

**ALGORITMOS E INTELIGENCIA
ARTIFICIAL EN LA SOCIEDAD DEL
SIGLO XXI**

Documentos Humboldt 20

*Asociación Alexander von Humboldt-
España (Editora)*

2024

Copyright 2024

© De la edición, Asociación Alexander von Humboldt de España
© De los textos y gráfico, sus autores

Todos los derechos reservados. Ni la totalidad ni parte de este libro puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética, o cualquier almacenamiento de información y sistema de recuperación sin permiso escrito de los titulares del copyright.

ISBN : 978-84-1071-445-8

Depósito legal: V-1792-2024

Índice

	<u>Pág.</u>
Presentación <i>Silvia Barona Vilar</i>	3
Palabras de Bienvenida del Representante de la Embajada Alemana en Madrid <i>Georg Schulze Zumkley</i>	16
Palabras de Bienvenida del Representante de la Fundación Alexander von Humboldt <i>Thomas Hesse</i>	22
Presentación de Paul Nemitz <i>Cristina Hermida del Llano</i>	27
“Macht, Recht, Forschung und Bildung im Zeitalter der Künstlichen Intelligenz” <i>Paul Nemitz</i>	31
Presentación de Nuría Oliver Ramírez Eduardo Demetrio Crespo.....	51

	<u>Pág.</u>
“Hacia una inteligencia artificial socialmente sostenible” <i>Nuria Oliver Ramírez</i>	55
Presentación de Daniel Innenarity <i>Elena Martínez García</i>	89
Hasta qué punto es democrática la democracia digital? <i>Daniel Innenarity</i>	94
Presentación de Azucena García Palacios <i>Sergio Rodríguez García</i>	135
“Inteligencia Artificial y salud mental” <i>Azucena García Palacios</i>	137

PRESENTACIÓN

Silvia Barona Vilar, Presidenta de la Asociación Alexander von Humboldt en España

Magnífica Sra. Rectora de la Universidad de Alicante, Prof. Amparo Navarro Faure,

Herr Dr. Thomas Hesse, Stellvertretender Generalsekretär, Alexander von Humboldt Stiftung

Herr Dr. Georg Schulze Zumkley, Consejero de Asuntos de Economía y Ciencia de la Embajada de Alemania en España,

Estimadas personalidades invitadas a este Encuentro

Estimadas Frau Luisa Rath, Referentin für Europa, Fundación Alexander von Humboldt y Frau Kristina Medjedovik, representante de la DAAD

Ilustres ponentes

Queridos compañeros y compañeras miembros de la Asociación Alexander von Humboldt España

Queridos colegas de la Universidad de Alicante y de otras Universidades que nos acompañan

Queridos estudiantes

Señoras y señores, meine Damen und Herren

Comenzamos este Encuentro sobre Algoritmos e Inteligencia Artificial en la sociedad del Siglo XXI en la Universidad de Alicante con enorme alegría; es el trigésimo Encuentro de la Asociación Alexander von Humboldt de España en esta maravillosa ciudad mediterránea y en esta estupenda Universidad.

Es por ello que vayan mis primeras palabras de agradecimiento a la Universidad de Alicante. Quiero expresar en nombre de la Asociación Alexander von Humboldt de España, a la Magnífica Rectora de la Universidad, nuestro más profundo agradecimiento por su hospitalidad y por hacer posible este Encuentro de nuestra Asociación en su casa.

Asimismo, un año más, quiero transmitir gratitud a la Fundación Alexander von Humboldt. Muchas gracias Herr Hesse por acompañarnos en este Encuentro.

Vielen Dank, Herr Hessen, dass Sie an diesem Treffen der Vereinigung an der Universität von Alicante teilnehmen. Für uns ist es absolut wichtig, die direkte Beteiligung unserer Mutter, der Alexander von Humboldt-Stiftung, an allem, was unsere Vereinigung tut, zu haben. Bitte danken Sie Herrn Dr. Robert Schölg für die guten Beziehungen zwischen der Stiftung und unserer Vereinigung.

Igualmente, aprovecho estas palabras iniciales para expresar públicamente la vinculación de la Asociación con la Embajada alemana. Tanto el embajador Dodl como la actual embajadora Frau Margarite Gosse, han abierto las puertas al diálogo estrecho con nuestra Asociación. Como Presidenta he participado en diversas actividades culturales propulsadas por la Agregada Cultural de la Embajada, Frau Isabela Neisinger.

En este sentido y ante la imposibilidad justificada de la Embajadora Gosse de asistir a este Encuentro, debido a las múltiples obligaciones que la Presidencia española en la Unión está ocasionando. En su lugar nos acompaña el Consejero de Asuntos de Economía y Ciencia de la Embajada Alemana, Herr Dr. Georg Schulze Zumkley,

Vielen Dank für Ihre Anwesenheit, Herr Dr. Schulze. Wir hoffen, dass die Beziehungen zwischen der Botschaft und der Vereinigung auch weiterhin so gut sein werden. Bitte danken Sie Botschafter Gosse für diese gute Freundschaft, die hoffentlich im Laufe der Zeit weiter wachsen wird.

Asimismo, agradezco a los ponentes, por sus generosas contribuciones en estos días, por haber hecho espacio en sus apretadas agendas y poder compartir con nosotros su expertisse en un tema tan complejo como el que nos ocupa en este Encuentro.

Es siempre difícil organizar un Congreso, un

Encuentro como el que realizamos. Este año las complejidades también han estado presentes, si bien en esa búsqueda de soluciones y en el apoyo recibido debo dar las gracias a dos personas en particular, por un lado, a Joaquín Silvestre, Tesorero de nuestra Junta y Catedrático de esta Universidad de Alicante, sin el cual este Encuentro difícilmente hubiera sido posible; gracias por apoyarme en todo momento. Y, en segundo lugar, debo citar en este acto la colaboración que desde el primer día recibí de nuestro compañero humboldtiano Daniel Innenarity, quien me ha guiado estupendamente en la organización del evento en momentos en que los acontecimientos nos venían desbordando. Gracias a los dos.

El tema que nos concita en este Encuentro nos va a permitir reflexionar acerca de la expansiva proyección de los algoritmos y la IA en la sociedad del siglo XXI. Un tema que genera sentimientos contradictorios, no dejando indiferentes. Es indudable que en la actualidad no puede negarse esta realidad y los avances se expanden sin fisuras a todos los ámbitos: consumo, economía, medicina, salud mental, ecología, agricultura, educación, gestión social, justicia, geografía humana, transporte-movilidad, etc., Todos los días escuchamos, leemos, oímos de nuevas herramientas, incluso de las inteligencias artificiales generativas, ChatGpt y similares.

Asistimos, por ende, a una cada vez más humanidad algorítmicamente asistida, una especie de proletariado digital que propaga

comportamientos globales de normalización y homogeneización, que no es sino lo que Byung-Chul HAN¹ ha considerado como una masa en forma de enjambre planetario. Y la razón no es otra que la aparición del homo digitalis que en palabras del filósofo coreano es cualquier cosa menos “nadie”.

No solo hay digitalización, ni siquiera podemos hablar tan solo de tecnología, sino de una fusión en estos últimos escalones de la tecnología con la medicina, la biología, la neurociencia, la lógica, la lingüística, la estadística, etc, que van a permitir agilizar una búsqueda primero, y una simplificación lógico-deductiva, después, para ofrecer respuestas automáticas (a través de fórmulas algorítmicas) que anteriormente eran fruto solo de la argumentación lógica del ser humano.

En el siglo XXI la sociedad ha mutado digitalmente. La tecnología irrumpió en las últimas décadas del siglo XX, de la mano de la revolución digital 3.0., un fenómeno desencadenado por el movimiento de informatización de la sociedad, iniciado en los años sesenta, impulsor de la germinación de la idea de que las máquinas de cálculo estaban dotadas de una enorme proyección de eficacia, que traería enormes beneficios al desarrollo económico y facilitaría la vida de las personas. Se nucleó sobre el internet, una especie de “bien

¹ HAN, B-Chul (2014), *En el enjambre*, Barcelona, Herder, p. 26.

global” de la sociedad, un instrumento que se presentaba como esencial para garantizar una sociedad acomodada, eficiente y ágil. Se convirtió en muy poco tiempo en un mantra social, como apunta Eric SADIN en una doxa, que propulsó una mirada acrítica de la revolución digital, y, con ella, el desarrollo de la digitalización integral del mundo², imprescindible compañera de viaje de la sociedad, mudando nuestro paisaje esféricamente, y propulsando la transformación del exhausto modelo revolucionario de contrato social –que ha servido durante tanto tiempo para conformar no solo el modelo de sociedad, sino, muy especialmente, el modelo de Estado- en un contrato social algorítmico.

Técnica, tecnología, ciencia, neurociencia, ética y derechos se van poco a poco integrando para presentar un entorno adecuado en un escenario preparado para el 4.0.: la globalización, la sociedad de masas y la sociedad de consumo³. En sus inicios esta revolución se objetivaba en el deseo del cambio del modus operandi empresarial, si bien se ha expandido a todos los sectores y ámbitos de la vida, convirtiéndose en una suerte de modus vivendi, transportando maneras de actuar del homo economicus a la cotidianeidad vital, pasando de lo que Norbert WIENER denominaba del

² SADIN, E. (2020), *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*, Buenos Aires, Caja Negra Editora, pp. 21-22.

³ *Ad extensum*, BARONA VILAR, S. (2021), *Algoritmización del Derecho y de la justicia*; cit., pp. 60-67.

Mensch al *Menschmaschine*⁴, y emergiendo un nuevo fenotipo denominado *Homo digitalis*⁵, de modo que efectivamente, como apunta Byung-Chul HAN, el hombre digital digita en el sentido de que cuenta y calcula constantemente⁶.

Así, el imparable desarrollo de la computación, vinculando además los desarrollos de la tecnología con los avances científicos, especialmente desde finales del siglo XX, fue generando una nueva etapa de industrialización, en la que se combinan digitalización, conectividad, automatización, robotización e inteligencia artificial. Surgió así la cuarta revolución industrial, o Industria 4.0. El origen del término Industria 4.0. se encuentra en un proyecto de estrategias de alta tecnología realizado por el gobierno alemán, cuyo objetivo era la creación de la fábrica inteligente o también conocida como Ciberfábrica, caracterizada por la gran interconexión entre máquinas automatizadas, la concurrencia de redes de comunicaciones, la integración de tecnologías avanzadas de procesamiento de datos, la robótica avanzada, la capacidad de autodiagnóstico de situaciones, el mejor intercambio de información

⁴ WIENER, N. (1952), *Mensch und Menschmaschine*, Frankfurt, Alfred Metzner Verlag, 4º ed., pp. 150-194.

⁵ MARTÍNEZ OJEDA, B.(2006), *Homo digitalis: etnografía de la cibercultura*, Bogotá, Universidad de los Andes, 2006; también puede verse LASALLE RUIZ, J.M. (2019), *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, Barcelona, Ed. Arpa, p. 42.

⁶ HAN, B-CH. (2014), *En el enjambre*, Barcelona, Ed. Herder, p. 60.

y una mayor eficiencia en la gestión de recursos naturales y humanos. La velocidad con la que camina este impulso volcánico es exponencial, favoreciendo la interconexión mundial, incidiendo no solo en el sector económico y de los negocios, sino en la sociedad en general y en las personas en particular, con un impacto transfronterizo sin delimitación territorial jamás visto⁷. Es la etapa de la mutación de la sociedad analógica hacia un mundo digital, innovativo y disruptivo, aunque con las falencias de todavía seguir pensando analógicamente y extrapolar el mundo analógico al mundo digital.

El tránsito tecnológico y digital ha permeado el mundo entero emergiendo herramientas algorítmicas que se expanden a todo y a todos sin exclusión, abrigando lo que MOROZOV denomina como el solucionismo tecnológico⁸, ofreciendo la pócima de la felicidad, adaptados a una vida híbrida, una integración de la Humanidad y la Tecnología, que destella con una luz que embarga y se expande a todos los ámbitos, a todas las áreas, un verdadero estilo de vida. Asistimos a una algoritmización de la vida, de la sociedad, del mundo, algo que ha propulsado, en palabras de Daniel INNENARITY Daniel INNENARITY al afirmar que “los algoritmos tienen una dimensión política en la medida que intervienen en el orden

⁷ SCHWAB, K.(2018), *La cuarta revolución Industrial*, Barcelona, Ed. Debate, 4 ed., p. 15.

⁸ Ideas que se encuentran en MOROZOV. E. (2015), *La locura del solucionismo tecnológico*, Madrid, Katx Editores.

social y estructuran nuestras decisiones”⁹.

Hablar de algoritmos e inteligencia artificial exige hoy un trabajo interdisciplinar donde la técnica, la neurociencia, la lingüística, la ética, la economía, la geografía, el derecho etc caminen de la mano, con simbiosis entre lo público y lo privado. Es un momento trascendental para el presente pero también para el futuro, en busca de un ecosistema entre los humanos y las máquinas que, allende posiciones extremistas como las transhumanistas, asumen que la criatura humana ha tocado techo y debe ser sustituida por la máquina, o los negacionistas, que mantienen una venda en su mirada, se busca la interacción, el impulso que beneficie a la Humanidad y no que la someta, todo y que los humanos nos hemos venido generando dependencias y esclavitudes tecnológicas que pueden provocar mentes líquidas, sin compromisos, sin profundizaciones, deshumanizando las relaciones. viviendo de los likes o de las redes sociales como muestrario de éxito y de vida. Basta pensar en la nomofobia, como patología generada por esa dependencia obsesiva el móvil.

Las bondades de los desarrollos de IA son exhibidas diariamente en los medios de comunicación y en los estudios científicos por sus desarrolladores, empero los riesgos también existen. Prueba de ello es el interés que desde el

⁹ INNENARITY, D. (2022), “Igualdad Algorítmica”, en *El País Semanal*, 13 de mayo.

seno de la UE se ha venido presentando, con código éticos y mas recientemente con la AI Act que está a punto de ser aprobada a finales de este año 2023, un esfuerzo regulatorio que incide en la protección de la privacidad amén de la delimitación de precauciones de estos sistemas. También en España los desarrollos normativos y también estructurales nos acompañan. El último eslabón ha sido el reciente RD de 22 de agosto de 2023 por el que se aprueba el Estatuto de la Agencia España de Supervisión de IA.

Ese interés en Europa por establecer límites éticos y jurídicos no cierra el círculo, porque con los desarrollos de IA se trasciende de la UE y se observa como en el Atlántico y en el Pacífico los desarrollos de IA están mucho menos limitados que en Europa, lo que genera asimetrías y diría que una suerte de entropía en el marco limitante o garantista, según se observe.

Con estas coordenadas de sociedad en estado de mudanza¹⁰, directamente manipulada hacia una sociedad del hiperconsumismo, totalmente conectada, el valor de la reflexión no existe, y se va generando un pensamiento acrítico, bajo control¹¹

¹⁰ BARONA VILAR, S. (2021), “Una justicia “digital” y “algorítmica” para una sociedad en estado de mudanza”, en la obra colectiva BARONA VILAR, S., *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, Valencia, Tirant lo Blanch, pp. 21 a 64.

¹¹ GABRIEL, M. (2020), *El sentido del pensamiento*, Barcelona, Pasado&Presente, 2 ed.. en sentido amplio.

y una liquidez¹² global de la vida, las sociedades, la cultura, la educación y en cierta medida también, en lo que a mi me atañe, de la Justicia.

La pandemia ha sido un gran aliado de esta reinención social, lo que ya se experimentó en otros momentos de la Historia de la humanidad, si bien en este momento con una característica indiscutible, la afectación a la sociedad global, a todo el planeta y sobre todo a alta velocidad, lo que sí es más nuevo. La soledad, la falta de comunicación directa, la melancolía, las falencias humanas y sociales encuentran su pócima en la tecnología, en las redes sociales, en lo que Byung-Chul Han denomina el enjambre digital¹³). La prueba más palmaria de cuanto hablamos es la enorme satisfacción que sienten una masa de humanos, los rebaños a que se refiere el autor coreano, por presentar en el escaparate de las redes sociales¹⁴ su vida privada, en muchas ocasiones más ficticia que real, al permitirnos ofrecer un personaje creado y maquillado,

¹² BAUMAN, Z. (2013), *Vida líquida*, Barcelona, Ed. Paidós, p. 109.

¹³ HAN, B-CH. (2020), *En el enjambre*, cit, pág. 26. Como apunta este autor, el modo de proceder de lo digital define un nuevo ser, básicamente porque, lejos de ser “singular” actúa de acuerdo con los parámetros de los otros, encantados de habitar en el panóptico digital, sintiéndose libres de exponer e iluminar con informaciones propias, narcisistas, ese mundo en el que se hallan, empero sin ser conscientes de que se hallan prisioneros de una sociedad de control. Puede verse, págs. 38, 90, 100, entre otras.

¹⁴ HABERMAS, K., *Problemas de legitimación del capitalismo tardío*, Ed. Cátedra, 1973.

expuesto en la red para obtener satisfacción de un ego que necesita de los likes para su autoestima.

Los retos y desafíos que los algoritmos y la IA presentan en la actualidad requieren de una mirada crítica, positiva y negativa, transversal, que permita abordar estos desafíos y el redimensionamiento del ser humano frente a las máquinas, a las que otorgar, si cabe, un status, de persona maquina de ser maquina, y que aborde el estadio de la Singularity (KURZWEIL¹⁵), de la humanidad aumentada¹⁶, suprahumanidad o transhumanismo, que ofrece una enorme mejora del ser humano desde los Cyborgs¹⁷, impulsada con el apoyo de los ideólogos de Silicon Valley¹⁸, un ideal constante que pretende garantizar los desarrollos tecnológicos adecuados para que sean capaces de gobernar la humanidad de forma más

¹⁵ KURZWEIL, R., *The Singularity is near*, New York, Penguin ed., 2005, págs. 28 a 35. Este autor, junto con los colaboradores de Silicon Valley, han sido los impulsores del transhumanismo, considerando que el ser humano, la criatura humana ha tocado techo y, por ello, el despertar del universo se producirá cuando la inteligencia humana se expanda por todo el universo en forma de energía y materia saturada.

¹⁶ SADIN, E., *La humanidad aumentada. La administración digital del mundo*, Buenos Aires, Ed. Caja Negra, 2017.

¹⁷ BENDEL, O., "Cyborg", *Gabler Wirtschaftslexikon. Das Wissen der Experten*, Springer Gabler, <https://wirtschaftslexikon.gabler.de/definition/cyborg-54197>.

¹⁸ LLANO ALONSO, F.H., *Homo excelsior. Los límites ético-jurídicos del trashumanismo*, Valencia, Tirant lo Blanch, 2018, pág. 97.

ágil, rápida e inteligente.

Cuestiones que van a ser planteadas a lo largo de mañana por los excelentes ponentes que nos van a acompañar, el Dr Paul Nemitz desde la Unión Europea, la Dra Nuria Oliver, el Prof. Dr Daniel Innenarity y la Prof Dra Azucena García Palacios. A todos ellos agradezco enormemente su participación, motivándonos a participar activamente en el coloquio. Un coloquio que promete ser interesante.

Un año más la Asociación de antiguos becarios y becarias de la Fundación Alexander von Humboldt y premiados por la Fundación nos reunimos, este año en la Universidad de Alicante y sobre una materia como es la que nos ocupa de algoritmos e inteligencia artificial, presentando esa misión que tenemos en el seno de la sociedad civil, científica e investigadora, mostrando el expertise amplio, heterogéneo y de diversos periodos generacionales de nuestros socios.

Y termino como empecé, reiterando mi agradecimiento a la Rectora de la Universidad de Alicante por hacerlo posible, gracias queridas autoridades presentes, compañeros y compañeras, amigos y amigas, por vuestra presencia y vuestro apoyo.

PALABRAS DE BIENVENIDA

Georg Schulze Zumkley

Leiter Wirtschafts- und Wissenschaftsreferat
Botschaft der Bundesrepublik Deutschland

Sra. Sra. Rectora de la Universidad de Alicante,
Dra. Amparo Navarro Faure,

Sr. Secretario General Adjunto de la Fundación
Alexander von Humboldt, Dr. Thomas Hesse

Sra. Consejera, María del Carmen de España
Menárguez,

Sra. Presidenta de la **Asociación Alexander von
Humboldt**, querida Silvia Barona,

Señoras y señores miembros de la AvH,

Meine Damen und Herren:

Es para mí una especial alegría poder darles hoy la bienvenida. Me gustaría dar las gracias especialmente a la Asociación AvH por la invitación.

- 1) Esta asamblea de “Humboldtianos” es una viva muestra de la **fortaleza de las relaciones científicas entre España y Alemania.**

Desde el establecimiento de la Fundación Alexander vom Humboldt en 1953, se han otorgado más de mil becas Humboldt a personas de España para realizar una estancia en Alemania. ¡Enhorabuena a los becarios de este año! **Esperamos que se conviertan en miembros de nuestra comunidad global** Alexander von Humboldt.

Se está desarrollando una **red en constante crecimiento de científicos de alto rango** con estrechos vínculos con Alemania como lugar de referencia para la educación superior y la investigación.

Desde la embajada, impulsamos el intercambio con la **Red de Científic@s Hispano-Alemanes**. Por ejemplo, los miembros de la red tuvieron intercambios online con la Secretaria de Estado Raquel Yotti sobre la nueva Ley de Ciencia¹⁹ y con José Manuel Pingarrón, Secretario General de Universidades sobre la “LOSU”²⁰. **No duden en hacerme saber si estan interesados en unirse a la Red de Científic@s Hispano-Alemanes.**²¹

Quisiera dar las gracias al equipo de la Fundación por sus excelentes esfuerzos bajo

¹⁹ Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Publicado en: «BOE» núm. 214, de 06/09/2022. Entrada en vigor: 07/09/2022

²⁰ Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario, publicado en: «BOE» núm. 70, de 23/03/2023. Entrada en vigor: 12/04/2023

²¹ Contacto: wi-100@madri.diplo.de

la gerencia de Silvia Barona. También quiero extender el agradecimiento de la Embajadora de Alemania, Maria Margarete Gosse, por la labor y la colaboración excelente con el equipo de la embajada durante el año pasado.

2) El **tema de la Jornada Científica** es pertinente, no solo para científicos, ingenieros o diplomáticos, sino también para la humanidad. Nos preguntamos si la **inteligencia artificial** será un **elemento que traerá esperanza** en una época de retos enormes para los europeos. Quizá voy a hablar de eso con mi nueva amiga **LuzIA**²². Lo cierto es que el año **2023** será recordado como el año de la inteligencia artificial.

3) Por su parte, **2022** será recordado por las generaciones futuras como el año en que el presidente ruso Putin trajo **la guerra** a Europa con la invasión de Ucrania. Putin convirtió en armas el gas, el trigo, el maíz – y la inteligencia artificial -, No es solo un ataque contra Ucrania, sino contra la Europa libre y democrática y sus valores. Para Alemania especialmente, pero también para la Unión Europea y la alianza de la OTAN, la guerra representa un cambio de época, eine **Zeitenwende**.

No es solo eine **La Unión Europea ha iniciado una nueva política de refuerzo y estabilidad**. El día 11 de marzo de 2022, en su

²² <https://app.luzia.chat/#/home>

reunión en Versailles, los líderes de la Unión Europea prometieron: „*Invertiremos en tecnologías digitales, entre las que se cuentan la inteligencia artificial (...)*“.²³

En 2023 la **Presidencia española del Consejo** llega en un momento de retos comunes.

- 4) En 2023, asimismo somos testigos de la **Zeitenwende de tecnología**. La edad de la inteligencia artificial ha arrancado.

The Age of AI has begun, tal y como escribió Bill Gates en sus famosos Gates Notes del día 21 de marzo de 2023.²⁴

Se refleja **esta** Zeitenwende en las actividades de mi equipo:

--En el mes de marzo, informamos a la embajadora acerca de los últimos avances de modelos como ChatGPT.

--En junio, organizamos un taller con la Cámara Alemana de Comercio. El CEO de GFT Technologies, una empresa del M-Dax alemán,

²³ Declaración de Versalles, 10 y 11 de marzo de 2022, sobre la agresión rusa contra Ucrania, así como sobre el fortalecimiento de las capacidades de defensa, la reducción de las dependencias energéticas y el desarrollo de una base económica más sólida. - <https://www.consilium.europa.eu/es/press/press-releases/2022/03/11/the-versailles-declaration-10-11-03-2022/>

²⁴ The Age of AI has begun - Artificial intelligence is as revolutionary as mobile phones and the Internet. By Bill Gates, March 21, 2023 - <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>

mostró el crecimiento el **ecosistema de los modelos de IA**. Se ve como un árbol que florece en primavera.

--Ahora Ustedes, ese grupo importante de la AvH, dedica su Jornada Científica al tema.

Los avances en el ámbito de la inteligencia artificial forman parte de una **cuestion estratégica clave**: ¿Cómo mantener y fortalecer la competitividad de nuestro continente Europa ante los desafíos internacionales en terminos de salud, de alimentación, de educación y formación, de costes de producción, disponibilidad de fuentes de energía económicas, demografía, capacidad de innovación y soberanía de datos?

Nosotros, los europeos, debemos avanzar en el ámbito de la inteligencia artificial, tanto en la innovación como ya avanzamos en la regulación: La Unión Europea está elaborando la primera ley integral sobre IA del mundo. Dentro de poco, el uso de la inteligencia artificial en la UE estará regulado por la Ley de Inteligencia Artificial – the **IA Act**.

Señoras y señores,

Hace poco tuve una **conversación con una joven llamada LuzIA**. Pregunté: “Which person will determine the technological development of AI?”

Me contestó mi nueva amiga: “The

development of AI is influenced by a **collective effort** involving researchers, engineers, policymakers, and various stakeholders in the field of technology.”

LuzIA es inteligente. LuzIA es útil. De hecho, LuzIA es una aplicación de inteligencia artificial.

LuzIA nos dice: “It’s a collective effort.” -
”Tenemos que **trabajar juntos** para ganar el reto de innovación y competitividad.”

Trabajemos y vivamos **todos juntos para que el árbol del futuro crezca en Europa.**

Según nuestro **principio rector de la unidad europea: «unidos en el interior, fuertes en el exterior».**

Les deseo a todos mucho éxito en esta reunión.

¡Muchas gracias!

PALABRAS DE BIENVENIDA

Thomas Hesse

Stellvertretender Generalsekretär del
Alexander von Humboldt Stiftung

Ilustrísima Professora Navarro Faure,

Sehr geehrter Herr Dr. Schulze Zumkley,

Iustrísima Señora Carmen de Espana

Liebe Frau Professorin Barona Vilar,

Liebe Humboldtianerinnen und
Humboldtianer, ladies and gentlemen,

It is a great pleasure for me to attend this year's annual meeting of *the Asociación Alexander von Humboldt de España*. I would like to thank the Spanish Humboldt Alumni Association most sincerely for their kind invitation and the *Universidad de Alicante*, represented by *la Señora Rectora Magnífica Navarro Faure*, for the wonderful hospitality. Please allow me to also emphasise that this year's meeting marks an anniversary celebration. It is truly fantastic that you are coming together in Alexander von Humboldt's

name for the 30th time to discuss the relevant issues of our time. I must therefore extend a special thank you to the Board of the *Asociación Alexander von Humboldt de España*, and its President, Silvia Barona Vilar. Through your unparalleled commitment to interdisciplinary and international exchange, you continuously succeed in carrying the spirit of Alexander von Humboldt to Spain, where it remains alive in a network of almost 800 academics from all research disciplines.

AAs a Humboldt Alumni Association, you play a key role in supporting regional and international scientific and cultural exchange between researchers from Spain and Germany. The collaborations that result from your commitment fulfil a special mission beyond academia as they contribute to a growing understanding between cultures and countries. Especially in times of political and social crises such as currently experienced in Europe – be it the war in Ukraine, rising nationalism in many EU member states or climate change – scientific dialogue is indispensable. We therefore consider ourselves very fortunate to have the *Asociación Alexander von Humboldt de España* as a strong partner by our side to promote research collaborations that are crucial for tackling scientific and societal challenges.

New research collaborations are created not least through the promotion of junior

scientists, many of whom are taking part in this event. Please be assured that we are always looking forward to welcoming new Humboldtians to our network, either through the Humboldt Research Fellowship Programme for highly qualified researchers or through the Hezekiah Wardwell Programme for highly talented musicians. Once selected, they all become members of our global Humboldt Family, consisting of more than 30,000 Humboldtians in over 140 countries, including 59 Nobel laureates. Such a network can proudly look at a remarkable scientific and cultural diversity, which makes it possible to jointly achieve solutions to the challenges of our time. Alexander von Humboldt himself already knew: "Ideas can only bear fruit if they take root in many minds". Since the founding of today's Alexander von Humboldt Foundation in 1953, more than 1100 excellent minds from the Spanish research landscape have received a research fellowship or award, mostly for projects in the natural sciences, closely followed by those in the humanities. The Alexander von Humboldt Professorship – Germany's most highly endowed science award – has been awarded to researchers from Spain three times already.

Humboldtians do not only benefit from an exceptional international network though. "Once a Humboldtian – always a Humboldtian" also means that the Alexander von Humboldt Foundation supports the individual life paths

of its alumni through a comprehensive alumni sponsorship portfolio. This includes renewed research stays to maintain, expand, or establish new contacts with colleagues in Germany, and financial support for regional conferences. We also encourage our alumni abroad to become hosts for German research fellows in the so-called Feodor Lynen Programme.

To keep our research fellowship programme attractive and competitive, we were recently able to increase the monthly fellowship rate and improve the flexibility of our guidelines. For example, in the foreseeable future it will be possible for our postdoctoral researchers to carry out their fellowships in several research stays – until now, this was offered exclusively to our experienced researchers. Furthermore, we will be able to pay family allowances for partners in addition to spouses and approve recreation periods based on the statutory holiday regulations.

But let us now turn our attention from future funding opportunities to the present event. You have come together here at the *Universidad de Alicante* to discuss algorithms and artificial intelligence in 21st century society – a topic that is not only pertinent to science. As the title of this conference already reveals, algorithms and artificial intelligence are socially relevant topics. Above all, these topics have long since arrived in society. Allow me to mention ChatGPT as an example: a chatbot that caused

a real hype only a few months ago. For the first time, non-experts were able to interact with AI. The enthusiasm for the performance of artificial intelligence was and still is huge. But so are the concerns about the risks of a technology that could spiral out of human control. The extent to which AI will revolutionise the world can hardly be estimated and is not very transparent to the public – but AI is here to stay. This is why we need a scientific dialogue that sets the course for a human-centred AI – a mission that the Alexander von Humboldt Foundation has recognized and thus introduced the Alexander von Humboldt Professorship for Artificial intelligence. Through this programme, we sponsor research on the social, legal, or ethical aspects of artificial intelligence. In addition, a podcast series called "AI and Us" was created. Experts from the Humboldt Network report on their research and answer the most important questions about artificial intelligence. What can AI achieve already and how will it benefit us in the future? What risks does it entail, and which questions does AI research have to face?

I am now very much looking forward to these and further questions also being discussed within and beyond the programme of this conference. I wish us all fruitful exchanges and a wonderful time here at the *Universidad de Alicante*.

PRESENTACIÓN DEL PONENTE PAUL NEMITZ

Cristina Hermida del Llano

Catedrática de Filosofía del Derecho,
Universidad Rey Juan Carlos (Madrid),
Humboldtiana

Queridos humboldtianos, distinguidos señoras
y señores:

Me gustaría comenzar agradeciendo a nuestra presidenta de la Asociación Alexander von Humboldt en España, Doña Silvia Barona, el honor de poder presentar al primer conferenciante de este XXX Encuentro anual de Humboldtianos, que este año tiene lugar en la bonita ciudad de Alicante.

Ayer estuvimos celebrando la sesión de apertura institucional en la sala “Rafael Altamira”, un personaje ilustre, muy querido para mí, no solo por haber sido un eminente jurista, historiador y pedagogo vinculado a los proyectos de la Institución Libre de Enseñanza, sino también por la relación que mantuve con su nieta, Pilar Altamira, escritora y bióloga de formación.

Hoy nos encontramos en el Salón de Grados de

la Facultad de Derecho de la Universidad de Alicante, abriendo la sesión científica de este Encuentro, con otro gran jurista de vertiente interdisciplinar.

Humboldt dijo que solo la ciencia que brota del interior puede cambiar el carácter, la personalidad de cada uno de nosotros. En cambio, hoy en día muchos consideran que la Inteligencia Artificial puede lograr lo mismo, haciéndonos mejores personas, y con ello además mejorar la sociedad. Para hablar de este complejo tema, y de cómo la falta de regulación en el ámbito de la Inteligencia Artificial no ayuda proteger el bien común, la seguridad jurídica y los derechos de los ciudadanos en democracia, tenemos hoy aquí a un experto internacional relevante en el campo de la Inteligencia Artificial como primer conferenciante:

El Sr. Paul Nemitz, a quien agradezco expresamente por su presencia.

Brevemente, resumiré su extraordinaria trayectoria profesional. Es director de Política de Justicia en la Comisión Europea; Profesor visitante de Derecho en el Colegio de Europa de Brujas; Miembro Honorario de la Universidad Libre de Bruselas (VUB).

Durante seis años ha sido director de Derechos Fundamentales y Derechos de los Ciudadanos en la misma Dirección General. Como director,

el Sr. Nemitz dirigió la reforma de la legislación sobre protección de datos en la UE (GDPR), las negociaciones del Escudo de Privacidad UE-EE.UU. y las negociaciones con las principales empresas de Internet estadounidenses del Código de Conducta de la UE contra la incitación a la violencia y la incitación al odio en Internet.

Además, el Sr. Nemitz es el coordinador del Grupo de Reflexión Transatlántico interdisciplinar sobre Democracia y Estado de Derecho en la era de la Inteligencia Artificial y ponente habitual en la Mesa Redonda sobre Inteligencia Artificial de Atenas. Fue miembro de la Comisión de Ética de Datos del Gobierno alemán hasta 2019.

Miembro de la Comisión de Medios de Comunicación y Política de Internet del Partido Socialdemócrata de Alemania (SPD), Berlín; miembro de la Junta de la Verein Gegen Vergessen, Berlín; Patrono del Instituto Leo Baeck, Nueva York; presidente del Patronato de la Fundación Arthur Langerman de la Universidad Técnica de Berlín (TU Berlín); miembro del Consejo Global de Inteligencia Extendida del IEEE y del MIT de Boston; miembro del Foro Global de Boston del Instituto Michael Dukakis. También forma parte del Consejo Asesor de la Trustless Computer Association de Zúrich y de la Association for Accountability and Internet Democracy, París.

Su intervención lleva por título “Macht, recht, Forchung und Bildung im Zeitalter der Künstlichen Intelligenz”. Tiene la palabra. Muchas gracias.

MACHT, RECHT, FORSCHUNG UND BILDUNG IM ZEITALTER DER KÜNSTLISCHEN INTELLIGENZ (*)

Paul Nemitz

Chefberater, European Commission

Muchas gracias. Hemos acordado que yo hable en inglés porque tenemos algunos oradores y ustedes no hablan alemán y quizás todos en la sala entiendan bien el inglés. Por supuesto, estaré encantado de aceptar preguntas y debatir más tarde.

Nuestro tema de hoy es un tema importante para todas las ciencias y también de relevancia activa en la política de la Unión Europea en Bruselas, que ahora están trabajando en lo que será la primera regulación democrática de esta tecnología. Las negociaciones legislativas entre el Parlamento Europeo y el Consejo de Ministros, que representan a los Estados de Europa, siempre se aprueban por doble mayoría.

(*) "Poder, Derecho, Investigación y Educación en la era de la Inteligencia Artificial". Artículo traducido por Miguel Angel Medina, Catedrático de Bioquímica y Biología Molecular, Universidad de Málaga. Humboldtiano.

Estas negociaciones legislativas se vieron realmente atrapadas por la aparición de ChatGPT y su mapa de éxito mundial. ChatGPT consiguió en pocos meses más de mil millones de euros para sus propietarios.

Algunos miembros del Parlamento presentaron discursos escritos por ChatGPT y, tras ser revisados de forma independiente, revelaron que ChatGPT había escrito el contenido de dichos discursos. Esto tuvo un enorme impacto en la negociación del acto legislativo y las últimas discusiones están tardando un poco más, pero confío en que a finales de año tendremos una ley en Europa que establezca las primeras normas básicas sobre inteligencia artificial.

Para simplificar y porque me dirijo a un público interdisciplinar, diría que estas normas básicas no son más que lo que un ingeniero responsable haría de todos modos. La ley obliga a los productores y usuarios de inteligencia artificial a realizar una evaluación de riesgos de la tecnología que comercializan, y la novedad es que esta evaluación de riesgos no sólo se refiere a la seguridad física de las máquinas, sino que también debe tener en

cuenta el impacto en la vida de las personas y en la democracia. ¿Por qué? Porque que esta tecnología entra en una nueva área de sustitución de las mentes, lo más simplemente describible como un paso de una situación en la que el sistema digital distribuye el contenido de la información a una nueva situación en la que los sistemas digitales lo producen. Eso es lo que ahora representa ChatGPT. La ley resalta el grupo de productores y usuarios de estos sistemas para averiguar si estos sistemas tienen muy alto, medio o bajo riesgo de acuerdo a como se describe en la legislación. Luego, de acuerdo con la clasificación de riesgo, la ley determina la profundidad de las obligaciones en términos de documentación del sistema para asegurarse de que funciona, en términos de si hace lo que se supone que debe hacer, en términos de monitorizar el sistema en la vida real, como una compañía farmacéutica tiene el deber de monitorizar un nuevo medicamento que se va a introducir en el mercado. El productor del medicamento tiene que seguir el impacto en la gente y si ve que la gente se está muriendo por tomar sus pastillas tiene que retirar la pastilla del mercado. Lo mismo con la Inteligencia Artificial, no sólo en relación con la muerte de personas, sino también en relación con el impacto negativo sobre los derechos fundamentales de la democracia. Es importantes instituir el deber de notificar a todo el mundo que está siendo abordado por la

inteligencia artificial. Es un asunto de dignidad humana que la gente sepa si se comunican con otro ser humano o si se comunican con la máquina, lo cual, por supuesto, ya no es evidente. Hoy en día, la tecnología está tan avanzada que puedes estar hablando con alguien por teléfono a través de internet y puedes pensar que estás hablando con un humano, pero estás hablando con una máquina, y esto afecta oralmente, pero también afecta por escrito. Así que es importante que la gente sepa que está hablando con una máquina; por supuesto, esto es especialmente importante cuando se refiere a los procesos democráticos. Si te levantas por la mañana y encuentras en tus redes sociales todos los argumentos a favor de un candidato y ya no puedes saber si esos argumentos vienen de la máquina o vienen de personas reales, eso significa realmente que la democracia no funciona para ti, porque quieres saber que el discurso sobre asuntos políticos tiene lugar entre seres humanos, que tienen derecho a voto, y no es algo que se modifique por la intervención de una máquina.

Así pues, esta es la idea rectora de esta legislación. Por supuesto, como cualquier legislación, no resolverá todos los problemas de este mundo, ni esta legislación se aplicará perfectamente. Ya está claro que el FSB ruso, por supuesto, seguirá haciendo propaganda, que parece humana, pero que es producida por máquinas. También está claro que habrá un

uso criminal de la IA, que no va a cumplir esta legislación. Sin embargo, yo diría que es un importante paso adelante también en términos de sentar el precedente de que la democracia actúa sobre estas tecnologías. Lo que me lleva a una cuestión que también me gustaría debatir con ustedes, porque sé que en la audiencia tenemos muchos filósofos, abogados y otros que pueden estar interesados en este tema: *La importancia de mantener la primacía de la democracia sobre la tecnología y los modelos de negocio*. Creo que esta es la gran cuestión de nuestro tiempo, y la Unión Europea es la única potencia líder que ha puesto o está poniendo en marcha una ley democrática sobre IA, que fue precedida por un estudio de alto nivel sobre ética de la IA. La sucesión de acontecimientos en la Unión Europea muestra exactamente de dónde venimos. Venimos de una era de 20 años bajo la regulación de Internet, inducida por una comprensión de la libertad que estaba muy moldeada por lo que yo llamaría la ideología californiana y lo que otros simplemente llaman el pensamiento americano moldeado, en particular por la *Declaración de la Independencia del Ciberespacio* de John Perry Barlow. Tal vez algunos de ustedes recuerden. La banda de rock Grateful Dead y John Perry Barlow en los años 90 escribieron algo que se llama la Declaración de Independencia del Ciberespacio. Y en esta declaración, con palabras muy poéticas, expuso la tesis, o la reivindicación, de que en este espacio de

Internet (que se suponía libre y sólo un espacio de la mente), la legislación, los parlamentos, los gobiernos no tienen legitimidad. Y dijo "*nosotros, los amos de Internet y del espacio digital, no os queremos aquí. Nos ocuparemos nosotros mismos de todos nuestros problemas*". Esto es algo que se leyó en el Foro Económico Mundial de Davos en aquel momento y tuvo un enorme impacto en la legislación estadounidense relativa a Internet, que luego se copió en la Unión Europea. Este tipo de pensamiento también condujo a la creación del grupo de ética de alto nivel sobre IA, o para el caso, el Comité Alemán de Ética de Datos, que fue nombrado por el gabinete de Merkel en Alemania. En ambos casos, se hizo evidente que estos grupos de ética (que son, por supuesto, un esfuerzo por despolitizar el debate sobre cómo abordar la política y las tecnologías del futuro) eran eminentemente políticos. En estos grupos, los informes se negociaban, como en una comisión parlamentaria. Ambos grupos, tanto el europeo como el alemán, presentaron informes consensuados con una diferencia. El Grupo Europeo, liderado por un representante de la industria (en concreto, el ex CEO de Nokia, de Finlandia) no se pronunció claramente sobre la necesidad de una ley para regular la IA. El grupo alemán, compuesto por ocho representantes nombrados por el Sr. Zee Hofer (a la sazón, ministro conservador del Interior de Alemania) y ocho representantes nombrados por el ministro de Justicia (a la

sazón, socialdemócrata), pertenecientes al gobierno de la gran Coalición, llegó a la conclusión de que es necesaria una legislación. Pero no sobre la base de consideraciones, digamos, filosóficas, sino más bien sobre la base de consideraciones de poder y oportunidad.

La cuestión es qué podemos dejar a la ética en relación a las normas para las nuevas tecnologías con buena conciencia. ¿Y qué tenemos que regular por ley para darle legitimidad democrática? La línea divisoria, la línea divisoria exacta, no se ha respondido en ninguno de estos informes. Pero los argumentos de oportunidad política del informe alemán fueron algo que luego convenció también a la Comisión Europea. Por supuesto, la Ley de IA sigue a toda una serie de otras leyes que hemos aprobado en la Unión Europea hasta ahora para regular la esfera digital, a saber, el *Acta de Mercados Digitales* (que refuerza la ley de competencia para intervenir cuando las empresas poderosas tratan de limitar la entrada en el mercado de nuevos competidores) y el *Acta de Servicios Digitales* (DSA, que regula cómo las plataformas tienen que comportarse comercialmente). Tomemos el caso de Amazon. ¿Puede Amazon, cuando se busca un producto, mostrar siempre como resultado de la búsqueda primero sus propios productos? ¿O tiene que comportarse como un mercado neutral? La respuesta del DSA es que tiene que comportarse como un

mercado neutral. Pero el DSA es también una normativa que reconoce, yo diría que por primera vez, que las plataformas digitales no son sólo lugares de comercio, no son como los mercados históricos de la Edad Media y hasta el siglo XX. También son lugares donde la gente construye su opinión política, por lo que son lugares de democracia. Por lo tanto, el DSA es una legislación que aborda tanto aspectos comerciales como aspectos importantes para el funcionamiento de la democracia. En concreto, obliga a las grandes plataformas, a las plataformas de control, a garantizar de manera estructural, mediante medidas organizativas, que las plataformas no se conviertan en caldo de cultivo de discursos de odio, de incitación a la violencia, de propaganda antidemocrática, de noticias falsas, etcétera. El que tanto el contenido ilegal como el legal, si se hace viral, si se presenta en masa, puede socavar la democracia, es algo que las plataformas tienen que abordar ahora a través de sus medidas estructurales. Lo que estamos intentando con el DSA en particular, combinado con otras leyes, es asegurar la esfera pública donde construimos nuestras opiniones. La gente construye sus opiniones hoy en día en Internet; eso está empíricamente muy claro: más del 50% de la gente ahora en Estados Unidos y Europa construye su opinión política en Internet, basándose en lo que consiguen ver de los algoritmos. Internet se ha convertido en la nueva esfera pública, y por eso es importante

que mantengamos en esta esfera pública unas condiciones que favorezcan la democracia. Una pregunta, por tanto, que me parece interesante que debatamos: ¿Dónde está esa línea divisoria? ¿Cuáles deberían ser las consideraciones? ¿Dónde deberíamos decir que la ética no es suficiente y que necesitamos la ley? Por supuesto, la ley es muy importante en Europa para garantizar la igualdad de condiciones entre las empresas de los 27 Estados miembros, la ley tiene legitimidad democrática y se puede hacer cumplir con el poder de los Estados. Y aquí volvemos a la cuestión del poder. Creo que es muy, muy importante que todo el pensamiento (incluido el pensamiento de los informáticos, o de la física, de las matemáticas) relacionado con la inteligencia artificial tenga siempre presente que se trata de una tecnología de poder. La inteligencia artificial nos permite recalibrar la distribución del poder entre los individuos y las máquinas. En muchos aspectos permite manipular a las personas si se combina con los perfiles que existen sobre nosotros en Internet basados en la recogida de nuestros datos personales de nuestro comportamiento. Si pensamos en un futuro en el que estos paquetes de información personal (que constan de cientos de parámetros sobre cada uno de nosotros) que muestran nuestro comportamiento en Internet y fuera de la red se utilicen y combinen con el poder de la inteligencia artificial, eso significa que no sólo

será más probable que se prediga nuestro comportamiento, sino que también será posible y más probable que ya hoy se manipule nuestro comportamiento. Estamos llegando a ver el mundo a través de una imagen compuesta por estos componentes, a saber, nuestro comportamiento anterior en los datos personales de los perfiles, y la inteligencia artificial, que calcula qué es lo más probable para mantenernos el mayor tiempo posible en la pantalla. Este es el mecanismo de la economía de la atención, que nos lleva a ciclos de radicalización. Hay un estudio famoso en Estados Unidos de Julia Angwin, *The Markup*, que era básicamente una recogida de datos de personas en época electoral, voluntarios que aceptaban añadir a su navegador una aplicación que recogía lo que llegaban a ver en la campaña electoral, lo que llegaban a ver en Facebook, lo que llegaban a ver en Twitter, en Instagram, etcétera, qué respuestas de búsqueda obtenían en Google. Estas personas habían declarado paralelamente a poner en estos navegadores qué opinión política tenían. Julia Angwin, *The Markup*, descubrió en este estudio que la visión del mundo que se presenta a los demócratas antes de las elecciones es completamente diferente de la visión del mundo que los algoritmos presentan a los republicanos antes de las elecciones. Esto es parte de la explicación de la cada vez mayor polarización e imposibilidad de ponerse de acuerdo en nada en Estados Unidos, es la base

de tu información desde el principio está filtrada según tus preferencias cognitivas por uno u otro candidato como presidente.

¿Cómo vas todavía a debatir con los demás? Falta una base común, lo que me lleva a una cuestión un poco alejada de la política cotidiana, pero que sigue siendo muy importante, a saber, la cuestión de la investigación y su deber para con la verdad. La investigación, por supuesto, es algo que en gran parte de la ciencia presupone la reproducibilidad de los resultados de los ensayos. Y que tiene al mismo tiempo la ambición (por cierto, como el periodismo) de alimentar en la sociedad los hallazgos comunes de verdades que son necesarias para que la sociedad encuentre el consenso en términos de seguir adelante y tomar decisiones razonables. Debemos preguntarnos: ¿Qué papel puede desempeñar en este proceso una inteligencia artificial en la que nunca la misma entrada en el sistema produce el mismo resultado? Yo diría que, del mismo modo que los juristas insistimos, si se utiliza la inteligencia artificial en la toma de decisiones de, por ejemplo, la administración pública, que afectan a los derechos individuales, esas decisiones deben seguir conteniendo motivaciones que permitan a un juez controlar si la decisión se ha tomado correctamente. Esto es algo que la IA hasta ahora no puede ofrecer. Yo diría que la ciencia debe insistir en esta prueba de reproducibilidad de los ensayos también,

cuando en los ensayos puede haber participado antes una IA. Digo esto porque hay una tendencia por parte de quienes quieren impulsar esta tecnología en el mercado, venderla a la administración pública, venderla a la ciencia, a socializarnos con un argumento que dice así: *"Si insistes en la capacidad explicativa, y reproductiva de los resultados que la IA ha producido, estás reduciendo el rendimiento de la IA porque estás limitando la capacidad de la IA para trabajar libremente, la IA entonces sólo será capaz de producir resultados que pueda explicar, y estos resultados te privarán de los hallazgos y resultados correctos del análisis de la IA en otras áreas, utilizando por ejemplo las redes neuronales profundas donde simplemente no podemos explicar cómo llegaron, pero las respuestas son correctas, por favor, créenos"*. Cuando se trata de la ley, cuando se trata del sistema legal, soy de la creencia muy estricta de que un principio básico del estado de derecho y de la libertad de las personas en las sociedades democráticas, es *el deber del estado de motivar sus decisiones*. Eso significa que no puede haber decisiones arbitrarias. Y la motivación debe ser de tal calidad que la decisión, cualquier decisión del Estado que afecte a nuestros derechos, debe poder ser controlada por un juez. Y si un juez considera que no existe una buena motivación, es su deber anular la decisión. Si aceptamos el argumento de que la inteligencia artificial es tan buena que deberíamos prescindir de la

necesidad de explicar cómo funciona, como en la ciencia, la necesidad de la capacidad reproductiva, creo que estamos perdiendo características básicas que hemos ganado en la ilustración y en el derecho; al menos una característica básica de un acuerdo constitucional que garantice que tenemos un sistema político, tenemos un sistema administrativo en el que hay un alto grado de control sobre la arbitrariedad y esto ocurre a través del deber de motivación y el juez.

Así que permítanme pasar a la última parte de mi presentación: la cuestión de si la IA puede beneficiarse de derechos que han sido escritos para los humanos. ¿De qué hablo? Hablo, por ejemplo, del derecho de propiedad intelectual (en alemán *geistiges Eigentum*). También hablo del derecho a la libertad de expresión, por ejemplo. Pensemos en esto un momento juntos. Hoy en día, la IA produce textos, produce imágenes que parecen arte. ¿Pueden estos productos producidos por la IA recibir una propiedad intelectual? Bueno, el término alemán *geistiges Eigentum* es muy informativo. Es quizá un poco más informativo que el término propiedad intelectual. *Geistiges Eigentum* presupone un *geist*, un espíritu. Así que la pregunta es: ¿hay un *geist*, un espíritu, en la IA? ¿Existen otras condiciones para conceder la propiedad intelectual? El derecho de propiedad intelectual está vinculado al ser humano, a su ingenio, a su *geist*. ¿Es un derecho que sólo puede

corresponder al ser humano? Se trata de una cuestión inminentemente importante en la relación de poder y en relación con lo que significa para los seres humanos de todo el mundo. Porque, seamos claros, si decimos "Fijémonos sólo en el impacto", si decimos "No, no es sólo para los humanos. Los textos de IA y el arte de IA pueden tener propiedad intelectual", probablemente signifique que millones de personas perderán su trabajo. Porque simplemente serán sustituidos como productores de propiedad intelectual y habrá un gran negocio que estará extremadamente centralizado porque los amos de la inteligencia artificial automatizada, que pretenden dar derechos de propiedad intelectual a la IA son aquellas empresas que ya eran poderosas en el dominio digital en los tiempos pre IA. Open AI ChatGPT comenzó como una aplicación supuestamente sin ánimo de lucro. Nunca estuve seguro de esto. Luego llegó una inversión de 10.000 millones de dólares por parte de Microsoft. Así que si dijéramos que los derechos de propiedad intelectual aumentan, significa un enorme cambio de derechos y un enorme cambio de riqueza para estas empresas. Pero yo diría que esa no es la consideración principal, aunque es una consideración importante. La consideración principal es la respuesta a la pregunta de si la propiedad intelectual (*geistiges Eigentum*) puede atribuirse a las máquinas. Y yo diría que no. Porque el derecho de *geistiges Eigentum* se

ha escrito para proteger el ingenio humano, la productividad humana. Y el hecho de que la IA haya sido invocada por alguien al principio con una pregunta con algunas aportaciones básicas no es suficiente en términos de aportación de ingenio para conceder un derecho de propiedad intelectual a la persona que ha hecho la máquina o a la persona que ha tecleado la pregunta sobre la que se ha producido el texto o ha puesto el elemento de partida para hacer una imagen. Pero también hay otra consideración. Me gustaría mantener este principio de que los derechos humanos no pueden corresponder a las máquinas como un principio general del derecho.

La siguiente pregunta es: "De acuerdo. De ninguna manera la IA puede tener propiedad intelectual. *¿Puede la IA ampararse en el privilegio de la libertad de expresión?*". Esta es de nuevo una pregunta muy, muy importante. *¿Supone alguna diferencia para nosotros que se nos grite agresivamente y se digan cosas equivocadas sobre nosotros, pero no cosas ilegales? Pero estas cosas vienen de una máquina, no vienen de un humano. El derecho a la libertad de expresión es un pilar esencial de la democracia y la democracia es cosa de humanos, no de máquinas. Si decimos: "El derecho a la libertad de expresión es algo que también corresponde a la máquina. Como la máquina no es más que una herramienta de una persona, atribuimos a la máquina el derecho de la persona que está detrás de la máquina*

tecleando", tendremos que vivir con estos hechos. El discurso agresivo, que está permitido en democracia, que no es ilegal. Mentir, que no es ilegal. Que todo esto se convierte en un producto de masas muy necesario. Así que yo diría que pensemos bien si esto es realmente hacia donde queremos ir. He probado esto en mí mismo. Me sorprendió mucho lo que Chat GPT escribió sobre mí y otros colegas. Y como el derecho a la libertad de expresión es un derecho humano y está relacionado con la democracia, no me gustaría que este derecho se atribuyera a máquinas que pueden producir cosas que contaminan el espacio público de tal manera que la democracia deja de funcionar. Así que *yo diría que el principio general de que los derechos humanos no pueden atribuirse a las máquinas también debería aplicarse al derecho a la libertad de expresión*. No hay derecho a la libertad de expresión de la IA, lo que significa tener muy claro que los que crearon ChatGPT, si ChatGPT sólo miente sobre ti, dice cosas erróneas sobre ti y demás, deberían tener una responsabilidad por esto, que es mayor que si otra persona dice lo mismo. Tendrán que disciplinar a la máquina más de lo que los humanos tienen que disciplinarse a sí mismos.

Permítanme pasar a la última cuestión que me gustaría debatir hoy, a saber, la cuestión de si una persona jurídica puede consistir únicamente en inteligencia artificial. Ya hemos visto en algunos países un debate sobre esto de

que un robot pueda ser una persona jurídica. En Estados Unidos hay tres estados que han adoptado leyes sobre organizaciones autónomas descentralizadas, digamos tecnológicas, organizaciones en las que el humano ya no desempeña un papel, no un papel importante. Y por supuesto, puedes imaginar el interés económico en hacer esto posible. Si todos los coches se conducen automáticamente por IA, en términos de liberar de responsabilidad a las compañías automovilísticas para decir "bueno, ya sabes, cada coche tiene su propia persona jurídica porque es una IA la que lo está dirigiendo". Sería fantástico. Podrías librarte de mucha responsabilidad por productos tecnológicos como productor de estos productos. Y de hecho la construcción de una persona jurídica en la ley es algo así como una ficción legal porque la persona jurídica, por supuesto, no es una persona. Es sólo una asamblea de personas y capital, y hay diferentes teorías en derecho de sociedades de responsabilidad limitada. Tradicionalmente, por supuesto, se crearon, digamos, como un impulso a los negocios, a la banca y demás, para compartir la responsabilidad y, por lo tanto, hacer posible ser más activo económicamente, asumir mayores riesgos, etcétera. También hay otra teoría que dice, bueno, necesitamos formas jurídicas para que las personas trabajen juntas, y esta unión tiene su propia relevancia y, por lo tanto, la unión se convierte en una persona

jurídica con la que se pueden tener derechos y obligaciones. Pero tenemos que preguntarnos: ¿qué problema estamos resolviendo? ¿Necesitamos dar una persona jurídica para resolver un problema importante de la sociedad en el interés común a la inteligencia artificial, o se trata de nuevo de una maniobra de los poderosos, de aquellos que tienen en sus manos una IA que funciona bien para ser aún más poderosos y protegerse de la responsabilidad? Así que yo diría aquí de nuevo, aunque la persona jurídica es una ficción, ha sido creada por ley y el legislador probablemente tendría de la misma forma el derecho de crear una persona jurídica donde un ser humano ya no juega ningún papel. No digo que jurídicamente no sea posible. Yo diría: No resuelve ningún problema en la sociedad. Crea más problemas. No lo necesitamos. Y diría también desde el punto de vista del principio general de que, a saber, el derecho, el concepto de derecho que hemos creado para que los seres humanos trabajen juntos. No digo que el derecho de sociedades sea un derecho humano. Pero estoy diciendo que hemos creado estas ficciones legales para hacer posibles las acciones humanas que van más allá. Y sólo el individuo puede ir. No deberíamos transmitir las. No deberíamos antropizar la inteligencia artificial. Y dar una persona jurídica a la inteligencia artificial iría en esta dirección. Así que esto es lo que me gustaría discutir con ustedes hoy al final de mi charla.

Hay dos grandes cuestiones; la primera, ¿somos capaces de mantener la primacía de la democracia sobre los modelos de negocio y la tecnología? Y la otra gran pregunta: ¿Debemos aceptar que en un futuro bastante cercano ya no haya nada especial en los humanos, que en derecho esté reservado a los humanos? Hay una elevación de la inteligencia artificial en términos de ser beneficiaria de derechos, que originalmente eran derechos humanos. Y puede que al final se produzca la singularidad, que haya realmente una fusión total de los humanos y la inteligencia artificial de las máquinas. Tal vez un día ya no podamos hacer ninguna diferencia. La cuestión aquí es: ¿Deberíamos adoptar una posición que diga, bueno, si tecnológicamente es posible, deberíamos hacerlo? ¿O deberíamos adoptar una posición normativa que mantenga al ser humano con sus debilidades, con sus sueños y aspiraciones, con su dignidad y su condición de persona en el centro del sistema de derecho, y que permita conceptos como la autodeterminación informativa de las personas, que permita un concepto de dignidad humana que incluya el principio de que los seres humanos no pueden ser objetos de las máquinas? En el sentido de la ilustración, utilizamos las máquinas como herramientas. No somos las herramientas de las máquinas. No estamos al mismo nivel. ¿Es ése un concepto que queremos mantener normativamente para el futuro? ¿Y cuáles son los argumentos

científicos y filosóficos para hacerlo? Creo que estas son las tareas que tenemos que abordar juntos y podemos basarnos en las filosofías de Hans Jonas. «*El principio de precaución*», su famoso libro, el principio de responsabilidad de 1979, quizá algunos de ustedes lo recuerden, fue la biblia del movimiento antiatómico de la época. Y podemos basarnos en filósofos como Hannah Arendt y Max Weber, o Heinrich Popitz, el sociólogo que dijo que el control del poder técnico será una función central de la democracia en el futuro. Esto ya lo dijo en los años 90. Así que, al final, la cuestión a la que nos enfrentamos es la relación de poder entre el ser humano y la máquina.

¿Nos dejamos llevar sólo por la tecnología y ése es nuestro destino, o damos forma como humanos, a través de un proceso democrático, a cómo vamos a vivir en el futuro y cuál será nuestra posición como humanos en este mundo tecnologizado?

Muchas gracias a todos.

PRESENTACIÓN DE LA PONENTE NURIA OLIVER

Eduardo Demetrio Crespo, Catedrático de Derecho Penal, Universidad de Castilla-La Mancha (Toledo), Humboldtiano

Estimadas/os Sras./Sres:

Constituye un verdadero honor y una alegría enorme participar en este Encuentro de la Asociación Alexander von Humboldt de España, que con tanto acierto preside mi querida amiga y admirada colega Prof.^a Dr.^a h. c. mult. Silvia Barona Vilar, a quien agradezco muy sinceramente su invitación para fungir como moderador en el día de hoy.

Asimismo, aprovecho para felicitarla por la elección del tema, que no podría ser de mayor actualidad e interés. En efecto, “Inteligencia Artificial” es, sin duda, la palabra del año, sobre ella aparecen todos los días novedades e informaciones que nos inquietan y ponen en cuestión prácticamente todo, al punto que se puede hablar de un “desafío epistémico” de dimensiones desconocidas hasta la fecha.

Una asociación científica como la nuestra debe ocuparse, sin duda, de los grandes desafíos que afronta la humanidad en todos los órdenes y más aún si se trata de uno, como es este el caso,

que requiere de un abordaje urgente y que atañe incluso al modo en el que el previsiblemente se seguirá produciendo el conocimiento en el futuro inmediato. Se ha hablado incluso del “fin de la teoría” porque, según parece, el planteamiento tradicional de la ciencia sobre la base de la elaboración de una hipótesis, modelo y prueba podría estar volviéndose obsoleto.

Sea como fuere, dicha evolución, cada vez de forma más clara, no podrá manejarse adecuadamente sin asumir en serio la interdisciplinariedad. No es el momento de caer en reduccionismos, sino de integrar las diferentes miradas y aproximaciones a los grandes problemas, aunando el talento que proviene tanto de las ciencias empíricas y de las “ciencias de la vida” como de las tradicionalmente llamadas “ciencias del espíritu” (*Geistenwissenschaften*), con sus diferentes métodos y objetos de estudio. El programa de estas jornadas es, sin duda, un buen ejemplo del camino que hay que recorrer.

Y nadie mejor que la Dr. ^a Nuria Oliver, cofundadora y directora del *Institute of Humanity-centric AI* de Alicante y cofundadora y vicepresidenta de ELLIS (*The European Laboratory for Learning and Intelligent Systems*), para introducirnos en el fascinante mundo de la inteligencia artificial, sus posibilidades y límites. Su ciertamente extraordinario *curriculum* la convierten en un

exponente internacional de primer nivel en este campo y nos hace sentir muy orgullosos por poder contar en nuestro país con una científica de tanto prestigio.

Por solo mencionar algunos de sus muchos méritos y reconocimientos, su trabajo en el modelado computacional del comportamiento humano utilizando técnicas de Inteligencia Artificial, interacción persona-ordenador, informática móvil y análisis de Big Data - especialmente para el Bien Social es bien conocido, con más de 170 publicaciones científicas que han recibido más de 24000 citas y 11 nominaciones y premios al mejor artículo. Según Research.com, la Dra. Oliver es la investigadora española en IA con mayor impacto (índice h de 71). Es coinventora de más de 40 patentes y ponente habitual en conferencias internacionales.

El trabajo de la Dra. Oliver ha sido reconocido internacionalmente con numerosos premios. Es la primera científica española en recibir el premio MIT TR100 (hoy TR35) Young Innovator Award (2004) y el premio Rising Talent del Foro de Mujeres para la Economía y la Sociedad (2009). Ha sido galardonada como Científica de Datos del Año en Europa (2019), Premio Ingeniera del Año por el COIT (2018), Medalla al Mérito Empresarial y Social de la Generalitat Valenciana (2018), Premio Mujer Digital Europea del Año (2016), Premio Nacional de Informática Ángela Robles (2016),

Premio Abie al Liderazgo Tecnológico (2021) y Premio Rey Jaime I de Nuevas Tecnologías (2021). Ha sido nombrada ChangeMaker por Forbes magazine (2023) y una de las 11 personas más influyentes en Inteligencia Artificial a nivel mundial por Pioneering Minds (2017).

Pero, más allá de todos estos datos, lo más importante tal vez sea destacar que su trabajo y esfuerzo se ponen al servicio de los demás y se dirigen a mejorar la calidad de vida de las personas tanto a nivel individual como colectivo. Es un buen exponente, por tanto, de aquello a lo que en buena lógica debería servir toda actividad científica.

HACIA UNA INTELIGENCIA ARTIFICIAL SOCIALMENTE SOSTENIBLE (*)

Nuria Oliver Ramírez

Fundación ELLIS (The European Laboratory for Learning and Intelligent Systems), Alicante

Para mí es un placer estar aquí para hablar de uno de los temas que más me apasiona y al que he dedicado mi vida profesional: ¿qué es la inteligencia artificial?

Antes de hablar de Inteligencia Artificial, resulta fundamental entender el contexto en el que estamos y los grandes retos a los que nos enfrentamos, porque uno de mis primeros mensajes es que necesitamos inteligencia artificial para poder abordar los grandes retos del siglo XXI. ¿Cuáles son estos retos?

El primero es el *envejecimiento* de la población: Europa es un continente envejecido; España es uno de los países más envejecidos del mundo: tenemos una de las tasas de natalidad más bajas del planeta.

(*) Transcrito en papel por: Miguel Angel Medina, Catedrático de Bioquímica y Biología Molecular, Universidad de Málaga. Humboldtiano.

Este envejecimiento genera no solamente un reto demográfico, sino también un reto económico, un reto médico, un reto social, que tenemos que abordar. Ya hay muchos sitios con más personas mayores de 65 años que niños menores de 5 años y se estima que no en mucho tiempo la mayoría de las personas no tendrá ningún familiar vivo durante parte de su existencia, lo cual es algo que nunca había pasado en la historia de la humanidad.

En segundo lugar, vivimos en un planeta con **recursos limitados**. Si continuamos consumiendo al ritmo al que lo estamos haciendo, pronto vamos a necesitar dos planetas Tierra, de los que obviamente no disponemos. Y, además, estos recursos no están distribuidos equitativamente en la sociedad. De hecho, en el siglo XXI estamos siendo testigos de una polarización en la acumulación de la riqueza. Por ejemplo, las 100 personas más ricas tienen más que los cuatro mil millones de personas más pobres. Esta polarización en la acumulación de la riqueza se piensa que es debida en parte al desarrollo tecnológico y a la inteligencia artificial.

Además, vivimos en un mundo con una gran **crisis energética**. Lo estamos experimentando en nuestros hogares, en nuestras instituciones. La demanda energética está creciendo, pero - además- la producción energética con energías

renovables no está teniendo el mismo ritmo que el de la demanda. Esto, entre otras cosas, a lo que lleva es al reto probablemente más relevante en nuestros tiempos: el **reto climático**. Sabemos que las temperaturas medias del planeta están aumentando. También están aumentando los niveles de los mares y los océanos y la frecuencia y la intensidad de los desastres naturales.

Y, cómo no, vivimos en un mundo susceptible a las **pandemias**. Estamos saliendo de una pandemia que ha paralizado el planeta durante 3 años.

Para todos estos retos vamos a necesitar la inteligencia artificial; esto es indiscutible. Desde un punto de vista tecnológico, vivimos un momento de grandísima oportunidad. En el contexto de la inteligencia artificial, cada día tenemos alguna gran noticia. Vivimos en un mundo en el que desde hace varios años hay más del doble de móviles que de humanos. El móvil es el dispositivo con mayor adopción en nuestra historia y modula nuestra experiencia vital porque siempre está con nosotros. Hasta el punto de que vamos a ver un concierto y no miramos el concierto: lo estamos grabando; o vamos a escuchar y a conocer a la primera y única candidata a la presidencia los Estados Unidos y no le miramos a la cara, sino que nos estamos haciendo un *selfie* con ella. Esto es un fenómeno que no es propio solamente de países desarrollados, es un fenómeno que se da

especialmente en países en desarrollo. Hay más personas en el mundo que tienen acceso a un móvil que personas que tienen acceso a agua potable o a retretes, por ejemplo.

Más allá de los móviles, vivimos en un mundo donde cada vez más vamos a convivir con coches autónomos, con robots sociales, donde en nuestro día a día interaccionamos con innumerables servicios de inteligencia artificial en nuestros móviles; un mundo donde cualquiera de los que estamos aquí vamos a poder viajar al espacio sin tener que ser astronautas y donde le hablamos a la tecnología, algo que formaba parte de la ciencia ficción hace 10 o 15 años: tenemos altavoces inteligentes, tenemos teléfonos inteligentes, coches a los que les hablamos y nos escuchan y nos entienden, sin teclado, sin pantalla. Esto es algo que era impensable hace 20 años y que lo tenemos gracias a la inteligencia artificial.

Vivimos en un mundo donde tenemos ciudades inteligentes, impresoras en tres dimensiones, donde no vamos a poder avanzar en las ciencias sin la ayuda de la inteligencia artificial y donde nosotros, el *Homo sapiens*, los humanos, también nos estamos aumentando con tecnología; ahora con tecnología que llevamos puesta, pero pronto con tecnología que llevaremos en nuestros propios cuerpos. Vivimos en un mundo donde de los 5 países más poblados 3 son países tales que no existían hace 20 años: *Facebook*, *Whatsapp* e *Instagram*;

países que, además, están gobernados por el mismo presidente no elegido democráticamente. Vivimos en un mundo donde no solamente los humanos estamos conectados sino también los objetos de nuestro día a día, el entorno, el planeta, el mundo físico está conectado y tiene incluso inteligencia: lo que se conoce como el internet de las cosas. Mientras en el año 2003 había solamente 0,08 dispositivos conectados por humano del planeta, en el año 2020 superamos los 6 dispositivos y medio por habitante, es decir, ha habido un crecimiento exponencial. Esta ubicuidad de la tecnología, de los móviles, del internet de las cosas, a lo que da lugar es a una generación de cantidades de datos ingentes. Es lo que se conoce como *Big Data*, datos a los que por su volumen, pero también por su naturaleza (porque son datos no estructurados, es decir, son vídeos, audios, imágenes, texto, datos de sensores), no podemos dar sentido y no podemos entender si no es con la ayuda de la inteligencia artificial. Esto también es muy importante entenderlo.

Vivimos también en un mundo donde no hay fronteras (no solamente digitalmente, sino también físicamente), tenemos más movilidad que nunca antes en nuestra historia y esta movilidad representa una vulnerabilidad, como hemos visto en el contexto de la pandemia, de las enfermedades infecciosas. Si el coronavirus, que apareció en China, se extendió tan rápidamente por todo el planeta

fue fruto de esta gran movilidad que tenemos.

En suma, nos encontramos inmersos en una revolución industrial que se conoce como la **cuarta revolución industrial**. Esta revolución industrial, igual que las revoluciones anteriores, está transformando profundamente todos los ámbitos de la sociedad. Representa una unión íntima, sin precedentes, entre el mundo físico, el mundo biológico y el mundo digital y está impulsada por disciplinas como la biotecnología, la nanotecnología, la ingeniería genética y la inteligencia artificial, que es mi especialidad.

Pero ¿qué es la inteligencia artificial? Yo creo que todos han oído hablar de ella, aunque pocos sepan exactamente lo que es. La **inteligencia artificial** es una disciplina que existe desde los años cincuenta y que tiene como objetivo la creación de sistemas computacionales, es decir, sistemas no biológicos que sean inteligentes tomando como referencia la inteligencia humana. La definición de la inteligencia artificial siempre ha venido acompañada de la definición de la inteligencia humana. Entonces, la primera pregunta que nos hacemos es: ¿qué es la inteligencia humana? Porque, si queremos emular la inteligencia humana, a la primera conclusión a la que llegamos es que no entendemos la inteligencia humana en su completitud y además es muy compleja. Efectivamente, los humanos tenemos muchas habilidades:

podemos percibir el entorno en el que nos encontramos, reconocer lo que estamos percibiendo y reaccionar en consecuencia, podemos resolver problemas, nos adaptamos (de hecho, si no lo hiciéramos, podríamos desaparecer como especie), aprendemos constantemente, tenemos múltiples inteligencias (emocional, social, musical, lógico-verbal, creatividad, etcétera). De la misma manera, dentro de la inteligencia artificial hay muchas áreas de conocimiento que intentan emular algunas de estas habilidades humanas; por ejemplo, el área dentro de la inteligencia artificial que intenta emular la capacidad que tenemos para ver sería la visión por ordenador, que es el área que nos permite entender imágenes, el área de procesamiento de imágenes, de reconocimiento de objetos, de reconocimiento de caras, todo a través del ordenador. El área de la inteligencia artificial que intenta emular la capacidad que tenemos los humanos para el lenguaje, para hablar y para comunicarnos sería por una parte el reconocimiento del habla y por otra el procesamiento del lenguaje natural.

Desde el punto de vista del nivel de competencia, los sistemas de inteligencia artificial se dividen en tres niveles. El primer nivel es el que tenemos hoy en día y se conoce como ***inteligencia artificial específica***. Son sistemas que son muy buenos haciendo una tarea, probablemente mejor que el mejor de los

humanos para esa tarea; por ejemplo, hace mucho tiempo que el mejor jugador de ajedrez no es humano, es un programa de ordenador; pero este programa solo sabe jugar al ajedrez y, de hecho, no sabe realmente lo que es el ajedrez y podría tener dificultades si cambiásemos las reglas de manera imprevista. El segundo nivel es la aspiración de la inteligencia artificial y se conoce como ***inteligencia artificial general***. Sería una inteligencia artificial tan competente como la de un humano, incluyendo conceptos tan difíciles como la autoconciencia. Yo creo que estamos muy lejos de tener inteligencia artificial general. Pero lo más importante, desde mi punto de vista, es entender que no necesitamos inteligencia artificial general para que la inteligencia artificial esté aquí ya, forme parte de nuestro día a día, represente un mercado de miles de millones de dólares a nivel mundial, y esté transformando la sociedad. Creo que este mensaje es importante porque muchas veces en los medios de comunicación se centra el foco en si alcanzaremos o no inteligencia artificial general algún día, como si solo cuando alcanzásemos inteligencia artificial general es cuando la inteligencia artificial debería preocuparnos o debería tener impacto. Y esto no es así. Independientemente de si algún día tendremos una inteligencia artificial general, la inteligencia artificial hoy está transformando el mundo y nos está impactando a todos y a todas.

Finalmente, hay autores que dicen: ¿por qué nos vamos a detener con la inteligencia humana? Si la capacidad de computación continúa creciendo exponencialmente, podríamos superar la inteligencia humana dando lugar al concepto de “**super inteligencia**”, que proponen autores como Nick Bostrom, de la Universidad de Oxford. Esta súper inteligencia es superior a la humana, por definición no la entendemos de la misma manera que una hormiga no entiende nuestra inteligencia y el momento en el tiempo en el que supuestamente superaríamos computacionalmente la inteligencia humana se conoce como la **singularidad**, de la que quizás habéis oído hablar o habéis leído en libros de ciencia ficción.

Históricamente ha habido dos grandes escuelas para conseguir inteligencia artificial y creo que es importante conocer esta perspectiva histórica para poder entender el momento en el que vivimos. La primera escuela es la escuela **top-down** o la escuela simbólico-lógica o la escuela clásica. Esta escuela parte de la siguiente idea: los humanos sabemos muchas cosas, tenemos mucho conocimiento; si programamos todo ese conocimiento en un ordenador usando símbolos y usamos reglas de la lógica para derivar conocimiento nuevo a partir de ese conocimiento base podríamos alcanzar la inteligencia artificial. Esta escuela representó, de hecho, la primera comercialización exitosa de la inteligencia

artificial desde mediados de los setenta hasta mediados de los noventa mediante los llamados sistemas expertos, de los que quizás habéis oído hablar y fueron el primer ejemplo de uso práctico de la inteligencia artificial. La segunda escuela, que es la escuela la que yo pertenezco, es la escuela **bottom-up** o la escuela conexionista. Esta escuela parte de otra observación distinta, que es la siguiente: los seres biológicos inteligentes aprendemos a partir de la experiencia, aprendemos a partir de nuestras interacciones con el entorno, aprendemos a partir de datos, de manera que si queremos crear inteligencia artificial deberíamos inventar algoritmos que aprendan a partir de datos. El ejemplo canónico de esta escuela son las redes neuronales de las que seguramente habéis oído hablar y que, de hecho, forman parte de la revolución que estamos viviendo en inteligencia artificial hoy en día. Pero, aunque hablemos ahora mucho de redes neuronales, quería compartir este artículo del *New York Times* de 1958, donde se habla de un dispositivo que inventó la armada americana que aprende haciendo. A lo que hace referencia el artículo es a la red neuronal más sencilla que hay, que es un **perceptrón**, que la propuso Frank Rosenblatt, quien ya en los años 50 decía “aprende como un humano; al principio se equivoca, pero conforme va teniendo más datos va mejorando”, e incluso se aventuró a predecir que en un futuro habría redes como esta que reconocerían las caras o

que traducirían de un idioma a otro. Efectivamente, los sistemas de hoy en día que reconocen caras o que traducen son, podríamos decir, redes neuronales en esteroides, mucho más complejas que esta, pero tienen el principio muy similar al del perceptrón.

Dentro de cada escuela, hay muchas áreas de conocimiento. Por ejemplo, en la escuela *top-down* tenemos la teoría de juegos, la lógica, la teoría del aprendizaje, la representación del conocimiento y en la escuela *bottom-up* tenemos todas las áreas relacionadas con la percepción computacional, el *machine learning* o aprendizaje automático; dentro del *machine learning* está el *deep learning* o aprendizaje profundo, todas las áreas relacionadas con la inteligencia emocional y social, etcétera. Aunque históricamente había un cierto enfrentamiento entre estas dos escuelas, en realidad cualquier sistema comercial de hoy en día tiene un componente muy grande de aprendizaje a partir de datos, pero también tiene un componente *top-down* de conocimiento del problema concreto para ayudar a que la interpretación y el análisis de estos datos tenga más sentido.

Hay muchos autores que equiparan el papel que jugó la electricidad en la segunda revolución industrial con el papel que hoy está jugando la inteligencia artificial en la cuarta revolución industrial porque la inteligencia

artificial tiene una serie de propiedades que son similares a las propiedades que tuvo la electricidad o que tiene la electricidad. En primer lugar, es una disciplina **transversal**, muchas veces se le llama una disciplina de propósito general. Eso quiere decir que la inteligencia artificial se puede aplicar a cualquier ámbito de la sociedad, a cualquier campo, sea la educación, la medicina, la meteorología, el transporte... Todas las áreas son susceptibles de ser impactadas por la inteligencia artificial. Además, la inteligencia artificial es **compleja**, es **escalable**, pero también es **actualizable** porque en realidad son programas de ordenador. Y tiene dos propiedades muy potentes: la primera es que no solamente permite interpretar el pasado o explicar el presente, sino que nos permite *predecir* el futuro; y, en segundo lugar (y esto es lo que estamos viviendo hoy en día), lo que más popularidad le está dando a la inteligencia artificial: no solamente puede interpretar datos o predecir datos, sino que puede **generar datos**, general textos, imágenes, vídeos, audios o datos de cualquier otra naturaleza. Esto es una propiedad extremadamente potente, pero también plantea muchos retos éticos.

Si la disciplina de la inteligencia artificial existe desde los años 50, ¿por qué se está produciendo hoy la cuarta revolución industrial? ¿Por qué no se produjo en los años 50, si la disciplina existe desde hace tanto tiempo? La contestación a estas preguntas es

que en los últimos 15 años se ha producido la confluencia de tres factores que han impulsado el desarrollo exponencial de las técnicas de inteligencia artificial *bottom-up*, las técnicas que se basan en el aprendizaje a partir de datos. ¿Cuáles son estos factores? El primero, obviamente, es el crecimiento exponencial en la disponibilidad de **datos**. Lo he dicho al principio: vivimos en un mundo de datos, estamos generando datos constantemente hasta el punto de que cada año o cada dos años generamos más datos que los 5000 años anteriores de la historia de la humanidad. Estos datos tienen un gran valor económico y solamente en Europa se estima que pronto superarán los mil billones de euros. El segundo factor es la disponibilidad de grandes **capacidades de computación** a bajo coste, gracias a la llamada Ley de Moore. Esta es una ley de crecimiento exponencial en la capacidad de computación por el mismo precio. Lo que vemos es que la capacidad de computación que se puede adquirir por un precio de referencia ha ido creciendo exponencialmente. Esta ley es la que hace que cuando nos compramos un móvil hoy, el mismo móvil el año que viene vale mucho más barato o por el mismo precio nos podemos comprar uno mucho mejor. A lo largo de la historia los sistemas de inteligencia artificial han ido aprovechando la computación del momento, desde los sistemas expertos de los años 70 a *DeepBlue* (que fue el sistema de IBM que ganó en 1997 al mejor jugador de

ajedrez del mundo en este momento, Gary Kasparov) o Watson (que fue el sistema que ganó en 2011 a los mejores jugadores de *Jeopardy* del mundo) o los sistemas actuales de aprendizaje profundo que tienen muchísima más capacidad de computación. De hecho, en el contexto de la inteligencia artificial se han desarrollado chips específicos para entrenar redes neuronales como las FPGA o las ASIC; la ASIC más popular es la TPU de *Google*, de la que igual habéis oído hablar. Tenemos datos, tenemos computación, y el tercer factor es el desarrollo de los **modelos de aprendizaje**. En realidad, los modelos que dominan hoy en día son modelos que están basados en la red neuronal de los años 50, el perceptrón de Rosenblatt, pero muchísimo más complejas; tan complejas como que pueden tener miles de millones de parámetros, cientos de capas, pero el principio básico matemático es el mismo. Estas redes son las que, entrenadas con cantidades ingentes de datos, gracias a la existencia de grandes capacidades de computación, están revolucionando la sociedad. Pero hay un cuarto ingrediente que también es muy importante, que es que se está produciendo una democratización en el acceso a las herramientas de inteligencia artificial. Hay muchísimas librerías hoy en día y muchísimas herramientas que generan un nivel de abstracción bastante alto, de manera que cualquier persona puede usar la inteligencia artificial. Algo que era totalmente inviable

cuando yo estaba haciendo el doctorado, cuando teníamos que programar las ecuaciones matemáticas para hacer que aprendiese algo el ordenador. Hoy en día es bastante fácil porque no tienes que programar ninguna ecuación matemática para conseguir que tu ordenador entienda o analice unos datos, o prediga.

Como he dicho, son estas redes neuronales profundas las que están detrás de esta gran revolución, de que podamos hablar de la tecnología, de que podamos traducir de unos lenguajes a otros, de que podamos reconocer caras, etcétera. Por estos avances en los últimos años hemos conseguido hitos que parecían inalcanzables. Como, por ejemplo, que en el año 2016 *AlphaGo* ganarse en el juego del go contra uno de los mejores jugadores de go del mundo. Con el juego de ajedrez, ya lo he dicho, se consiguió en el 1997 con inteligencia artificial, pero el juego del go se pensaba que iba a ser más difícil de conseguir porque es un juego donde hay muchas más posibilidades de movimiento y por tanto era mucho más complejo; sin embargo, en el 2016 con técnicas de aprendizaje profundo de redes neuronales profundas conseguimos ganar a uno de los mejores de los humanos. En 2017, *AI Libratus* ganó, de paliza además, contra algunos de los mejores jugadores de póker del mundo. El póker es un juego de información incompleta; a diferencia del ajedrez o del go, en el póker yo no sé qué cartas tienen mis oponentes, y

además mis oponentes pueden no decir la verdad, se pueden echar faroles, algo que se pensaba que iba a ser muy difícil para la inteligencia artificial y sin embargo en 2017 *AI Libratus* ganó apabullantemente contra los jugadores de póker humanos. En 2018, *AlphaZero* ganó contra el mejor jugador de ajedrez del mundo (que es otro programa de ordenador) pero habiendo aprendido a jugar al ajedrez por sí mismo. A diferencia de estos otros programas que habían sido entrenados por humanos, *AlphaZero* aprendió a jugar jugando contra sí mismo millones de veces. O, más recientemente, los avances que *AlphaFold* ha conseguido en la inferencia de la estructura en tres dimensiones de las proteínas, un problema que se consideraba no resolubles computacionalmente si no hubiese ordenadores cuánticos y que lo hemos conseguido con técnicas de aprendizaje profundo.

Otra revolución que se está produciendo es el uso de estas técnicas de aprendizaje profundo en el contexto de los que se conoce como modelos generativos de inteligencia artificial, es decir, modelos que generan contenido, no solamente que interpretan, sino que generan. El más popular es *ChatGPT*. Supongo que todos aquí conocéis *ChatGPT* o habéis oído hablar al menos de *ChatGPT* y lo habéis utilizado. No podía dar hoy esta charla sin pedirle a *ChatGPT* que me escribiese un poema sobre la *Asociación Alexander von Humboldt de España*

para ver lo que me decía y esto es lo que me ha dicho [recita el poema].

Lo más importante no es el poema, que no sé si tiene alta calidad poética o no; es que este poema lo ha generado en un segundo y podemos generar un millón de poemas simplemente apretando el botón y ahora mismo hay cientos de millones de personas que están utilizando *ChatGPT*. El nivel de competencia que estos sistemas están alcanzando no ha tenido precedentes en nuestra historia y por primera vez en la historia de la humanidad un ente no humano tiene la capacidad de generar contenido que solo antes los humanos éramos capaces de generar. De manera que, a partir de ahora, lo que leamos, lo que veamos, lo que escuchemos no va a estar generado por humanos con alta probabilidad sin que nosotros lo sepamos. Eso es algo que representa un reto social increíble.

Estas técnicas generativas también están revolucionando el contexto de la imagen y también de la música. Ninguna de las imágenes que aquí muestro ha sido generada por ningún humano, son generadas por algoritmos en cuestión de segundos. El impacto que esto tiene en los artistas es enorme porque estos algoritmos han sido entrenados con todos los textos que han sido escritos y que están disponibles en Internet. En el caso de imágenes, han sido entrenados con cientos o millones de millones de imágenes que están

disponibles, sin pedirle permiso a los autores de imágenes o a los autores de esos textos si permiten que se utilicen sus datos para entrenar estos sistemas. Hay un movimiento mundial de artistas contemporáneos que no quieren que sus obras se usen para entrenar estos sistemas porque tú le puedes pedir que te haga una imagen en el estilo de quien tú quieras; si es un autor contemporáneo, genera imágenes que parece que las ha hecho esa persona, pero la persona no las ha hecho. El impacto que puede tener evidentemente en la capacidad que tengan los artistas para vender sus obras, si se pueden hacer automática y gratuitamente, es muy grande.

Más allá de estos ejemplos, en nuestro día a día interaccionamos con sistemas de inteligencia artificial. Son algoritmos de inteligencia artificial los que deciden qué amigos tenemos, qué actualizaciones vemos de nuestros amigos, qué noticias leemos, qué información encontramos en Internet, qué libros compramos, qué películas vemos, qué música escuchamos. Algoritmos de inteligencia artificial permiten que le hablemos a nuestros móviles, son nuestros altavoces. También deciden por dónde vamos: cada vez más los programas de mapas utilizan inteligencia artificial para tener en cuenta el tráfico, las condiciones meteorológicas o para también servir a los intereses de quienes hagan el programa para a lo mejor hacernos ir por caminos dónde hay ciertos comercios que les

han pagado por la publicidad. Hace mucho tiempo que las transacciones en bolsa y los procesos de manufactura están automatizados. Como he dicho, no podremos tener coches autónomos sin inteligencia artificial y tampoco podremos tener avances en las ciencias sin la inteligencia artificial.

También cada vez más son algoritmos de inteligencia artificial los que van a decidir si nos dan un trabajo o no, van a decidir si nos aceptan en una universidad o no, van a decidir qué tratamiento médico se nos aplica, qué diagnóstico se nos determina, o qué sentencia judicial se nos dictamina y estos son ejemplos reales. Por eso, la inteligencia artificial ha dejado de ser una disciplina científico-tecnológica y ha pasado a ser un tema político como ilustra esta transparencia que muestra el número de veces que el término inteligencia artificial en azul o el término *machine learning* o aprendizaje automático en rojo ha sido mencionado en el Parlamento británico y en el Congreso americano. Como podéis ver, es como si no existiese hasta el año 2016, y a partir de ese año crece exponencialmente el interés, hasta el punto de que no es sorprendente que haya más de una cincuentena de países del mundo que hayan diseñado y publicado sus estrategias de inteligencia artificial. Aquí en Europa tenemos la estrategia europea que se publicó hace un par de años y que tiene como elemento angular lo que se conoce como una inteligencia artificial confiable, es decir, una

inteligencia artificial que sea por y para las personas. Más acá del contexto europeo, en España hemos tenido un fenómeno un poco curioso porque la estrategia de inteligencia artificial estatal se publicó en diciembre del 2020, más tarde que dos comunidades autónomas (Cataluña y la Comunidad Valenciana), que publicaron sus estrategias regionales más de un año antes de que se publicase la estrategia estatal. Esto es interesante porque, de alguna manera, refleja la preocupación a todos los niveles por el impacto que la inteligencia artificial está teniendo en la sociedad. Como quizás hayáis oído, en abril del 21 (durante la pandemia) la Comisión Europea publicó un comunicado donde se proponía una regulación europea a la inteligencia artificial que ahora mismo está en trámite parlamentario y se anticipa que entrará en vigor este año 2023. Una de las acciones que están incluidas en esta regulación europea es la supervisión de que se cumple la regulación, donde se anticipa que habrá un órgano supervisor a nivel europeo, pero también habrá agencias supervisoras en cada país miembro. Aquí en España hubo una convocatoria competitiva para determinar dónde estaría la agencia española de supervisión de inteligencia artificial y finalmente va a estar en Galicia, en La Coruña.

Desde mi perspectiva, cualquier desarrollo, cualquier uso, cualquier investigación en inteligencia artificial debería considerar estas

cinco dimensiones: 1) la tecnológica; 2) la del marco de referencia regulador y legal; 3) la ética; 4) la de los ciudadanos, la sociedad y la educación; y 6) la de la economía, la productividad, el trabajo y el mercado. No tengo tiempo de explicarlas todas en detalle, pero quería simplemente dar unas pinceladas de un par de ellas. Desde el punto de vista económico es indiscutible, no hay ningún analista que no prediga que la inteligencia artificial va a tener un impacto económico inmenso y va a generar riqueza. Otra cosa es que esa riqueza esté distribuida de manera homogénea o no, pero riqueza está generando y va a generar. En este contexto, como europeos, uno de los grandes retos que tenemos es que Europa no lidera la inteligencia artificial, es que esta cuarta revolución industrial de la que he estado hablando no está siendo liderada por Europa, es que ninguno de los ejemplos que he dado de ecosistemas de inteligencia artificial es europeo. Hay un gran liderazgo ahora mismo en inteligencia artificial por parte de Norteamérica y por parte de China, y Europa se ha quedado muy relegada y esto es altamente no deseable como europeos porque estamos hablando de una disciplina absolutamente estratégica, una disciplina que está transformando la sociedad y esa transformación no la estamos diseñando, ni decidiendo los europeos, la estamos usando, pero no la estamos inventando.

Para intentar combatir esta situación los

científicos europeos nos movilizamos en 2018 y creamos **ELLIS**, que es el acrónimo de *European Lab for Learning & Intelligent Systems* y que tiene como objetivo contribuir a la soberanía tecnológica europea en investigación en inteligencia artificial para evitar la fuga de talento. Se está produciendo y se va produciendo durante bastantes años una masiva fuga de talento investigador excelente europeo fuera de Europa y yo creo que la gota que colmó el vaso fue cuando en 2018 los ganadores del Premio Turing (que es el equivalente al Premio Nobel, no hay premio Nobel en informática) fueron Yann LeCun, Geoffrey Hinton y Yoshua Bengio, tres científicos de origen europeo, ninguno de los cuales está en Europa ahora mismo y que son los padres del aprendizaje profundo y de las redes neuronales profundas; es decir, la revolución de la inteligencia artificial sí tiene un origen europeo y sus inventores son europeos, pero ninguno está en Europa. Los tres están en Norteamérica. Este hecho, como europeos, debería entristecernos. Con ELLIS lo que queremos es conseguir atraer y retener las mejores mentes y también crear una red, porque entendemos que diferentes regiones van a hacer esfuerzos para atraer y retener estas mejores mentes humanas y crear una red que tenga la masa crítica suficiente como para ser relevantes a nivel mundial. En ELLIS estamos haciendo muchas acciones. Yo soy cofundadora y vicepresidenta de ELLIS. Quizás

una de las iniciativas más interesantes es que estamos creando unidades ELLIS, que son grupos excelentes de investigación que se comprometen a crear un entorno que sea lo suficientemente atractivo para atraer y retener el talento y que también se comprometen a contribuir con la misión y la visión de ELLIS. Ahora mismo hay 39 unidades ELLIS en 14 países de Europa e Israel.

Una de estas unidades es la unidad ELLIS de Alicante, que creamos hace un par de años. Esta es una unidad especial porque es la única unidad que ha sido creada desde cero. Es una fundación de investigación en inteligencia artificial y además es especial porque tiene una temática clara, que es inteligencia artificial ética, responsable y para el bien social. Es la única unidad de toda la red que está dedicada a esta temática.

Desde punto de vista laboral, cualquier revolución industrial ha conllevado una profunda transformación del mercado laboral. Eso siempre ha sucedido y está sucediendo con esta revolución. Hay muchos trabajos que existían cuando mi abuela era pequeña que hoy no existen; por ejemplo: telefonista, sereno, lamparero, despertador humano... Todas estas profesiones ya no existen, pero también hay muchas profesiones hoy que no existían hace 20 años. Por ejemplo, todas las relacionadas con *smartphones* porque no había *smartphone*, todas relacionadas con las redes sociales

porque no había redes sociales, todas las relacionadas con la nube porque no había nube, o incluso uno de los trabajos más recientes ahora y que tiene mucha demanda y se conoce como *AI prompt engineer*, que son los humanos que le dicen a los sistemas generativos de inteligencia artificial qué es lo que tienen que generar, es decir, personas que le dicen a *ChatGPT* lo que tiene que generar o que le dicen a *DALL-E* qué tipo de imagen tiene que generar porque no es trivial saberle preguntar a la inteligencia artificial para que te genere lo que tú quieres. Este nuevo trabajo no consiste en otra cosa que ser un humano que le está pidiendo cosas a la inteligencia artificial. Sin duda, va a haber una gran generación de puestos de trabajo y, sin duda, estos trabajos van a estar relacionados con la tecnología ampliamente hablando, y específicamente con la inteligencia artificial. Una de las preguntas que deberíamos hacernos como sociedad y como institución es si nos estamos preparando para esta gran transformación. Porque sabemos que va a haber muchos nuevos puestos de trabajo, pero también sabemos que va a haber muchos trabajos que se van a ver desplazados. ¿Qué van a hacer todas esas personas cuyos trabajos se ven desplazados? ¿Estamos invirtiendo en formación a los profesionales? Yo creo que no, y creo que deberíamos.

Otra dimensión que quería destacar, porque es quizás menos conocida, de la inteligencia

artificial es la dimensión social. Está muy relacionada con mi primer mensaje, en que declaré que sin la inteligencia artificial no vamos a poder abordar los grandes retos del siglo XXI. Y es porque la inteligencia artificial tiene una vertiente y una potencialidad inmensa para tener impacto social positivo. En este contexto, a nivel mundial quizás una de las opciones más claras es con relación a los objetivos de desarrollo sostenible. Cuando en 2015 los Objetivos del Milenio llegaban a su fin, el Secretario General de Naciones Unidas hizo un llamamiento a lo que llamaron una *revolución de los datos*. Se dieron cuenta de que los datos y la inteligencia artificial podrían jugar un papel muy importante para tener impacto social positivo. De hecho, encargaron un informe que se llama *Data Revolution Report* a un conjunto de expertos incluyendo el director de mi tesis, donde se esbozaba una visión de cómo la inteligencia artificial y los datos nos podrían ayudar en el contexto de esos nuevos objetivos que se estaban definiendo y que fueron los 17 objetivos de desarrollo sostenible. De manera que hay un movimiento mundial, al que pertenecemos desde ELLIS Alicante, de cómo los datos y la inteligencia artificial nos pueden ayudar a conseguir los 17 objetivos de desarrollo sostenible, cómo nos pueden ayudar a medir, cómo nos pueden ayudar a acelerar su consecución, cómo nos pueden ayudar a mitigar barreras que nos estén impidiendo

conseguirlos, etcétera. Si os interesa, en la página web de ELLIS Alicante hay un artículo sobre este tema.

Algunas áreas claras de impacto en las que he trabajado personalmente son el tema de la **gestión de crisis**, relacionadas con desastres naturales o crisis humanitarias, para poder de manera muy rápida determinar cuántas personas se han visto afectadas, dónde están las personas, si hay migraciones como consecuencia de las crisis, incluyendo migraciones como consecuencia del del cambio climático. Otra área de mucho potencial es la **inclusión financiera** y el desarrollo económico. Lo he dicho al principio, no podemos tener **ciudades inteligentes**, ni **transporte inteligente** sin inteligencia artificial, con lo cual es absolutamente necesaria en estos contextos. Un área en la que ha habido mucho trabajo las últimas dos décadas y que ha sido muy relevante en los últimos 3 años es el uso de los datos por la inteligencia artificial para **luchar contra enfermedades infecciosas y pandemias** y hablaré un poquito de lo que hicimos en Valencia. No podemos tener tampoco **energía inteligente** (*smart grid*) sin inteligencia artificial. A nivel de los **censos**, probablemente lo sepáis, prácticamente todos los institutos de estadística del mundo están utilizando fuentes alternativas de datos y algoritmos de inteligencia artificial para hacer censos que sean más económicos y frecuentes. Tampoco

podemos tener una ***agricultura de precisión*** y una agricultura inteligente sin inteligencia artificial.

Si os interesa el tema, hay muchas publicaciones que podéis leer y esta es una de las áreas en las que trabajamos en el ELLIS Alicante (Oliver et al. Sci. Adv. DOI: 10.1126.sciadv.abc0764, 2020). Solamente por dar una pincelada, durante la pandemia lideré un equipo de más de una veintena de investigadores e investigadoras de toda la Comunidad Valenciana y trabajamos para suplir está brecha que hay entre dónde están los datos (si entendemos que los datos son una representación digital de una realidad subyacente) y dónde están los tomadores de decisiones. Lo que queremos es que estas decisiones estén basadas en la evidencia, que estén basadas en el conocimiento científico, pero no es obvio cómo pasar de los datos a los tomadores de decisiones. Para ello, lo que hicimos durante la pandemia fue estructurar nuestro trabajo en cuatro áreas, modelamos la movilidad humana, desarrollamos modelos epidemiológicos computacionales de inteligencia artificial que predecían la curva pandémica y el número de casos bajo distintas posibles medidas de confinamiento, modelos predictivos de ocupación hospitalaria, y una gran encuesta ciudadana (que igual alguno de vosotros contestó) que se llama *covid-19 impactsurvey*, que tiene más de 720000 respuestas y es una de las más grandes del

mundo. Este trabajo ha sido pionero a nivel mundial, pudimos ayudar para tomar mejores decisiones para contribuir a salvar vidas, y también nuestros modelos resultaron los ganadores mundiales de una competición que se llama *X Prize Pandemic Response Challenge*, siendo la primera vez que un equipo español ganaba una competición de la Fundación X Prize.

Para terminar, quería dar unas pinceladas sobre los retos éticos. Por una parte, sabemos que necesitamos inteligencia artificial, pero por otra parte no está garantizado en absoluto que el impacto de la inteligencia artificial vaya a ser positivo. Tenemos que abordar las limitaciones que tienen hoy y los retos que nos presenta y estos son algunos de los más importantes. En primer lugar, la violación computacional de la ***privacidad***. Lo he mencionado un poquito en el contexto de los modelos generativos y cómo se utilizan datos sin pedir permiso. Pero con modelos de inteligencia artificial podemos inferir atributos muy personales de las personas: su orientación sexual, su estado civil, su religión, obviamente su sexo, etcétera, a partir de datos no personales. Por ejemplo, solamente mirando el patrón de lo que le des *like* o no le des *like* en una red social se puede inferir mucha información sobre una persona. Esto es algo que tenemos que controlar. En segundo lugar, tenemos el reto de la ***discriminación algorítmica*** y los sesgos. He comentado que

cada vez más algoritmos van a decidir si nos dan un trabajo, si nos determinan un tratamiento para una enfermedad, etcétera. La pregunta quizás es: ¿y por qué usar algoritmos si los humanos llevamos tomando estas decisiones desde que existimos? La respuesta es porque los humanos no somos perfectos, los humanos tenemos unos 200 sesgos cognitivos, los humanos tenemos amigos y enemigos, somos susceptibles a la corrupción, tenemos hambre, tenemos sueño, nos cansamos. Es obvio que nuestras decisiones no son perfectas. La idea y el ideal es decir: bueno, si los datos son un reflejo de la realidad y los algoritmos se han entrenado con esos datos, dado que los algoritmos no tienen sueño, ni hambre, ni amigos, ni enemigos, ni son susceptibles a la corrupción, en teoría las decisiones que estos algoritmos podrían darnos deberían ser mejores que las decisiones de los humanos. Esa es la aspiración, pero lo que estamos encontrando es que los algoritmos sí que tienen sesgos, que pueden venir de los propios datos porque los datos estén sesgados, porque la sociedad tiene patrones de discriminación, pero también sesgos porque los algoritmos estén mal hechos o porque se utilicen mal. De hecho, hay un área entera, que es la de la justicia algorítmica y es un área en la que estamos trabajando en ELLIS Alicante. Otro gran reto es el reto de la **opacidad**. Estos sistemas son extremadamente complejos. Por daros una cifra, ¿alguien sabe cuántos

parámetros tiene *ChatGPT3*, que es el modelo de mensaje de ChatGPT? Así, a ojo de buen cubero, ¿nadie se atreve a decir cuántos parámetros tiene: 1000, 10000, 1000000? 175 mil millones. No hay quien entienda un modelo con 175 mil millones de parámetros, pero necesitamos entender cómo funciona. Porque si a mi la mejor red neuronal del mundo me dice que tengo cáncer, pero no me dice por qué, pues no sé si me lo voy a creer. Necesitaría poder explicarlo. Entonces, hay un área entera que se conoce como inteligencia artificial explicable y es un área en la que trabajamos en ELLIS Alicante. Otra línea en la que estamos trabajando que es menos conocida es la manipulación subliminal del comportamiento humano. Todas y todos los que estamos aquí, vemos nuestro comportamiento manipulado cada día por algoritmos de inteligencia artificial que intentan captar nuestra atención, que hacen que “cliquemos” aquí o allá, que quieren que sigamos utilizando la aplicación para monetizar nuestros datos y nuestro tiempo. Entonces, tenemos que crear protecciones ante esa manipulación, que sólo para aumentar. Porque imaginad el poder que tiene un *chatbot* como *ChatGPT*, con el que hablamos y nos contesta como un humano; imaginad el poder que puede tener para manipularnos. Otro gran reto es el de la **falta de veracidad**: como os he dicho, con los modelos generativos ya no sabemos si lo que leemos y lo que vemos, salvo lo que vemos en

el mundo físico delante de nosotros, es real o no, si se corresponde con una realidad subyacente o no. Hay un área entera de cómo usar la inteligencia artificial para poder determinar lo que es verdad y lo que no es verdad. No lo he comentado hasta ahora, pero otro gran reto es que estos modelos tienen una inmensa **huella de carbono**. Entrenar una única red neuronal de este volumen, de este tamaño, como GP3 consume más energía que el consumo energético de muchos países del mundo y esto es algo que no podemos obviar. Usar *ChatGPT*, cada vez que le preguntamos algo, tiene un gran consumo energético, porque contestar (correr la red para contestar) también tiene un gran consumo energético. Hay una línea entera que se conoce como *inteligencia artificial verde*, que es cómo conseguir que estos sistemas no tengan tanto consumo energético, cómo hacerlos más pequeños y más eficientes. Otro gran reto es la **falta de diversidad**. Falta de diversidad en la disciplina: expertas técnicamente en la disciplina probablemente haya menos de un 12%; pero también falta de diversidad en los algoritmos. Los algoritmos tienden a encasillarnos en ciertos estereotipos. Si vemos películas de terror", ya sólo nos recomiendan películas de terror; si leemos noticias de una cierta orientación política, ya solo vemos noticias de esa orientación política. Esto lo que genera son lo que se conoce como *cámaras de resonancia* y *filtros burbuja*, que se piensa que

están contribuyendo en parte a la polarización de la sociedad. Además, tenemos una situación de **asimetría**, de la que yo he hablado en el contexto de la polarización geográfica con Norteamérica y China en cuanto al liderazgo; pero hay otra gran asimetría, y es que datos muy valiosos para poder tener impacto social positivo no son datos públicos, son datos privados, fundamentalmente de las grandes de las tecnológicas. Entonces, muy poca gente puede realizar este potencial que tiene la inteligencia artificial por esta asimetría. Finalmente, la inteligencia artificial no es perfecta y es **engañable**, es decir, tú puedes completamente engañarla. Si hay un sistema que detecta caras y yo no quiero que me reconozca el sistema, me pongo unas gafas de una manera, o me pongo un puntito o me pongo una camiseta. Hay algo que consigue engañar a la inteligencia artificial y entonces ya soy invisible para ese algoritmo. Esto se conoce como *inteligencia artificial adversarial* y es otra gran línea de investigación, que es cómo engañar a los sistemas de inteligencia artificial. Los sistemas de inteligencia artificial intentan no ser engañados y entonces se establece una especie de carrera. En ELLIS Alicante también trabajamos en estos temas.

Para recordar un poco los principios que como sociedad creo que deberíamos exigir a los sistemas de inteligencia artificial, me gusta utilizar este acrónimo en inglés que se conoce como FATEN y creo que, todos y todas que

estamos aquí, deberíamos demandar como sociedad, debería haber un movimiento social que acompañe al desarrollo tecnológico y que exija el tipo de tecnología que queremos. Yo creo que la tecnología que queremos, que deberíamos querer, debería cumplir estos principios FATEN. En primer lugar, la F es de ***Fear***, es decir, *justa*: deberíamos tener garantías de no discriminación. En segundo lugar, la A es de ***Autonomía***, es decir, debería preservar el principio de la autonomía, debería no manipularnos; la A también es de ***Accountability***, es decir, debería de haber una clara atribución de responsabilidad de las consecuencias del uso de estos sistemas; y la A también es de ***Aumento de la inteligencia***; la visión debería ser la de aumentarnos a nosotros, no de sustituirnos, de cómo la inteligencia artificial me puede a mí realizar mi potencial, no de cómo me puede eliminar, sustituir y no sé qué hacer con mi vida. La T es de ***Trust***, de confianza y de transparencia; lo he comentado en el contexto de la opacidad. La E, y aquí hago un poco de trampa, es del principio de la ***bEneficencia***, es decir, maximizando el impacto positivo con diversidad, con sostenibilidad; y, además, la E es de la necesaria inversión en ***Educación*** y también es la E de ***Equidad***. La N es del principio de ***No maleficencia***, es decir, minimizando el impacto negativo con fiabilidad, seguridad, reproducibilidad y preservando siempre la privacidad.

Concluyo con esta cita de Stephen Hawking, que no dijo mucho, pero a mí me inspiró. Él dijo: *“la inteligencia artificial seguramente va a ser lo mejor o lo peor que le pase a la humanidad”*. Os invito a que os unáis a nuestra causa, que es conseguir que sea lo mejor que nos ha pasado. ¡Muchas gracias!

PRESENTACIÓN DEL PONENTE DANIEL INNENARITY

Elena Martínez García, Catedrática de Derecho Procesal, Universitat de València, Humboldtiana

Todas las personas presentes hemos pasado muchas horas de estudio leyendo a autores que un día tenemos la suerte de conocer; este es mi caso con el profesor Innerarity.

Daniel Innerarity es catedrático de filosofía política, investigador Ikerbasque en la Universidad del País Vasco y Titular de la Cátedra Inteligencia Artificial y Democracia en el Instituto Europeo de Florencia. Ha sido profesor invitado en diversas universidades europeas y americanas. Sus últimos libros son *Una teoría de la democracia compleja* y *La libertad democrática*.

Ha obtenido, entre otros, el premio Euskadi de Ensayo, el Premio Nacional de Ensayo, el Premio Príncipe de Viana de la Cultura y el Premio Nacional de Investigación en Humanidades 2022.

Igualmente, es miembro del Consejo del Institut für Sozialforschung, el centro en el que

nació la llamada Escuela de Frankfurt. Allí conoció y se hizo discípulo de Ulrich Beck y Jürgen Habermas.

Fue becario de la Fundación en los años 80 en Munich.

Además, es colaborador habitual de opinión en El País y otros medios de comunicación.

Hasta aquí un muy breve resumen de sus méritos. Sin embargo, me gustaría señalar que el profesor Innerarity es una rara avis. Por qué digo esto? No es solo por tener un apellido peculiar y muy poco común, como acredita el CIS, tampoco por ser filósofo o por haber sido discípulo de Ulrich Beck y Jürgen Habermas; incluso tampoco por ser becario Humboldt o Premio Nacional.

En mi opinión, lo que le hace tan especial al profesor Innerarity, es el hecho de que propone en sus ensayos conceptos como la autocontención, el diálogo como forma de integrar la diversidad, porque reconoce que la libertad, para ser democrática, debe tener límites desde lo personal y desde el Estado para que funciones la democracia.

Nos recuerda, además, que estamos ante un cambio de paradigma, donde eso de ser de izquierdas o de derechas ya no es único o lo principal, porque en la sociedad de desconocimiento (una de sus obras) el género,

el sexo, la raza, la clase, la globalización y la inteligencia artificial son intersecciones que delimitan lo que pensamos sobre la libertad y la justicia. No es tan sencillo como sentirse de izquierdas o de derechas....

Es singular porque reconoce que entramos en la era de la interdependencia de las personas y la ecoddependencia de la naturaleza. Recuerdo precisamente en tiempo de de Covid escucharte en una conferencia decir que la pandemia no era el fin del mundo, sino el fin de un mundo: aquel en el que nos creímos independientes y autónomos y ajenos a los problemas climáticos y fruto del extractivismo enloquecido. Ahora ya sabemos por la pandemia que estamos más unidos de lo que creemos, que mi libertad deambulatoria pasa por la salud del colectivo, que las personas salvaron a las personas, que dejamos de hacer y el clima respiró.

Los mensajes y experiencias post covid está ya en el aire. Podemos hacer caso omiso o ponernos a trabajar. Porque con la IA la democracia va a ser sometida a fuertes pruebas de resistencia como la vivida con la pandemia. “Los buenos y malos gobernantes se serán más visibles, según preparen la infraestructura del Estado para generar confianza en la democracia o no”, afirma el profesor Innerarity. Y sobre esta confianza en el Estado y sus instituciones, nos corresponde colaborar con cambios internos, en nuestra forma de

vida, consumo y ocio. Nuestro concepto de felicidad.

Y el profesor Daniel Innerarity es rara avis, porque no le gusta jugar al pollito inglés. ¿recordáis el juego? Es un juego infantil donde sabes que en algún momento vas a morir, pero sigues avanzando hacia adelante pensando que no te va a tocar, haces como si no ves.

Daniel Innerarity denuncia la desigualdad, la pobreza, la crisis de las mujeres y los cuidados, la crisis climática como retos planetarios, a los que la IA no puede dar una solución, si las personas que están detrás de este sistema no abren los ojos a todo lo que estamos diciendo y los valores que traen consigo. No jugar a pollito inglés, sino hacer un buen diagnóstico global y compartir retos y soluciones. Esta es su propuesta.

Recientemente viene publicando una serie de libros que son claves para entender las sociedades modernas y los dilemas políticos que enfrentan. Yo diría que es una voz imprescindible en el actual contexto de incertidumbre, descontento político, polarización y falta de concordia en la vida pública.

La Academia tiene un gran reto por delante. Gracias Daniel por ser faro.

Esperamos como Humboldtiano verte por aquí

y seguir desde la filosofía política moviendo
nuestra conciencia.

¿HASTA QUÉ PUNTO ES DEMOCRÁTICA LA DEMOCRACIA DIGITAL?

Daniel Innenarity

Catedrático de Filosofía Política, investigador Ikerbasque en la Universidad del País Vasco y titular de la Cátedra Inteligencia Artificial y Democracia del Instituto Europeo de Florencia.

La transformación digital está suscitando preguntas no menores y a partir de inquietudes de signo opuesto. Por un lado, el temor de que sean los algoritmos los enterradores de la democracia. ¿Estamos entrando en un nuevo totalitarismo de la mano de la ideología de la optimización? ¿Siguen teniendo sentido la información razonada, la decisión propia, el autogobierno democrático en esos nuevos entornos tecnológicos?

Está claro que vivimos hoy en un momento histórico por lo que se refiere a la inteligencia artificial y, particularmente, en cuanto a sus efectos sobre la democracia. No saldremos de esta situación de perplejidad mientras no analicemos con precisión los términos del problema. Lo que aquí me planteo es examinar la identificación y la agregación de preferencias como uno de los principales asuntos en los que se juega ese impacto de la digitalización en la

democracia. Solo así podremos aventurar algo razonable sobre la supervivencia de la democracia en medio de la gobernanza algorítmica.

1.- LA IDENTIFICACIÓN DE LAS PREFERENCIAS EN LA GOBERNANZA ALGORÍTMICA

Si la democracia es un sistema de agregación de preferencias, se requiere previamente indentificarlas bien, para lo que los diversos procedimientos de análisis de datos se presentan como una gran ayuda. Identificación y agregación de preferencias son dos prestaciones diferentes, cada una con sus dificultades propias y que merecen un examen específico.

El mundo imaginado por la razón algorítmica se rige por la promesa de satisfacer nuestras preferencias, una vez que se supone capaz de identificarlas con exactitud, sin ninguna voluntad de prescripción paternalista. Al examinar esta promesa surgen al menos dos interrogantes: si se trata propiamente de nuestras preferencias (¿de *quién* son las preferencias?) y, desde el punto de vista del tiempo, acerca del momento en que se formularon (¿de *cuándo* son las preferencias?). Estas dos preguntas nos conducen a indagar, en primer lugar, acerca de la posible intromisión de otros en su constitución y, seguidamente, sobre un posible reduccionismo

de nuestras preferencias a las del pasado. El sueño de una democracia que obrara de conformidad con lo que queremos, de un sistema de satisfacción de necesidades, aun dando por válido que la democracia no sea más que eso, tendría que demostrar que las preferencias son realmente *nuestras* y que son *todas* las nuestras, es decir, incluidas las preferencias posibles del futuro. Y a este respecto tengo una doble sospecha: que la racionalidad algorítmica suponga de hecho una intromisión indebida y un recorte también injustificado, que en nuestra voluntad política así concebida sean otros los que deciden qué hemos de preferir y que se de por sentado que solo podemos preferir lo que hemos preferido en el pasado. De ser así, habría en nuestras preferencias algorítmicamente identificadas menos subjetividad y menos futuro de lo que debería corresponder al ideal de ciudadanía democrática.

a).-La construcción algorítmica de las preferencias

Las promesas de exactitud algorítmica se han hecho valer contra las inexactitudes de la representación analógica, la granularidad de la sociedad digital contra la generalidad de la democracia representativa. Frente a la idea de que nuestros representantes no nos representan adecuadamente, la lógica de los *big data* se corresponde con una “sociedad de las singularidades” (Reckwitz 2017). La

sociedad no es observada desde categorías en las que encajarían los individuos sino a partir de las huellas que realmente dejan y que los singularizan. La digitalización funciona como la infraestructura que hace posible esta economía de las singularidades y el estilo de vida singular, unas prácticas que obedecen menos a la reducción de la complejidad social que al sentido individual. Una de las funciones de la capacidad computacional es descubrir lo que la gente quiere o necesita sin necesidad de preguntárselo, a partir de lo que dicen en las redes sociales, su movilidad, su consumo o la presión arterial registrada. Hay un montón de tecnologías haciendo esta función, mediante el procesamiento del lenguaje natural, la lingüística computacional, la biométrica, el reconocimiento facial, la geolocalización y otros tipos de análisis que sirven para determinar quiénes somos, qué queremos y cómo nos sentimos. ¿Podríamos haber encontrado aquí el verdadero cumplimiento de la promesa democrática de autodeterminación? Si es cierto que mucha gente no vota cuando se les da la oportunidad de opinar directamente sobre cuestiones políticas, parece entonces razonable defender que el "sentiment analysis" y otros métodos de registrar las preferencias ciudadanas —la "subconscious governance"— serían instrumentos eficaces de lo que puede llamarse "indirect direct democracy" (Leighninger / Moore-Vissing 2018). Si a esto se añade el

convencimiento de que los datos son un lenguaje no distorsionado y que representa objetivamente la realidad, dispondríamos entonces de unas tecnologías con unas potencialidades democráticas indiscutibles, pero igualmente capaces de una gran eficacia de gobierno. Este entorno tecnológico es el que permitiría tanto hacer valer la propia singularidad como el *microtargeting* político que incide sobre ella, la soberanía democrática del usuario y el control estatal particularizado. Nos encontramos así con una de las paradojas de la tecnología de lo singular: que sirve tanto para el respeto a la diferencia específica como para la intromisión y el avasallamiento.

Ahora bien, la primera pregunta que deberíamos hacernos es si se trata propiamente de nuestra singularidad, si esas preferencias son realmente nuestras y si las recomendaciones personalizadas, las necesidades satisfechas, las políticas granularizadas responden a lo que nos interesa o a lo que otros consideran que nos debería interesar. Se podría sintetizar todo esto en dos tipos de problemas: uno de objetividad y otro de artificialidad.

El primero de ellos tiene que ver con lo que nuestras preferencias revelan de nosotros y si eso que revelan es verdaderamente lo que preferimos. Quiero advertir con ello que es cuestionable que nuestras preferencias se revelen en nuestros comportamientos, como

también lo es la suposición de que nuestros comportamientos estén completamente determinados por nuestras preferencias. Un análisis más sutil de nuestras acciones pone de manifiesto que no hay siempre una coherencia entre todo ello, de modo que podemos actuar contra nuestras preferencias, por virtud o por torpeza. Cuando se trata de preferencias todavía no plenamente manifestadas, incipientes o futuras, el problema de la objetividad se agrava todavía más. En tales casos, los algoritmos aplican parámetros que no se basan totalmente en los comportamientos y las preferencias de hecho, sino que generan una simulación para predecir las preferencias futuras, lo cual nos introduce en un espacio mucho menos objetivo de lo que se presume.

El otro conjunto de problemas hace referencia a si las preferencias son identificadas o más bien creadas. ¿Están ofreciéndonos lo que queremos o terminamos queriendo lo que nos ofrecen? Hay una dimensión de construcción de nuestras preferencias por los algoritmos; aunque los algoritmos se presenten como quien meramente identifica las preferencias, pueden estar induciéndolas en una cierta medida. Sabemos bien que el capitalismo es un sistema no solo de satisfacción sino de creación de necesidades. En la medida en que las preferencias son endógenas —configuradas por los sistemas que aseguran únicamente responder a ellas— la satisfacción de

preferencias no puede servir como un criterio normativo independiente. Cuando los algoritmos que se presentan como la respuesta a nuestros deseos están en realidad sintonizados para optimizar (y por lo tanto monetizar) nuestro "compromiso" con la plataforma, entonces que el patrón de consumo resultante es cualquier cosa menos democrático.

De este modo, la publicidad personalizada puede llevarnos a perder buena parte del control sobre nosotros mismos debido a que no podemos centrarnos en lo que más valoramos porque nuestra atención ha sido capturada por las plataformas (Vallor 2015). De todas maneras, cuando se denuncian críticamente este tipo de mecanismos conviene no subrayar en exceso el lado intencional y manipulador porque en ocasiones el condicionamiento es más banal y se desprende de la naturaleza misma de los artefactos tecnológicos. Cuando hablamos concretamente de la democracia, la insistencia en su *hackeabilidad* o injerencia electoral por parte de "malign foreign actors" (Geller 2018) puede hacernos perder de vista que el deterioro habitual de la democracia no procede tanto de una manipulación extranjera como del condicionamiento sutil que ejerce la racionalidad algorítmica sobre la conversación pública y las decisiones colectivas.

Se supone que los algoritmos de recomendación están a nuestro servicio

porque no hacen otra cosa que ayudarnos a identificar las fuentes de satisfacción sin imponérselas, pero en realidad socavan nuestra autonomía de tres maneras: 1. porque están concebidos de manera que captan nuestra atención y desarmen nuestras defensas contra el diluvio de información, 2. porque no se limitan a respetar nuestras preferencias sino que las modifican poniéndonos poco a poco en una dirección que no hemos elegido y con tanta profusión que nos impide reflexionar y 3. porque creen poder identificar nuestras preferencias no teniendo en cuenta más que una observación superficial de nuestro comportamiento.

Este procedimiento está fuera del control de los usuarios; se basa en el comportamiento no reflexivo más que en las preferencias expresadas u objeto de deliberación. Forma parte de esa tendencia a la automatización de la acción política que es diametralmente opuesta a la experiencia de la libertad política (Züger et al. 2017, 275). La gubernamentalidad algorítmica es altamente despolitizadora. Prescinde de los sujetos humanos reflexivos al producir modos de comportamiento supraindividuales, que no requieren que los sujetos den cuenta de sí mismos (Rouvroy / Berns 2013). La política en la era digital tiene que considerar a los usuarios como sujetos políticos (Fuchs 2015), para lo cual debemos examinar el peso del pasado en la gobernanza algorítmica y cómo proteger la indeterminación del futuro.

b).- El poder de las preferencias del pasado

¿Cuál es entonces el valor democrático de datos, recomendaciones y predicciones? Hay quien dirá que todo eso son nuestras decisiones libres del pasado, invitaciones a decidir en el presente o apuestas acerca de cómo decidiremos en el futuro, es decir, se trata de nuestras decisiones en cualquier caso. Desde este punto de vista no habría ninguna tensión entre gobernanza basada en los datos y democracia, sino todo lo contrario. Pero la democracia no es la traducción inmediata y agregada de lo que decidimos individualmente en el pasado; el carácter dinámico y transformador de la vida democrática incluye un elemento de cambio, descubrimiento y emergencia para lo que no sirve un sistema pensado para hacernos descubrir únicamente lo que ya sabemos. Hoy por hoy la inteligencia artificial no parece apropiada para realizar esta voluntad de transformación, que constituye un elemento esencial de nuestra decisión democrática.

El problema estriba en que hacen predicciones que reflejan patrones del pasado. Los sistemas de aprendizaje automático funcionan de tal manera que los patrones encontrados en los datos subyacentes se convierten en reglas para la producción futura de resultados. Esto significa que tales sistemas solo pueden producir conocimientos que se refieran al pasado, porque los datos solo pueden describir

el pasado. La mayor parte de las previsiones algorítmicas se basan en la idea de que el futuro será lo más parecido al pasado, de que nuestras preferencias futuras representarán una continuidad de nuestro comportamiento anterior tal y como queda registrado en los datos de nuestra movilidad o consumo. Este modelo es especialmente inadecuado para aquellas actividades que tienen un propósito de intervenir en el mundo con el objetivo de cambiarlo. Los humanos, y especialmente nuestra actividad política, no aspiramos solo a reflejar lo que hay sino a cambiar ciertas cosas de un modo intencional. La gobernanza algorítmica podría ser muy útil para una concepción meramente agregativa de la democracia, pero parece limitada si tenemos una idea más deliberativa de la vida política, es decir, un proceso de mediación en el que nuestra voluntad política puede distanciarse de nuestras preferencias iniciales y donde los intereses sean entendidos más como algo que se descubre en los procesos de interacción comunicativa que como puntos de partida para el *bargaining* político.

Estos sistemas reproducen las desigualdades en tres pasos: 1. la realidad de las desventajas sociales queda plasmada en los datos, ya sea mediante un sesgo distorsionador o mediante una representación correcta de las desigualdades, 2. en un segundo paso, se refuerza normativamente como una situación de hecho supuestamente objetiva y 3. por

último, se refuerza por medio de las decisiones que se deducen de tal análisis, ya sea una contratación, una admisión o una condena. El criterio de lo que ocurrió en el pasado (quién fue contratado, admitido o condenado) se convierte en una receta para la discriminación algorítmica, es decir, para replicar la marginación del pasado. Las bases de datos con las que se entrenan los algoritmos contienen el sesgo agregado de la cultura penal, del racismo o del machismo estructural, que es precisamente lo que pretendemos superar. Si los algoritmos se basan en el pasado para decidir el futuro, estamos configurando un sistema que perpetúa un pasado del que no podemos escapar, un sistema que aumenta la dependencia del pasado, reduce la elección individual y del que resulta una “desventaja acumulativa” (Gandy 2009). De este modo, el futuro parece perder el carácter abierto de un reino de posibilidades para convertirse en un espacio cerrado, una extrapolación ininterrumpida del presente, donde cada sujeto estaría congelado en el eterno ahora de sus propias preferencias y desventajas estructurales (Bridle 2019).

Quisiera mencionar como contra-ejemplo de la voluntad de ruptura frente a esa dependencia respecto del pasado el tema de la historia salarial. El salario pasado determina la diferencia futura y, a su vez, afecta a las negociaciones salariales. A la hora de fijar el salario, un proceso cognitivo común

denominado sesgo de anclaje ("anchoring bias") puede hacer que nos basemos demasiado en los datos iniciales en detrimento de una valoración más racional de la valía de una persona. En 2016, Massachusetts se convirtió en el primer estado americano en aprobar una ley que prohíbe a los empleadores preguntar a los candidatos sobre su historial salarial. Desde entonces, más de una docena de estados han seguido su ejemplo. Las nuevas leyes también obligan a los empleadores a no basarse tanto en los niveles salariales anteriores, que pueden estar contaminados por la desigualdad sistémica (Lobel 2022). Es una manera de liberar al presente y al futuro del poder del pasado.

Tal vez lo más insatisfactorio de esta revolución de los cálculos es que no es nada revolucionaria. El análisis de datos actúa como un dispositivo de registro, hasta el punto de tener grandes dificultades para identificar lo que en la realidad hay de aspiración, deseo o contradicción. La ideología de esta sedicente superación de toda ideología es un "conductismo radical" (2010) o un "*data behaviourism*" (Rouvroy 2013): por un lado, nos pensamos como sujetos emancipados de toda determinación, pero continuamos siendo, en una medida mayor de lo que desearíamos, seres previsibles al alcance de los calculadores. No es verdad que dejarlo todo en manos de nuestra decisión —como consumidores o votantes— entronice nuestra libre decisión,

aunque solo sea por el hecho de que incluso cuando tenemos la sensación de tomar decisiones singulares, nuestros comportamientos obedecen a los hábitos inscritos en nuestra socialización. Pero es que, además, si hemos de tomarnos nuestra libertad en serio, también forma parte de ella nuestra aspiración de modificar lo que hemos sido, dando así lugar a situaciones hasta cierto punto impredecibles. Y a este respecto los algoritmos que se dicen predictivos son muy conservadores. Son predictivos porque formulan continuamente la hipótesis de que nuestro futuro será una reproducción de nuestro pasado, pero no entran en la compleja subjetividad de las personas y de las sociedades, donde también se plantean deseos y aspiraciones. Aquí es donde tiene todo el sentido denunciar a la inteligencia artificial como algo diseñado para consolidar los intereses dominantes, como un registro del poder (Crawford 2021, 8). Nos encontraríamos ante una nueva versión de aquella "fuerza normativa de lo fáctico" de la que hablaba hace muchos años el jurista Georg Jellinek (1914, 337). ¿Cómo queremos entender la realidad de nuestras sociedades si no introducimos en nuestros análisis, además de nuestros comportamientos de hecho, las enormes asimetrías en términos de poder, las injusticias de este mundo y nuestras aspiraciones de cambiarlo?

Las más sofisticadas redes neuronales solo

detectan regularidades y patrones basados en datos que proceden del pasado; no se hacen cargo de nuestra verdadera identidad, como prometen, sino de la figura estadísticamente generada a partir de nuestros perfiles pasados. Para muchas situaciones la extrapolación del pasado es insuficiente o solo proporciona soluciones estandarizadas que no tienen en cuenta la dimensión de apertura e indeterminación que caracteriza al futuro humano. El valor de las predicciones y la utilidad de la extrapolación se reduce especialmente en entornos inestables o momentos de crisis en los que disminuye la pertinencia de lo que ha estado vigente en el pasado. ¿Estamos dispuestos a que los datos que alimentan los algoritmos conviertan a nuestro pasado en nuestro futuro?

c).-La protección de las preferencias futuras

Hay diversos autores que han criticado la gobernanza algorítmica como una amenaza para la democracia, pero casi todos lo han hecho porque condiciona nuestras decisiones presentes y muy pocos han advertido que el problema más importante es hasta qué punto se desentiende de nuestras decisiones futuras. La democracia no es hacer lo que queremos sino, muchas veces, poder cambiar lo que queremos. ¿Conocen realmente los algoritmos nuestra voluntad profunda o solo su dimensión más superficial, las rutinas más que los anhelos? En este sentido la tesis del

capitalismo de vigilancia (Zuboff 2018) no me parece acertada porque da a entender que se nos conoce demasiado y el problema es que la observación algorítmica nos conoce demasiado poco y minusvalora ciertos aspectos de nuestra voluntad que no quedan registrados en nuestros datos pasados.

La gobernanza algorítmica estrecha el "*aspirational self*" (Reviglio / Agosti 2020, 2), esa dimensión que las máquinas no son capaces de anticipar y que sucede más allá de las huellas que dejamos en los registros y las bases de datos, aunque con frecuencia esas huellas se nos asignen como un destino. Una de las condiciones necesarias para el desarrollo de la autonomía es "la posibilidad de que el individuo conciba su existencia no como la confirmación o repetición de sus propias huellas, sino como la posibilidad de cambiar de rumbo, de explorar nuevas formas de vida y de ser, en una palabra, de ir a donde no se espera que vaya" (Rouvroy 2008, 252). Los seres humanos no podríamos autodeterminarnos si no hubiera un espacio de indeterminación, una falta de continuidad entre lo que éramos y lo que podemos llegar a ser, una posibilidad, aunque sea muy leve, de quebrar el poder del pasado sobre el presente y el futuro, algo que hemos llamado en ocasiones rebelión, conversión o transformación, que designa una misma disposición sobre sí mismo en términos de ruptura de la previsibilidad.

Nuestra pulsión de libertad se vería quebrada si creyéramos que las predicciones son el único futuro que tenemos. La política no es tanto un instrumento para la satisfacción mecánica de nuestras necesidades como un medio para combatir nuestra perplejidad personal y colectiva acerca de cuáles son y como satisfacerlas. Que los humanos no sepamos propiamente lo que queremos es una experiencia antropológica básica y ancestral; que compensemos esa ignorancia con sistemas inteligentes y algoritmos es una experiencia ultramoderna. La cuestión es si lo uno y lo otro son un apoyo para la propia reflexividad o su aniquilación. El problema es cómo conseguimos que los algoritmos razonen de una manera contrafáctica, introduciendo un factor de contingencia que altere ese continuismo histórico, abriendo así la posibilidad de actuar de modo diferente. Mientras la inteligencia artificial no conozca esta diferencia entre lo fáctico y lo posible de la que es consciente cualquier razón natural tendrá muchas limitaciones a la hora de hacerse cargo de decisiones humanas (Pearl / Mackenzie 2018).

La política no es una simple administración continuista del pasado sino la posibilidad siempre abierta de quebrar la inercia del pasado. De esa fuerza carecen los algoritmos en la medida en que hacen predicciones sin otra referencia que el pasado. ¿Cómo especificamos nuestros objetivos de modo que las máquinas

no tengan que hacer otra cosa que perseguirlos eficientemente? ¿Sabemos lo que verdaderamente deseamos? ¿Estamos seguros de que lo que deseamos ahora será lo que deseemos en el futuro? Los algoritmos del machine learning pueden anticipar nuestros "future propensities" y amenazar así que hagamos posibles "alternative political futures" (Amoore 2020). Shoshana Zuboff formulaba acertadamente esta tendencia cuando escribía que en la era digital está en juego un derecho al futuro (2018, 595). Tocamos aquí un elemento nuclear de nuestra identidad. Nietzsche sentenciaba que el ser humano se caracteriza por ser "un primer movimiento" (1967, 207). En esa línea, Hannah Arendt definía la capacidad humana de actuar, como la posibilidad de comenzar, es decir, de realizar algo imprevisto. Una propiedad que también llamaba "natalidad" (1998) o "impredecibilidad" (1993) del comportamiento humano.

En contraste con ello, el riesgo de la gobernanza algorítmica es que preconfigure el campo de lo posible a partir de los perfiles elaborados sobre la base de la pertenencia a determinado patrón o grupo social, sin dejar espacio a lo inesperable. En última instancia, la inteligencia y el pensamiento no pueden modelizarse como una mera secuencia de operaciones aritméticas estandarizadas que representen todas las conclusiones posibles. La inteligencia se manifiesta precisamente en

contextos imprevisibles que requieren vías de solución inéditas. Una operación realizada dentro del lenguaje simbólico de un algoritmo sólo puede procesar problemas predefinidos. Lo inesperado —un suceso fuera del marco de referencia con pretensión de validez— permanece invisible dentro de la lógica del sistema. Esto tiene una gran significación democrática porque, en palabras de Luhmann, "la democracia es un inusual mantener abiertas las posibilidades de elección futura" (2018, 133).

2.-LA AGREGACIÓN DE LAS PREFERENCIAS EN LA GOBERNANZA ALGORÍTMICA

Si en el apartado anterior hemos analizado la capacidad de identificar correctamente las preferencias, ahora nos corresponde hacerlo con la agregación de las preferencias; si lo anterior era más bien una tarea epistemológica, entramos ahora en un horizonte normativo y práctico: ¿cómo compatibilizar preferencias distintas? ¿Qué procedimiento proporciona a este respecto un resultado que podemos calificar como más justo y democrático?

a).-La agregación imposible

Supongamos que estuviéramos de acuerdo en la concepción de lo equitativo y que sólo habría que establecer un procedimiento para agregar nuestros diferentes intereses. ¿Sería esto

posible? ¿Es razonable buscar un procedimiento técnico que determinara la resultante equilibrada de nuestras distintas preferencias?

Esta aspiración se encuentra de entrada con la dificultad de que no estamos ante un asunto que tenga una solución “técnica”, si por tal entendemos algo que nos ahorre juicios de valor y desactive el carácter controvertido de cualquier decisión pública. La fairness no consiste en un technical debiasing, sino que incluye un amplio análisis social sobre el modo como es usada la inteligencia artificial en un contexto dado, de manera que sea posible una mejor auditoría de los sesgos. Al igual que la fairness no es una propiedad de los algoritmos, sino más bien de las decisiones que contienen (en su diseño, análisis o aplicación) (Ochigame / Barabas / Dinakar / Virza / Ito 2018, 4), la discriminación no es solo una cuestión algorítmica. Todo lo que tiene que ver con la justicia y la discriminación es tan contextual y controvertido que no siempre se presta a formalismos matemáticos (Selbst / Boyd / Friedler / Venkatasubramanian / Vertesi 2019).

La justicia algorítmica no puede ser una implementación que satisfaga ciertos "fairness measures" (indicadores de igualdad) incontrovertidos. Qué idea de fairness y qué otros valores deben ser considerados en un algoritmo supone un desafío de naturaleza

política, no simplemente técnica, que requiere acomodar diversos intereses en conflicto; es una tarea política que no puede ser realizada por unos técnicos o por unos algoritmos que no tuvieran necesidad de contar con la opinión ciudadana, es decir, que debe llevarse a cabo democráticamente, abriendo estas definiciones a la pública discusión.

Si hubiera acuerdo acerca de qué significa "fairness", entonces el algoritmo desarrollaría una tarea puramente técnica; se trataría nada más que de encontrar el mejor modo de operacionalizar esa idea de justicia. Pero hay ciertas decisiones difíciles (hard choices) acerca de cómo medir la justicia que hay que tomar antes de que comience el trabajo técnico de detectar y mitigar la injusticia. ¿Podemos determinar qué significa "exactitud" y cómo se mide sin hacer algún tipo de juicio ético-político sobre los tipos de errores que pensamos que es más urgente evitar o cuál es el objetivo final de una organización? Parece claro que se trata de asuntos que deben ser objeto de discusión política y no de agregación algorítmica. El problema es que la idea de justicia es un verdadero campo de batalla democrático, un concepto elevadamente controvertido en cualquier sociedad plural. Hay más desacuerdos acerca de los valores en sí mismos que sobre los medios de conseguirlos. Ningún dispositivo tecnológico puede ahorrarnos el trabajo de discusión democrática en torno a los fines, aunque pueda

facilitarnos enormemente la tarea de implementación de los objetivos que democráticamente hemos decidido perseguir.

El carácter controvertido, político y no meramente técnico de la fairness plantea otro problema adicional. Además de la dificultad de ponerse de acuerdo en torno a una idea de justicia, está la imposibilidad de satisfacer igualmente y simultáneamente esa diversidad de aspiraciones de justicia. “Las constricciones prácticas y sociales impiden que todas las preferencias sean satisfechas simultáneamente, lo que significa que los robots tienen que mediar entre preferencias conflictivas, algo por lo que filósofos y científicos sociales han luchado durante milenios” (Russell 2019, 32). Más difícil que identificar preferencias particulares es agregarlas y hacerlas compatibles (Züger / Milan, / Tanczer 2017). Imaginemos que la tecnología nos ha permitido identificar todos los deseos, preferencias y decisiones individuales, ¿habríamos hecho innecesario cualquier elemento de mediación para la configuración de la voluntad popular? ¿Nos bastaría agregar sin deliberación las decisiones así registradas?

Nos encontramos ante una variante digital del llamado “impossibility theorem” de Arrow (1950) por el que se declaraba como algo imposible satisfacer valores distintos (Friedler / Scheidegger / Venkatasubramanian 2016;

Berk / Heidari / Jabbari / Kearns / Roth 2017; Miconi 2017). Se formula así la idea de que es matemáticamente imposible que un algoritmo satisfaga simultáneamente las diversas ideas de justicia que sostenemos. No es posible recoger las diferentes preocupaciones de justicia que tenemos en una sociedad plural, ni resulta verosímil que lleguemos a un entendimiento pleno acerca de ese valor. Además, el valor de la justicia está relacionado con otros valores como la seguridad o la libertad, por lo que su formalización técnica resulta todavía más inverosímil.

Que una parte de nuestros desacuerdos sea irresoluble técnicamente y tenga una naturaleza política no es necesariamente una mala noticia. Se trata de una imposibilidad que nos obliga a explorar un modelo de decisión que tal vez tenga una gran fuerza democratizadora. El carácter controvertible de ciertos asuntos, su ambigüedad, tiene un valor político en la medida en que obliga a negociar y buscar compromisos una vez que los procedimientos de técnica algorítmica nos han dejado tirados (Coyle / Weller 2020). Se podría hablar incluso de una cierta incompatibilidad entre la lógica de los algoritmos y la de la política. El *machine learning* optimiza la consecución de objetivos una vez que estos han sido explícitamente formulados. La política, por el contrario, se basa en una cierta ambigüedad en relación con los objetivos, gracias a la cual hay un espacio para lograr

compromisos. La política es con mucha frecuencia un *trade-off* entre preferencias e intereses distintos e incluso contrapuestos. Los algoritmos son optimizadores de una determinada decisión, pero no toleran la ambigüedad. Por eso la justicia algorítmica no vendrá de que mejoremos los datos o desprejuiciemos los algoritmos sino de que sustituyamos un procedimiento de agregación de intereses y preferencias por uno de deliberación.

b).-La autodeterminación deliberativa

La concepción deliberativa de la democracia parte del supuesto de que, si bien es cierto que la política está para satisfacer los intereses de las personas, esos intereses no se determinan con independencia de la reflexión sobre ellos y su compatibilidad con los de los demás. La democracia no es tanto que se tenga en cuenta nuestra opinión o se satisfaga nuestro interés como que dispongamos de un espacio público en el que configurar nuestra opinión e identificar nuestros intereses teniendo en cuenta los de los demás. Ni el interés individual está plenamente fijado, ni el interés colectivo están dados de antemano o pueden confiarse a una mera agregación de los intereses individuales; ambos tienen una dimensión de construcción pública. Hace falta establecer un marco de diálogo y negociación que permita la construcción equitativa de esa voluntad general. El problema de la gobernanza

algorítmica es que registra nuestros intereses, pero no los convierte en objeto de reflexión.

Los automatismos son procesos que funcionan precisamente porque no obligan a tematizar los presupuestos sobre los que discurren. Los seres humanos, tanto en el plano personal como en el colectivo, realizamos tareas mecánicas y vivimos sin cuestionar las prioridades que una vez establecimos, pero hemos de estar abiertos a otro tipo de situaciones en las que se requiere de nosotros un examen de las rutinas y una reorientación hacia objetivos nuevos. Una de las revitalizaciones de la democracia a finales del siglo pasado vino precisamente del concepto de “democracia reflexiva” (Beck / Lash / Giddens 1994) con el que diversos pensadores defendían la interrupción crítica de una modernización irreflexiva. De este modo no hacían otra cosa que acentuar una propiedad de la política como actividad que se interroga por los fines frente a las rutinas administrativas. Pues bien, la gobernanza algorítmica carece por sí misma de la capacidad de cuestionarse sus objetivos o lo hace —en virtud de procesos de aprendizaje— dentro de un marco que no se ha dado a sí misma y que por ello no puede cuestionar radicalmente.

La reflexividad es lo que hace posible la deliberación democrática, es decir, aquella forma de interacción que no es solo una

negociación de nuestras preferencias e intereses, sino que permite incluso su revisión y ponderación reflexiva. El sentido de las instituciones de la mediación en una democracia consiste en establecer una distancia entre la voluntad inmediata y la decisión política. El procedimiento para ello es la apertura de espacios en los que sea posible algo así como una desaceleración de las decisiones para permitir el libre intercambio de las opiniones y los puntos de vista. Una democracia requiere esta capacidad cuando se trata de satisfacer preferencias e intereses diversos, que no pocas veces plantean exigencias disparatadas.

A este respecto, la presencia del pueblo en la democracia algorítmica es más de *volonté de tous* que de *volonté générale*, por utilizar la terminología de Rousseau, más de agregación que de configuración, de soberanía que de democracia: nuestras preferencias de partida son tomadas en consideración, por supuesto, pero se nos priva del momento de construcción deliberativa en el que esas preferencias ya no son meramente agregadas, sino que interaccionan con otras. De este modo no se abre ese espacio de indeterminación que permitiría una reformulación transformadora de tales preferencias atendiendo a su (in)compatibilidad con las de otros. Estos sistemas no contemplan otro modelo que el de unos individuos maximizando su utilidad. El problema de la gobernanza algorítmica es que

gracias a los algoritmos intervenimos en la expresión de preferencias e intereses, pero no en la construcción de una totalidad social deseable que nos habría permitido eventualmente modificarlos. Nuestra presencia en el proceso democrático algorítmico sería la de poner nuestros rastros y huellas a disposición de los sistemas de decisión, pero no la de intervenir en el diálogo en el que se ponderan esos datos y se delibera acerca de la idea de sociedad deseable a partir de ellos. En una democracia algorítmica ser ciudadano consistiría en tener el derecho a emitir deseos, pero no a ponderarlos con los de otros e incluso modificar esos deseos propios. La ciudadanía se reduciría a la generación de datos. Este modelo de gobernanza tiene al menos estas debilidades desde el punto de vista democrático: 1) que pensemos que al emitir señales digitales ya hemos expresado suficientemente lo que queremos; 2) que lo hayamos hecho sin interiorizar explícitamente la compatibilidad de nuestra voluntad con la de otros; y 3) que de este modo nos creamos eximidos de pensar en qué tipo de sociedad resultante queremos.

Nos encontramos ante dos tipos diferentes de racionalidad y sus correspondientes modelos de gobernanza. Una “gubernamentabilidad algorítmica” que es implícita y automática, con criterios emanados en tiempo real de la realidad digitalizada y una “gubernamentabilidad política” explícita que

resulta de una deliberación que requiere reflexión y consume tiempo (Rouvroy 2013, 66). La primera de ellas, impulsada tecnológicamente, parece desafiar la interrogación, el análisis y la rendición de cuentas, conduciendo así a configurar un entorno político e institucional sin un debate significativo ni oportunidades de impugnación (Gree / Hu 2018; Waldman 2019, 72). Las psicotecnologías automatizadas de la digitalización pueden debilitar la democracia en la medida en que dificulten realizar su dimensión deliberativa. En vez de constituir sujetos políticos que estén dispuestos y sean capaces de entrar en un proceso de reflexión conjunta sobre la configuración del bien común, pueden estar generando sujetos apolíticos a quienes la idea misma de una negociación democrática de intereses y preferencias les resulte incomprensible.

Pensar adecuadamente la democracia en un entorno algorítmico requiere entender en qué consiste la voluntad política, que no es la afirmación solipsista de lo que yo quiero ni la simple agregación de voluntades configuradas de un modo a-político. La paradoja que planteo es que precisamente cuando estamos tratando de establecer un marco conceptual para pensar la igualdad, la no discriminación, hace falta revisar el papel que desempeña el individuo en una democracia. La democracia necesita que posibilitemos aquellas experiencias de comunicación, contestación y conflicto que nos

hacen a los humanos seres sociales. Y puede estar ocurriendo que la personalización algorítmica en sus diversas formas (granularización, mercantilización, microtargeting...) esté limitando la diversidad de información, la exposición a puntos de vista alternativos y dificultar el descubrimiento de posibles preferencias.

Tal vez la igualdad esté exigiendo otros objetivos que posiblemente estén en conflicto con la personalización. Solo una concepción extremadamente individualista de lo social puede consagrar el interés individual hasta el punto de considerarlo la última palabra y hacerlo en nombre de la justicia. ¿Qué hacemos cuando aquello que el usuario quiere contribuye a la injusticia? El concepto de “post-userism” es un planteamiento teórico que cuestiona el foco que ha dominado la interacción entre los humanos y las máquinas proponiendo la conveniencia de considerar un marco más amplio (Baumer / Brubaker 2017). En vez de centrarse en el individuo, habría que entender la justicia en términos de distribución: cómo está distribuido el daño o el beneficio entre los diferentes individuos y grupos. La elicitación de preferencias a partir de un conjunto fijo de alternativas es a menudo insuficiente e injusto, dado que estas preferencias reflejan los sesgos y las desigualdades existentes en una sociedad y dado que esos métodos no vienen acompañados de una deliberación democrática

significativa (Robertson / Salehi 2020; Martí 2021).

El tránsito hacia un modelo deliberativo implica también un cambio en cuanto al modo de considerar el punto de partida (los intereses o preferencias individuales), que pasarían a ser entendidos como algo indeterminado, flexible, dinámico y cambiante. Podría conseguirse así una cierta convergencia entre la ciencia computacional y la teoría deliberativa de la democracia. Dice Stuart Russell que la inteligencia artificial ha prestado poca atención a la incertidumbre, como si hubiera siempre un perfecto conocimiento del objetivo. Esto puede valer para determinados juegos, pero para otro tipo de problemas las preferencias relevantes no son inicialmente conocidas. La idea de cómo tomar decisiones con objetivos abiertos e indeterminados es un desafío tanto para la computación como para la teoría de la democracia. Hablar de intereses y preferencias como si fueran evidencias y además de fácil implementación es una simpleza incompatible con la complejidad de los humanos y de nuestras sociedades (Innerarity 2019 y 2023). ¿Cómo conseguimos que un robot aprenda a entender las preferencias subyacentes en el comportamiento humano, que es “irracional, inconsistente, de voluntad débil y computacionalmente limitado, por lo que sus acciones no siempre reflejan sus verdaderas preferencias” (Russell 2019, 32).

La cuestión de los sesgos y la equidad algorítmica suele plantearse como si la justicia consistiera en respetar unas propiedades o intereses que las personas o grupos *tienen* y no como la generación de un marco en el que esas personas o grupos puedan relacionarse reflexivamente con sus propiedades o intereses. En una democracia deliberativa se trataría de decidir de acuerdo con unas preferencias humanas cuya plasticidad permite que vayan cambiando con el tiempo y especialmente en el diálogo y conflicto con las de otros (Pettigrew 2020). No estamos solo ante la exigencia de autogobernarnos sino ante la posibilidad de cambiarnos. Para ello se requiere un entorno algorítmico que no se limite a registrar lo que fácticamente revelamos querer, sino que permita una autocontrolada capacidad de desafiar esa facticidad y modificarla. Y aquí la consecución de un equilibrado sistema algorítmico es de la mayor importancia ya que tenemos que reconsiderar cómo gestionamos el posible conflicto entre la satisfacción de nuestras preferencias inmediatas y nuestra capacidad de configurar esas preferencias en el largo plazo. Los algoritmos pueden dar más peso a las preferencias de largo plazo y a las preferencias racionales sobre las inmediatas y emocionales, pero también puede ocurrir exactamente lo contrario: que un sistema algorítmico de mero registro de nuestro comportamiento impida la consideración de

futuros alternativos. Un nuevo giro deliberativo de la democracia en la era de la inteligencia artificial corregiría el “hedonismo psicológico” (Gal 2017) al que se reduce la democracia cuando se limita a la satisfacción digital de las preferencias individuales. La democracia requiere un cierto grado de “incomodidad”, por ejemplo, para limitar los deseos individuales cuando afectan negativamente al conjunto de la sociedad, para asegurar la autonomía personal o para introducir consideraciones de largo plazo que puedan estar en conflicto con los intereses inmediatos. No hay verdadera autodeterminación si no podemos pensar más allá de nosotros mismos y de la actual configuración de la sociedad. Solo esta capacidad asegura la vitalidad de una sociedad democrática.

REFERENCIAS:

Amoore, Louise (2020), *Cloud Ethics. Algorithms and the Attributes of Ourselves and Others*, Durham and London: Duke University Press.

Arendt, Hannah (1998), *The Human Condition*, The University of Chicago Press.

Arendt, Hannah / Jaspers, Karl (1993), *Correspondence 1926-1969*, London: Harvest Books.

Arrow, Kenneth J. (1950), "A Difficulty in the Concept of Social Welfare", *Journal of Political Economy* 58, 328–346.

Baumer, Eric / Brubaker, Jed (2017). "Post-userism", en Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, 6291–6303.

Beck, Ulrich / Lash, Scott / Giddens, Anthony (1994), *Reflexive Modernization*, Cambridge: Polity Press.

Berk, Richard / Heidari, Hoda / Jabbari, Shahin / Kearns, Michael / Roth, Aaron (2017), "Fairness in Criminal Justice Risk Assessments: The State of the Art", *arXiv preprint*, arXiv:1703.09207

Binns, Reuben (2018a), "Fairness in Machine Learning: Lessons from Political Philosophy", *Proceedings of Machine Learning Research* 81, 149-159.

— (2018b), "Algorithmic Accountability and Public Reason", *Philosophy & Technology* 31, 543-556.

Bridle, James (2019), *New Dark Age. Der Sieg der Technologie und das Ende der Zukunft*, München: C.H. Beck.

Brooks, Frederick P. (1975), *The Mythical Man-Month: Essays on Software Engineering*,

Massachusetts: Addison-Wesley.

Christian, Brian (2020), *The Alignment Problem: Machine Learning and Human Values*, New York: Norton & Company.

Coglianesi, Cary / Lai, Alicia, Algorithm vs. Algorithm (2022). Duke Law Journal, Vol. 72, University of Pennsylvania Law School, Public Law Research Paper No. 22-11, Available at SSRN: <https://ssrn.com/abstract=4026207>

Collins, Patricia Hill (2002), *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, New York: Routledge.

Coyle, Diane / Weller, Adrian (2020), "Explaining Machine Learning Reveals Policy Challenges", *Science* 386 / 6498, 1433-1434.

Crenshaw, Kimberlé (ed.) (2019), *Seeing Race Again: Countering Colorblindness across the Disciplines*, Berkeley: University of California Press.

Dieterich, William / Mendoza, Christina / Brennan, Tim (2016), "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", Northpoint Inc. Available Online at: [http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica Commentary Final 070616.pdf](http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica%20Commentary%20Final%2070616.pdf)

Dwork, Cynthia / Hardt, Moritz / Pitassi, Toniann / Reingold, Omer / Zemel, Richard

(2012), "Fairness through awareness", *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214-226.

Friedler, Sorelle / Scheidegger, Carlos / Venkatasubramanian, Suresh (2016), "On the (Im)possibility of Fairness", *arXiv preprint*, arXiv:1609.07236.

Fuchs, Christian (2015), "Towards the public service internet as alternative to the commercial internet", *OFF Texte*, 20, 43-50.

Gal, Michal (2017), "Algorithmic challenges to autonomous choice", *Michigan Telecommunications and Technology Law Review* 25, 59-104.

Gandy, Oscar (2010), "Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems," *Ethics and Information Technology* 12/1, 29-42.

Geller, Eric (2018), "Despite Trump's assurances, states struggling to protect 2020 election"
<https://www.politico.com/story/2018/07/27/trump-election-security-2020-states-714777>

Green, Been / Hu, Lily (2018), "The myth in the methodology: Towards a recontextualization of fairness in machine learning", *Machine Learning: The Debates workshop at the 35th*

International Conference on Machine Learning.
<https://www.benzevgreen.com/wpcontent/uploads/2019/02/18-icmldebates.pdf>

Hanna, Alex / Denton, Emily / Smart, Andrew / Smith-Loud, Jamila (2020), “Towards a critical race methodology in algorithmic fairness”, Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.
<https://doi.org/10.1145/3351095.3372826>

Hardt, Moritz (2014), “How Big Data Is Unfair: Understanding Unintended Sources of Unfairness in Data Driven Decision Making”, Medium, September 26, 2014.
<https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>

Innerarity, Daniel (2019), “Democratic equality: an egalitarian defense of political mediation”, *Constellations. An International Journal of Critical and Democratic Theory*, 26/4, 513-524.

— (2023), *A theory of complex democracy. Governing in the Twenty-first century*, London: Bloomsbury.

Jellinek, Georg (1914), *Allgemeine Staatslehre*, Berlin: Haring.

Jolls, Chistine / Sunstein, Cass R. / Thaler, Richard (1998.), “A Behavioral Approach to Law and Economics”, *Stanford Law Review* 50 (5), 1471-550.

Kahneman, Daniel (2011), *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux.

Kasy, Maximilian / Abebe, Rediet (2021), "Fairness, equality, and power in algorithmic decision-making", en Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 576-586.

Kleinberg, Jon / Ludwig, Jens / Mullainathan, Sendhil / Sunstein, Cass R. (2018), "Discrimination in the Age of Algorithms", *Journal of Legal Analysis* 10, 113-174.

Lai, Alicia (2018), *Brain Bait: Effects of Cognitive Biases on Scientific Evidence in Legal Decision-Making*, A.B. thesis, Princeton University.

Leighninger, Matt / Moore-Vissing, Quixada (2018), *Rewiring democracy: Subconscious Technologies, Conscious Engagement, and the Future of Politics*, Public Agenda: San Francisco.

Lobel, Orly (2022), *The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future*, New York: PublicAffairs.

Luhmann, Niklas (2018), *Soziologische Aufklärung 4. Beiträge zur funktionellen Differenzierung der Gesellschaft*, Wiessbaden: Verlag für Sozialwissenschaften.

Majone, Giandomenico (1989), *Evidence, Argument, & Persuasion in the Policy Process*,

New Haven and London: Yale University Press.

Malik, Momin M. (2020) "A hierarchy of limitations in machine learning", arXiv preprint.

<https://arxiv.org/pdf/2002.05193.pdf>

Martí, José Luis (2021), "New Technologies at the Service of Deliberative Democracy" en Amato, Guiliano / Barbisan, Benedetta / Pinelli, Cesar (eds.), *Rule of Law vs Majoritarian Democracy*, New York: Bloomsbury, 199-220.

Miconi, Thomas (2017), "The Impossibility of "Fairness": A Generalized Impossibility Result for Decisions", arXiv preprint, arXiv:1707.01195

Mitchell, Shira / Potash, Eric / Barocas, Solon / D'Amour, Alexander / Lum, Kristian (2021), "Algorithmic fairness: Choices, assumptions, and definitions", *Annual Review of Statistics and Its Application*, 8, 141-163.
<https://doi.org/10.1146/annurev-statistics-042720-125902>

Mittelstadt, Brent / Allo, Patrick / Taddeo, Mariarosaria / Wachter, Sandra / Floridi, Luciano (2016), "The ethics of algorithms: Mapping the debate", *Big Data & Society*, 3/2, July-December 2016.

Nietzsche, Friedrich (1967), *Nachgelassene Fragmente 1882-1884*, en *Sämtliche Werke 10, Kritische Studienaufgabe*, ed. Colli-Montinari,

Berlin/New York: de Gruyter.

Ochigame, Rodrigo / Barabas, Chelsea / Dinakar, Karthik / Virza, Madars / Ito, Joichi (2018), “Beyond legitimation: Rethinking fairness, interpretability, and accuracy in machine learning”, en *The Debates*, at the 35th International Conference on Machine Learning.

Parnas, David Lorge (1985), “Software Aspects of Strategic Defense Systems”, *American Scientist*, September-October 1985, 432-440.

Pearl, Judea / Mackenzie, Dana (2018), *The Book of Why: The New Science of Cause and Effect*, London: Allen Lane.

Pettigrew, Richard (2020), *Choosing for Changing Selves*, Oxford University Press.

Robertson, Samantha / Salehi, Niloufar (2020), “What if I don’t like any of the choices? The limits of preference elicitation for participatory algorithm design”, en *Participatory Approaches to Machine Learning Workshop*, ICML 2020.
<https://arxiv.org/pdf/2007.06718.pdf>

Reckwitz, Andreas (2017), *Die Gesellschaft der Singularitäten: Zum Strukturwandel der Moderne*, Berlin: Suhrkamp.

Reviglio, Urbano / Agosti, Claudio (2020), “Thinking Outside Black-Box: The Case for “Algorithmic Sovereignty”, *Social Media &*

Society, August-June, 1-12.

Rouvroy, Antoinette (2008), "Reinventer l'art d'oublier et de se faire oublier dans la société de l'information?", en Lacour, Stéphanie (2008), *La sécurité de l'individu numérisé. Reflexions prospectives et internationales*, Paris: L'Harmattan, 249-278.

— (2013), "The End(s) of Critique: Data Behaviourism versus Due Process", en Hildebrand, Mireille / de Vries, Katja (eds.), *Privacy Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, New York: Routledge, 143-167.

Rouvroy, Antoinette / Berns, Thomas (2013), "Algorithmic governmentality and prospects of emancipation. Disparateness as a precondition for individuation through relationships?", *Réseaux* 177, 1, 163-196.

Russell, Stuart (2019), "The purpose put into the machine", Brockman, John (ed.), *Possible Minds. 25 Ways of Looking at AI*, New York: Penguin, 20-32.

Selbst, Andrew D. / Boyd, Danah / Friedler, Sorelle / A., Venkatasubramanian, Suresh / Vertesi, Janet (2019). "Fairness and abstraction in sociotechnical systems", en Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.

<https://doi.org/10.1145/3287560.3287598>

Thaler, Richard H. (2015), *Misbehaving: The making of behavioral economics*, New York: Norton & Co.

Vallor, Shannon (2015), "Moral deskilling and upskilling in a new machine age", *Philosophy & Technology*, 28(1), 107–124.

Waldman, Ari Ezra (2019), "Power, Process, and Automated Decision-Making", 88 *Fordham Law Review* 613. Available at: <https://ir.lawnet.fordham.edu/flr/vol88/iss2/9>

Wang, Annie J., (2018), "Procedural justice and risk-assessment algorithms". *SSRN Electronic Journal* 2018: 1–31. DOI: 10.2139/ssrn.3170136.

Whittaker, Meredith /, Alper, Meryl / Bennett, Cynthia L., / Hendren, Sara / Kaziunas, Liz / Mills, Mara / Morris, Meredith / Rankin, Joy / Rogers, Emily / Salas, Marcel / West, Sarah M. (2019), "Disability Bias, and AI", <https://ainowinstitute.org/disabilitybiasai-2019.pdf>

Zarsky, Tal (2016), "The Trouble with Algorithmic Decisions An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making", *Science, Technology & Human Values* 41, 118–132.

Zuboff, Shoshana (2018), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs.

Züger, Theresa / Milan, Stefania / Tanczer, Leonie Maria (2017), "Sand im Getriebe der Informationsgesellschaft: Wie digitale Technologien die Paradigmen des Zivilen Ungehorsams herausfordern und verändern", in Jacob, Daniel / Thiel, Thorsten (eds.), *Politische Theorie und Digitalisierung*, Baden-Baden: Nomos, 265-296.

PRESENTACIÓN DE LA PONENTE AZUCENA GARCÍA PALACIOS

Sergio Rodríguez García, Catedrático de Paleontología, Universidad Complutense de Madrid, Humboldtiano

Azucena García Palacios es Catedrática de Psicopatología y directora del equipo de investigación **Labpsitec** de la Universitat Jaume I de Castellón de la Plana.

La profesora García Palacios ha participado como investigadora en más de 30 proyectos de investigación financiados por entidades nacionales y por la Unión Europea. Es autora de más de 200 artículos publicados en revistas de prestigio. Y de al menos 3 libros,

- Terapia dialéctica comportamental

-Emociones ¿Como Gestionarlas?
(COMPORTAMIENTO HUMANO)

-Manual de prácticas de psicopatología

Ha llevado a cabo estancias en las universidades de Harvard y Washington, y ha impartido diversos cursos dirigidos a profesionales de la salud mental y en postgrados universitarios.

La línea de investigación principal de su grupo es la Psicopatología y la mejora de los tratamientos psicológicos, principalmente para los trastornos emocionales, los trastornos de personalidad y el dolor crónico.

El grupo de investigación Labpsitec es líder a nivel internacional en la aplicación de las Tecnologías de la Información y la Comunicación en el ámbito de la Psicología Clínica y de la Salud. Su actual línea de investigación incluye explorar el potencial de la Inteligencia Artificial para la mejora de la detección temprana de problemas de salud mental y la personalización de los tratamientos psicológicos.

También ha explorado **el uso de la realidad virtual** y realidad aumentada para superar las fobias y programas de tratamiento online para trastornos emocionales (ansiedad, depresión), trastornos adaptativos o juego patológico.

INTELIGENCIA ARTIFICIAL Y SALUD MENTAL

Azucena García Palacios

Catedrática de Psicopatología, Universitat Jaume I Castellón

1.- Introducción

“No tengo fuerzas para nada, ni para trabajar ni para ver a mis amigos, lo intento, pero no puedo, siento un peso enorme en el pecho continuamente. Creo que he decepcionado a todos, que no merezco que me quieran y da igual lo que haga, no se trata de que ya no pueda sentir alegría, es que tampoco puedo sentir ya ni tristeza, no siento nada” (Juan, 34 años).

¿Qué va a pasar cuando no yo no esté? ¿quién va a cuidar a mi hijo? ¿quién va a conseguir que quiera salir de casa e ir al trabajo? ¿quién va a estar pendiente de detectar los primeros síntomas del episodio y llevarle al psiquiatra”. (Luisa, 69 años).

Estas son ejemplos de breves descripciones de lo que supone para los afectados y sus familias

vivir con un problema de salud mental.

La Organización Mundial de la Salud (OMS) define la salud mental como:

“Un estado de bienestar en el cual cada individuo desarrolla su potencial, puede afrontar las tensiones de la vida, puede trabajar de forma productiva y fructífera, y puede aportar algo a su comunidad.” (WHO, 2013).

Muchas personas en el mundo no alcanzan ese estado de bienestar o, aunque lo hagan, en algún momento de sus vidas se ve afectada su salud emocional. Y cuando eso sucede puede tomar la forma de un trastorno mental: Trastorno depresivo mayor, trastorno de ansiedad generalizada, trastorno de estrés postraumático, esquizofrenia, trastorno de la personalidad, trastornos por uso de sustancias, etc.

Un número importante de personas en el mundo, alrededor del 12%, vive en este momento con un trastorno mental y si tenemos en cuenta toda la vida de una persona, alrededor del 25% tendrán un trastorno mental a lo largo de su vida (Rehm & Shield, 2019), siendo estos trastornos una de las principales causas de discapacidad. En el estudio sobre la carga total de enfermedad, teniendo en cuenta todas las enfermedades, no sólo las consideradas mentales, las autolesiones, la depresión y los trastornos de

ansiedad constituyen la 3^a, 4^a y 5^a causa de discapacidad en niños, adolescentes y adultos jóvenes de entre 10 y 24 años; y la depresión es la 6^a causa de discapacidad en los adultos (Vos et al., 2020). Los trastornos mentales también constituyen un factor de riesgo para una muerte prematura (Charlston et al., 2015). Las personas con trastornos mentales graves mueren de media de 10 a 20 años antes que la población general. Además, el problema afecta de forma sustancial a la economía, ya que muchas de las bajas laborales por enfermedad están causadas por los trastornos mentales y un número importante de las personas que reciben ayudas por discapacidad tienen un trastorno mental.

Estos datos epidemiológicos a nivel global indican la magnitud del problema, que se ha agravado aún más con el impacto de la pandemia de COVID-19; por ejemplo, la ansiedad y depresión aumentaron más de un 25% en el primer año de la pandemia y se ha observado que el impacto ha sido mayor en grupos específicos como los adolescentes y adultos jóvenes (WHO, 2022). Por otro lado, como aspecto positivo, una de las consecuencias de la pandemia ha sido el aumento de la conciencia de la importancia de la salud mental, lo que está provocando un movimiento para incrementar la inversión en salud mental como una prioridad de las políticas públicas (p.ej., Patel et al., 2018). Buenos ejemplos son el Plan de Acción Integral

sobre Salud Mental 2023-2030 publicado por la OMS en 2021 (WHO, 2021) y en nuestro país la reciente Estrategia de Salud Mental del Sistema Nacional de Salud que generó el Plan de Acción de Salud Mental 2022-2024 en colaboración con las Comunidades Autónomas. La salud mental se plantea también como una prioridad en la tarea de aumentar el bienestar y la felicidad de las naciones como indica el World Happiness Report (Helliwell et al., 2023).

Pese a que la conciencia es mayor y pese a que se disponen de tratamientos farmacológicos y psicológicos basados en la evidencia que pueden aliviar significativamente los síntomas de muchos de los trastornos mentales, los sistemas de salud no están respondiendo adecuadamente a las personas con trastornos mentales. La mayoría de las personas que mueren por suicidio no han tenido contacto con los servicios de salud en los 12 meses previos. En un estudio en el que se explora la atención a la salud mental en 84 países, incluida España, (Moitra et al., 2022) se encontró que solo el 23% de las personas con depresión en los países con altos ingresos y el 3% en los países con ingresos medios y bajos recibían un tratamiento mínimamente adecuado para su problema. Comparar estos datos con otras condiciones médicas nos indica la poca atención prestada a la salud mental. En los países de altos ingresos (incluida España) el 62% de las personas con hipertensión o

diabetes reciben un tratamiento adecuado (Zhou et al., 2019).

El problema va más allá de un tema sanitario, se trata de un problema social y ético, ya que sin un tratamiento adecuado muchos de los trastornos mentales tienen consecuencias muy negativas que reducen significativamente el bienestar, la satisfacción y la calidad de vida y causan desigualdades sociales; por ejemplo, en España 8 de cada 10 personas con problemas de salud mental no tienen empleo, según el Instituto nacional de Estadística (INE, 2019). En definitiva, podemos considerar la problemática de la salud mental como una de las mayores fuentes de desigualdad e injusticia (Layard & Clark, 2014).

2.- TICs y salud mental

Como hemos comentado, disponemos de tratamientos farmacológicos y psicológicos que han demostrado eficacia en el tratamiento de los síntomas y consecuencias de sufrir un problema de salud mental. Sin embargo, pese a ello, tenemos un **problema de accesibilidad**, ya que los tratamientos no llegan a las personas que los necesitan. La falta de disponibilidad de las intervenciones para mejorar la salud mental es muy preocupante. Voces autorizadas como las de Kazdin nos indican algunas de las barreras: la falta de profesionales en los sistemas de salud (por ejemplo, en España sólo hay 5 psicólogos clínicos por cada 100.000

habitantes, muy por debajo de la media europea de 18) y los altos costes que tiene la forma de administrar los tratamientos psicológicos (sesiones individuales semanales de unos 50 minutos). Entre las posibles soluciones se encuentra aumentar el número de profesionales y organizar de forma más eficiente la provisión de servicios desde la prevención a la atención especializada, pasando por la atención primaria (Kazdin & Blase, 2011). Esto puede realizarse utilizando una estrategia por pasos en función de la gravedad como se ha hecho en Inglaterra con el programa *Improving Access to Psychological Therapies* con buenos resultados (*National Collaborating Centre for Mental Health, 2021*). Otras acciones incluyen innovar respecto a ir más allá de la terapia individual cara a cara aumentando la diversidad en las formas de administrar psicoterapia, siendo flexibles y ampliando la oferta de servicios. Es necesario diseñar intervenciones y métodos para aumentar el alcance y la escalabilidad de los tratamientos, que estén más integrados en la vida cotidiana de la persona, ampliando los entornos en los que se proporcionen intervenciones, contando con la aceptabilidad de los implicados (personas con trastornos mentales, familiares, profesionales de la salud mental, etc.). Estas innovaciones tienen que estar basadas en la evidencia científica y ser asequibles para los sistemas de salud (Kazdin, 2015).

La tarea es muy compleja, pero tenemos la responsabilidad de acometerla.

Una de las propuestas para hacer más accesibles las intervenciones psicológicas es el uso de las **Tecnologías de la Información y la Comunicación (TICs)**, y más concretamente Internet. Las ventajas incluyen poder llegar a un mayor número de personas que la terapia tradicional cara a cara. También está indicado para garantizar el acceso a personas para las que la terapia tradicional puede ser inaccesible (por distancia geográfica o dificultades de movilidad).

En los últimos 20 años, el interés de la investigación por las **intervenciones psicológicas online** se ha traducido en la publicación de un número creciente de estudios en distintas partes del mundo que avalan la eficacia de dichos tratamientos en distintos problemas psicológicos, sobre todo en los trastornos más prevalentes, la depresión y la ansiedad. Andersson et al., (2019) realizaron un estudio de metaanálisis en el que se incluían 166 ensayos clínicos que señalaban que la terapia online era tan eficaz como la terapia cara a cara. La forma más adecuada de recibir un tratamiento online es mediante la prescripción por parte de un profesional de la salud que tras una evaluación determina que el paciente puede beneficiarse del mismo y le ofrece esa posibilidad. El paciente puede entonces acceder mediante un usuario y

contraseña a un programa en una web en la que hay distintos módulos terapéuticos que le proporcionan aprendizaje y aplicación de estrategias para mejorar su estado clínico. Es posible combinar el tratamiento online con sesiones cara a cara; por ejemplo, el paciente va una vez al mes a su centro de salud y durante las restantes 3 semanas realiza el tratamiento online. Esas sesiones cara a cara se pueden realizar presencialmente o por videoconferencia. También es posible que el tratamiento sea autoaplicado (sin visitas cara a cara) y que el profesional haga una llamada corta de unos 10 minutos una vez a la semana para resolver dudas. Los estudios indican que para la depresión y ansiedad de gravedad leve y moderada este último formato es eficaz (p.ej., Domhardt et al., 2019). Introducir estos tratamientos como un servicio más en nuestros centros de salud en contextos de atención primaria y especializada puede contribuir ampliamente a aumentar el número de personas que se pueden beneficiar. No se trata de sustituir a los profesionales. La terapia online está indicada sobre todo para trastornos de ansiedad y depresivos de gravedad leve o moderada. Si se ofrecen estos tratamientos a ese perfil de pacientes, los profesionales tendrán más tiempo para poder ofrecer sesiones cara a cara más frecuentes a pacientes con problemas más graves. Se trata de hacer el sistema más eficiente. En la tabla 1 se ofrece un resumen de las posibilidades de aplicación de

los tratamientos administrados por Internet y los ámbitos de aplicación.

Tabla 1. Posibilidades de los tratamientos psicológicos administrados por Internet.

Forma de aplicación	Descripción	Ámbito de aplicación
Autoaplicado	El paciente accede al programa y se lo administra sin apoyo de un profesional.	Se puede utilizar mientras la persona se encuentra en lista de espera o en personas con síntomas de ansiedad y depresión subclínicos.
Autoaplicado + apoyo telefónico una vez por semana (10 minutos)	El paciente accede al programa y se lo administra con apoyo de un profesional que le contacta una vez por semana por teléfono (llamadas de 10 minutos para resolver dudas)	Trastornos de ansiedad y depresivos con gravedad leve-moderada. Ámbito de atención primaria y atención especializada.
Combinado (sesiones presenciales + online)	El paciente acude a sesiones presenciales con el profesional,	Trastornos de ansiedad y depresivos con gravedad

	<p>pero estas se reducen porque puede trabajar también en casa en el programa online (p.ej., en un mes recibe una sesión presencial cara a cara y las tres semanas restantes trabaja en casa con el programa online).</p>	<p>moderada. Ámbito de atención primaria y atención especializada.</p>
<p>Combinado online (sesiones por videoconferencia + online)</p>	<p>Lo mismo que el anterior pero las sesiones cara a cara se realizan por medio de videoconferencia.</p>	<p>Trastornos de ansiedad y depresivos con gravedad moderada. Ámbito de atención primaria y atención especializada.</p>
<p>Tratamiento online como tarea para casa</p>	<p>El paciente recibe un tratamiento tradicional cara a cara y el tratamiento online sólo sirve como tarea para casa en las sesiones presenciales</p>	<p>Trastornos de ansiedad y depresivos con gravedad moderada-grave. Ámbito de atención primaria y atención</p>

Los resultados de las investigaciones sobre el uso de TICs como forma de mejorar el acceso a las intervenciones son muy prometedores. Sin embargo, la posibilidad de que estas herramientas se implementen, es decir, que se ofrezcan de forma rutinaria, no es tan fácil.

En general, existe un gran problema en la translación de los estudios de eficacia de intervenciones basadas en la evidencia a la práctica clínica diaria. Se estima que el lapso es de 17-20 años (*Institute of Medicine, 2001*). Tradicionalmente, la investigación en la eficacia de las intervenciones se realiza con recursos de investigación humanos y monetarios externos a las infraestructuras clínicas en las que los pacientes reciben tratamientos, por lo que, una vez realizado el estudio, resulta difícil que se continúe realizando la intervención. Y cuando se consigue implementar la intervención, el lapso es excesivo. Es por eso por lo que ha surgido una línea de trabajo denominada **ciencia de la implementación** que está desarrollando métodos y procedimientos para conseguir que las intervenciones basadas en la evidencia se incorporen más rápidamente a la práctica clínica diaria. La ciencia de la implementación se define como “el estudio científico de métodos para promover la incorporación de

resultados de investigación y otras prácticas basadas en la evidencia en la práctica diaria, y, así mejorar la calidad y eficiencia de los servicios de salud” (Bauer et al., 2015).

El Protocolo Transdiagnóstico UJI online, desarrollado por nuestro grupo de investigación, LabPsittec en la Universitat Jaume I de Castellón es un buen ejemplo de un tratamiento psicológico online. Es una intervención dirigida a los trastornos emocionales: trastorno depresivo mayor, trastorno depresivo persistente o distimia, trastorno depresivo no especificado, trastorno de pánico, agorafobia, trastorno de ansiedad social, trastorno de ansiedad generalizada y trastorno obsesivo compulsivo. El protocolo ha sido adaptado a partir del Protocolo Unificado (Barlow et al., 2011) y de procedimientos de aceptación y atención plena de la Terapia Dialéctica Comportamental (Linehan, 2015). El paciente aprende diversas técnicas y estrategias que le permitan mejorar la capacidad para regular las emociones y reducir el uso de estrategias desadaptativas como la evitación o la falta de actividades significativas. El protocolo incluye componentes terapéuticos tradicionales basados en la evidencia para el tratamiento de los trastornos emocionales como la psicoeducación, la motivación para el cambio y la prevención de recaídas. También incluye componentes dirigidos a la regulación del afecto negativo: a) Atención plena centrada en el momento presente y aceptación, b)

flexibilidad cognitiva, c) conciencia de la evitación emocional y comportamental y d) exposición a las experiencias emocionales que se evitan porque causan malestar. Finalmente, y como una innovación con respecto a otros programas, se incluye un componente específico para la regulación del afecto positivo, con el objetivo de promover las fortalezas psicológicas y mejorar el bienestar.

Este protocolo se ha puesto a prueba en dos ensayos clínicos controlados, uno realizado en población diagnosticada con trastornos de ansiedad y depresivos que acudía a atención especializada en la sanidad pública en la Comunidad Valenciana y en Cataluña (González-Robles et al., 2020) y otro con personas con esos mismos diagnósticos, pero en un contexto comunitario, es decir, personas que sufrían esos trastornos, pero no estaban recibiendo tratamiento en ningún servicio de salud (Díaz-García et al., 2021). Fueron tratadas 394 personas en un formato autoaplicado complementado con llamadas telefónicas semanales de 5-10 minutos (el tiempo total de contacto telefónico no superó los 60 minutos de media, considerando el tratamiento completo). Los resultados de los dos ensayos indicaron que el tratamiento online tuvo una buena aceptación por parte de los pacientes y profesionales y que, respecto a su eficacia, fue superior a una condición de lista de espera o al tratamiento habitual. Y lo que es muy importante, se demostró que era posible

aplicar un tratamiento online en contextos en los que existen pocos recursos, indicando que este protocolo puede ser una herramienta útil para aumentar la accesibilidad a los tratamientos psicológicos en nuestro contexto más próximo.

Tras ofrecer el protocolo en distintos ámbitos para que sea utilizado en la práctica clínica diaria, por fin hemos podido transferirlo a un área de salud (en este caso en Cataluña) que comenzará a utilizarlo como un servicio más en su oferta de servicios de salud mental.

Nuestro grupo de investigación desarrolló su primer programa online “Háblame” para el miedo a hablar en público en el año 2000. Cumpliendo lo que señala la literatura, hemos tardado unos 20 años en conseguir llevar nuestros resultados de investigación a la práctica clínica diaria. Estamos muy satisfechos, pero esperamos que la mayor sensibilidad de toda la sociedad hacia la salud mental y hacia la potencialidad de las TICs haga posible que no tengamos que esperar todo ese tiempo para conseguir que todas las personas que lo necesiten se puedan beneficiar.

Otra innovación que complementa e incluso optimiza el uso de los tratamientos online es la inclusión de los **teléfonos inteligentes como apoyo a la evaluación, la detección, el diagnóstico y el tratamiento**. Mediante una App la persona puede recibir notificaciones

una o dos veces al día para informar de algunas variables relevantes como la intensidad de su tristeza o el impulso a consumir alcohol. A lo largo de los días podemos disponer de una evaluación más precisa ya que se realiza en el momento en el que sucede y en el lugar en el que sucede (en vez de como lo hacemos tradicionalmente, pasando un cuestionario retrospectivo en la consulta sobre lo que sintió la semana pasada). Es lo que llamamos **Evaluación Ecológica Momentánea** (EMA por sus siglas en inglés) y ha supuesto una revolución en el campo de la evaluación psicológica (Shiffman, Stone & Hufford, 2008). Y no sólo sirve para evaluar, sino que podemos enviar indicaciones a través del móvil en ese preciso instante para ayudar a la persona a manejar sus emociones o sus impulsos. Es lo que denominamos **Intervención Ecológica Momentánea** (EMI por sus siglas en inglés). Además, los datos recogidos por la EMA pueden complementarse con otros datos más objetivos mediante sensores (por ejemplo, tasa cardíaca o actividad física), mejorando todavía más la precisión de la evaluación y pudiendo ofrecer indicaciones más adecuadas y personalizadas (Mohr, Zhang & Schueller, 2017). Por ejemplo, si la persona nos informa mediante la App que su tristeza en los últimos tres días ha sido mayor de 8 y su acelerómetro nos indica que su movimiento ha disminuido significativamente podemos enviarle una indicación terapéutica adecuada. El uso de

EMIs se está revelando como una herramienta muy útil en otro de los aspectos propuestos para mejorar la atención a las personas con problemas de salud mental, recibir intervenciones personalizadas, adecuadas al estado clínico de la persona (Heron & Smyth, 2010). Los componentes terapéuticos no cambian, pero se administran en el momento en que la persona los necesita.

Nuestro grupo de investigación está trabajando en este ámbito también. Hemos desarrollado varias Apps para la evaluación ecológica momentánea de variables relevantes en una condición que tiene un impacto negativo en la salud mental como es el dolor crónico. El diseño, desarrollo y validación de la App Monitor de Dolor dio lugar a un procedimiento EMA validado psicométricamente que sirve para evaluar hasta 34 variables como la intensidad de la fatiga, el estado de ánimo, la catastrofización, la calidad del sueño o el nivel de actividad entre otros. Los resultados de nuestras investigaciones señalan que el uso de la App puede resultar muy beneficioso no sólo para una evaluación más precisa, diaria y realizada en el contexto natural, sino para, en función de la programación de alarmas poder dar indicaciones terapéuticas, por ejemplo, en el tratamiento de condiciones de dolor crónico como lumbalgia crónica o fibromialgia (Susó-Ribera et al., 2018) o en el manejo del estrés en profesionales sanitarios (Castilla et al., 2022).

3.- IA y salud mental

Tanto los tratamientos online como el uso de EMA y EMI conllevan la recogida de datos continuos que pueden ser muy relevantes para comprender mejor los problemas mentales. Es aquí donde está introduciéndose la gran revolución tecnológica de los últimos años que tanto nos atrae, pero también nos inquieta, la **Inteligencia Artificial**.

El término Inteligencia Artificial (IA) fue acuñado por McCarthy en la convención de Dartmouth en 1956 para referirse a la “disciplina de la informática o la ingeniería que se ocupa del diseño de sistemas inteligentes”.

La IA tiene potencial por lo que se refiere a la salud mental. Entre las posibilidades se indican las siguientes (p.ej., Brinker et al., 2019).

--Posibilidad de analizar datos de forma rápida y manejando una gran cantidad y variedad de datos (biológicos, psicológicos, sociales; medidas subjetivas, biomarcadores).

--Formular modelos de riesgo para determinar la predisposición a manifestar un problema mental.

--Desarrollar herramientas de screening para detectar primeros signos de problemas mentales.

--Uso de agentes conversacionales para

detección precoz, diagnóstico y tratamiento

La aplicación de la IA en salud mental es más compleja porque los profesionales de la salud mental se apoyan en datos de distintas fuentes para realizar los diagnósticos y tomar las decisiones terapéuticas. De hecho, no disponemos de ninguna prueba objetiva (por ejemplo, un indicador biológico medible) para determinar el diagnóstico de ninguno de los trastornos mentales. La evaluación se apoya bastante en las denominadas *soft-skills* como crear una relación terapéutica, la observación directa de las conductas del paciente y la propia visión del paciente de experiencias subjetivas.

Y es precisamente por esto por lo que la IA ofrece un gran potencial para redefinir el diagnóstico y la comprensión de los problemas mentales. De hecho, la IA se presenta no solo como una oportunidad para el diagnóstico sino como una herramienta de prescreening que puede ayudar de forma significativa a la prevención y detección temprana de problemas y a implementar intervenciones mucho más personalizadas en función de las necesidades de cada persona. La IA también permitiría poder tomar decisiones clínicas durante el tratamiento y ajustarlo, es decir, se trataría de una personalización no sólo teniendo en cuenta el perfil inicial del usuario sino su respuesta al tratamiento (Graham et al., 2020).

Existen ya experiencias en la utilización de la IA en salud mental (Le Glaz et al., 2021; Higgins et al., 2023). Los datos que se han introducido en la IA incluyen datos de salud en formato electrónico, medidas de autoinforme (p.ej., estado de ánimo percibido), datos de resonancia cerebral, datos de uso de smartphone, datos de comportamientos grabados en vídeo y datos del uso de redes sociales. Los problemas mentales en los que más se ha introducido la IA incluyen desde trastorno mental grave como la esquizofrenia hasta trastornos muy prevalentes como la depresión, pasando por problemas que generan gran alarma social como el suicidio. Los resultados son prometedores indicando que los algoritmos de machine learning (ML) tienen un potencial importante para detectar estos problemas de forma precisa.

En la investigación sobre la utilidad de la IA en salud mental también se está contemplando el **proceso de introducción de la IA en los servicios de salud mental**. Como hemos dicho existe una brecha temporal importante en medicina entre el desarrollo y validación de procedimientos basados en la evidencia y su adopción en la clínica diaria. Con el fin de que esa brecha se reduzca lo más posible con relación a la adopción de la IA en salud mental se recomienda sobre todo que se expliciten con la mayor transparencia posible los algoritmos utilizados a todos los agentes implicados (personas afectadas por problemas de salud

mental, profesionales de la salud mental, gestores de salud y sociedad en general) (Lu et al., 2023).

Es decir, pese a que el impacto potencial de la IA en la salud mental es incuestionable, los retos son también grandes. Los retos éticos como el uso de datos sesgados y garantizar que se respeten los principios de autonomía, beneficencia y justicia que deben guiar el desarrollo de la investigación en IA y salud. De Choudhury y Kiciman (2018) resumen las tareas a realizar:

- 1.-Integrar la IA con la inteligencia humana para asegurar validez de constructo,
- 2.-Detectar factores no observados en los datos,
- 3.-Evaluar el impacto de los datos sesgados
- 4.-Identificar de forma proactiva los errores que cometa la IA.
- 5.-Mitigar los errores encontrados.

Existe ya una disciplina específica para tratar estos aspectos en cualquier desarrollo de IA, se denomina **IA responsable**

Estos aspectos éticos no sólo se tienen en cuenta cuando se utiliza la IA sino en cualquier aplicación de la tecnología que hemos comentado. Una de las mejores garantías

cuando uno se aproxima a las soluciones TICs en salud es comprobar que las aplicaciones han sido validadas científicamente. Parece obvio, pero no es así. En una excelente revisión realizada por el grupo INIT de la UJI en la que nuestro grupo colaboró (Miralles et al., 2020) se encontró que menos de un 1% de las Apps para salud mental que se pueden descargar en nuestros móviles habían sido validadas mediante un ensayo clínico controlado. Es imperativo avanzar en la regulación de los requisitos que deben cumplir las soluciones TICs para salud en general y para salud mental en particular antes de que salgan al mercado.

4.- Conclusiones

La Organización Mundial de la Salud (OMS) define la salud mental como:

“Un estado de bienestar en el cual cada individuo desarrolla su potencial, puede afrontar las tensiones de la vida, puede trabajar de forma productiva y fructífera, y puede aportar algo a su comunidad.” (WHO, 2013).

Pese a las altas tasas de prevalencia de los trastornos mentales y la gran carga de enfermedad que suponen, y pese a que disponemos de tratamientos eficaces para estos problemas, las personas que los necesitan no los están recibiendo. Es urgente y una obligación ética y de justicia social mejorar el acceso a los tratamientos para los trastornos

mentales.

El uso de tecnologías de la información y la comunicación (TICs) como los tratamientos administrados por Internet y el uso de evaluación ecológica momentánea e intervención psicológica momentánea mediante teléfonos móviles pueden ayudar en la tarea de que los tratamientos sean más accesibles.

No se trata de sustituir a los profesionales. La terapia administrada mediante TICs está indicada sobre todo para trastornos de ansiedad y depresivos de gravedad leve o moderada. Si se ofrecen estos tratamientos a ese perfil de pacientes, los profesionales tendrán más tiempo para poder ofrecer sesiones cara a cara más frecuentes a pacientes con problemas más graves.

La inclusión de la Inteligencia Artificial (IA) puede mejorar aún más algunos aspectos relacionados con la mejora de la atención a los problemas de salud mental:

--Posibilidad de analizar datos de forma rápida y manejando una gran cantidad y variedad de datos (biológicos, psicológicos, sociales; medidas subjetivas, biomarcadores).

--Formular modelos de riesgo para determinar la predisposición a manifestar un problema mental.

--Desarrollar herramientas de screening para detectar primeros signos de problemas mentales.

--Uso de agentes conversacionales para detección precoz, diagnóstico y tratamiento

Nos encontramos en un momento en el que todavía hay mucho desconocimiento de las TICs y el potencial de la IA en el sector sanitario y en los usuarios en general, por lo que es necesaria la reestructuración del sistema sanitario para incluir las TICs y la IA, lo que conlleva una mayor inversión del gasto sanitario en tecnologías.

Para impulsar el proceso de adopción de las TICs y la IA en salud mental se recomienda que participen en ese proceso equipos interdisciplinarios: ciencias de la salud (psicología), ingenieros, derecho, economía, sociología, etc. También es importante incluir a usuarios (profesionales de la salud mental, gestores de la salud mental y personas afectadas por problemas de salud mental). Y todo ello con la guía de la IA responsable para asegurar los aspectos éticos.

REFERENCIAS

Andersson, G., Carlbring, P., Titov, N., & Lindefors, N. (2019). Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-

analyses. *Canadian Journal of Psychiatry*, 64(7), 465-470. doi: 10.1177/0706743719839381.

Barlow, D. H., Ellard, K. K., Fairholme, C. P., Farchione, T. J., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. (2011). *The unified protocol for transdiagnostic treatment of emotional disorders: Client and therapist workbooks*. New York: Oxford University Press.

Bauer, M. S., Damschroder, L., Hagedom, H., Smith, J., & Kilbourne, M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*, 3, 32. doi: 10.1186/s40359-015-0089-9.

Brinker T. J., Hekler A., Hauschild A., Berking C., Schilling, B., Enk, A. H., et al. (2019). Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. *European Journal of Cancer*, 111, 30–37. doi: 10.1016/j.ejca.2018.12.016.

Castilla, D., Navarro-Haro, M. V., Suso-Ribera, c., Díaz-García, A., Zaragoza, I. & Garcia-Palacios, a. (2022). Ecological Momentary Intervention for emotion regulation with healthcare workers via smartphone: A randomized controlled trial protocol. *BMC psychiatry*, 22, 164. doi: 10.1186/s12888-022-03800-x

Charlson, F. J., Baxter, A. J., Dua, T., Degenhardt, L., Whiteford, H. A., & Vos, T. (2015). Excess

mortality from mental, neurological and substance use disorders in the global burden of disease study 2010. *Epidemiology and Psychiatric Sciences*, 24, 121–140. doi: 10.1017/S2045796014000687.

De Choudhury, M., & Kiciman, E. (2018). Integrating Artificial and Human Intelligence in Complex, Sensitive Problem Domains: Experiences from Mental Health. *AI Magazine*, 39(3), 69-80. doi: 10.1609/aimag.v39i3.2815.

Diaz-Garcia, A., Gonzalez-Robles, A., Garcia-Palacios, A., Fernandez-Alvarez, J., Castilla, D., Breton, J., Baños, R. Quero, S., & Botella, C. (2021). Negative and positive affect regulation in a transdiagnostic Internet-based protocol for emotional disorders: Randomized controlled trial. *Journal of Medical Internet Research*, 23(2), e21335. doi: 10.2196/21335.

Domhardt, M., GeBlein, H., von Rezori, R. E., & Baumeister, H. (2019). Internet- and mobile-based interventions for anxiety disorders: A meta-analytic review of intervention components. *Depression and Anxiety*, 36(3), 213-224. doi: 10.1002/da.22860.

Graham, S., Depp, C., Lee, E. E., et al. (2020). Artificial Intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, 21(11), 116. doi:10.1007/s11920-019-1094-0.

Gonzalez-Robles, A., Diaz-Garcia, A., Garcia-Palacios, A., Roca, P., Ramos-Quiroga, J. A., & Botella, C. (2020). Effectiveness of a transdiagnostic guided Internet-delivered protocol for emotional disorders versus treatment as usual in specialized care: Randomized controlled trial. *Journal of Medical Internet Research*, 22(7), e18220. doi: 10.2196/18220.

Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., De Neve, J.-E., & Wang, S. (Eds.). (2023). *World Happiness Report 2023* (11th ed.). Sustainable Development Solutions Network.

Heron, K.E., & Smyth, J.M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, 15, 1–39. doi:10.1348/135910709X466063.

Higgins, O. & Brooke, L. (2023). Artificial Intelligence and machine learning based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, 32, 966-978. doi: 10.1111/inm.13114

Instituto Nacional de Estadística (2019). *El empleo de las personas con discapacidad*.

Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the*

21st Century.

Kazdin, A. E. (2015). Technology-based interventions and reducing the burdens of mental illness: Perspectives and comments on the special series. *Cognitive and Behavioral Practice*, 22(3), 359–366. doi: 10.1016/j.cbpra.2015.04.004.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives in Psychological Science*, 6(1), 21-37. doi: 10.1177/1745691610393527.

Layard, R., & Clark, D. M. (2014). *Thrive. The power of evidence-based psychological therapies*. Allen Lane: Penguin Books.

Le Glaz et al. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of Medical Internet Research*, 23: e15708. doi: 10.2196/15708

Linehan, M. M. (2015). *Skills training manual for treating borderline personality disorder*. 2nd edition. New York, NY: Guilford Press.

Lu, T. et al (2023). Bridging the gap between artificial intelligence and mental health. *Science Bulletin*, 68, 1606-1610. Doi: 10.1016/j.scib.2023.07.015

Miralles, I., Granell, C., Díaz-Sanahuja, L., Van Woensel, W., Bretón-López, J., Mira, A., Castilla,

D., & Casteleyn, S. (2020). Smartphone Apps for the Treatment of Mental Disorders: Systematic Review. *Journal of Medical Internet Research Mhealth Uhealth*, 8(4), e14897. doi: 10.2196/14897.

Mohr, D.C., Zhang, M., & Schueller, S.M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, 13, 23–47, doi:10.1146/annurev-clinpsy-032816-044949.

Moitra, M., Santomauro, D., Collins, P. Y., Vos, T., Whiteford, H., Saxena S, et al. (2022). The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: A systematic review and Bayesian meta-regression analysis. *PLoS Medicine*, 19(2), e1003901. doi: 10.1371/journal.pmed.1003901.

National Collaborating Centre for Mental Health (2021). *The improving access to psychological therapies manual*.

Patel, V., Saxena, S., Lund, C., et al. (2018). The Lancet commission on global mental health and sustainable development. *Lancet*, 392, 1553–1598. doi: 10.1016/S0140-6736(18)31612-X.

Plan de Acción de Salud Mental 2022-2024. Ministerio de Sanidad.

[https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/docs/saludmental/PLAN ACCION SALUD MENTAL 2022-2024.pdf](https://www.sanidad.gob.es/organizacion/sns/planCalidadSNS/docs/saludmental/PLAN_ACCION_SALUD_MENTAL_2022-2024.pdf)

Rehm J., & Shield, K.D. (2019). Global burden of disease and the impact of mental and addictive disorders. *Current Psychiatry Report*, 2, 10. doi: 10.1007/s11920-019-0997-0.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4, 1–32, doi:10.1146/annurev.clinpsy.3.022806.091415.

Suso-Ribera, C., Castilla, D., Zaragoza, I., Ribera-Canudas, M.V., Botella, C., & García-Palacios, A. (2018). Validity, reliability, feasibility, and usefulness of pain monitor. *The Clinical Journal of Pain*, 34(10), 900-908. doi: 10.1097/AJP.0000000000000618.

Vos, T., Lim, S. S., Abbafati, C., et al. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*, 396, 1204–22. doi: 10.1016/S0140-6736(20)30925-9.

World Health Organization (2013). *Comprehensive Mental Health Action Plan 2013–2020*. Geneva: WHO.

World Health Organization (2021). *Mental health atlas 2020*. Geneva: WHO.

World Health Organization (2022). Mental Health and COVID-19: Early evidence of the pandemic's impact. Geneva: WHO.

Zhou, B., Danaei, G., Stevens, G. A., et al. (2019). Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys. *Lancet*, 394, 639-651. doi: 10.1016/S0140-6736(19)31145-6.

