordenador - BCI*).

La tecnología avanza a una velocidad tan vertiginosa que (WILLIAMS, 41, 125 y 147) el proceso de adaptación y aprendizaje constantes al que nos somete hace que no podamos evaluar los riesgos implícitos que conlleva ni tampoco evitar sus influencias perniciosas. Esta aceleración también dificulta que la investigación y evaluación de su impacto pueda seguirle el ritmo y, de forma derivada, puede condenar a todo intento de regulación jurídica a un eterno rebufa. No cabe duda que, al mirar hacia el futuro (TETLOCK, 69), es obvio que «todavía no sabemos lo que tenemos que explicar». Y, por este motivo, «Necesitamos estar alerta respecto a la multiplicidad de caminos a través de los cuales las causas potenciales podrían producir una multiplicidad de potenciales efectos». En estas condiciones, aunque es claro que no podemos aspirar a tener hojas de ruta clarividentes que nos anticipen el devenir futuro, sí deberíamos aunar esfuerzos para diseñar normas proactivas que, a modo de *faros*, permitan orientarnos en

la niebla. Especialmente porque una estrategia meramente reactiva, probablemente siempre se quedaría corta y/o llegaría tarde y, por este motivo, podría cobrarse un alto precio.

En cualquier caso, el reto al que nos enfrentamos es hercúleo por varios motivos (al menos, tres):

- En primer lugar, porque, seguramente, estamos hablando de un riesgo *invisible*, bien porque todavía no se ha corporizado (o no todavía en toda su potencialidad), o bien, porque si lo ha hecho, quienes lo explotan (o están en condiciones de hacerlo) es muy probable que opten por la máxima discreción posible (y aspiren a que opere *en la sombra*).

- En segundo lugar, porque este desconocimiento y, especialmente, el hecho de que sea un fenómeno que no tenga precedentes, hace que sea imposible saber en qué sentido difiere de todo lo conocido anterior. En efecto, parafraseando a WILLIAMS (62 y 63 – que toma las palabras de WITTGENSTEIN⁵⁸ y de JUDT⁵⁹) es posible que nuestro diálogo social y político carezca todavía de un lenguaje preciso para abordar el desafío jurídico que tenemos por delante y esto nos genere una cierta incapacidad para enfrentarnos a él. En la medida que seguimos operando con categorías semánticas propias del mundo consciente, quizás, estos contornos conceptuales se erijan en límites para abordar los retos jurídicos del mundo preconscious (si no somos capaces de nombrar los problemas, difícilmente podremos solucionarlos). De ahí que, en este momento, parece que sólo estamos en condiciones de *intuir* la *posible* existencia de una (eventual) amenaza, pero todavía no podemos calibrar su naturaleza ni su intensidad. Por este motivo, el diseño de mecanismos jurídicos de prevención y reacción es especialmente complejo. Y, una derivada de lo anterior, sea que la sociedad todavía no vea la necesidad de ocuparse de este problema *en este momento*; o bien, de forma insuficiente, sólo se exija para aquellas dimensiones que conoce⁶⁰.

- En tercer lugar, porque, como ya ha sucedido con la implantación de otras tecnologías extractivas (por ejemplo, entre otros muchos, *Street View* de Google⁶¹), el capitalismo de la vigilancia aunará esfuerzos para neutralizar cualquier resistencia a su ímpetu extractivo; y para, en paralelo, alimentar la consolidación y habituación de estas tecnologías con el propósito de normalizar la apropiación del excedente conductual (en este sentido, nuevamente, es ilustrativo de esta estrategia la campaña publicitaria de la empresa *Meta* – exaltando las *virtudes* que el metaverso aportará a la *vida real* y, por ejemplo, al ámbito de la salud y a los tratamientos médicos).

⁵⁸ «Los límites de mi lenguaje significan los límites de mi mundo».

⁵⁹ «uno tiene que ser capaz de nombrar un problema si quiere solucionarlo».

⁶⁰ Con este propósito *descriptivo*, es claro que las referencias a la *conciencia* en los arts. 20.1.d y 30.2 CE, salvo que fueran objeto de una profunda redefinición hermenéutica, quedaría lejos del riesgo que se ha descrito en estas páginas. Y, en paralelo, aunque comparto que el acceso a estos estratos mentales bien podría identificarse con lo que MERCADER UGUINA (142) califica como la «intimidad de la intimidad» (y, por ende, podría quedar subsumido en el art. 18.1 y 4 CE), creo que la naturaleza pluriofensiva de la amenaza sobrepasa esta dimensión.

⁶¹ Y que (que describe ZUBOFF) forman parte de la una secuencia predecible de fases para normalizar la extracción del excedente: la incursión, la habituación, la adaptación y la redirección.

En cualquier caso, para completar esta aproximación, creo que también deben destacarse algunas reflexiones concluyentes más (también tres):

- En primer lugar, creo que la existencia de algoritmos que sean capaces de detectar patrones de conducta profundos (y, por lo tanto, contribuyan a desvelar la dimensión preconsciente humana) no parece que sea una innovación que deba ser calificada, en sí misma, como negativa⁶². Haciendo un paralelismo con otros avances en el ámbito de la salud, creo que estos descubrimientos podrían ser equiparables a la detección de nuevos indicadores en las analíticas de sangre y la consiguiente mejora en los diagnósticos de los pacientes. Si los algoritmos nos permiten acceder a lo que ha estado oculto tras el *velo de la evolución* no deberíamos denostarlo. Especialmente porque, la amenaza no radica tanto en el potencial de la herramienta, sino en el uso que se hace de la información a la que da acceso. Si el conocimiento de esta dimensión quedara reservado a la esfera estrictamente privada y, por ejemplo, se garantizara una encriptación *segura* (y, en paralelo, si se declara su absoluta inalienabilidad), la amenaza podría disiparse notablemente.

- En segundo lugar, es claro que las *fuerzas del mercado* apostarán (de hecho, ya lo están haciendo) por el acceso a nuestro preconsciente y lo harán a partir de la acumulación de nuestros *datos en la sombra*. Los modos de vida individuales y colectivos (SADIN, 2020, 30, 37, 52, 67, 75 y 112) están llamados a orientarse cada vez más por sistemas que, lejos de meramente informarnos, se están convirtiendo instancias dirigidas a conducirnos. Y, al hacerlo, lógicamente, se apropian de nuestra facultad de juicio (sin que, además, su prisma ético sea escrutable). En definitiva, están tratando de asumir el monopolio de la *racionalidad*. Especialmente, porque sus impulsores estiman que estas *tecnologías de la perfección*, a partir de la idea de que el comportamiento humano es optimizable si se modifica su entorno, pueden contribuir «a erigir un mejor orden general de los asuntos humanos». La sumisión a este *gobierno de los números* (empleando la construcción de SUPLOT), en virtud de la cual la normatividad se proyecta como un acto de programación (o computación), ajeno al arbitrio de la voluntad humana, es la renuncia que debemos asumir si queremos mitigar la incertidumbre y el peaje necesario para liberar a la humanidad del pesado fardo de la responsabilidad anudada a la toma de decisiones.

De ahí que deberíamos prevenirnos frente a cualquier correlación no espuria que la estadística computacional sea capaz de desvelar y que opere como factor precipitador previo, inmediato y decisivo de una posterior acción humana por debajo del nivel consciente. Aunque la monitorización ubicua en el ámbito profesional es todavía incipiente, en la medida que ya se están dando algunos ejemplos y algunos sectores son más proclives que otros, convendría hacer un seguimiento exhaustivo de su evolución (pues, podrían ser, empleando la metáfora de ZUBOFF, nuestros *canarios en la mina*). Desde este punto de vista, el art. 5.1 del proyecto de Reglamento de AI es un primer paso muy importante, pues, como muro de contención, debería erigirse en uno de nuestros principales *faros* para avanzar en este futuro incierto.

⁶² Así también se manifiesta WILLIAMS (56) cuando, a propósito del diseño de la arquitectura computacional, afirma que «el diseño persuasivo no es nocivo en sí mismo, por supuesto, por más que se valga de nuestros sesgos cognitivos. De hecho, podría usarse en nuestro provecho».

- En tercer lugar, debemos aspirar a alcanzar un equilibrio que trate de compensar las limitaciones humanas y también las que padecen las máquinas. Lejos del determinismo tecnológico, no deberíamos olvidar que, hasta las máquinas no sean capaces de reproducirse a sí mismas (CARR, 2011, 61 y 64), «toda tecnología es expresión de la voluntad humana». Sin menospreciar la inercia tecnológica (o *ímpetu tecnológico*) que dificulta los cambios de orientación (CARR, 2016, 200), en la medida que todavía estamos en disposición de poder elegir, (CARR, 2016, 187 a 193) debemos abogar por una *automatización centrada en los humanos* (y no en la actualmente predominante *automatización centrada en la tecnología*). Debemos primar y potenciar los sistemas que, lejos de suplantar el criterio humano, lo suplementen.

En definitiva, debemos tratar de prevenirnos frente a nuestras propias limitaciones y, sin duda, la tecnología puede sernos de ayuda frente a nosotros mismos. Aunque también es verdad que, si la empleamos estúpidamente, podría acabar acelerando nuestra extinción como especie...

7. Bibliografía citada

- George A. AKERLOF y Robert J. SHILLER (2015), *La economía de la manipulación*, Deusto.
- George A. AKERLOF y Robert J. SHILLER (2009), *Animal Spirits*, Gestión 2000, Madrid.
- Dan ARIELY (2008), *Las trampas del deseo*. Ariel, Madrid.
- Ignasi BELTRAN DE HEREDIA (2022), «Nadie da duros a cuatro pesetas (transparencia algorítmica y representantes de los trabajadores: el nuevo art. 64.4.d ET)». En *XXXII Congreso Nacional Asociación Española de Derecho del Trabajo y de la Seguridad Social*, Ministerio de Trabajo y Economía Social. Pendiente de Publicación (987 – 1001).
- Ignasi BELTRAN DE HEREDIA (2020), «Automatización y obsolescencia humana», En Peguera y Cerrillo (Coords.), *Retos jurídicos de la inteligencia artificial*, Aranzadi. (113 – 124).
- James BRIDLE (2020), *La nueva edad oscura*, Debate.
- Jan Christoph BUBLITZ (2022), «Novel Neurorights: From Nonsense to Substance», *Neuroethics*, 15:7, <https://link.springer.com/article/10.1007/s12152-022-09481-3>
- Nicholas CARR (2016), *Atrapados*, Taurus.
- Nicholas CARR (2011), *Superficiales*, Taurus.
- Antonio DIÉGUEZ (2019), *Transhumanismo*, Herder.
- Marcus DU SAUTOY (2020), *Programados para crear*, Acantilado.
- Marcus DU SAUTOY (2018), *Lo que no podemos saber*, Acantilado.
- Lisa FELDMAN BARRETT (2018), *La vida secreta del cerebro*, Paidós.
- Hannah FRY (2019), *Hola Mundo*, Blackie Books.
- Michael GAZZANIGA (1993), *El cerebro social*, Alianza.
- Jared GENSER, Stephanie HERRMANN y Rafael YUSTE (2022), *International Human Rights Protection Gaps in the Age of Neurotechnology*, NeuroRights Foundation, <https://ntc.columbia.edu/wp-content/uploads/2022/05/NeuroRightsFoundationPUBLICAnalysis5.6.22.pdf>
- Daniel GILBERT (2017), *Tropezar con la felicidad*, Ariel.

- Ruth W. GRANT (2021), *Los hilos que nos mueven*, Avarigani.
- Jonathan HAIDT (2019), *La mente de los justos*, Deusto.
- Byung-Chul HAN (2022), *Infocracia*, Taurus.
- Yuval-N. HARARI, (2016), *Homo Deus*, Debate.
- Philip N. JOHNSON-LAIRD (1990), *El ordenador y la mente*, Paidós.
- Daniel KAHNEMAN (2012), *Pensar rápido, pensar despacio*, Debolsillo.
- Daniel KAHNEMAN, Olivier SIBONY y Cass R. SUNSTEIN (2021), *Ruido*, Debate.
- Michael LEWIS (2003), *Moneyball, The Art of Winning an Unfair Game*, WW Norton & Co.
- Michael J. MAUBOUSSIN (2013), *La ecuación del éxito*, Empresa activa.
- Viktor MAYER-SCHÖNBERGER y Kenneth CUKIER (2013), *Big Data*, Turner Noema.
- Lee McINTYRE (2020), *La actitud científica*, Cátedra.
- Jesús MERCADER UGUINA (2022), «En busca del empleador invisible: algoritmos e inteligencia artificial en el derecho digital del trabajo», *El Cronista*, núm. 100 (136 – 145)
- Leonard MLODINOW (2008), *El andar del borracho*, Drakontos.
- Evgeny MOROZOV (2015), *La locura del solucionismo tecnológico*, Katz.
- José M. MUÑOZ y José A. MORINARO (2022), «The Rise of Neurotechnology: Reconceptualizing Human Rights», *CINET*, <https://inecenter.com/the-rise-of-neurotechnology-reconceptualizing-human-rights/>
- Ana B. MUÑOZ RUIZ (2022), «Inteligencia artificial y uso de algoritmos para gestionar el trabajo: la deshumanización de los trabajadores», *EuSocialCit*, <https://www.eusocialcit.eu/artificial-intelligence-and-the-use-of-algorithms-to-manage-work-blog/>
- Cathy O’NEIL (2017), *Armas de destrucción matemática*, Capitán Swing.
- Vance PACKARD (1973), *Las formas ocultas de propaganda*, Editorial Sudamericana.
- Judea PEARL y Dana MACKENZIE (2020), *El libro del porqué*, Pasado y Presente.
- Marta PEIRANO (2019), *El enemigo conoce el sistema*, Debate.
- Alex PENTLAND (2010), *Señales honestas*, Milrazones.
- Steven PINKER (2021), *Racionalidad*, Paidós.
- Eric SADIN (2022), *La edad del individuo tirano*, Editorial Caja Negra.
- Eric SADIN (2020), *La inteligencia artificial o el desafío del siglo: anatomía de un antihumanismo radical*, Editorial Caja Negra.
- Michael J. SANDEL (2016), *Justicia*, Debolsillo.
- Nate SILVER (2014), *La señal y el ruido*, Península.
- Barrhus F. SKINNER (1980), *Más allá de la libertad y la dignidad*, Editorial Fontanella.
- Cass R. SUNSTEIN (2017), *Paternalismo libertario: ¿por qué un empujoncito?*, Herder Editorial
- Cass R. SUNSTEIN (2012), «Análisis conductual del derecho», *Themis-Revista de Derecho*, núm. 62.
- Cass R. SUNSTEIN (2009), *Leyes de miedo*. Katz, Buenos Aires.
- Cass R. SUNSTEIN y Richard THALER (2009), *Un pequeño empujón*, Taurus.
- Dick SWAAB (2014), *Somos nuestro cerebro*, Plataforma Actual.

- Alain SUPIOT (2015), *La gouvernance par les nombres*, Fayard.
- Nassim N. TALEB (2012), *El cisne negro*, Booket.
- Philip TETLOCK (2016), *El juicio político de los expertos*, Capitán Swing.
- Richard THALER (2016), *Todo lo que he aprendido sobre la psicología económica*, Deusto.
- James WILLIAMS (2021), *Clics contra la humanidad*, Gatopardo Ensayo.
- Tim WU (2020), *Comerciantes de atención*, Capitán Swing.
- Rafael YUSTE y Sara GOERING (2017), «Four ethical priorities for neurotechnologies and AI», *Nature*, vol. 551.
- Shoshana ZUBOFF (2020), *La era del capitalismo de la vigilancia*, Paidós.

Última consulta de todos los enlaces: 13/11/2022