SISTEMAS DE CONTROL DELEGADO (CoDe)

DELEGATED CONTROL SYSTEMS (DeCo)

Alfredo Marcos
Universidad de Valladolid
amarcos@fyl.uva.es; www.fyl.uva.es/~wfilosof/webMarcos/

1. Acción e interacción

La experiencia humana es interactiva[1]. Actuamos sobre los sistemas naturales y sociales. Dejamos en ellos nuestra huella. Por ejemplo, en forma de huella ecológica o de memoria social de nuestros actos. También actuamos sobre los sistemas técnicos. Para empezar, los producimos. Y este actuar siempre se vuelve sobre quien actúa. La interacción modifica los dos polos, no solo uno de ellos. Dejamos nuestra huella sobre las tecnologías digitales tanto como estas nos afectan a nosotros. Nuestra forma de tejer lo técnico nos esculpe también como seres humanos, para bien o para mal. Huella digital deja cada golpe de tecla, cada dato que confío a una red social, huella que queda impresa en el soporte tecnológico y lo va configurando, que queda impresa en el propio sujeto humano y también contribuye a conformarlo. Es así, hemos dicho, para bien o para mal. Lo cual nos empuja de bruces al tema de la neutralidad de lo técnico.

2. ¿Neutralidad de lo técnico?

Lo técnico resulta imprescindible para la vida humana. En este sentido no es neutral, es simplemente necesario y, por lo tanto, bueno para nosotros, como ha señalado Ignacio Quintanilla. Nuestra nutrición depende de las herramientas líticas que complementan nuestro sistema masticatorio. Nosotros hemos hecho y usado esas herramientas, hemos dejado sobre el sílex nuestra huella. Y los útiles la han dejado incluso sobre nuestra anatomía, que ha co-evolucionado con la técnica.

Por otro lado, una vez que tenemos un cuchillo, es cierto que podemos usarlo para trocear la comida o para asesinar al vecino, para bien o para mal. En este segundo sentido, la técnica es neutral, su bondad o maldad depende del uso que le demos. Aunque hay elementos técnicos que ya están orientados a ciertos usos. Por ejemplo, la bomba atómica es una tecnología orientada a la destrucción. Pensemos ahora en la clonación humana. No se me ocurre en qué sentido podría ser una tecnología neutral. Parece que ya desde su mismo concepto ataca a la dignidad de las personas. Pero si se entiende la clonación en un sentido más amplio, como una técnica aplicable a cualquier ser vivo, entonces podríamos suponer que tiene usos beneficiosos, por ejemplo en el terreno de la agricultura o de la ganadería o de la gestión ambiental, mientras que otros, señaladamente los que se ciernen sobre el ser humano, son claramente perjudiciales.

Por último, en el desarrollo tecnológico se producen continuamente elecciones, optamos por investigar y potenciar ciertas tecnologías en detrimento de otras alternativas. A veces se justifica la sumisión a la llamada inteligencia artificial (IA) argumentando que es un desarrollo tecnológico inevitable. Si, en efecto, la tecnología estuviera marcada por un designio inexorable, entonces, todo debate o reflexión sobre la misma estaría de más. Pero no es así, podemos, dentro de ciertos márgenes, elegir, y el esfuerzo que se pone en una dirección se escatima en

[1] John Dewey, *El arte como experiencia*, FCE, CdMx, 1949.

otra. En este sentido tampoco hay neutralidad. No es lo mismo investigar e invertir en tecnologías limpias que en tecnologías contaminantes, por ejemplo. Téngase en cuenta que cada tecnología por la que optamos cambia -en el mejor de los casos, amplía- nuestras capacidades, pero también nuestras necesidades. La telefonía móvil nos dota de nuevas capacidades, pero nos ata a nuevas necesidades, como por ejemplo a las que se refieren a la fabricación de baterías, a la obtención de los componentes de las mismas y a la gestión de los correspondientes desechos, por no hablar de las mil y una ataduras que nos imponen las compañías. Tampoco en este sentido hay neutralidad, cada servicio que la técnica nos ofrece lleva asociadas servidumbres, cada nueva capacidad, nuevas necesidades.

Hechas estas matizaciones respecto de los múltiples sentidos de la neutralidad (o falta de neutralidad) de la técnica, pasemos al caso que más directamente nos ocupa, el de la huella digital, que efectivamente puede resultarnos servicial, pero también puede someternos a servidumbre. Todo dependerá de la dirección en la que desarrollemos las tecnologías de la información y también del significado que les demos. En este caso es crucial la interpretación que hagamos de los sistemas digitales y, en especial, de los llamados sistemas de IA. Una mala interpretación de los mismos puede contribuir a devastar la esencia humana –por usar aquí una expresión de Heidegger[2]-, mientras que, bajo una correcta interpretación de su naturaleza y funciones, los sistemas llamados de IA pueden aportar ventajas a nuestra vida.

Aquí, la cuestión de la neutralidad (o falta de ella) se presenta en estos términos: una mala interpretación de los sistemas digitales conduce a un desarrollo desacertado de los mismos, devastador en el peor de los casos. Los sistemas producidos bajo una interpretación incorrecta de su ontología ya no resultan neutrales, sino que inevitablemente causan daño a la vida humana. Lo contrario sucede con los sistemas cuya concepción, diseño, desarrollo y aplicación vienen guiadas por una ontología ajustada y una antropología sensata. En otras palabras, nos conviene una actitud de desasimiento hacia lo digital, una actitud que Heidegger llamaba *Gelassenheit* y que nosotros solemos traducir por serenidad, es decir, una actitud que nos permita decir *sí* y *no* a lo técnico, a lo digital en este caso. *Sí*, cuando viene a mejorar la vida de las personas; *no*, cuando contribuye a devastar la esencia del ser humano. Veamos si podemos discernir cuándo es uno y cuándo lo otro.

3. Sistemas de IA. Una reflexión ontológica

Lo más elemental que podemos preguntarnos en clave ontológica es si la inteligencia artificial existe o no. Luc Julia, quien dirigió el equipo de desarrollo de SIRI en Apple, titula un libro reciente de manera provocadora: *L'intelligence artificielle n'existe pas*[3]. Por su parte, Elik J. Larson acaba de publicar un libro, en parecida línea, titulado *The Myth of Artificial Intelligence Why Computers Can't Think the Way We Do*[4]. En el avance del libro podemos leer: "Inductive AI

[2] Martin Heidegger, "Serenidad", *Revista Colombiana de Psicología*, 3: 22-28, p. 27, 1994, traducción de Antonio de Zubiaurre, disponible en: https://revistas.unal.edu.co/index.php/psicologia/article/view/15808/16639.

[3] Luc Julia, *L'intelligence artificielle n'existe pas*, Éditions First, París, 2019.

[4] Elik J. Larson, *The Myth of Artificial Intelligence Why Computers Can't Think the Way We Do*, Harvard University Press, 2021; avance disponible en: https://www.hup.harvard.edu/catalog.php?isbn=9780674983519&fbclid=IwAR1Ip4reXUrekQjdpCxpxiEm093dSssuplolS9d4SWQqLVZN5zGcJ92zaHA]. Véase también, en un sentido muy similar: Marie David y Cédric Sauviat, *Intelligence Artificielle. La nouvelle barbarie*, Éditions du Rocher, Mónaco, 2019 y Richard Benjamins e Idoia Salazar, *El mito del algoritmo*, Anaya, Madrid, 2020.

will continue to improve at narrow tasks, but if we want to make real progress, we will need to start by more fully appreciating the only true intelligence we know—our own".

Me sumo a este tipo de tesis, siempre que sean bien comprendidas. No es que no exista *algo* a lo que llamamos inteligencia artificial, es que la denominación es desorientadora, tal vez simplemente errónea. "El término inteligencia artificial surgió en los años cincuenta -afirma Katharina Zweig-, cuando los científicos querían recaudar dinero para sus investigaciones. Pensaron que sonaba a algo que el Estado fomentaría de buen grado. Y ahora pendemos de este nombre. La mayoría de los científicos informáticos lo encuentran inapropiado"[5]. Según Luc Julia, en 1956, durante la famosa conferencia de Dartmouth, John McCarthy convenció a sus colegas para llamar IA a algo que no tiene nada que ver con la inteligencia.

Yo matizaría: un sistema de IA sí tiene algo que ver con la inteligencia, pero lo que tiene de inteligente lo pone el ser humano, no es artificial, y lo que tiene de artificial no es en absoluto inteligente[6]. Habría que proponer, pues, una denominación mejor, que no condujese a confusión. *Machine learning* o *deep learning* son denominaciones igualmente confusas, especialmente si hemos de entender que es la propia máquina la que aprende. Todas estas denominaciones –señuelos, habría que decir- tienen una función comercial, publicitaria, incluso propagandística, pero no responden a la verdad de la cosa. Inmediatamente entran en resonancia con la ciencia ficción y con los titulares de los medios. Y así, ensoñaciones y terrores futuroscópicos comienzan a crecer como bola de nieve. Pero ninguna máquina entiende, ni conoce, ni aprende, ni cuenta hasta dos siquiera. Sí las personas, con ayuda, a veces, de las máquinas. Por ello se han propuesto términos como *inteligencia asistida*, *inteligencia ampliada* o bien *inteligencia artificial centrada en humanos* (*Human-Centered AI* es el nombre de un instituto de investigación recientemente creado en Stanford). Son nombres más apropiados, pues indican que el sujeto inteligente es una persona, mientras que la máquina puede asistir o ampliar la inteligencia de dicho sujeto. También podríamos hablar, y creo que sería lo más ajustado, de *sistemas de control delegado* (CoDe). Y si hablamos de "sistemas", antes de dirimir la cuestión del nombre tenemos que profundizar un poco en la configuración sistémica de lo técnico.

Según Miguel Ángel Quintanilla[7], la tecnología es un sistema de acciones intencionales. Aquí la clave está en el enfoque sistémico. El cambio de perspectiva que necesitamos podría resumirse en pocas palabras: hay que pasar de considerar los sistemas de IA como "sistemas técnicos con consecuencias sociales" a considerarlos como "sistemas sociales técnicamente implementados"[8].

Un sistema de IA incluye muchos elementos. También un sistema de lavado: lavadora, detergente, electricidad, toma de agua y desagüe, ropa… y un usuario con ciertos conocimientos e intenciones. Fuera de este sistema, una lavadora deja de ser una lavadora para convertirse en una simple masa metálica de unos setenta kilogramos. Un sistema de IA exige también todo un entramado de nexos entre el sistema eléctrico, las computadoras y su software, así como una

[5] Anna Von Hopffgarten entrevista a Katharina Zweig (directora del Laboratorio de Responsabilidad Algorítmica de la Universidad de Kaiserslautern), "La inteligencia artificial carece de la flexibilidad de decisión humana", *Mente y Cerebro*, 106: 66-69, enero/febrero, 2021.
[6] Alfredo Marcos, "Información e Inteligencia Artificial", *Apeiron. Estudios de Filosofía*, 12: 73-82, 2020.
[7] Miguel Ángel Quintanilla, *Tecnología: Un enfoque filosófico y otros ensayos de filosofía de la tecnología*, CdMx, FCE, 2005.
[8] R. Hirschheim, H.K. Klein y K. Lyytinen, *Information Systems Development and Data Modelling. Conceptual and Philosophical Foundations*, Cambridge University Press, Cambridge, 1995, p. 1.

red de carácter social urdida por personas que diseñan, mantienen y utilizan el sistema de IA, que configuran un marco legal e institucional en el que dicho sistema se asienta y cobra sentido… De nuevo: una computadora, con sus algoritmos y datos a cuestas, pero aislada de las personas, no es ya una computadora, pasa a ser una masa de equis kilogramos dentro de la cual se producen cambios electromagnéticos. Aislada de una fuente de energía no es ni siquiera esto, pues las actuales máquinas de IA consumen ingentes cantidades de energía eléctrica. O sea, las personas también forman parte de los sistemas de IA, como diseñadores, mantenedores, reguladores, usuarios, supervisores… Es en estas personas, y no en la parte artificial, en las que reside la inteligencia de un sistema de IA. Las máquinas no pueden ser inteligentes. Esta limitación no responde a un problema técnico que pueda ser subsanado, sino a una cuestión ontológica de base.

Pensemos por un momento qué significa inteligencia. El significado de este término no podemos estipularlo sin más arbitrariamente, sino que tenemos que tomarlo de su uso común. Partamos, pues, de la etimología. *Intelligere*, en latín consta del prefijo *intus* (dentro) o bien *inter* (entre), además de la palabra *legere*, relacionada con el verbo griego *λέγειν*. Este último significa hablar, decir, relatar. De la misma raíz viene *λόγος*. En griego arcaico, como por ejemplo en Homero, *λέγειν* significa seleccionar, recolectar o enumerar. Procede probablemente de la raíz indoeuropea, *leg-* (elegir, diferenciar). Ya tenemos algunas ideas sobre la palabra española "inteligencia". Dicho término sugiere significados como leer-dentro, ir a la esencia y a las causas de las cosas, separar lo distinto, elegir, ligar o unir lo semejante. Parece que en el fondo pragmático de la palabra hay una metáfora de tipo agrícola, que se proyecta sobre el ámbito del conocimiento. Separamos el trigo de la cizaña y agavillamos el trigo, separamos lo diferente y después reunimos lo semejante. También en el plano epistémico separamos y reunimos. Reunimos lo semejante en el concepto y, en última instancia, en la unidad de la conciencia.

Nuestra segunda fuente será el diccionario. Podemos acudir a las acepciones pertinentes del *DRAE*: inteligencia es "capacidad de entender o comprender […] capacidad de resolver problemas". Sabemos que la parte artificial de un sistema IA es incapaz por sí sola de entender o de comprender. Ni siquiera se puede decir con propiedad que una máquina cuente o compute. Contar implica reunir dos (o más) momentos, y mantenerlos como tales, sin confundirlos en uno, en una sola e idéntica representación consciente[9], entendiendo al mismo tiempo la semejanza y la diferencia entre ambos, cosa que una máquina no hace. Sí es cierto, en cambio, que la IA puede ayudarnos a resolver múltiples problemas (cómputo, geolocalización, logística, asistencia telefónica, asistencia al diagnóstico médico, a la publicidad y un largo etcétera). Pero estos problemas no lo son para la parte artificial del sistema, sino para el diseñador o para el usuario del mismo. Para una máquina de reconocimiento facial, el reconocer o no a un delincuente no supone un problema. Es un problema para la seguridad de las personas, y el sistema puede ayudarnos a afrontarlo. Por supuesto, el mismo sistema puede servir para controlar a la población de un país y para facilitar allí la represión política. Pero esto tampoco es un problema para las cámaras o para el software implicado. Lo es, indudablemente, para los sufridos súbditos del país en cuestión. Ni los martillos, ni los ábacos, ni las computadoras más avanzadas tienen problemas. Los problemas los tenemos nosotros, como seres humanos. Solo un ser vivo puede morir o sufrir, solo una persona puede preguntarse por el sentido de su vida. Esos son

[9] Trato este asunto por extenso en A. Marcos, "La relación de semejanza como principio de inteligibilidad de la naturaleza", en F. Rodríguez Valls (ed.), *La inteligencia en la naturaleza*, Biblioteca Nueva, Madrid, 2012, pp. 73-94.

problemas. Y tanto un martillo como una red informática, cada uno a su modo, pueden ayudarnos a afrontarlos (o a empeorarlos). Pero esto no los convierte en inteligentes.

Podemos verlo también si atacamos la cuestión desde otro ángulo. A veces se caracteriza la llamada IA por su capacidad de simulación[10]. Apelando al tema clásico de *The Platters*, podríamos decir que la IA es *The Great Pretender*. Simula funciones propias de la inteligencia humana, se dice. Pero simular no es ser. Simular la inteligencia no es lo mismo que ser inteligente. La simulación, además, consta solo como tal para el ser humano que la observa, no para la máquina. Por otro lado, la propia noción de función remite inexorablemente a la de un ser para la cual un efecto dado es funcional. Aquí, los sistemas artificiales dependen también de la funcionalidad que puedan tener para el ser humano. Fuera del marco humano, las luces que se encienden y apagan en una pantalla o los movimientos de un robot son meros efectos, no cumplen funciones.

Es muy ilustrativo al respecto el relato de Miguel de Unamuno titulado *Mecanópolis*[11]. En esta distopía, el protagonista alcanza una ciudad perfectamente mecanizada, pero carente por completo de habitantes. Con ello, el autor introduce en la ciudad mecanizada un punto de vista, el del protagonista humano, lo cual le permite describir la urbe en términos inteligibles, como algo más que un amasijo de materiales en movimiento. El día en que el protagonista aparece por la cuidad, los movimientos de las máquinas, simples efectos hasta entonces, comienzan a ser funcionales. El punto de vista humano cambia incluso su ontología: un pedazo de metal que gira sobre otro, por ejemplo, pasa a ser la rueda de un tranvía.

Porque la cuestión es en el fondo –insistamos- de carácter ontológico. Los artefactos, en la tradición aristotélica, son sustancias solo en un sentido accidental. Los seres vivos, y en especial los seres humanos, es decir, las personas, sí son sustancias en sentido propio y paradigmático. Al tratarse de una diferencia ontológica, la esperanza (o la amenaza) de anularla por la vía de la sofisticación tecnológica resulta ilusoria, un mero error categorial[12].

Lo dicho hasta aquí afecta a cualquier sistema tecnológico (de lavado, de transporte, de producción de energía, de comunicación…). Todos ellos, si se colocan al margen de lo humano pierden la funcionalidad, pasan a ser meros sistemas de efectos físicos. Como quiera que su ontología viene dada por su función, pierden también su rango ontológico, dejan de ser lo que eran. Por seguir con el ejemplo: una lavadora en Neptuno no es ya una lavadora. Pero la ontología de los sistemas llamados de IA depende aun más intensamente de la mirada humana, pues se sitúan en el ámbito de lo intencional, es decir, de lo semiótico. En este ámbito las

[10] Véase Stuart Russell and Peter Norvig, *Artificial Intelligence. A Modern Approach* (4th ed.), Pearson, Boston, 2021. El libro está estructurado según las cuatro combinaciones posibles de dos pares de conceptos: *thought*-*behaviour* y *rational*-*human*. Salen, así, cuatro tipos de IA. *i*) La IA que trata de pensar (*thought*) según una teoría general de la racionalidad (*rational*), *ii*) la que trata de pensar (*thought*) imitando al ser humano (*human*), *iii*) la que busca un comportamiento inteligente (*behaviour*) según una teoría general de la razón práctica y *iv*) la que busca un comportamiento inteligente (*behaviour*) a imitación de la praxis humana (*human*). En dos de estas variantes se entiende la IA como simulación de funciones humanas. Por otro lado, es fácil proyectar el primer par sobre las acepciones de diccionario: *thought*-entender, *behaviour*-resolver problemas.

[11] Este relato ha sido recientemente comentado desde un punto de vista filosófico por Alicia Villar y Mario Ramos, "Mecanópolis: una distopía de Miguel de Unamuno", *Pensamiento*, 75: 321-343, 2019.

[12] Es interesante al respecto el libro de David J. Gunkel, *Robot Rights*, MIT Press, 2018. En su último capítulo cuestiona precisamente la relación de precedencia de la ontología sobre la ética. Apela, para ello, a Emmanuel Lévinas. Lo cierto es que Lévinas habla de la constitución del yo a partir de la mirada del otro. Se sobreentiende que el otro es una persona, un ser humano, y no una máquina.

entidades se sostienen sobre tres apoyos. Si retiramos uno de ellos se vienen abajo, como les pasa a los taburetes, si se me permite la imagen. Charles S. Peirce lo expone con claridad: "All dynamical action, or action of brute force […] takes place between two subjects […] But by semiosis I mean, on the contrary, an action or influence which is or involves a cooperation of three subjects, such as a sign, its object and its interpretant, this three-relative influence not being in any way resolvable into actions between pairs[13].

¿Qué queremos decir cuando afirmamos que una máquina almacena o procesa mis datos financieros o médicos? Decimos que cierto estado electromagnético de la máquina (*sign*) se relaciona con mi nómina o con mi tensión arterial (*object*). Obviamente no hay entre ellos una relación física, sino una relación semiótica que se establece a través de una persona (*interpretant*) capaz de entender –con la ayuda de ciertos interfaces- estados electromagnéticos como ingresos o como presión sanguínea. Análogamente, la máquina solo juega al ajedrez o al *go* si una persona puede relacionar los estados físicos de la misma con estos juegos tradicionalmente ejecutados por humanos. El caso del ajedrez es muy ilustrativo: cuando una máquina cumple por fin ciertas expectativas, queda al mismo tiempo privada del aura mitológica que la envolvía, rebajada al nivel de lo prosaico, desprovista de fantasma y de glamour. En 1997, como es sabido, una máquina de IBM, *Deep Blue* –de nuevo, un nombre pensado para el estrellato de los medios-, doblegaba al campeón del mundo de ajedrez, Gary Kasparov. Fue toda una sensación… que inmediatamente decayó en trivialidad. Hoy los ajedrecistas emplean rutinariamente máquinas para entrenarse, como módulos de análisis que les ayudan a saber si una posición es o no *jugable*. Y todos nos hemos enterado, por fin, de que un robot ajedrecista es aproximadamente tan interesante como un robot aspirador. Sin *interpretant*, la máquina solo cambia de un estado físico a otro. Ya no es parte de un sistema de IA. Es solo un trozo de materia, como una lavadora en Neptuno.

Solemos imaginar, en cambio, que en nuestra ausencia las cosas siguen teniendo la misma entidad que en nuestra presencia. Así, imaginamos que una máquina que forma parte de un sistema de IA, junto con ciertas personas, sigue siendo inteligente, sigue constituyendo un sistema de IA, aunque prescindamos de la mirada, de la presencia, de esas personas. Pero erramos. Y no por un exceso de imaginación, sino por una falta de imaginación. No es fácil imaginar cómo se ve el mundo cuando el mundo no es visto. Antonio Machado escribió en su día estos versos: "Dijo Dios: Brote la nada. / Y alzó la mano derecha, / hasta ocultar su mirada. / Y quedó la nada hecha" [14]. O bien, reinterpretando a Cervantes: "Fuese, y no hubo nada"[15]. La mirada de Dios sostiene al mundo en el ser. La del ser humano sostiene el ser de lo artificial y de lo semiótico. Sin la mirada de Dios todo es aniquilado. Sin la mirada de una persona, lo semiótico no pasa a la nada, pues nuestro poder no es tanto, pero sí se aplana, queda en pura realidad física. De ahí la dificultad de imaginar. Es más fácil soñar que todo sigue igual cuando dejo de mirar. El niño cierra la puerta del frigorífico e imagina su interior iluminado. No por una imaginación desbocada, sino por incapacidad para imaginar la oscuridad que nunca ve, pues cuando lo vuelve a abrir "ahí sigue" la luz. Es lo que podríamos llamar *efecto Toy Story*. La presencia, la mano y el ojo del niño convierten un pedazo de plástico verde en un dinosaurio tímido. El niño imagina que cuando él sale de la habitación ahí sigue el dinosaurio. No es capaz de imaginarlo como el pedazo de plástico verde que es una vez que el niño deja el cuarto de

[13] Charles S. Peirce, *Collected Papers*, Harvard University Press, Cambridge (MA), vol. 5. p. 484, 1931-1935.
[14] Antonio Machado, *Juan de Mairena I, recopilación póstuma de textos del apócrifo de Antonio Machado*, Losada, Buenos Aires, p. 139. 1943.
[15] Miguel de Cervantes, Soneto *Al túmulo del rey Felipe II en Sevilla*, 1598.

juegos o se duerme. No puede imaginarlo porque cuando despierta, el dinosaurio "todavía está" ahí[16]. De hecho, la trama de *Toy Story* (USA, 1995) no ocurre en un cuarto de juegos, sino en la (pobre) imaginación de Andy.

¿Qué imaginamos que sucedería si el ser humano saliese de la habitación, si quedase al margen de los sistemas de IA? Para algunos, esto ocurrirá a partir del punto que llaman *singularidad*[17]. Desde ahí, las máquinas generarían otras máquinas mejores y más listas. Un mundo post-humano controlado por robots dotados de IA se abriría paso. Pero quizá podríamos imaginar, por el contrario, que las máquinas dejadas a sí mismas pronto producirían fallos de funcionamiento en virtud de la tendencia general a la entropía, de los defectos de diseño y construcción, así como de la dificultad para obtener energía estable; decaerían y se verían incorporadas a los procesos naturales de tipo físico, como la erosión, químico, como la oxidación o biológico. El paisaje post-humano más probable no es el de la Tierra gobernada por robots inteligentes, como tantas veces hemos visto en el cine, sino el de una frondosa jungla que esconde en sus entrañas, junto con las piedras de antiguos templos, auténticas cochambres de silicio, plástico y metal. De hecho, toda máquina, y más aun las más sofisticadas, ha de ser mantenida. O sea, "man-tenida", es decir, llevada de la mano por las personas. No hay nadie más popular y buscado en una corporación que el personal de mantenimiento informático. De hecho, el personal de mantenimiento informático suele hacer turnos de guardia para que las máquinas, inteligentes y autónomas, no se queden ni un instante sin alguien que les eche un mano. Todo sistema de IA requiere mantenimiento. Y los más sofisticados requieren más mantenimiento, no menos.

En suma, los datos son datos acerca de algo, la inteligencia lo es de algo, la información lo es sobre algo. Son entidades de la *terceridad*, semióticas, intencionales. El estado electromagnético (o cuántico) de un computador no es de por sí un dato, a menos que un interpretante lo conecte con un objeto. Además, no hay dato aislado, que no esté integrado en un espacio de posibilidades; la tinta sobre el papel, por ejemplo, tiene una cierta forma, pero podría tener otra dentro de un cierto espacio de posibilidades que el lector conoce. La dependencia de los datos respecto de una conciencia que los convierta en tales resulta extrema en el caso de los llamados datos sintéticos, cuya relación con la verdad es difícil de establecer. Se trata de conjuntos de datos (¿?) no tomados de la realidad, sino generados mediante un modelo que reproduce las propiedades estadísticas del conjunto de datos reales. Se usan para anonimizar información médica de pacientes o de clientes del sector financiero, para entrenar algoritmos a menor coste que el que supone adquirir datos reales; también se pueden obtener datos sintéticos para diseñar coches autónomos mediante pruebas previas en carreteras virtuales.

Cuando la mirada humana, cuando la mano de la persona, sale de escena, Mecanópolis no es ya una ciudad de máquinas, sino un montón de materia metálica en movimiento. Y un sistema de IA deja inmediatamente de ser inteligente. Ya no hay datos. Ya no entiende nada. Ya no simula nada. Ya no cumple función alguna. Sus problemas se han acabado al fin. Lo que llamábamos información se diluye. Una decisión deja de serlo. Añadamos que no hay realidad virtual, sino representación digital de la realidad, de la única realidad que existe, dentro de la cual lo virtual tiene cabida como representación, pero decae cuando no hay nadie ante quien representar algo.

[16] Parafraseando el famoso micro-relato de Augusto Monterroso: "Cuando despertó, el dinosaurio todavía estaba allí" (*Obras completas y otros cuentos*, 1990).
[17] Ray Kurzweil, *La singularidad está cerca*, Lola Books, Berlín, 2019.

En este sentido -coincido con Luc Julia-, la IA no existe. Es el mero efecto de una pobre imaginación metafísica, puesta al servicio a veces de la ambición pecuniaria o política.

Existen, eso sí, sistemas de IA de los cuales forman parte también las personas y que, a fin de evitar equívocos, deberían recibir un nombre menos confuso. Entre otras cosas porque la perturbación que genera un mal nombre acaba por proyectarse sobre la propia antropología, sobre la idea que el ser humano tiene de sí mismo. Así, la inteligencia humana es degradada o reducida a un juego de interacciones físicas, y el propio ser humano pasa a ser comprendido en términos dualistas, como una suerte de lugar de encuentro fortuito entre un hardware corporal y un software mental susceptible de migración[18]. Además, al negarle inteligencia a los artefactos, estamos protegiendo la valoración social de la propia inteligencia, pues, cuando se atribuye inteligencia a una cosa, se activa en nosotros "un mecanismo de autoprotección consistente en devaluar la deseabilidad de […] la inteligencia"[19].

4. Sistemas de IA. Una reflexión epistemológica

El recorrido ontológico nos ha permitido establecer que la IA, como tal, no existe, y que los llamados sistemas de IA son más bien sistemas de ayuda a la inteligencia humana[20]. ¿Pero qué tipo de ayuda pueden prestar a nuestra inteligencia? Pisamos ya el terreno de la epistemología y vemos que los algoritmos de la llamada IA son buenos para detectar líneas de correlación entre enormes cantidades de datos. Pero las correlaciones no son todavía relaciones causales, no nos permiten entender el fenómeno al que nos enfrentamos ni dar explicación del mismo. Para hacerlo necesitaríamos teorías que conjeturen relaciones causales. Es cierto que la detección de correlaciones en bruto nos pone ya en un cierto camino, nos sugiere hipótesis, nos sirve como herramienta heurística. Y, en este sentido, puede ser de gran ayuda para llegar a entender una determinada parte de la realidad, siempre que no se pretenda una completa automatización de la búsqueda científica.

Téngase en cuenta que existe ya una cierta tendencia a la deshumanización de la investigación científica. Las prácticas tecnocientíficas se realizan en un contexto cada vez más automatizado. Los propios objetivos de la tecnociencia parecen haber girado, desde la intelección del mundo, hacia la obtención y procesamiento de grandes cantidades de datos (*big data*). Las ciencias ómicas (*omics sciences*) constituyen un claro ejemplo reciente de esta tendencia, si bien los antecedentes de la misma pueden rastrearse muy atrás: la idea de *ars magna*, propuesta por Ramón Llull, la *característica universalis*, de Leibniz, o la *construcción lógica del mundo* defendida por Carnap se encaminaban ya hacia la mecanización del conocimiento, entendido este como una combinatoria automatizable. Se podrían añadir más ítems a esta lista, como por ejemplo el caso histórico de la química relacional del siglo XVIII[21]. Se intentó entonces la construcción de exhaustivas tablas de relaciones; se trataba de registrar *todas* las posibles relaciones entre sustancias. La química tradicional, en cambio, podía conformarse con el

[18] Marie David y Cédric Sauviat, *Intelligence Artificielle. La nouvelle barbarie*, Éditions du Rocher, Mónaco, 2019, pp. 203-204.

[19] Manuel Carabantes, "Inteligencia artificial lingüística perfecta: efectos sobre la autopercepción del ser humano", *Scio*, 18: 207-234, 2020, p. 218.

[20] Queda abierta una cuestión ontológica: ¿qué implicaciones tiene el reemplazo de una ontología de sustancias por una de sistemas?

[21] Isabelle Stengers, "La afinidad ambigua: el sueño newtoniano de la química del siglo XVIII", en Michel Serres, *Historia de las ciencias*, Cátedra, Madrid, 1991, pp. 337-362.

conocimiento de las reacciones consideradas más relevantes, es decir, más desveladoras de las propiedades de cada sustancia, sin necesidad de emprender una compilación exhaustiva. Cito este caso histórico por las lecciones que podemos extraer de él para nuestra situación actual. En el momento en que se prescinde de una estimación prudencial –humana, por tanto- de la relevancia y del sentido, todo dato reviste la misma importancia que cualquier otro y se ha de emprender una búsqueda exhaustiva, combinatoria y automática.

Pensemos en lo que está sucediendo hoy en las ciencias llamadas *ómicas* (genómica, epigenómica, proteinómica, metabolómica, estudios del conectoma cerebral…, y así una pléyade de ellas, hasta la exposómica), que han nacido como secuela del PGH. El sufijo *-oma* procede del griego y sugiere la noción de totalidad. Las ciencias que lo adoptan tratan, con un fuerte apoyo computacional, de listar la totalidad de los elementos de un dominio dado (todos los genes, todas las proteínas, todas las rutas metabólicas o de conexión neuronal…, todos los factores ambientales a los que estamos expuestos). Es significativo que el sufijo más tradicional para las ciencias venga del término griego *logos* (biología, geología…), que hace referencia a un saber inteligente, y no meramente enciclopédico. Nada hay que objetar a las ciencias *ómicas*, que pueden resultar muy valiosas, pero sí al intento de reducir toda ciencia a un ejercicio de secuenciación y búsqueda de correlaciones. Esta forma de entender la tecnociencia desplazaría a los márgenes de la ciencia lo propiamente humano, lo que va más allá de un mero registro y combinación de elementos, es decir, lo que tiene que ver con la genuina creatividad, con la esfera emocional, con las intuiciones y la experiencia vivida, con el sentido, la relevancia y los valores, incluidos los de carácter moral y estético, con la reflexión y con la conversación. La recolección de datos y búsqueda de correlaciones constituye un paso importante para el avance de la ciencia, claro está, pero una ciencia reducida a la mera acumulación y combinatoria de datos supondría un esfuerzo tan caro como estéril para nuestra intelección del mundo. A la hora de enfrentarnos a fenómeno complejos, tenemos que echar mano, sí, de la fuerza bruta de computación, pero también de toda la imaginación, creatividad, intuición y prudencia de que sea capaz el espíritu humano. De hecho, el más abarcador "método" científico, el que regula la aplicación de todos los demás, incluidos los automatizables, es la prudencia[22].

Recordemos ahora que además de la inteligencia como comprensión, está la inteligencia para afrontar problemas. Aquí también son de gran ayuda los algoritmos, capaces de manejar grandes cantidades de datos. Gracias al repunte de la investigación en IA en la última década se han logrado sistemas muy útiles y precisos en diversos ámbitos. Siempre se trata de sistemas especializados en una cierta función. Han mejorado, indudablemente, los sistemas de traducción automática, los de reconocimiento facial y por voz, los que simulan visión, los sistemas especializados en medicina, publicidad o finanzas, los sistemas logísticos, de conducción automática, de pilotaje o de combate, los robots que simulan habilidades conversacionales (*chatbots*) o buscan la persuasión, los módulos de análisis para el ajedrez o el go… y así sucesivamente. Cada uno de ellos, a su modo, podría ayudarnos a resolver problemas.

En muchos casos se trata de ceder el control de algunas acciones a la parte artificial de un sistema de IA. Dicho sistema funciona gracias al histórico de datos que se le suministra. En realidad hay aquí una suerte de inferencia inductiva, tocada, como no podía ser de otra forma, por la debilidad propia de las inferencias ampliativas. El sistema no es estático, es capaz de

[22] En esta línea, podemos leer: "Código de conducta articulado en torno a seis puntos: 1. Prudencia…" *Declaración de Barcelona para el desarrollo y uso adecuado de la IA en Europa*, 2017.

asumir modificaciones (lo que algunos llaman aprendizaje), pero siempre a partir de una ciertas expectativas determinadas por el histórico de datos.

Los algoritmos parten de un histórico de datos, crean a partir del mismo un espacio n-dimensional y lanzan, dentro de este espacio, numerosísimos cortes de n-1 dimensiones. Buscan, entonces, la posible correlación de cada uno de dichos cortes con los valores de la dimensión restante, que es la que queremos controlar. Generan, así, un sistema de expectativas. Cuando disponemos de muchos datos, por ejemplo, sobre el solicitante de un crédito, podemos establecer con cierta probabilidad si el crédito resultaría fallido o no. Contando con dicha información, podemos decidir entonces la concesión o denegación del crédito. Los datos de que disponemos sobre el solicitante configuran una sección de n-1 dimensiones en el espacio n-dimensional, sección que estará, en el mejor de los casos, fuertemente correlacionada con la dimensión que nos interesa controlar, a saber, la probabilidad de *default*.

Hasta donde se me alcanza, el conocimiento que este tipo de algoritmos nos aporta está tocado por las limitaciones clásicamente asociadas a la inferencia inductiva en lo que a la relación con la verdad se refiere[23]. Las premisas de una inferencia inductiva pueden ser todas ellas verdaderas y aun así puede ser falsa la conclusión. De modo análogo, pueden ser correctos todos los datos del histórico con que nutrimos un algoritmo y aun así puede fallar el sistema de expectativas que genera. Dicho de otro modo, no podemos esperar de nuestros algoritmos predicciones seguras de futuro, sino modelos condicionales falibles[24]. Los sistemas de IA no hablan en futuro, sino en condicional: "comprobados todos los datos de que disponemos respecto del solicitante, la probabilidad de default *sería X*, si se cumpliese *Y*". Ahora bien, *Y* puede englobar numerosísimas condiciones de contexto, pero en esencia podríamos resumir su contenido con las cláusula latinas *rebus sic stantibus* o *ceteris paribus*. O sea, la probabilidad sería *X* si todo siguiese siendo, más o menos, como viene siendo. La inferencia inductiva se puede convertir en una inferencia que transmita la verdad de modo seguro si incluimos una premisa que indique que todo seguirá siendo, más o menos, como viene siendo. Pero esta premisa no puede ser obtenida inductivamente. Newton ya se percató de este problema y lo abordó en sus "Reglas para filosofar". Antes de recomendar la inducción como método científico, en su cuarta regla, se asegura de haber introducido otra, la segunda, que recomienda investigar la naturaleza como si fuese regular[25]. Los algoritmos serían seguros si no hubiese novedades importantes en el universo, si el futuro no estuviese abierto[26], si todo sucediese con exquisita regularidad.

Pero, por lo que sabemos, el universo no es una suerte de reloj eterno, sino un acontecimiento único, histórico. Está dotado este hecho único de un cierto entramado de regularidades no estrictas, suficiente para posibilitar la vida y la intelección, pero compensado con novedades impredecibles. Es así en cualquier escala (y hoy es menos necesario ejemplificar al respecto de lo que lo era antes de la pandemia de COVID-19). Esta distribución tan peculiar de la constancia y la ruptura afecta tanto al discurrir de los astros (recordemos los problemas históricos que la humanidad ha tenido y tiene para establecer un calendario), como a nuestra vida cotidiana, hecha de ciclos imperfectos, de ritmos circadianos, de costumbre y sobresalto. Solo una inteligencia viva, sentiente, situada, una inteligencia prudencial como la humana puede entenderse con esta desconcertante textura del universo. Aprendemos de la experiencia, pero

[23] David Hume, *Tratado de la naturaleza humana*, Tecnos, Madrid, 1980, sección IV.
[24] Marie David y Cédric Sauviat, *Intelligence Artificielle. La nouvelle barbarie*, Éditions du Rocher, Mónaco, 2019, pp. 182-188.
[25] Isaac Newton, *Principios matemáticos de la filosofía natural*, Editora Nacional, Madrid, 1982.
[26] Karl R. Popper y Konrad Lorenz, *El porvenir está abierto*, Tusquets, Barcelona, 1992.

sabemos a un tiempo que no hay ninguna garantía de que las cosas sigan siendo como venían siendo. Es más, según Hans Jonas: "Nosotros sabemos —y tal vez es lo único que sabemos— que la mayoría de las cosas serán distintas […], que hemos de contar siempre con la novedad, pero que no sabemos calcularla"[27]. De ahí la conveniencia de la humildad intelectual, que se ha vestido a lo largo del tiempo de actitud socrática, de prudencia aristotélica, de docta ignorancia, de falibilismo actual.

Un sistema de IA genera unas expectativas. Coloca un punto en un espacio n-dimensional construido a partir de un histórico de datos, y, en función de ello, nos dice qué se puede esperar del objeto representado por ese punto. Pero el sistema puede colapsar cuando registra la ocurrencia de algo cuya posibilidad ni siquiera había sido considerada[28]. Cuando esto ocurre, el propio sistema se queda sin capacidad de adaptación, no puede aprender de esta experiencia. Un sistema de IA orientado a la concesión de créditos –por seguir con el ejemplo- puede asumir que alguno de los préstamos concedidos resulten fallidos, y puede integrar estos nuevos datos en el histórico, aprender de ello y reorganizar su geometría. Lo que no puede asumir es el colapso repentino, el fallo abrupto, de todos los créditos en vigor, aun de los tenidos por más solventes. Si esto sucede, no son los algoritmos los que deben reaccionar, sino los responsables del banco en cuestión. Y reaccionarán, en primer lugar, cambiando drásticamente las expectativas. Lo pueden hacer dado que no son artefactos, sino personas conscientes que pueden llegar a entender el fenómeno, cosa que no se espera de la máquina, y que pueden poner en marcha su creatividad para generar en adelante mejores expectativas con o sin ayuda mecánica. Pensemos que el colapso de todos los créditos podría haberse debido a un fenómeno sísmico o climático o astronómico, pero también a una moda cultural, a un movimiento político o a una pandemia… Una persona puede llegar a entender lo que sucedió. Es la persona, el ser humano que dispone de una inteligencia general, quien puede reemplazar el sistema colapsado por otro. Solo una persona puede conectar y reconectar intencionalmente el plano lógico (digital) con el físico (analógico).

Cuando nuestro sistema de expectativas colapsa, podemos sobrevivir gracias a que podemos pasar a otro. Y este paso no tiene por qué ser puramente arbitrario, azaroso o irracional, sino que, en algún sentido, está guidado por un saber práctico y social que Aristóteles llamó *phrónesis*, prudencia. Dicho saber nos facilita la constitución integradora de la experiencia, la gestión de las emociones vinculadas a la frustración de expectativas, la propedéutica del momento creativo y el filtrado crítico de los sistemas de expectativas emergentes

En términos lógicos, podemos construir expectativas por inducción, generalizando correlaciones pasadas. Esto lo hacen muy bien las personas y el resto de los seres vivos, incluso a partir de muy pocos casos. También los sistemas recientes de IA son excelentes en este tipo de tareas, aunque requieren masivos aportes de datos históricos. Pero cuando fallan las expectativas, tenemos que intentar entender lo que está pasando. Lo hacemos creando hipótesis que, de ser

[27] Hans Jonas, *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*, Herder, Barcelona, 1995, p. 200.

[28] Hay un problema insoslayable que los especialistas llaman de "problema de cola larga", es el que hace que siempre queden fuera de consideración un número indefinido de situaciones posibles: "a menudo hay una gran 'cola' de situaciones […] que tienen una probabilidad muy pequeña de ocurrir. Eso provoca que tales situaciones no aparezcan casi nunca en los datos de entrenamiento, por lo que un sistema de aprendizaje supervisado errará estrepitosamente ante ellas". Ramón López de Mántaras (fundador del Instituto de Investigación en IA del CSIC), "El nuevo traje de la IA", *Investigación y Ciencia*, 526: 50-59, 2020, p. 54.

correctas, explicarían la situación. Este tipo de inferencia se denomina abducción[29]. La abducción es creativa, nos obliga a salir de los sistemas de expectativas dados. Y aquí las personas son imprescindibles, con su peculiar ontología, con su conciencia, intencionalidad e interacción con el mundo. Fue Charles S. Peirce quien primero estudió a fondo este tipo de inferencias imprescindibles para enfrentarnos a lo inesperado.

En consecuencia, un sistema de IA que pretenda sustituir a la prudencia humana simplemente estaría fuera de lugar, fuera del universo que nos alberga. Por el contrario, un sistema de IA inscrito en el marco de la inteligencia prudencial humana estará en el lugar que le corresponde y podrá cumplir funciones de gran valor al servicio de la vida humana.

5. Sistema de IA. Cuestiones prácticas (ético-políticas)

Podemos volver ahora a la cuestión del nombre, de la cual partíamos. Hemos visto que no deberíamos descartarla como una cuestión *meramente* nominal. Si se hubiera inventado un nombre arbitrario, sin previa carga semántica, no tendríamos nada que decir al respecto. Pero la noción de inteligencia, que se pretende desplazar al ámbito de lo artificial, acarrea ya consigo una fuerte carga semántica previa. Esta carga semántica no dice verdad respecto de la realidad de la cosa, pero resulta útil para otros usos. Es decir, el nombre, aquí, ejerce una función deliberadamente propagandística que nos conduce a una mala interpretación de la realidad, tanto de la realidad de los algoritmos como de la realidad humana. En mi opinión, a los llamados sistemas de IA sería mejor denominarlos sistemas de control delegado (CoDe). Esta denominación hace más justicia a su verdadera ontología y, sobre todo, nos conduce de modo más directo a las cuestiones prácticas realmente importantes.

Empezamos a ver cuál ha de ser la posición relativa del ser humano y de las máquinas llamadas de IA. No se trata de deformar al primero para que encaje en un mundo presuntamente dominado por inteligencias mecánicas, como en una especie de lecho de Procusto, sino de poner las segundas en el marco de la vida humana, fuera del cual dejan de funcionar (porque pierden sus funciones y porque no hay quien las mantenga). Dejan de ser. ¿Qué pueden aportar a la vida humana los algoritmos llamados de IA? Nos permiten delegar ciertos procesos de control. Solo las personas pueden tomar decisiones. El mismo concepto de decisión es ajeno a lo mecánico. Lo que llamamos "decisión" en un sistema de IA lo será solo en la medida en que un ser humano haya tomado la genuina decisión de delegar alguna acción en el sistema, de ponerla bajo su control, es decir, de automatizarla. La responsabilidad última, salgan bien o mal las cosas, solo puede ser de un ser humano. Y es responsabilidad nuestra, por cierto, apoyarnos para tomar decisiones en los mejores sistemas CoDe disponibles.

Ahora podemos plantear con más propiedad las cuestiones prácticas pertinentes, de tipo ético, político, educativo, legal… ¿Quién delega?, ¿está legitimado para hacerlo? ¿En qué sistemas CoDe delega?, ¿son los más apropiados? ¿Qué tipo de acciones son delegadas?, ¿resultan, de verdad, delegables? ¿Por cuánto tiempo se cede el control?, ¿es sensato hacerlo durante tanto tiempo? ¿Es reversible la delegación? ¿Qué procedimientos de supervisión o evaluación existen?, ¿son suficientes? ¿Qué riesgos se asumen en caso de fallo?, ¿es prudente asumirlos? ¿Qué ventajas se obtienen para la vida humana con la delegación de control?, ¿qué se pierde a cambio?...

[29] Atocha Aliseda, *Abductive Reasoning*, Springer, Dordrecht, 2006.

Los problemas prácticos realmente importantes no tienen que ver con un supuesto futuro post-humano de máquinas inteligentes. Dejemos estas cuestiones para la ficción recreativa. Lo crucial tiene que ver con el presente[30], con el modo en que se están empleando ya los sistemas de CoDe, con la responsabilidad que sobre ello tienen ciertas personas, empresas y gobiernos, así como con el impacto que este empleo tiene sobre la vida de todos nosotros. Por otro lado, nos damos cuenta de que la evaluación de los sistemas CoDe ha de hacerse caso por caso. Habrá que decir, en la línea indicada por Heidegger, sí y no a la llamada IA. No creo que se pueda hacer un juicio conjunto de plano rechazo o de sumiso acatamiento. Y, en cada caso, habrá que juzgar desde una actitud prudencial respecto a todos los aspectos de la delegación de control.

Los sistemas de CoDe han de funcionar en muchos casos como cajas negras. Quizá solo podamos llegar a auditar algunos especialmente críticos. La idea de que todos pueden ser auditados minuciosamente, de que deberían resultar transparentes, es, a mi modo de ver ilusoria, aunque constituya el primer y más instintivo refugio para ciertas instituciones[31]. Pero si pretendiésemos auditar todos los sistemas de CoDe, estos perderían la mayor parte de las ventajas que nos aportan, en especial las que se refieren al factor tiempo, que marca un límite teórico para cualquier auditoría humana de un sistema de CoDe. Salvo, claro está, que delegásemos la auditoría en otros sistemas de meta-CoDe, que no harían sino reproducir el problema en un nivel superior. No es esta la vía por la que podremos reconducir los problemas que el CoDe pudiera producirnos. Hace falta, como decía, un abordaje prudencial.

Veamos algunos ejemplos, sin ninguna pretensión de exhaustividad ni de sistematicidad, sino más bien a título meramente ilustrativo. Los sistemas de geolocalización son de gran utilidad, pero hemos de aprender a usarlos de un modo en el que no resulten dañadas nuestras capacidades más elementales de orientación. En general, hay que preservar las habilidades, capacidades y facultades de las personas, sobre todo las más básicas y genéricas, como, por ejemplo, las habilidades lingüísticas, de cálculo o de orientación. El CoDe puede ser –ya es- de gran ayuda en medicina, pero hay que emplearlo de manera que el médico conserve cierta autonomía y toda la capacidad de decisión, hay que proteger legalmente dicha autonomía y formar en las facultades y hospitales a los nuevos profesionales para que sepan ejercerla en cooperación con los mejores sistemas tecnológicos disponibles. Las tecnologías aplicadas a la publicidad y a la persuasión pueden dinamizar la economía y facilitar la vida cotidiana, pero han de ser juzgadas como lo haríamos con cualquier recurso retórico, que estimamos positivo si

---

[30] Véase a este respecto: Marie David y Cédric Sauviat, *Intelligence Artificielle. La nouvelle barbarie*, Éditions du Rocher, Mónaco, 2019. Ramón López de Mántaras, en la misma línea, afirma: "Lo que debería aterrorizarnos no es un futuro dominado por una hipotética IA superior […] Lo que realmente debería preocuparnos es la situación presente, en la que estamos delegando cada vez más tareas en una IA tan limitada como la actual". "El nuevo traje de la IA", *Investigación y Ciencia*, 526: 50-59, 2020, p. 59.

[31] Otro tic muy común en muchos organismos políticos consiste en la reducción de todos los problemas (que son, como hemos visto, fundamentalmente metafísicos) a cuestiones éticas, aparentemente más tratables desde la idea de consenso. La UNESCO busca agrupar sus recomendaciones en la llamada *Declaración Universal sobre la IA* (https://elpais.com/tecnologia/2020-09-17/la-unesco-pone-los-cimientos-de-la-declaracion-universal-de-la-inteligencia-artificial.html?rel=lom&fbclid=IwAR02dsUWwW1rY6EtvHtYActsssZSoIaZoyMzOPRD1vyZ8nevDXVKLVXEeSk). La UNESCO ha publicado ya lo que se ha denominado -irónicamente, hemos de suponer- el *Consenso de Beijing sobre IA y educación* (https://unesdoc.unesco.org/ark:/48223/pf0000368303). Por su parte, la UE también prepara directrices éticas y legales para la IA a través del Parlamento Europeo (https://www.europarl.europa.eu/news/es/headlines/society/20201015STO89417/regulacion-de-la-inteligencia-artificial-en-la-ue-la-propuesta-del-parlamento) y de la Comisión Europea (The European Commission's High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI Working Document for Stakeholders' Consultation*, Brussels, 18 December 2018).

facilita la comunicación de la verdad, pero negativo si se pone al servicio del engaño. Algo análogo se puede decir del problema de la distancia moral, que aparece ligado al CoDe de armamento. Puede verse al respecto la película titulada *Espías desde el cielo* (*Eye in the sky*, UK, 2015). Presenta de forma muy vívida los retos morales a los que nos aboca la delegación de control en situaciones de guerra o similares. Si algo se hace obvio es que no podemos renunciar a un margen de decisión humana iluminada por la prudencia. Del mismo modo, la persona que pilota un avión o conduce un coche sabe hasta qué punto puede delegar el control del mismo en una máquina, durante cuánto tiempo y en qué circunstancias, así como cuándo ha de recuperarlo. El software de traducción automática es utilísimo para pasar texto desde la lengua materna a otra conocida por el usuario, quien puede, así, valorar la fidelidad de la traducción y retocarla. Por otra parte, resulta arriesgado y poco aconsejable el confiar en sistemas de traducción para una lengua de destino completamente desconocida por el usuario. La vigilancia del sistema se ejerce mediante retro-traducción o volcado a una tercera lengua. Nadie pretende vigilar la calidad de una traducción auditando los algoritmos que la producen.

En el plano educativo, es, desde luego, imprescindible formar en la virtud a las personas, más que instruir a las máquinas en la ideología de lo políticamente correcto. Los sesgos injustos no serán reconducidos mediante la reprogramación de los algoritmos, sino mediante el desarrollo de personas virtuosas. En concreto, hablo de virtudes como el desasimiento, la creatividad, la laboriosidad y la prudencia. Dichas virtudes han de ser educadas en el contexto tecnológico actual. Se logran mediante un proceso educativo en la práctica de las mismas. Por ejemplo, se puede cultivar el desasimiento (serenidad, *Gelanssenheit*) mediante prácticas de silencio tecnológico[32]. El ejercicio de la prudencia respecto del CoDe requiere una cierta familiaridad con los sistemas tecnológicos que lo ejercen. La creatividad humana en nuestros días también se ejerce en un contexto tecnológico respecto del cual necesitamos distancia crítica y claridad de objetivos. Gracias a la creatividad y a la laboriosidad humana es factible la orientación del desarrollo tecnológico hacia la genuina realización humana.

En el plano político, el enfoque de CoDe nos permite identificar inmediatamente el déficit de legitimidad y el riesgo para la libertad de las personas. Así, la mayor parte de la las tecnologías de CoDe están en manos, en última instancia, del Partido Comunista Chino o de unas pocas grandes corporaciones norteamericanas. Europa tiene poco que hacer al respecto. Los grandes de la llamada IA los conocemos mnemotécnicamente por sus siglas. Son, del lado occidental, los GAFAM (Google, Amazon, Facebook, Appel y Microsoft) y, del lado chino, los BATX (Baidu, Alibaba, Tencent, Xiaomi). La primera impresión que obtenemos es que hay demasiado poder en muy pocas manos.

Podemos empezar a preguntarnos si es legítimo el poder que detentan. Si no lo es, tampoco lo será que lo deleguen en los algoritmos al uso. Es curioso que un régimen como el chino haya logrado desviar el descontento de la población hacia el funcionamiento de los algoritmos[33]. Se entiende que, después de Tiananmen, en un país que mantiene la pena de muerte –y la aplica, según datos de Amnistía Internacional[34], cada año a miles de personas-, cualquier subterfugio

---

[32] A. Marcos, "Silencio tecnológico", *Scio*, 15: 157-176, 2018.

[33] Christina Larson, "La tecnogobernanza china: ¿quién necesita democracia si tiene datos?", *MIT Technology Review*, 31 Agosto, 2018. Disponible en: https://www.technologyreview.es/s/10481/la-tecnogobernanza-china-quien-necesita-democracia-si-tiene-datos.

[34] https://www.es.amnesty.org/en-que-estamos/noticias/noticia/articulo/pena-de-muerte-china-el-mayor-verdugo-del-mundo-debe-reconocer-el-nivel-grotesco-del-uso-que-ha/?gclid=EAIaIQobChMIj7zirpKC7gIVWO3tCh1ODQ8OEAAYASAAEgIKbPD_BwE.

para la crítica sea bienvenido como válvula de escape. Los algoritmos desempeñan entonces el papel de un muñeco de vudú contra el que sí está permitido cargar. Lo que se entiende peor son coincidencias como esta: "los principios chinos [de la *Academia de Inteligencia Artificial de Pekín*] están estrechamente alineados con las recomendaciones de Unicef" respecto al impacto de la IA sobre los niños[35]. El Partido Comunista Chino, que siempre ha sido una desgracia para el pueblo chino, constituye, ahora, con la llamada IA en sus manos, una amenaza para la humanidad entera. La crítica a la llamada IA debería empezar por ahí. Cualesquiera recomendaciones sensatas al respecto deberían incluir una seria impugnación del uso que el Partido Comunista Chino hace de los sistemas de CoDe en un país sin contrapoderes, sin oposición, sin prensa libre, sin jueces independientes. Si obviamos esta cuestión, cualquier documento de recomendaciones sobre huella digital, por bienintencionado que sea, quedará cubierto por el descrédito, si no por el oprobio. En vista de ello, tanto Unesco como Unicef deberían reconsiderar lo que están haciendo al respecto (y dejando de hacer).

Y algo análogo cabe decir respecto a las grandes corporaciones norteamericanas que controlan el resto del pastel de la llamada IA. Es perentorio, por la salud de las democracias y por la libertad de las personas, que el poder desmedido de dichas corporaciones sea disuelto. Es decir, que sean fraccionadas las empresas en cuestión y que la población cobre conciencia de lo que puede hacer al respecto. Un consumo responsable, un uso lúcido de las redes sociales y una gestión prudencial de los propios datos resultarían ya de ayuda. Pero además sería pertinente una decidida presión política de los ciudadanos y de la opinión pública a favor de la disolución de núcleos de poder abusivo[36].

Con lo dicho no deseo ni siquiera insinuar que haya simetría entre el caso chino y el americano, ni que Europa deba colocarse en una posición de equidistancia en materia de IA. En el primer caso, el uso de sistemas de CoDe para el sometimiento de la población debería ser simplemente combatido por Europa. En el segundo caso, el del oligopolio del CoDe detentado por unas pocas empresas norteamericanas, bastaría con apoyar desde Europa las acciones a favor de la libre competencia que ya están siendo sugeridas en USA por los dos partidos mayoritarios, acciones que van, desde luego, en la dirección de un fraccionamiento de las corporaciones GAFAM. Los Estados Unidos todavía pueden ser para nosotros, europeos, un aliado en el desarrollo de sistemas de IA decentes, lo cual no puede decirse del partido que actualmente tiraniza a China.

Dentro de las cuestiones prácticas (es decir, ético-políticas) quiero referirme, por último, al tema de los presuntos derechos digitales. Se especula actualmente con la proclamación de una nueva generación de derechos humanos, los llamados derechos digitales. Y vamos ya por la cuarta o quinta generación de supuestos derechos humanos. Según algunos, la quinta debería ya rebasar el ámbito de lo propiamente humano, para ser de aplicación también a los robots y al software inteligente. Debería estar claro ya a estas alturas que los robots no tienen derechos (ni son responsables de nada ni han de pagar impuestos), tienen cierto valor y, en función del mismo han de ser mantenidos como haríamos con cualquier otro artefacto valioso mientras lo sea. También debería estar claro que no hay software inteligente, sino uso inteligente del software,

---

[35] Karen Hao, "El mundo empieza a sentir la urgencia de proteger a los niños de la IA", *MIT Technology Review*, 25 Septiembre, 2020. Disponible en: https://www.technologyreview.es/s/12643/el-mundo-empieza-sentir-la-urgencia-de-proteger-los-ninos-de-la-ia.

[36] A este respecto, resulta interesantísimo y muy claro el libro de Carissa Véliz, *Privacy is Power. Why and How You Should Take Back Control of Your Data*, Transworld Digital, Londres, 2020.

de donde se sigue que es absurdo hablar de los derechos del software. Estamos, de nuevo, ante un error categorial más bien elemental.

Tampoco me parece muy perspicaz ni muy útil solicitar para los seres humanos una cuarta generación de derechos, esta vez –digamos, por resumir- digitales. Inventar derechos humanos nuevos debilita la propia idea misma de "derechos humanos". La enorme fuerza retórica que (¿todavía?) tiene esta figura ha servido durante décadas, desde la II Guerra Mundial, para proteger a las personas de los abusos del poder. Los derechos humanos se basan en la dignidad de la persona y en su pertenencia a la familia humana. Es siempre tentador emplear esta fórmula para proteger cualquier bien tenido por valioso. Pero, en la medida en que se amplía el campo al cual se aplica, inexorablemente merma la fuerza protectora de la misma.

Si entendemos por derechos humanos los de la primera generación, fundamentalmente el derecho a la vida y a la libertad, entonces la acusación de haber violado los derechos humanos es una acusación de enorme gravedad. Pero si incluimos, no solo los derechos de segunda y tercera generación, que ya son de algún modo clásicos, sino también los nuevos derechos digitales y los neuroderechos, más los derechos "humanos" de los robots o de los animales, entonces la acusación de violación de un derecho se vuelve apenas turbadora. ¿Nos afectará profundamente la desconexión de un robot o la negativa a borrar alguno de nuestros datos de las redes?, ¿estaremos motivados para luchar denodadamente contra alguna de estas violaciones de presuntos derechos? Mucho más perspicaz y útil sería conectar lo digital con los derechos humanos básicos, mostrar cómo afecta a los mismos, sin inventar nuevas e imaginativas listas de supuestos derechos.

Lo grave de algunos sistemas de CoDe no es que violen nuestros presuntos derechos digitales, es que amenazan nuestra vida o comprometen nuestra libertad. Las leyes que rijan en lo digital deberían derivarse, casi como teoremas, de los derechos humanos más básicos, que juegan aquí como axiomas en todos los sentidos (lógico y axiológico). Como apreciaría cualquier matemático, es más elegante un sistema de pocos axiomas independientes, fértil en consecuencias, que uno al cual haya que añadir axiomas cada dos por tres. Creo, pues, que no debería emprenderse la redacción de ninguna carta de supuestos derechos digitales, sino la revisión crítica de los usos que damos a lo digital en conexión con nuestros más básicos derechos (vida y libertad). Este enfoque es suficiente para discernir las tecnologías CoDe que resultan beneficiosas para las personas y para las sociedades de las que tienden a devastar la esencia humana.

6. Resumen conclusivo

- Nuestra acción es siempre interacción, revierte sobre nosotros. Si actuamos sobre la tecnología la tecnología nos cambia, para bien o para mal.

- Lo técnico no es neutral. En general es necesario para la vida humana. En concreto, hay tecnologías mejores que otras, y hay que optar. Cada tecnología tiene riesgos y beneficios, aporta un servicio y supone una servidumbre, y hay que sopesar.

- ¿Cómo podemos valorar las tecnologías llamadas de IA? Hay que establecer su ontología, su epistemología y –solo después- sus consecuencias prácticas (ético-políticas).

- Ontología: La IA no existe. Sin la mirada humana cualquier parte de la realidad pierde su aspecto intencional y colapsa en simple sistema físico. Existen sistemas de IA de los cuales

forman parte las personas. Lo que tienen de inteligente no es artificial y lo que tienen de artificial no es inteligente. El nombre, por lo tanto es desorientador. Sería más acertado hablar de sistemas de sistemas CoDe.

- Epistemología: La aplicación de un enfoque automatista a la investigación científica tiende a empobrecerla si no se hace de modo crítico. Los sistemas CoDe son falibles, se ven afectados por las limitaciones clásicas de la inferencia inductiva, no predicen con seguridad, sino de forma condicional, es decir, no predicen, sino que proponen modelos condicionales. No podemos prescindir, pues, de la prudencia y creatividad humanas.

- Aspectos prácticos: Los sistemas CoDe ya están aportando ventajas y causando problemas hoy. No es pertinente discutir sus aspectos prácticos mirando a un lejano futuro post-humano, sino al presente de nuestras vidas. La mayor parte de la llamada IA depende del Partido Comunista Chino y de unas pocas corporaciones norteamericanas. El control que estos polos de poder ejercen sobre nuestras vidas, gracias a los sistemas de CoDe que emplean, es ilegítimo en el primer caso, abusivo en el segundo. Aun así, no es adecuada la promoción de nuevos derechos *ad hoc*. La protección frente a este tipo de control, así como el cultivo de buenas prácticas, ha de derivar de la educación en virtudes (prudencia, serenidad, creatividad, laboriosidad…) y del respecto a los derechos humanos fundamentales, con muy especial atención a las derechos básicos de las personas más vulnerables.