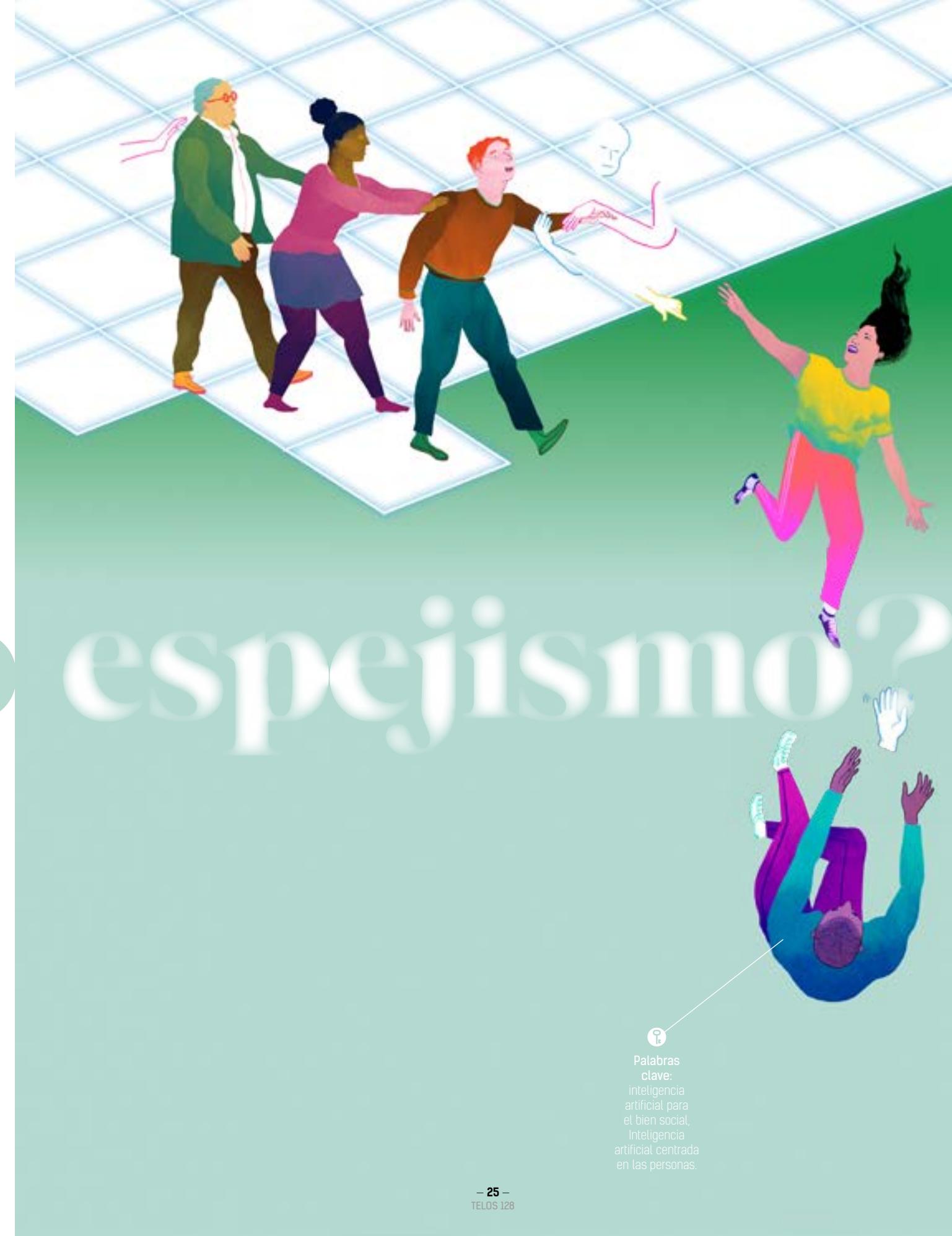


UNA IA SOSTENIBLE, EQUITATIVA
Y QUE AYUDE A ALCANZAR NUESTRO
MÁXIMO POTENCIAL COMO SERES
HUMANOS ES POSIBLE

Inteligencia artificial y derechos digitales: ¿progreso o espejismo?

Ni es más inteligente que los humanos, ni sirve para todo. Pero, si somos conscientes de sus limitaciones e implantamos controles que mitiguen sus riesgos, la inteligencia artificial puede convertirse en una aliada que nos permita vivir mejor y, sobre todo, sin perder lo que nos define como personas.

TEXTO: NURIA OLIVER
ILUSTRACIÓN: KEVIN WARD



Palabras
clave:
inteligencia
artificial para
el bien social,
Inteligencia
artificial centrada
en las personas.

La inteligencia artificial (IA) hoy está presente, de manera invisible, en prácticamente todos los ámbitos de nuestra sociedad. Su adopción, a escala masiva, no tiene precedentes y el potencial que nos brinda para tener impacto social positivo es inmenso. De hecho, necesitamos apoyarnos en la IA para abordar los grandes retos del siglo XXI, desde el cambio climático al envejecimiento de la población. Nos puede ayudar, por ejemplo, a detectar tumores antes de que sean visibles al ojo humano en pruebas radiológicas, a reducir el consumo energético de nuestras ciudades, hogares y empresas, o a encontrar patrones en datos científicos que a los humanos nos llevaría décadas descubrir.

Una de las aspiraciones es que la IA nos haga más humanos al liberarnos de tareas repetitivas, mecánicas y de poco valor. Dicha aspiración es posible, pero su consecución no está garantizada. De hecho, si dejamos que la IA tome

decisiones por nosotros en exceso, corremos el riesgo de atrofiar nuestras propias habilidades: el pensamiento crítico, la empatía, la creatividad, la comprensión lectora, la capacidad de escribir... Esto puede conducirnos no solo a una dependencia silenciosa, sino a la pérdida de habilidades que nos definen como humanos, como nuestra capacidad para utilizar el lenguaje.

No es la solución para todo

Además, vivimos en un momento de cierto *solucionismo tecnológico* que nos empuja a recurrir a la IA incluso donde una solución más simple sería más eficaz, entendible y sostenible y también menos arriesgada. Por ejemplo, ¿realmente necesitamos un asistente inteligente para encender las luces de nuestra casa o un modelo complejo

para organizar un calendario que podríamos gestionar con una sencilla aplicación? Innovar no es acumular tecnología, sino mejorar de verdad la vida de las personas. En muchos casos, un buen diseño de procesos o una interfaz clara vuelven innecesarios millones de parámetros entrenados con cantidades ingentes de datos y con un consumo energético inmenso.

Una pregunta que surge con frecuencia es si la IA es más inteligente que nosotros, los humanos. Sin duda, no lo es. Los sistemas de IA actuales son excelentes realizando tareas muy concretas como jugar al ajedrez, clasificar imágenes, traducir de un idioma a otro, predecir patrones y generar texto plausible. Probablemente, sean mejores que el mejor de los humanos, pero solo saben hacer dicha tarea: es lo que se conoce como "inteligencia artificial específica". Esos mismos sistemas que tanto nos impresionan carecen, entre otras cosas, de comprensión del contexto, de modelo del mundo, de sentido común, de emociones, de capacidad para aprender constantemente y adaptarse, de conciencia... Son programas de ordenador, sin cuerpo y sin experiencia vital en el mundo físico. Y aunque sintamos la necesidad de compararnos, y en algunos casos nos parezca que su inteligencia supera a la nuestra, es importante recordarnos que dicha comparación no tiene sentido. Un debate más productivo es preguntarnos cómo podemos aprovechar las innegables capacidades de la IA para que los humanos desarrollemos todo nuestro potencial.

Original, pero no creativa

El debate se intensifica con la irrupción de las técnicas de IA generativa, capaces de crear imágenes, vídeos, música, texto, código o incluso moléculas de manera automática y casi

instantánea. ¿Puede ser la inteligencia artificial creativa? Ese contenido puede ser original y sorprendente, pero no nace de experiencias, emociones o intenciones propias. Es el resultado de procesar y recombinar patrones aprendidos de millones de obras previas. Un pintor humano no puede decidir romper con todas sus influencias por una convicción íntima; un modelo de IA, por definición, no tiene convicciones. Eso no significa que no pueda ser útil en procesos creativos: puede inspirar, explorar variaciones y desbloquear ideas, pero su creatividad es instrumental. La creatividad humana, en cambio, es vivencial: lleva consigo el peso de nuestras historias, ideas, emociones, deseos o miedos. No olvidemos que la expresión creativa humana es mucho más que la obra creativa, producto de dicha creación. La expresión creativa es

también el proceso, el mensaje, la intencionalidad, el contexto... elementos fundamentales del proceso que no existen cuando un algoritmo de IA generativa produce un poema, canción o diseño gráfico, por muy sofisticados que sean.

El entusiasmo colectivo con el que hemos adoptado la inteligencia artificial en nuestra sociedad viene acompañado a menudo por una peligrosa idea: que la IA es siempre sinónimo de progreso, entendido como mejora de la calidad de vida de las personas (de todas, no solo de algunas), así como del resto de seres vivos del planeta y el planeta en sí mismo. Una IA que optimiza la logística de un almacén puede resultar útil, pero si para lograrlo recurre a una vigilancia excesiva, consume cantidades ingentes de agua potable y electricidad y precariza el trabajo humano, ¿podemos llamarlo progreso? ■■■



Si dejamos que la inteligencia artificial tome decisiones por nosotros en exceso, corremos el riesgo de atrofiar nuestras propias habilidades

No debemos olvidar que no todo desarrollo tecnológico es progreso

En última instancia, la IA debería evaluarse no por lo impresionante que pueda ser su rendimiento, sino por su impacto en la sociedad.

Además, los sistemas de inteligencia artificial no son perfectos. Adolecen de limitaciones importantes que, entre otras cosas, impactan directamente en los derechos fundamentales de las personas, como son la discriminación, los sesgos y la estereotipación algorítmica; la falta de transparencia, diversidad y veracidad; la violación de la privacidad; la excesiva huella de carbono; la manipulación subliminal de nuestro comportamiento; la difícil reproducibilidad; y las vulnerabilidades de seguridad.

Tomemos, como ejemplo, el reto de la discriminación. Los sistemas de IA actuales están basados en métodos de aprendizaje estadístico que necesitan ser entrenados con cantidades ingentes de datos, datos que en muchos casos reflejan nuestras desigualdades y prejuicios. Por tanto, si no corregimos los sesgos, estereotipos y patrones de discriminación existentes en la sociedad, los algoritmos de IA, entrenados con datos humanos, no solo los replicarán, sino que en muchos casos los amplificarán. Así darán lugar, por ejemplo, a sistemas de contratación que penalizan sin querer a mujeres o minorías, aplicaciones de reconocimiento facial que fallan más en personas con ciertos tonos de piel o sistemas de

recomendación que nos encierran en burbujas de información y cámaras de resonancia que refuerzan nuestras propias creencias y prejuicios.

La discriminación automatizada no es más aceptable que la humana; de hecho, puede ser más peligrosa por su inmensa escalabilidad y su aparente —que no real— objetividad matemática. Además, las consecuencias negativas fruto del uso de la IA no solo son resultado de una negligencia o diseño deficiente, sino que también pueden ser intencionales, como cuando se usa la IA para manipular elecciones o vigilar masivamente a la población. Pero en todos los casos, el daño es real.

El código FATE

El carácter transversal de la IA, es decir, su capacidad para ser aplicada prácticamente en todos los ámbitos y sectores de la sociedad, la dota de un inmenso poder, lo que nos obliga a preguntarnos cómo mitigar sus riesgos, invirtiendo en investigación en inteligencia artificial responsable, acompañando su despliegue en la sociedad con regulaciones sólidas, auditorías independientes y un compromiso ético firme e integrando los derechos digitales desde su concepción. Para recordar qué características debemos exigir a los sistemas de IA me gusta utilizar el acrónimo en inglés FATE:

1. F de *fairness* o justicia. Exigir garantías de no discriminación, sesgos o estereotipación fruto del uso de sistemas de IA.

2. A de *autonomy* o autonomía. Según este valor central en la ética occidental, cada persona debería tener la capacidad de decidir sus propios pensamientos y acciones, asegurando por tanto la libre elección, la libertad de pensamiento y de acción. Sin embargo, hoy en día podemos construir modelos computacionales de nuestros deseos, necesidades, personalidad y comportamiento con la capacidad de influir en nuestras decisiones y acciones de manera subliminal, como ha quedado patente en los procesos electorales de Estados Unidos y el Reino Unido. Por ello, deberíamos garantizar que los sistemas inteligentes tomen las decisiones preservando siempre la autonomía y la dignidad humanas.

La A también es de ***accountability*** o atribución de responsabilidad. Es decir, se trata de tener claridad a la hora de atribuir responsabilidad de las consecuencias de las decisiones algorítmicas. Y de ***augmentation*** o aumento, de manera que los sistemas de IA se utilicen para aumentar o complementar la inteligencia humana, no para reemplazarla.

3. T de *trust* o confianza. La confianza es un pilar básico en las relaciones entre humanos e instituciones. La tecnología necesita un entorno de confianza con sus usuarios que cada



vez delegan (delegamos) más nuestras vidas a servicios digitales. Para que exista confianza, han de cumplirse tres condiciones: (1) la ***competencia***, es decir, la habilidad para realizar con solvencia la tarea comprometida; (2) la ***fiabilidad***, es decir, la competencia sostenida en el tiempo; y (3) la ***honestidad y transparencia***.

Por ello, la T también es de ***transparency*** o transparencia. Hace referencia a la calidad de poder entender un modelo o un proceso computacional. Estos serán transparentes si una persona puede observarlos y entenderlos con facilidad. Las decisiones algorítmicas pueden ser opacas por tres motivos: ***intencionalmente***, para proteger la propiedad intelectual de los creadores de dichos algoritmos; debido a una ***falta de conocimiento*** por parte de los usuarios que les impida entender cómo funcionan los algoritmos y modelos computacionales construidos a partir de los datos; e ***intrínsecamente***, dado que ciertos métodos de aprendizaje por ordenador (por ejemplo, modelos de *deep learning* o aprendizaje profundo) son extremadamente complejos. Asimismo, es imprescindible que los sistemas de inteligencia artificial sean transparentes no solo con relación a qué datos captan y analizan sobre el comportamiento humano y para qué propósitos, sino también respecto a en qué situaciones los humanos están interaccionando con sistemas artificiales (por ejemplo, *chatbots*) o con otros humanos. ■■■

Es posible que el contenido creado por una IA sea original y sorprendente, pero no nace de experiencias, emociones o intenciones propias



La discriminación automatizada puede ser más peligrosa que la humana por su inmensa escalabilidad y su aparente (que no real) objetividad matemática

4. E de education o educación. Es decir, invertir en educación a todos los niveles, empezando por la educación obligatoria. En el libro *Los nativos digitales no existen. Cómo educar a tus hijos para un mundo digital* (Deusto, 2017) escribí un capítulo llamado “Eruditos Digitales” donde enfatizo la necesidad de enseñar tanto pensamiento computacional desde primero de primaria, como de desarrollar el pensamiento crítico, la creatividad y habilidades de las inteligencias social y emocional. Hoy en día no estamos desarrollando estas facultades, pero cada vez van a resultar más importantes para nuestra salud mental y nuestra coexistencia pacífica y armoniosa tanto con la tecnología como con otros humanos y con nuestro planeta. La adopción masiva de herramientas de inteligencia artificial generativa de textos, música, audios o vídeos cuestiona los modelos tradicionales de enseñanza y evaluación, dado que cualquier estudiante puede, en cuestión de segundos, producir redacciones, comentarios, resúmenes o artículos

—veraces o no— sobre prácticamente cualquier tema y en cualquier idioma.

El nivel de competencia de estos sistemas nos ha sorprendido a todos y todas, expertos y legos en la materia. Su falta de rigor y veracidad hace, hoy más que nunca, relevante la necesidad de desarrollar el espíritu crítico y de validar y contrastar la información con fuentes reputadas y solventes. Sin duda, cada vez más a menudo, no podremos ni deberemos creer todo lo que leemos, escuchamos o vemos en el mundo digital. La educación también es necesaria para la ciudadanía, los profesionales —sobre todo a aquellos cuyas profesiones están siendo transformadas por la tecnología—, los trabajadores del sector público y nuestros representantes políticos. Una apuesta ambiciosa por la educación en competencias tecnológicas es vital para poder reducir la situación de *asimetría* en la que nos encontramos inmersos hoy en día: asimetría con respecto al acceso a los datos y, aún más importante, con respecto al acceso al conocimiento experto para saber

qué hacer con dichos datos. Comparto las palabras de Marie Curie: “Nada en la vida debería temerse, sino entenderse. Ahora es momento de entender más para así temer menos”.

La E también es de **bEneficence** o **bEneficencia**, es decir, maximizar el impacto positivo del uso de la inteligencia artificial, con *sostenibilidad, diversidad, honestidad y veracidad*. Porque no olvidemos que no todo desarrollo tecnológico es progreso. Y a lo que deberíamos aspirar y en lo que deberíamos invertir es en el progreso, entendido como una mejora en la calidad de vida de las personas, del resto de seres vivos y de nuestro planeta.

Y de **Equidad**. El desarrollo y crecimiento de internet y de la World Wide Web durante la Tercera y Cuarta Revolución Industrial ha sido sin duda clave para la democratización del acceso al conocimiento. Sin embargo, los principios de universalización del conocimiento y democratización del acceso a la tecnología están siendo cuestionados hoy en día en gran parte por la situación de dominancia extrema de las grandes empresas tecnológicas americanas y chinas. Es el fenómeno llamado “*winner takes all*” (el ganador se lo lleva todo). Juntos, estos gigantes tecnológicos tienen un valor de mercado de más de 5 millones de millones (billones) de euros y unas cuotas de mercado en EE. UU., de más de un 90 % en las búsquedas de internet (Google), de un 70 % de las redes sociales (Facebook) o de un 50 % del comercio electrónico (Amazon).

En consecuencia, un elevado porcentaje de los datos sobre el comportamiento humano existentes hoy en día son datos privados, captados, analizados y explotados por estas grandes empresas tecnológicas que conocen no solo nuestros hábitos, necesidades,

intereses o relaciones sociales, sino también nuestra orientación sexual o política, nuestros niveles de felicidad, de educación o de salud mental. Por tanto, si queremos maximizar el impacto positivo del desarrollo tecnológico y, en particular, de la inteligencia artificial en la sociedad —dado que dicha inteligencia necesita datos para poder aprender— deberíamos plantearnos nuevos modelos de propiedad, gestión y regulación de los datos. Sin embargo, la complejidad en su aplicación práctica pone de manifiesto la dificultad para definir e implementar el concepto de “propiedad” cuando hablamos de un bien intangible, distribuido, variado, creciente, dinámico y replicable infinitas veces a coste prácticamente cero.

5. N de non-maleficence o no maleficencia. Se trata de minimizar el impacto negativo que pueda derivarse del uso de las decisiones algorítmicas. Para ello, es importante aplicar un principio de *prudencia*, así como *garantizar la seguridad, fiabilidad y reproducibilidad* de los sistemas, preservando siempre la *privacidad* de las personas.

Solo cuando respetemos estos requisitos seremos capaces de avanzar y conseguir una inteligencia artificial socialmente sostenible, por y para las personas, ayudándonos a desarrollar nuestro potencial y salvaguardando lo que no queremos perder. Porque el verdadero avance no será el que nos deslumbre con algoritmos más rápidos o modelos más grandes, sino el que nos permita vivir mejor, con más justicia, más libertad, prosperidad y, sobre todo, sin perder lo que nos define como seres humanos.

Precisamente, para contribuir a hacer realidad esa visión es para lo que hemos creado ELLIS Alicante. Les invito a unirse a nuestra causa, apoyando el trabajo de nuestra fundación.

Autora



NURIA OLIVER

Doctora en IA por el Instituto Tecnológico de Massachusetts (MIT) e ingeniera de Telecomunicación por la Universidad Politécnica de Madrid (UPM). Ha sido directora de Investigación en Telefónica I+D y Vodafone. Actualmente dirige ELLIS Alicante, fundación dedicada a la investigación en IA ética, responsable y para el bien social.

English

A sustainable, equitable AI that helps us reach our full potential as human beings is possible.

ARTIFICIAL INTELLIGENCE AND DIGITAL RIGHTS: PROGRESS OR MIRAGE?

It is neither smarter than humans nor useful for everything. However, if we are aware of its limitations and implement controls to mitigate its risks, artificial intelligence can become an ally that allows us to live better and, above all, without losing what defines us as people.

Keywords: human-centric AI, AI for social good.

Bibliografía

Baidal, M., Derner, E., Oliver, N. *Guardians of Trust: Risks and Opportunities for LLMs in Mental Health*. Vienna, Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI), 2025. Disponible en: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=VJCMGYAAAAJ&sortby=pubdate&citation_for_view=VJCMGYAAAAJ:NxmKEeNBbOMC

Favero, L., Pérez-Ortiz, J. A., Käser, T., Oliver, N. “Do AI tutors empower or enslave learners? Toward a critical use of AI in education” en *arXiv preprint* (2025). Disponible en: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=VJCMGYAAAAJ&sortby=pubdate&citation_for_view=VJCMGYAAAAJ:EseWqaRxkBgC

Favero, L., Pérez-Ortiz, J. A., Käser, T., Oliver, N. “Enhancing critical thinking in education by means of a Socratic chatbot” en *International Workshop on AI in Education and Educational Research* (2024, pp. 17-32). Disponible en: https://www.researchgate.net/publication/383911327_Enhancing_CriticalThinking_in_Education_by_means_of_a_Socratic_Chatbot

Lepri, B., Oliver, N., Pentland, A. “Ethical machines: The human-centric use of artificial intelligence” en *iScience* (2021, 24(3), p. 102249). Disponible en: <https://ellisalicante.org/publications/lepri2021ethical-en/>

Oliver, N. “Governance in the era of data-driven decision-making algorithms”. En: González, A. & Jansen, M. (Eds.). *Women Shaping Global Economic Governance*. London: Centre for Economic Policy Research, 2019. Disponible en: https://cepr.org/system/files/publication-files/60131-women_shaping_global_economic_governance.pdf

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deleataille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., Saramäki, J., Scarpino, S. V., Tizzoni, M., Verhulst, S., Vincik, P. “Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle” en *Science Advances* (2020, 6(23), eabc0764). Disponible en: <https://ellisalicante.org/publications/oliver-2020mobile2-en/>