

**NADIE DA DUROS A CUATRO PESETAS
(TRANSPARENCIA ALGORÍTMICA Y REPRESENTANTES DE LOS
TRABAJADORES: EL NUEVO ART. 64.4.d ET)**

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

1. NADIE DA DUROS A CUATRO PESETAS (Y LA BASURA ES BASURA)

El propósito de esta comunicación es analizar la obligación de transparencia algorítmica que se desprende del nuevo art. 64.4.d ET en la redacción dada por la Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales.

No obstante, antes de abordar su contenido, conviene llevar a cabo una aproximación con carácter propedéutico. Al hablar de algoritmos (sin perjuicio de la definición a la que se hará referencia posteriormente) convendría tener muy presente qué es lo que, a fecha de hoy, la automatización (o “inteligencia artificial”) puede hacer y qué no. Y, en este sentido (siguiendo a PEARL/MACKENZIE, 376), debe advertirse que “desde luego por ahora no hemos hecho máquinas que piensen en ninguna interpretación suficientemente humana de la palabra. Hasta ahora solo podemos simular el pensamiento humano en dominios de definición muy reducida, que únicamente cuentan con las estructuras causales más primitivas. En este ámbito podemos crear máquinas cuyo rendimiento supera incluso al humano, porque hablamos de dominios que recompensan lo único que los ordenadores saben hacer bien: computar”.

A pesar de esta limitación (¿temporal?) de las máquinas, sí estamos experimentando un cambio de paradigma de una enorme trascendencia. En efecto, lejos todavía de la “inteligencia” artificial, estamos viviendo una revolución de la estadística computacional (FRY, 14) que está posibilitando el tránsito de una lógica predictiva basada en la modelación y la causalidad a otra fundamentada, meramente, en las correlaciones.

Si pretendemos medir algo (sintetizando el fantástico trabajo de MAYER/CUKIER) es obvio que lo mejor es disponer de todos los datos. Sin embargo, si la escala es vasta, este método no es práctico y, por este motivo, se ha acudido al muestreo. Así pues, partiendo de este escenario, en un entorno de “datos escasos”, se colige que conviene tratar de medir, de la forma más precisa posible, aquello que se pretende cuantificar. De hecho, ante la dificultad para recoger, organizar, almacenar y analizar la información, durante décadas, se ha

recurrido a la estadística para tratar confirmar el mejor resultado al alcance con la menor cantidad de datos posible.

Aunque pueda resultar contraintuitivo, la cuestión es que la precisión de una muestra aumenta (acusadamente) con su aleatoriedad (y no con el mayor tamaño de la misma). Además, en la medida que sólo se recopilan ciertos y limitados datos, junto a la aleatoriedad de la muestra apuntada, es esencial asegurarse de la exactitud y calidad de los mismos (para evitar desviaciones y errores). Por otra parte, es necesario acudir a hipótesis basadas en teorías para tratar de hacer una aproximación que ayude a determinar qué variables deben ser examinadas y recoger datos en consecuencia (los cuales contribuyen a desmentirlas o confirmarlas).

A partir de todos estos elementos, se espera detectar relaciones de causalidad que contribuyan a detectar patrones y hacer predicciones. Sin embargo, lo cierto es que, como apuntan MAYER/CUKIER (86), este método tampoco es infalible, pues, es difícil encontrar relaciones causales concluyentes: “la causalidad rara vez – o ninguna – puede ser demostrada, solo mostrada con un elevado grado de probabilidad”. De modo que, si se quiere dar validez a una teoría, se ha propuesto adoptar la lógica de la refutación (POPPER, 149), en vez de la mera búsqueda de factores de confirmación de la misma.

La era de los datos masivos (siguiendo con MAYER/CUKIER) está alterando profundamente la realidad analítica descrita. Especialmente porque, en la medida que es posible procesar ingentes cantidades de datos, ya no tiene sentido recurrir a una muestra. De hecho, de lo que se trata (41) es de recopilar todos los datos posibles y, cuando sea factible, absolutamente todos (incluso los erróneos). Esto es: “N = todo”.

Y, la carrera hacia la datificación masiva (o “minería de la realidad”) está, en estos momentos, en plena ebullición. Lo que, en el marco del “capitalismo de la vigilancia” (ZUBOFF), ha propiciado una voraz lucha para conquistar todos los espacios susceptibles de extracción del excedente conductual humano. Esta “minería humana”, en definitiva (MAYER/CUKIER, 142), se ha convertido en el nuevo polvo de oro con el que fundir lingotes. De hecho, ya se habla de la “economía de datos” (MAYER-SCHÖNBERGER/RAMGE) y, desde este punto de vista, el denominado “solucionismo tecnológico” (MOROZOV) – la hiperconexión de todo – no dejaría de ser un “pretexto” para expandir la extracción de datos a nuevos espacios vírgenes. Y en esta “caza” (por motivos diversos) ya participan entes privados como públicos.

En todo caso, este método supone un cambio radical con respecto al anterior. Especialmente porque, aunque puede traducirse en un aumento de la imprecisión, tiene la ventaja de que la suma puede contribuir a mejorar el pronóstico (probabilísticamente hablando). En el fondo (MAYER/CUKIER, 57), “más [datos] es mejor que mejor[/es datos]”. Para ilustrar: si sólo tenemos un termómetro para medir la temperatura de un viñedo, cuanto más preciso, mejor. Ahora bien, si podemos disponer de un sensor de temperatura para cada cepa, quizás, podamos renunciar a un poco de exactitud, a cambio de acceder a un “mapeo” de la temperatura general y/o a su progreso continuo en el tiempo.

De hecho, de esta forma también es posible abordar el análisis de subcategorías o detalles de fenómenos concurrentes (de más difícil acceso a través de las muestras). Especialmente porque (42 y 43) “el empleo de la totalidad de los datos hace posible advertir conexiones y detalles que de otro modo quedan oscurecidos en la vastedad de la información”. De todo ello se extrae que (52) “los datos masivos convierten los cálculos aritméticos en algo más probabilístico que preciso”. Y, en este sentido, la verdadera revolución que el procesamiento de datos masivos provoca radica en la proliferación de la analítica predictiva.

Y esto, describiendo el cambio de paradigma que se ha avanzado, se traduce en el hecho de que no es necesario saber el por qué de las cosas, bastando con saber el qué. Este método, en definitiva, consiste en averiguar (a través de correlaciones) que algo pasa, aunque no se sepa exactamente qué es lo que lo provoca (esto es, cuáles son sus mecanismos internos). Aunque no puede excluirse que dos cosas se comporten del mismo modo por pura coincidencia, lo cierto es que una correlación puede facilitar una alta relación estadística entre dos valores de datos.

Y, llegados a este punto, es conveniente detenerse en la descripción de lo que es una correlación (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”. Un ejemplo, quizás, ayude de nuevo a ilustrar lo anterior: si los datos muestran que un zumbido precede a la ruptura de una determinada pieza de una máquina, basta estar pendiente del patrón (el zumbido) para anticiparse al cambio de la misma (aunque no se tenga ni remota idea de qué es lo que acaba provocando la ruptura). Aunque conviene reiterar que (SILVER, 228) la existencia a de “una relación estadística entre dos variables no significa que una influya en la otra, ni viceversa”.

En definitiva, repárese que se renuncia al conocimiento de las causas internas (a la causalidad) en favor de una (mera) aproximación (en el mejor de los casos, altamente probable). Y la acumulación de datos masivos es lo que, precisamente, permite que la probabilidad de que exista un vínculo fuerte entre dos acontecimientos se eleve (aunque como se apuntará, puede no ser así). Téngase presente, por consiguiente, que ya no es necesario describir una hipótesis. Basta dejar que los datos, a través de un análisis de correlación, “hablen” y permitan hacer predicciones (en definitiva, saquen “sus propias” conclusiones).

La subsistencia del ser humano probablemente ha sido posible, entre otros motivos, a la capacidad de anticipar el futuro. 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. Nuestro lóbulo frontal (GILBERT, 35 y 43) se ha erigido en nuestro “previsiómetro” durante miles de años, utilizando la información del pasado (y tratando de detectar las relaciones de causalidad), para proyectarla hacia el futuro (para anticiparnos y evitar riesgos y/o peligros). Y, de hecho, 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)¹. No obstante, 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”. Desde “hace poco”, el uso de “muestras” aleatorias de datos (lo más precisos posibles), como se ha expuesto, ha sido el principal instrumento para combatir este sesgo (tan) humano. Y, la datificación y las correlaciones están sentando las bases de lo que probablemente será el nuevo paradigma.

Como expone DU SAUTOY (86 y 87), la nueva era del aprendizaje automático está estrechamente ligada a la descomunal proliferación de datos. Es la “chispa” que ha desatado su explosión. De modo que “la idea central del aprendizaje automático es que se puede crear un algoritmo que sea capaz de plantearse nuevas preguntas si ve que algo no va bien. Que aprenda de sus errores. Con cada error, el algoritmo retoca sus ecuaciones, de modo que la próxima vez actuará de forma diferente y no tropezará con la misma piedra. Por eso es tan importante el acceso a los datos: cuantos más ejemplos tenga estos algoritmos inteligentes para ejercitarse, más experiencia acumularán y más se perfeccionarán con cada retoque”. Este tipo de algoritmo, a su vez, tiene la particularidad (DU SAUTOY, 92) de que los cambios en la programación no son de “arriba abajo”, esto es, a través de las órdenes comunicadas por un programador, sino que se trata de un proceso de “abajo arriba”: es el algoritmo el que se transforma desde abajo (él sólo), interactuando cada vez con más datos. De ahí que se hable de “aprendizaje automático”.

¹ Y siguiendo con GAZZANIGA (118) – en una cita extensa, pero muy ilustrativa: “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”. Así pues, (260) en el seno de la organización medular de nuestro sistema cognitivo, tenemos una capacidad especial que interpreta constantemente teorías sobre las “relaciones causales entre los acontecimientos elementales que ocurren en el interior y en el exterior de nuestras cabezas”. En definitiva, el hemisferio izquierdo es el que posibilita la capacidad de hacer inferencias.

Este *big data* en expansión es el combustible que alimenta el “entrenamientos de datos” que da pie a la iteración que necesitan los algoritmos para poder identificar las correlaciones (en el mejor de los casos, fuertes) que lubrican la nueva analítica predictiva (gracias al papel de las redes bayesianas²). En definitiva, describen la nueva (en términos de ZUBOFF) “utopía de certeza”.

No obstante, este avance revolucionario también tiene sus limitaciones severas y proyecta escenarios inquietantemente sombríos. Permítanme sintetizárselos:

- En primer lugar, de forma progresiva, la “intuición” humana va quedando relegada a un plano marginal, frente a la “certeza” del orden algorítmico (SADIN)³. Porque lo cierto es que, de algún modo, están contribuyendo a descifrar ciertos patrones de nuestro comportamiento. En efecto, “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” (SAUTOY, 110 y 111). En definitiva, la comprensión del cerebro humano se ha erigido en el último bastión a conquistar. En una nueva versión de “conductismo severo digital”, se aspira a conducir el comportamiento y controlar los deseos mediante “estimulaciones” externas sofisticadas. De ahí que se haya expandido el neuromarketing, la neuroeducación, la neuropolítica, etc. Y, frente a estos avances, quizás, deberíamos empezar a hablar de derechos fundamentales de “quinta generación” (y que – con una denominación que podría mejorarse – se engloban en lo que se conoce genéricamente como “neuroderecho”).

- En segundo lugar, es difícil que podamos codificar todos los datos y/o bien que las máquinas sepan interpretarlos adecuadamente (por ejemplo, tienen muchos problemas para comprender la ambigüedad y el contexto del lenguaje natural), sin olvidar 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 que débiles. En efecto, (MAYER/CUKIER, 74) debe tenerse en cuenta que “cuando el número de datos crece en orden de magnitud, también se aprecian más correlaciones espurias, fenómenos que parecen estar conectados aun cuando no sea así”. Esta confusión entre la señal y el ruido puede llevar a pronósticos desviados. Especialmente porque (241) en la medida que “nunca podemos disponer de información perfecta, nuestras predicciones resultan inherentemente falibles”. Es lo que BRIDLE (96), también críticamente, describe como la “magia (o la falacia) de los macrodatos”. Como apunta DU SAUTOY (120) aunque los algoritmos pueden analizar ingentes cantidades de datos y darles un sentido, son propensos a omitir las irregularidades (esto es, las

² Sobre el Teorema de Bayes y una posible aplicación en el Derecho del Trabajo y, en concreto, el despido preventivo, véase, BELTRAN DE HEREDIA RUIZ (2021, 1).

³ Entre otros muchos efectos, esto provoca un bloqueo de la creatividad. De hecho (siguiendo con BECK, 295 y 299), “un mundo sin fallos no debería parecernos muy progresista, sino todo lo contrario. Sería un mundo estático, estable y enemigo del progreso, porque sin el riesgo de cometer un error tampoco existe el valor para descubrir algo nuevo (...). El nuevo conocimiento no se encuentra en los libros, sino que debe crearse antes. Pero esto solo es posible si se asume el riesgo de cometer un fallo”. Además, hay evidencias claras (CARR, 25) de que el trabajo creativo y analítico de toda clase (incluido el altamente cualificado) se ve mediatizado por el software.

situaciones fuera de lo normal). Sin olvidar que, en ocasiones, deben hacer valoraciones sobre hechos que están “fuera de muestra” (SILVER, 62 y 63), es decir, sobre acontecimientos sobre los que no hay antecedentes y, por este motivo, el margen de error podría ser muy elevado.

En el fondo (y justificando el título de este breve ensayo), siguiendo de nuevo con DU SAUTOY (120 y 121), el teorema *No free lunch* (que vendría a ser: “nadie da duros a 4 pesetas”), “prueba que no hay un algoritmo de aprendizaje universal que pueda predecir los resultados correctamente en cualquier circunstancia. El teorema prueba que, aunque se muestre al algoritmo de aprendizaje la mitad de los datos, siempre es posible reorganizar el resto de los datos, los que están ocultos, de modo que el algoritmo sea capaz de producir una buena predicción para los datos conocidos, pero que no encaje con el resto de los datos. Los datos nunca serán suficientes por sí mismos. Tienen que conjugarse con conocimiento. Aquí es donde la programación humana parece mejor adaptada para tener en cuenta el contexto y una visión de conjunto, al menos por ahora”. Además, “No está nada claro que todos los problemas a los que debe enfrentarse un ente inteligente sean computables (...) y, por tanto, que una máquina inteligente pueda resolverlos operando sólo con algoritmos” (DÍEGUEZ, 80). O, al menos (si se quiere alcanzar una IA “fuerte”), difícilmente será capaz de hacerlo hasta que no “entienda” la causalidad de los fenómenos y los contrafactuales (véase, PEARL/MACKENZIE).

- En tercer lugar, téngase presente que (DU SAUTOY, 119), ya hay numerosas pruebas de que los algoritmos cobijan “sesgos ocultos” (una preocupante derivada del fenómeno que se sintetiza con la expresión: “entra basura, sale basura”; o “*garbage in, garbage out*” – “GIGO”). O, 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”. Siempre que haya un sesgo en los datos de entrenamiento (KAHNEMAN/SIBONY/SUNSTEIN, 369 y 370) es muy posible que se diseñe, de forma deliberada o no, un algoritmo que codifique la discriminación. En el fondo (CARR, 144 a 147; y DÍEGUEZ, 81), los algoritmos son “sabio-idiotas”, pues, si bien es cierto que un sistema experto puede superar la capacidad de un ser humano (de hecho, una simple calculadora ya lo hace), no han podido superar la “estrechez de percepción” que los caracteriza (pues, fuera de su ámbito de aplicación son perfectamente inútiles). Además, aunque son altamente prácticos y productivos, carecen absolutamente de “curiosidad, imaginación y sentido real del mundo”, mostrando todavía una “falta espeluznante de sentido común” (y son capaces de repetir un error exponencialmente).

– En cuarto lugar, el carácter oscuro de las lógicas internas de algunos algoritmos (como los de aprendizaje profundo). En estas “cajas negras” (O’NEIL), el rastreo y monitorización de las decisiones adoptadas es, en el mejor de los casos, “dificultoso” (¡incluso para sus propios creadores!). De hecho, BODEN (68), refiriéndose a las redes neuronales, afirma que es la forma en que funcionan lo que las hace intrigantes. 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 bien, abundando en esta idea, siguiendo con SADIN (76), “hay un nuevo tipo de ‘caja negra’ en el horizonte, y no es la generada por las bases de datos o los algoritmos que impiden a los usuarios, de facto, por su opacidad estructural, apresar su constitución, sino otra caja negra producto del encadenamiento de caracteres cuya evolución combinatoria se nos volvería cada vez más oscura: ‘una vez que la red de neuronas aprendió a reconocer algo, un desarrollador no se puede dar cuenta de cuánto éxito ha tenido. Es como en el cerebro: usted no puede cortar una cabeza y mirar cómo funciona”. Como apunta DU SAUTOY (177 a 179), la dificultad de trazabilidad de los algoritmos que “aprenden” y “cambian” constantemente, hace que sólo seamos capaces de conocer cómo funcionan de forma indirecta. Por ejemplo, el Proyecto *DeepDream* trata de descifrar el “razonamiento” de un algoritmo de reconocimiento facial a través de las imágenes que genera. Lo que no deja de ser paradójico, pues, del mismo modo que el ser humano ha utilizado el arte para exteriorizar los aspectos más íntimos y profundos de su ser, parece que ahora estamos haciendo lo mismo con estas máquinas. Y, LATORRE (211) advierte que el “problema de la trazabilidad podría escapar de nuestras manos si las propias inteligencias artificiales pasan a corregirse a sí mismas. La velocidad de cambio de un código fuente que se mejora a sí mismo puede hacerse vertiginosa, de forma que todo el proceso escape a las capacidades humanas”.

Paradójicamente, a la luz de todo lo expuesto, adviértase que estamos sustituyendo las incógnitas derivadas de nuestra (errática y humana) intuición causal por otras igualmente indescifrables. Por todo ello, más allá de dar a conocer los algoritmos (como, por ejemplo, establece el nuevo art. 64.4.d ET), es exigible un mecanismo que permita su “traducción” humana. Sin esta transparencia y comprensibilidad, especialmente exigible en los algoritmos de aprendizaje profundo (DU SAUTOY – 120 – habla de “metalenguaje”), difícilmente, seremos capaces de “entender” sus decisiones (y acabar siendo víctimas de las mismas – O’NEIL o FRY).

– Y, en quinto lugar, porque (MAYER/CUKIER, 188, 216 y 217) podemos llegar a articular sanciones (o castigos) a partir de meras propensiones basadas en correlaciones y no en causalidades. 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 replanteamiento profundo sobre la responsabilidad individual (y, condicionaría de forma determinante nuestra libertad). Además, reparen que la tarea de refutar lo predicho, no sólo nos exigirá dar luz sobre lo insondable, sino que, además (MAYER/CUKIER, 236) es “axiomáticamente imposible demostrar la falsedad de una predicción” (esto es, de algo que no ha pasado).

Es obvio que los estos instrumentos pueden contribuir a suplir muchas de las carencias humanas y, por consiguiente, pueden complementar nuestras capacidades. No obstante, a la luz de lo expuesto, deben ser empleados con (mucho) cautela, pues, tienen una capacidad “destruktiva” (O’NEIL) que no puede desdeñarse en absoluto. Esto justifica que la Ley 12/2021 (derivada del RD Ley 9/2021) haya conferido un derecho de información al respecto a la representación unitaria. Extremo que será objeto de estudio a continuación.

2. LA TRANSPARENCIA ALGORÍTMICA AL ALCANCE DE LA REPRESENTACIÓN DE LOS TRABAJADORES: EL NUEVO ART. 64.4.D ET

La predicción sobre la base de *big data* persigue una quimera, pues, es ilusorio pretender neutralizar la incertidumbre (y aspirar a anticipar con “certeza” el futuro). El conocido como “problema de la inducción”/“problema del pavo” (RUSSELL/TALEB) no se disipa (completamente) con más datos.

Desde este punto de vista (y teniendo en cuenta el concepto *GIGO*), las voces de alarma frente a las injusticias, equidad y margen de error de las decisiones basadas en estos instrumentos están plenamente justificadas (algunos ejemplos, entre otros, en FRY u O’NEIL). Y las medidas legales dirigidas a dar más transparencia a estos instrumentos (empleados por el sector privado y el público) son absolutamente oportunas, necesarias y no deberían postergarse.

En este contexto (y tras la exposición – propedéutica – del anterior apartado) se incardina el nuevo art. 64.4.d ET que reza como sigue: “*El comité de empresa, con la periodicidad que proceda en cada caso, tendrá derecho a: d) Ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles*”. Esta disposición complementa el derecho de alcance individual recogido en el RGPD y, al dotarla de una dimensión colectiva, contribuye a reforzarlo.

El ámbito de aplicación de esta obligación está delimitado por tres dimensiones: subjetiva, objetiva y temporal:

- *En el primer plano*, es claro que el sujeto obligado es el empresario (de cualquier sector - y no sólo de las plataformas digitales); y los receptores son el comité de empresa, los delegados de personal, la representación sindical y, según la materia y las circunstancias, podría extenderse a los delegados de prevención.

- *En el plano objetivo*, la definición de “algoritmo” y de “sistema de inteligencia artificial” se erigen en el primero de los “escollos” a superar para poder determinar el ámbito de aplicación de esta nueva obligación de transparencia. Lejos todavía de la existencia de una “inteligencia mecánica”, la clave del precepto es que el término “algoritmo” puede estar referido a decisiones

automatizadas o no⁴. Lo que, ciertamente, hace que el apartado d) describa un ámbito de aplicación mayor que el del art. 22 RGPD (circunscrito a las “decisiones individuales automatizadas”⁵). En efecto, téngase presente que la empresa está obligada a facilitar la información sobre estos instrumentos siempre que “afecten a la toma de decisiones” (sobre ciertas materias). Y, por consiguiente, no está exigiendo que la obligación nazca sólo si la decisión “ha sido tomada” por el programa (término este último que, a los efectos expositivos, se va a emplear como equivalente al de “algoritmo” y “software”⁶). Lo que podría plantear algunos problemas, pues, podría ser difícil deslindar el funcionamiento de la “máquina” de los sesgos y heurísticas propios del empresario que ha acabado tomando la decisión a partir de la información suministrada por la misma (como, por ejemplo, lo que se conoce como “complacencia automatizada” y/o “sesgo de la automatización”⁷, sin olvidar los problemas “humanos” para comprender el sentido de la estadística y de la probabilidad⁸).

El listado de materias podría erigirse en un elemento que contribuyera a esta acotación si no fuera porque incluye todo el “ciclo vital” contractual: condiciones de trabajo, el acceso y mantenimiento del empleo (lo que incluye, obviamente, a la flexibilidad interna y a la externa) y la elaboración de perfiles.

Así pues, con estos mimbres, es difícil delimitar qué queda fuera de este deber de información empresarial (por ejemplo, una fórmula de Excel podría quedar subsumida en este precepto, pero también reglas matemáticas y/o estadísticas

⁴ Un algoritmo (siguiendo a FRY, p. 8) es “un conjunto ordenado y finito de operaciones que permite hallar la solución a un problema”. Es (siguiendo con la misma autora), “simplemente una serie de instrucciones lógicas que muestran cómo realizar una tarea de principio a fin. En esta amplia definición, la receta de un pastel es un algoritmo, como lo es la lista de indicaciones que podríamos darle a un extraño que se ha perdido (...), [E]n teoría cualquier lista autónoma de instrucciones para lograr un objetivo concreto y definido podría calificarse de algoritmo”. En definitiva (HARARI, 100) es “un conjunto metódico de pasos que pueden emplearse para hacerse cálculos, resolver problemas y alcanzar decisiones”.

⁵ Y que debe complementarse con el contenido de los arts. 5, 13.2 f) y 14.2 g) RGPD.

⁶ De hecho, “algoritmo” y “programa” se han empleado indistintamente como equivalentes (JOHNSON-LAIRD, 42). Y, probablemente, podría extenderse a “software”, pues, el art. 3.1 de la propuesta de Reglamento de IA, describe un sistema de IA como “*El software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa*”.

⁷ Siguiendo a CARR (85 a 88) la complacencia automatizada “tiene lugar cuando un ordenador nos atonta en una falsa sensación de seguridad. Estamos tan confiados en que la máquina trabajará inmaculadamente y solucionará cualquier imprevisto que dejamos nuestra atención a la deriva. Nos desenganchamos de nuestro trabajo o al menos de la parte de él que maneja el software, y podemos como resultado de ello perdernos señales de que algo va mal”. De modo que “cuanto más sofisticado es un ordenador (...) más empezamos a asumir que está corrigiendo tus errores, más empezamos a creer que lo que sale de la máquina es exactamente como debería ser. Es algo visceral”. Y el sesgo de la automatización “surge cuando las personas dan un peso excesivo a la información que aparece en sus monitores. La creen incluso cuando la información es errónea o engañosa. Su confianza en el software se vuelve tan intensa que ignoran o desechan otras fuentes de información, incluidos sus propios sentidos”. Como apunta TALEB (232) ciertas predicciones, en el momento que aparecen en una pantalla, la proyección cobra vida propia, pierde su vaguedad y abstracción y se convierte en lo que los filósofos llaman reificación, algo investido de la calidad de concreto; así adquiere una vida nueva como objeto tangible”.

⁸ Extensamente al respecto, BELTRAN DE HEREDIA RUIZ (2021,2).

sencillas y, obviamente, también las más complejas); y, por este motivo, todo apunta a que acabarán siendo los Tribunales los que establezcan los límites – probablemente, ante la negativa empresarial a satisfacer ciertas peticiones (siendo preferible que sean acotados por la negociación colectiva ex art. 64.9 ET).

En paralelo, teniendo en cuenta el principio *G/GO* anteriormente expuesto, es esencial saber si la norma extiende la obligación de información a los datos que alimentan al programa en cuestión. Especialmente, porque si la respuesta fuera negativa podría limitar la efectividad del propósito que el precepto persigue, pues, el software podría haberse configurado de forma neutra y, en cambio, fuera la base de datos que lo alimenta la que podría contener la desviación que se pretende evitar. Sobre esta cuestión, atendiendo a la literalidad de la norma, parece que la respuesta debería ser positiva, pues, explícitamente debe informarse sobre los “parámetros, reglas o instrucciones” en los que se fundamentan estos programas. Y, aunque esta enumeración está formulada en términos ejemplificativos (PÉREZ AMORÓS, 186), la mención del término “parámetro” podría ser determinante si se acude a su definición: “Dato o factor que se toma como necesario para analizar o valorar una situación”. Por consiguiente, parece que debería informarse sobre los datos y el “entrenamiento” que se alcance con ellos si fuera el caso. Si el software es capaz de “aprender” a partir de iteraciones, es obvio que en aras a la transparencia deberá facilitarse toda esta información (lo que, como se apuntará, en el caso de las “cajas negras” podría acabar siendo una tarea muy compleja). En cuanto a las “reglas” e “instrucciones” son términos que no parecen exigir un detalle técnico exhaustivo y/o profundo (especialmente porque, como también se apuntará a continuación, según los casos y la complejidad que atesoren, podrían erigirse en una vía para, precisamente, “ofuscar” la información solicitada).

A su vez, en la medida que es posible que pocas empresas tengan capacidad para desarrollar internamente su propio software, en muchas ocasiones, operarán conforme a productos adquiridos en el mercado y que estén protegidos por leyes de propiedad industrial e intelectual, sin olvidar que parte de esta información también podría quedar amparada por la Ley 1/2019, de Secretos empresariales.

Finalmente, en el caso de que ninguno de los límites anteriores impida el acceso a la información, también pueden plantearse dificultades para su comprensión. En la medida que no puede exigirse a quien haga la petición que tenga un nivel técnico experto, es obvio que la obligación de información empresarial debería incluir su procesamiento para que sea “consumible” para poder ser valorada y evaluada. Si no se incluyera esta dimensión, la transmisión de la información “en bruto” podría erigirse en una barrera insuperable para comprender su funcionamiento. Lo que, quizás, debería ser calificado como un incumplimiento de la Ley. En definitiva (PÉREZ AMORÓS, 187), el “derecho de información sobre algoritmos ha de ser suficiente y oportuno”. Motivo por el que desplaza a las empresas la obligación de conocer qué software están contratando y cuál es su funcionamiento (especialmente si no quieren incurrir en la responsabilidad que el incumplimiento del art. 64.4.d ET podría acarrear).

Relacionado con lo anterior, una de las cuestiones que podrían plantear una importante controversia es determinar si la obligación que se deriva del citado ap. d) podría exigir que las empresas hagan una “auditoría algorítmica”. En la medida que la transmisión de información “en bruto” impida acceder a un mínimo nivel de transparencia sobre el funcionamiento del software (como, por ejemplo, si se llegan a emplear “cajas negras”) no parece descabellado pensar que estén obligadas a dar cuenta de los instrumentos de que se han dotado. En algunos casos, parece que sería la opción más accesible (y/o única posible). Es obvio que esta cuestión podría acarrear un (elevado) coste económico. No obstante, esto podría crear la presión necesaria para que las compañías que programan estos productos facilitaran algo parecido a un “prospecto” sobre el funcionamiento del software (algo similar a lo que tienen que hacer las farmacéuticas con sus medicamentos para venderlos; o que, incluso, como sucede en este sector también, en función de la materia a la que van dirigidos⁹, tuvieran que estar sometidos a una autorización previa)¹⁰.

A la luz de todo lo expuesto, puede vislumbrarse con mayor claridad que la cuestión inicial sobre qué debe subsumirse en el concepto de “algoritmo” es absolutamente medular, pues, medirá la carga a la que está sometidas las empresas.

- Y, finalmente, *en el plano temporal*, un aspecto que no queda explicitado es la periodicidad con la que la empresa debe dar cuenta de esta información. Sobre esta cuestión deben realizarse dos aproximaciones: el momento y la frecuencia. En cuanto a la segunda, el art. 64.4 ET, de forma genérica, somete esta obligación a la “periodicidad que proceda en cada caso”. En este sentido, parece

⁹ Recuérdese que en la propuesta de Reglamento Europeo de IA (abril 2021), en el Anexo III se describen como sistema de IA de alto riesgo los relacionados con el Empleo, gestión de los trabajadores y acceso al autoempleo. Y, en concreto, los siguientes: “a) sistemas de IA destinados a utilizarse para la contratación o selección de personas físicas, especialmente para anunciar puestos vacantes, clasificar y filtrar solicitudes o evaluar a candidatos en el transcurso de entrevistas o pruebas; b) IA destinada a utilizarse para tomar decisiones relativas a la promoción y resolución de relaciones contractuales de índole laboral, a la asignación de tareas y al seguimiento y evaluación del rendimiento y la conducta de las personas en el marco de dichas relaciones”.

¹⁰ De hecho, la propuesta de Reglamento de IA anteriormente citado establece lo siguiente (ap. 5.2.3 de la “Explicación detallada de las disposiciones específicas de la propuesta”): “se establecerá un nuevo sistema de cumplimiento y aplicación para los sistemas de IA de alto riesgo independientes que se mencionan en el anexo III. Este seguirá el modelo de la legislación del nuevo marco legislativo que los proveedores aplicarán mediante controles internos, con la salvedad de los sistemas de identificación biométrica remota, que se someterán a evaluaciones de la conformidad efectuadas por terceros. Una solución efectiva y razonable para dichos sistemas podría consistir en realizar una evaluación integral de la conformidad ex ante mediante controles internos, combinada con una supervisión ex post estricta, dado que la intervención reguladora se encuentra en una fase temprana, que el sector de la IA es muy innovador y que apenas están empezando a acumularse los conocimientos necesarios para llevar a cabo auditorías. Para evaluar los sistemas de IA de alto riesgo «independientes» mediante controles internos, sería necesario cumplir ex ante de manera plena, efectiva y debidamente documentada todos los requisitos del Reglamento, así como los sólidos sistemas de gestión de la calidad y los riesgos y el seguimiento posterior a la comercialización. Una vez que el proveedor haya llevado a cabo la evaluación de la conformidad oportuna, deberá registrar dichos sistemas de IA de alto riesgo independientes en una base de datos de la UE que la Comisión gestionará con el propósito de redoblar la transparencia y la vigilancia públicas y de fortalecer la supervisión ex post por parte de las autoridades competentes”.

que, como mínimo, deberá informarse en el momento de su implantación y también en el que se introduzca algún cambio que altere su funcionamiento. El problema es que, si el aprendizaje es iterativo a partir del entrenamiento de datos, a medida que el algoritmo “aprenda”, su funcionamiento podría verse modificado (e, incluso, recordando las palabras de LATORRE, podría hacerlo “vertiginosamente”). Y, para el caso de que se empleen algoritmos de “abajo arriba” (DU SAUTOY), sin exigirse una información en tiempo real (pues, podría suceder que no sea materialmente posible), sí parece que debería exigirse una periodicidad mucho más alta. Otra derivada de la dimensión temporal es si la empresa también está obligada a suministrar la información una vez que la decisión se ha tomado (hacerlo con carácter previo es indudable). Salvo en los supuestos en los que sea de aplicación el art. 22 RGPD, también dependerá del tipo de software, pues, es posible que sólo sea posible conocer el funcionamiento del algoritmo (o a alcanzar una mejor comprensión sobre su funcionamiento) a partir de los resultados. En definitiva, parece que dependerá de la configuración de cada uno de los programas.

3. Valoración final

A la luz de todo lo expuesto, podrían extraerse las siguientes valoraciones finales (dos):

- *Primera valoración final:* siguiendo a DU SAUTOY (366), debe advertirse que “hay mucho bombo sobre la inteligencia artificial. Existen demasiadas iniciativas que se anuncian con la etiqueta de la inteligencia artificial, pero que son poco más que estadística o ciencia de datos”. A pesar de ello (y de sus profundas limitaciones), no podemos olvidar que los algoritmos son una poderosa herramienta para complementar nuestras limitaciones. La cuestión clave es si, en contextos de incertidumbre, pueden adoptar mejores decisiones que los humanos. Y, ciertamente, con las debidas cautelas y exigiendo la transparencia oportuna, en algunos ámbitos importantes podemos pensar que sí.

De hecho, tengan en cuenta que el juicio humano no sólo puede ser mejorado por decisiones adoptadas por algoritmos, sino que muchos instrumentos de predicción “mecánica” basados en reglas estadísticas sencillas (y, obviamente, también por otras más complejas), son capaces de mejorar nuestro “rendimiento” a la hora de adoptar decisiones.

Desde este punto de vista, permítanme que comparta las siguientes reflexiones expuestas por KAHNEMAN/SIBONY/SUNSTEIN, en favor de una predicción “mecánica” (basada en reglas sencillas, estadísticas o no, o en sofisticados modelos de IA) con carácter preferente a un juicio humano:

Primera: “la resistencia a los algoritmos, o ‘aversión a los algoritmos’, no siempre se manifiesta en un rechazo generalizado a la adopción de nuevas herramientas de apoyo a la toma de decisiones. Más a menudo, la gente está dispuesta a dar una oportunidad a un algoritmo, pero deja de confiar en él en cuanto ve que comete errores” (153).

Segunda: “a pesar de todas las pruebas a favor de los métodos de predicción mecánicos y algorítmicos, y a pesar del cálculo racional que muestra con claridad el valor de las mejoras incrementales en la exactitud de la predicción, muchos responsables de la toma de decisiones rechazarán los métodos de la toma de decisiones que les privan de la capacidad de ejercer su intuición” (164).

Tercera: “los modelos son sistemáticamente mejores que las personas, pero no 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” (160).

Cuarta: “los algoritmos no son, no serán pronto, un sustituto universal del juicio humano” (247).

Quinta: “un algoritmo puede ser más transparente que los seres humanos” y “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”; y “si los algoritmos cometen menos errores que los expertos humanos y, sin embargo, tenemos una preferencia intuitiva por las personas, nuestras preferencias intuitivas deberían ser examinadas cuidadosamente” (370 y 371).

- *Segunda valoración final*: aunque todavía estamos en una fase incipiente, es muy probable que el mundo del trabajo se verá profundamente afectado por esta tecnología (y también que lo hará aceleradamente). Desde este punto de vista, el contenido del art. 64.4.d ET, a pesar de su falta de concreción, tiene la virtud de sentar unas bases claras a partir de una formulación simple y con una afectación ambiciosa y profunda.

El carácter críptico de esta tecnología para el ciudadano medio hace de la transparencia uno de los mayores retos que la sociedad debe afrontar y, en particular, las relaciones laborales. Sin duda, nos jugamos mucho, pues, la información ininteligible es tan opaca como el mayor de los silencios. Los empresarios deben ser plenamente conscientes de los productos que contratan y/o del software que sea capaz de desarrollar y, especialmente, de la responsabilidad que puede derivarse del uso de los mismos. Y deberían aunarse esfuerzos para garantizar la comprensibilidad de estos procesos y cualquier interpretación que la obstaculizara podría ser muy contraproducente¹¹. Especialmente porque (SADIN, 77), a medida que estos instrumentos puedan perfeccionarse indefinidamente, se irán alejando del control de quienes los han concebido.

No creo que, como humanos, nos convenga que se vayan acercando a la idea de un “agente computacional autónomo” sin nuestra supervisión.

¹¹ Ni la propiedad industrial e intelectual ni tampoco el secreto empresarial deberían amparar la vulneración de la primera ley de la robótica de I. ASIMOV (“Un robot no puede dañar a un ser humano ni, por inacción, permitir que un ser humano sufra daño”).

4. Bibliografía

- Henning BECK (2019), *Errar es útil*, Ariel.
- Ignasi BELTRAN HEREDIA RUIZ (2021), “Despido económico preventivo y teorema de Bayes”, *IUSLabor*, núm. 1, pp. 66 – 98.
- BELTRAN HEREDIA RUIZ, I. (2021), “Riesgo laboral grave e inminente, juicio de probabilidad y paralización de la actividad empresarial”, *IUSLabor*, núm. 2, pp. 198 – 234.
- Margaret A. BODEN (2016), *Inteligencia artificial*, Turner Noema.
- James BRIDLE (2020), *La nueva edad oscura*, Debate.
- Nicholas CARR (2014), *Atrapados*, Taurus.
- Marcus DU SAUTOY (2020), *Programados para crear*, Acantilado.
- Hannah FRY (2019), *Hola Mundo*. Blackie Books.
- Michael S. GAZZANIGA, M. (1993), *El cerebro social*, Alianza.
- Daniel GILBERT (2006), *Tropezar con la felicidad*, Destino.
- Yuval N. HARARI (2016), *Homo deus*, Penguin Random House.
- 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.
- José I. LATORRE (2019), *Ética para máquinas*, Ariel.
- Viktor MAYER-SCHÖNBERGER y Kenneth CUKIER (2013), *Big Data*, Turner Noema.
- Viktor MAYER-SCHÖNBERGER y Thomas RAMGE (2019), *La reinención de la economía*, Turner Noema.
- Evgeny MOROZOV (2015), *La locura del solucionismo tecnológico*, Katz.
- Cathy O’NEIL (2017), *Armas de destrucción matemática*, Capitán Swing.
- Judea PEARL y Dana MACKENZIE (2020), *El libro del porqué*, Pasado y Presente.
- Francisco PÉREZ AMOROS (2021), “¿Quién vigila al algoritmo?: los derechos de información de los representantes de los repartidores en la empresa sobre los algoritmos de las plataformas de reparto”, *e-Revista Internacional de la Protección Social*, Vol. VI, núm. 1, pp. 173 – 187.
- Karl R. POPPER (1984), *La miseria del historicismo*, Alianza Taurus.
- Eric SADIN (2020), *La inteligencia artificial o el desafío del siglo*, Caja Negra.
- Nate SILVER (2014), *La señal y el ruido*, Península.
- Nassim N. TALEB (2011), *El cisne negro*, Paidós.
- Shoshana ZUBOFF (2020), *La era del capitalismo de la vigilancia*, Paidós.