

Jornada: "Inteligencia artificial: retos, riesgos y oportunidades"

Organiza la Dirección General de Desarrollo Estatutario, Departamento de Presidencia, Interior y Cultura del Gobierno de Aragón. 23 de noviembre a las 16:45 horas, Sala de la Corona del Edificio Pignatelli

Inteligencia artificial ¿Sueño o pesadilla?

Miguel L. Lacruz Mantecón – Facultad de Derecho - Unizar

Cuando nos enfrentamos a una tecnología realmente innovadora, como lo es la de la IA, cualquier intento de predicción resulta simplemente una apuesta, y muy posiblemente esto es lo que ocurra con las afirmaciones que yo pueda hacer aquí hoy. Por eso creo que hay que ser prudentes, y hacer un abundante empleo del tiempo verbal futuro imperfecto, y del modo subjuntivo, que es el que utilizan quienes no están muy seguros de lo que afirman.

La IA es ciertamente una tecnología innovadora, tanto que, como nos dice el profesor de Oxford Luciano Floridi¹, inaugura un nuevo capítulo de la historia de la humanidad: “La revolución digital sólo sucederá una vez, y esa vez es ahora. Hemos pasado una página en la historia humana y un nuevo capítulo ha comenzado. Las futuras generaciones no sabrán lo que era una realidad analógica, predigital y off-line. Somos la última generación en experimentarla”. Lo que significa, en definitiva, dos cosas: que somos dinosaurios, y que ya no hay vuelta atrás, que no podemos “desinventar” la IA.

Con estas prevenciones, la relación entre IA y Derecho se puede abordar atendiendo a las transformaciones que esta tecnología puede causar en la sociedad, procurando acomodar la regulación jurídica a las nuevas realidades, e intentando que los derechos de las personas no se vean lesionados. Naturalmente, habrá intereses que salgan perjudicados, y frente a esto poco o nada se puede hacer. Cuando se popularizaron los automóviles y los camiones, todo el sector del transporte mediante caballerías y carros sufrió, y miles de personas tuvieron que cambiar de trabajo. En una conferencia sobre IA que tuvo lugar el pasado martes en Ibercaja, el investigador del CSIC Pedro Meseguer trajo como icono de la IA una portada de *The Economist*, en la que las letras del acrónimo AI aparecían una con una aureola de ángel y otra con cuernos y rabo de demonio, para reflejar que la nueva tecnología traerá tanto bienes como males. De esto les voy a hablar, pero además intentaré advertir que los mayores males que trae la IA derivan precisamente de su mayor bien, de su funcionamiento perfecto.

Desde el Derecho privado, la incidencia de la IA se advierte en ámbitos como el de la subjetividad y personalidad jurídicas, en el que quieren aparecer los robots como nuevos sujetos o agentes en la vida jurídica, y relacionado con ello, el de la responsabilidad civil por daños. También el ámbito de la contratación electrónica, el de las propiedades intelectual e industrial, en el que aparecen nuevos autores e inventores no humanos, o el de los derechos al honor, intimidad personal y familiar, y otros derechos fundamentales como el de la libertad de opinión e información, en los que se están produciendo alteraciones que no encuentran

¹ FLORIDI, Luciano, *The Ethics of Artificial intelligence*, Oxford University Press, Oxford, 2023, p. xi.

solución en la ciencia jurídica. A mi juicio, la incidencia más profunda se va a producir en dos ámbitos.

Primero, el de la subjetividad de los sistemas inteligentes, especialmente si se trata de robots androides. Esta subjetividad, pese a lo que se dijo desde la archiconocida Resolución UE de 16 de febrero de 2017, y su tan repetida personalidad electrónica para robots, no creo que venga por la concesión de una personalidad a estos entes, como reconocimiento a su inteligencia. Son máquinas, y aunque realizan tareas inteligentes, su inteligencia no es humana, y cometeremos un grave error si nos confundimos en este punto. Lo que ocurre es que esta inteligencia por un lado produce beneficios, y por otro crea en los humanos que interactúan con los robots un espejismo de humanidad que les impide tratarlos como máquinas. El reconocimiento de una cierta subjetividad de los robots vendrá impuesta por razones de conveniencia, y por nuestra propia ética, no por la inexistente humanidad de las máquinas.

Segundo ámbito de incidencia: el de los derechos de la personalidad. En este ámbito, la IA puede introducir enormes disfunciones, que se relacionan con el campo de los daños y la responsabilidad, pero que va más allá del simple mal funcionamiento del sistema que produce daños que deben ser reparados. La cuestión de los daños producidos por la IA se ha abordado al modo clásico, como la determinación e imputación de la obligación de indemnizar derivada de una utilización negligente o malintencionada del sistema, siguiendo una línea de culpa subjetiva, o imponiendo una responsabilidad objetiva o cuasiobjetiva. Estas soluciones clásicas son adaptables a la reparación de daños producidos por el mal funcionamiento de sistemas inteligentes: errores, sesgos, pérdida de beneficios, daños materiales, etc.

Pero, por otra parte, tenemos los daños causados no por el mal funcionamiento del sistema, sino los que se derivan de un correcto funcionamiento, pero que no fueron tenidos en cuenta. Vengo últimamente repitiendo hasta hacerme pesado que los daños más importantes que puede causar la IA no se producirán porque funcione mal, sino al contrario, porque funciona demasiado bien.

Como ejemplo de esto tenemos la diferenciación recogida en el art. 5 de la llamada Ley de Inteligencia Artificial (LIA, en realidad la *Propuesta de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de IA*, de 2021) entre las técnicas de inteligencia artificial *prohibidos* y las que «solamente» son *de alto riesgo*. Entre los prohibidos se cuentan las técnicas de persuasión y alteración subliminal de comportamientos (como los *nudges*), las de perfilado de sujetos y su calificación social (como el sistema de crédito social chino, o los sistemas de control de fronteras *iBorder ctrl* o *AVATAR*). También los sistemas de identificación biométrica remota «en tiempo real», como los sistemas policiales SARI (italiano) o el sistema inglés AFR (*Automated Facial Recognition*). Tras estos sistemas prohibidos, la Ley de la IA refiere una serie de sistemas que son calificados como de alto riesgo: sistemas para la gestión de la sanidad, funcionamiento del tráfico rodado y el suministro de agua, gas, calefacción y electricidad, sistemas de selección para el acceso a instituciones educativas y acceso al empleo, etc.

Comparando estos dos tipos de sistemas, vemos que los prohibidos lo son porque su mismo funcionamiento causa daños en derechos fundamentales de las personas, derechos a la

intimidad, a la libertad de información y a la libre formación y manifestación de opinión, etc. Y este daño a los derechos se produce, precisamente, porque la perfección del sistema va más allá de la falibilidad e imperfección humanas. En cambio, los sistemas de alto riesgo causan daños cuando su funcionamiento no es todo lo perfecto que debe ser, sobre todo en ámbitos especialmente sensibles como el de la sanidad y medicina, la justicia, el tráfico y las comunicaciones, o los suministros básicos para la vida humana.

La LIA no incide apenas en un tipo de sistemas inteligentes que a mi juicio pueden ser especialmente dañinos. Se trata de los sistemas de IA generalistas, o «modelos fundacionales», aquellos que no tienen una tarea predefinida, sino que asumen los encargos que se les haga. Ejemplo de estos sistemas es el famoso *ChatGpt*, o en el diseño, *Dall-E*.

Pues bien, para tratar de la peligrosidad de la IA, y las posibles actuaciones al respecto, se ha celebrado los días 1 y 2 de noviembre de este año una importante cumbre en Bletchley Park, Inglaterra, lugar mítico por ser dónde Alan Turing descifró el código *Enigma*. En esta cumbre, y en un diálogo retransmitido entre el Premier Rishi Sunak y el magnate Elon Musk, éste declaró que la IA era «una de las mayores amenazas para la humanidad», y esto cuando precisamente el lema de la reunión era «Cumbre por la seguridad de la IA».

En la Declaración de esta Cumbre de Bletchley Park se consideran especialmente peligrosos estos sistemas de IA generalistas, o «modelos fundacionales», respecto de los cuales se señala que pueden surgir riesgos sustanciales debido a un posible uso malicioso o a problemas de control. «...Existe la posibilidad de que se produzcan daños graves, incluso catastróficos, ya sean deliberados o no».

¿Por qué son estos sistemas generalistas o modelos fundacionales tan peligrosos? A mi juicio, la peligrosidad de estos sistemas inteligentes deriva de algo en principio, bueno: de su enorme capacidad creativa, que ya alcanza a la creatividad humana. Los nuevos sistemas de IA pueden mimetizar la capacidad creativa humana y producir obras artísticas, gráficas, pictóricas, obras intelectuales como novelas o monografías de investigación y obras musicales, de manera que en mi opinión hay que hablar ya de un “autor cibernético”. En el campo del arte se anuncian producciones de nuevas obras de pintores clásicos, fallecidos hace siglos, cuya obra se crea hoy con su estilo y su arte. En cinematografía se «resucita» a actores fallecidos, y se les hace actuar junto a los vivos, o se rejuvenece a los actores ya ancianos, para hacer efecto de regresión en el tiempo. También mimetizan la capacidad de invención del ser humano, y en investigación genética y biotecnológica se prevén grandes avances con el uso de sistemas inteligentes.

Esta capacidad creativa de la IA produce sin embargo dos consecuencias negativas. La primera deriva de la capacidad de la IA para imitar la realidad creando imágenes de calidad fotográfica, y vídeos hasta el punto que deja de ser posible distinguir la imagen real de la inventada. Estas creaciones o imitaciones de la realidad, pueden incidir en la lesión de algunos de los derechos fundamentales, pero en mi opinión el daño es más importante, porque se falsifica la realidad, produciendo una pérdida de los referentes reales y del marco auténtico de la vida humana. Corremos un serio peligro de perder la certeza acerca de nuestros registros de la realidad, con el resultado de producirse una desinformación general y una

inverificabilidad de hechos y pruebas. En medios periodísticos e informativos ya no son novedad las intoxicaciones o *deep fakes* y las falsas noticias o *fake news*.

La segunda consecuencia negativa está en la capacidad de la IA de imitar los resultados inteligentes del esfuerzo humano. Esto a la larga nos puede llevar a delegar en la máquina la realización de todo tipo de esfuerzos intelectuales, con la consecuencia de que perderemos destreza intelectual. Esto ya ha ocurrido: desde los años 70 del pasado siglo con la popularización de las calculadoras de bolsillo, nuestras capacidades de cálculo se han visto desplazadas por las máquinas. En la educación, corremos el riesgo de lo que yo llamo «deslocalizar» nuestra inteligencia, al trasladar la producción de resultados inteligentes de nuestro propio cerebro al de la máquina, y encima engañarnos adjudicándonos unos resultados que en realidad no son nuestros. Cada vez es más difícil distinguir la labor intelectual humana de la de la máquina.

Un dato objetivo que refleja esta pérdida de capacidades es la llamada “reversión del efecto Flynn”, que describe un continuo descenso del coeficiente de inteligencia o IQ de los estudiantes en general desde mediados de los años 90, cambio que según el propio James Flynn² ha tenido lugar conforme los niños han estado más inmersos en la cultura electrónica visual y auditiva moderna: televisión, videojuegos y teléfonos móviles, que alteran la capacidad de pensar en términos de abstracciones y desarrollar la lógica deductiva y la planificación sistemática. Y para el filósofo Carabantes López³ la culpable es directamente la IA y la tecnología informática, afirmando que desde 1995: «...se ha producido un descenso en la demanda social de habilidades intelectuales... Al disminuir la demanda de inteligencia, los individuos, lamarckianamente, han dejado de cultivarla ...la causa principal ha sido la penetración de la informática en la vida cotidiana». La realidad nos está demostrando que si la inteligencia se puede obtener de fuentes no humanas a bajo coste, ésta pasa a ser una capacidad menos valorada entre los humanos, igual que la fuerza física de un cargador de puerto es menos valorada cuando se tienen carretillas elevadoras y montacargas: se produce una *devaluación social* de la inteligencia. Y lo que es peor, se produce lo que el informe *Future Risks of Frontier AI*, que se divulgó en la cumbre de Bletchley Park, denomina «Dependencia humana excesiva de los sistemas de inteligencia artificial», que posteriormente no se puede eliminar, dependencia que ya tenemos con las calculadoras digitales o con Google para recordar cualquier dato. Todos hemos oído la frase. “Para qué te lo vas a aprender de memoria, si lo tienes todo en Internet”.

Con todo, ya se aprecia una cierta reacción por parte de los responsables gubernamentales, al menos respecto a la falsificación de la realidad. Así en la UE tenemos la Recomendación (UE) 2023/2425 de la Comisión, de 20 de octubre de 2023, que se emite para luchar contra la difusión en línea de contenidos ilícitos y nocivos, o de desinformación y *fake news*, en relación con las crisis internacionales por la agresión de Rusia a Ucrania y el ataque terrorista de Hamás a Israel. Y la francesa *Proposition de loi n°1630 visant à encadrer l'intelligence artificielle par le droit d'auteur*, de 12 de septiembre de este año 2023, nos pone en guardia en su exposición

² FLYNN James R./SHAYER, Michael, “IQ decline and Piaget: Does the rot start at the top?”, *Intelligence* (2017), <https://doi.org/10.1016/j.intell.2017.11.010>.

³ CARABANTES LÓPEZ, Manuel, «Inteligencia artificial lingüística perfecta: efectos sobre la autopercepción del ser humano», *SCIO. Revista de Filosofía*, n.º 18, Julio de 2020, pp. 224 y ss.

de motivos sobre las fotografías, textos y vídeos creados digitalmente, que, aparte de problemas de Derecho de autor, plantean cuestiones de ética, entorpecen nuestro libre albedrío e incluso plantean problemas para la supervivencia de nuestra creatividad humana.

Por su parte, la Declaración de la Cumbre de Bletchley Park alerta específicamente sobre «...el potencial de riesgos imprevistos derivados de la capacidad de manipular contenido o generar contenido engañoso», riesgos especialmente importantes en los sistemas inteligentes de propósito general, por su facilidad de uso.

Frente a estos problemas, ¿qué puede hacer el Derecho? Desde la UE se propone una regulación jurídica que acote el uso de la IA para fines beneficiosos, una especie de principio de no maleficencia. Por supuesto, hay que alejar esta tecnología de las etapas escolares de la educación, y la prohibición de las aplicaciones que sean lesivas para los derechos de las personas y el control de la ciudadanía. Todo esto lo digo, además, con pocas ilusiones, y teniendo muy presente que lo deseable sería un control mundial de la IA, similar al de la energía atómica.