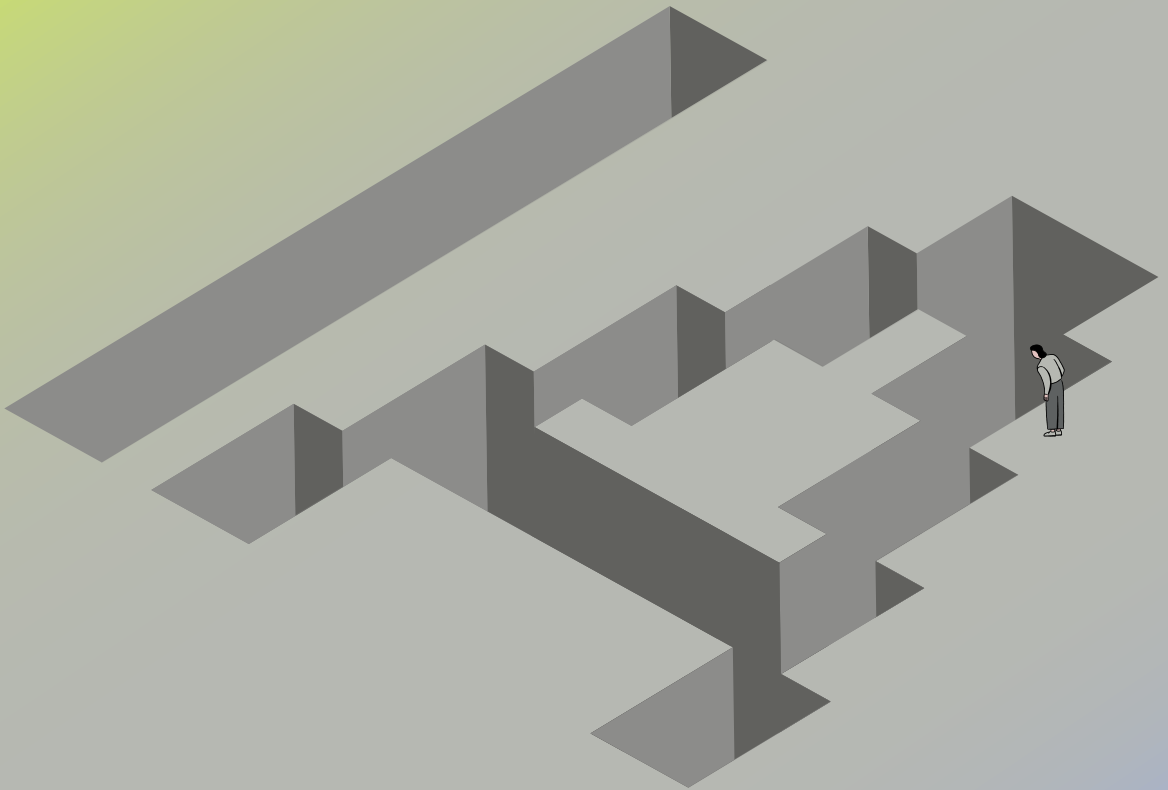


¿Quién acciona la máquina?

La IA, el periodismo y (el agotamiento de) los datos en la era de la *enshittification* de internet

INFORME EMILI GISBERT



Edita

Unió de Periodistes Valencians

Autora

Sara Calvo Tarancón

Diseño y maquetación

Sud+Sud

Impresión

Toniburguera slu

Depósito legal

V-3925-2024

Prólogo	8
Abstract	10
Prefacio	12
12	Un sondeo sobre inteligencia y artificialidad
14	La <i>enshittification</i> de nuestro todo: internet muere donde la IA nace
Introducción	18
18	Ni tan inteligente ni tan artificial
20	El misterio de la explicabilidad: ¿Cómo funciona la IA generativa?
Ética y tecnología	24
24	La conciencia (¿ya no?) Es patrimonio de la humanidad
26	La creatividad hace humana a la máquina: antropomorfización y AGI (Artificial General Intelligence)
Historia	30
30	Antecedentes, teoría e historia: Ada Lovelace, Turing, Altman
32	Todo lo que necesitas es atención: neuronas sintéticas, inteligencia y máquinas
IA y periodismo	36
36	Radiografía del sistema actual: gloria y decadencia, éxitos y despidos, y la reinención constante de las redacciones
38	La sentencia de muerte del gran buscador
40	El caso de los medios valencianos y el ejemplo de las corporaciones públicas
41	El periodismo local en À Punt y las voces sintéticas
44	La excepcionalidad del New York Times: la denuncia a las tecnológicas por “usar su contenido sin pagar”
46	Si hay dinero, no hay pleitos
47	Los datos y el fin del petróleo
50	De Gizmodo a La Nación: IA generativa como problema y también como solución
54	Protocolos para un uso de la IA responsable
53	La difícil convivencia entre la IA y los derechos de autor
54	Quién expolia a quién: la propiedad intelectual en la era de internet
57	La transparencia del algoritmo: cuestiones sobre la ética de la IA, los vikingos negros y el sobreajuste de los modelos
59	Dispensadores de contenidos a la carta en época electoral: el papel de las plataformas de difusión en la contención de la desinformación

Legislación y marco regulatorio 64

- 64 Mecanismos legales de control de la desinformación
- 66 IA para desenmascarar a la IA o el robot que caza robots
- 67 Herramientas de verificación disponibles y fiables
- 68 Legislar a priori: AI Act, la primera ley que pone coto a la Inteligencia Artificial
- 70 Rojo, amarillo y verde: la división legal según el riesgo de cada producto con IA
- 72 Privacidad y seguridad en las excepciones de la AI Act: el debate sobre la captación biométrica en las calles y en las fronteras
- 75 El caso de iBorder Control: automatismos en la gestión fronteriza

Transparencia, tecnosolucionismo y cambio climático 76

- 76 Dinero público, código público: el poder institucional y los algoritmos automatizados
- 77 Del tecnosolucionismo al incremento de la crisis climática
- 79 La importancia del código de la IA
- 80 Inteligente, artificial y de código abierto
- 83 El bien común, la privacidad y el capitalismo de plataformas

Una forma de concluir 86

- 86 Entre la amenaza colosal y la oportunidad
- 86 “Somos inteligentes porque vivimos”

Bibliografía 91

Glosario de entrevistas 92

Glosario de términos 93

La IA que no podrá sustituir al periodismo (de verdad)

La Unió de Periodistes Valencians decidió encargar este texto ante el auge y la incidencia, cada vez mayor, de la Inteligencia Artificial en los procesos informativos. Un reto que sabíamos que quedaría desactualizado en poco tiempo debido a la velocidad con que evoluciona, pero que pretendía sentar una base para entender la actualidad y poseer una visión de sus inicios en el futuro.

Una de las definiciones del Periodismo asegura que su función es sacar a la luz lo que alguien no quiere que se sepa. Para ello hay que construir relaciones, generar confianza, disponer de informaciones de diversas fuentes y actuar casi en la clandestinidad. Casi todo (si exceptuamos el tercer apunte) únicamente atribuible a los seres humanos.

Escuchamos en 2023 a Julia Otero en Onda Cero leer una noticia sobre cómo la tecnología destruiría puestos de trabajo, en un texto que hubiera pasado por actual si al finalizarlo no hubiera dicho la presentadora que pertenecía a la época de la Revolución Industrial.

Desde entonces, y las cifras son incontestables, el número de personas empleadas es el mayor de la historia. Y con cada revolución técnica ha ocurrido exactamente la misma circunstancia, pese a posarse en la sociedad el mismo tipo de miedo.

Los ERE realizados desde 2007 en los medios de comunicación no los han hecho máquinas, sino personas. Quienes, seguramente, de haber contado con más datos o alternativas quizá hubieran tomado decisiones diferentes. O adelantado otras que hoy demuestran que un medio puede vivir sin depender de la publicidad.

Y no solo un medio. Cuando algunos miembros de la Unió comenzaron a estudiar la carrera de Ciencias de la Información, cerca del año 2000, únicamente podían plantearse un horizonte laboral en radios, televisiones y periódicos clásicos. Hoy, pese a los despidos, existen perfiles que viven de comunicar en sus redes sociales, en Twitch, en Youtube, en un podcast o a través de una newsletter.

¿Puede la Inteligencia Artificial suponer un riesgo para los periodistas? Sin duda. En primer lugar, porque muchas empresas buscarán la optimización de noticias (que no de informaciones. El matiz es fundamental) en base a esta tecnología. Y en segundo, porque para el ejercicio que estamos realizando hoy día donde se premia el copia y pega y el clickbait, poca especialización es requerida.

Sin embargo, vayamos al lado opuesto. ¿Puede la IA contribuir a mejorar las prestaciones de una redacción pequeña? ¿Es (ya sabemos que sí) capaz de encontrar datos y patrones que supongan noticiabilidad y que el ojo humano tardaría días o incluso meses en descubrir? Con un uso generalizado, ¿permitirá acceder a informaciones que de otro modo permanecerían ocultas?

Volvamos a la definición de Periodismo, que también engloba saber hacer las preguntas correctas a las personas adecuadas. O, en este caso, a las máquinas adecuadas.

Si estamos apostando por un Periodismo de soluciones, hagámoslo también con la Inteligencia Artificial. Formémonos. Veamos en qué puede ayudarnos. Y aprovechemos una era donde, si somos capaces de dominar la tecnología antes de que ella nos domine a nosotros, es posible que consigamos dar (de nuevo) un salto hacia adelante en una profesión que necesita recuperar la credibilidad. Sobre todo, porque pronto será difícil discernir lo que es verdad de lo que no lo es.

Y ahí, siempre, debe haber un periodista vigilante.

Abstract

La irrelevancia del periodismo es una amenaza mayor que cualquier supercomputadora *omnimodal* de avanzadas redes neuronales sintéticas. De la profesión dependerá diferenciarse de la *enshittification* o buscar inteligentes excusas artificiales con las que justificar sus prácticas poco éticas y de dudosa calidad informativa. Nadie –ni los medios– se salva de convivir con nuevos modelos y herramientas de inteligencia artificial que van incorporando como parte de su día a día. Esto es algo que ha pasado históricamente con la tecnología, y para el neurocientífico Mariano Sigman, la clave es que la novedad no nos reemplace sino que nos multiplique.

Esta puede ser –o está siendo ya– la enésima revolución que lo cambiará todo. Nuestra investigación es una foto fija del arranque del año 2024, una narración coral en la que participan por supuesto periodistas, tanto de València como de fuera, perfiles técnicos, expertas en desinformación, programadores informáticos, investigadoras en el campo de la tecnología y los derechos humanos, abogadas, juristas, jefes de sección, responsables de informativos, redactoras, etc. A lo largo de estas páginas tratamos de sumar opiniones que ayuden a redefinir una profesión periodística muy precarizada y a ratos denostada, pero que sigue siendo extremadamente necesaria. Rigor, credibilidad, confianza, veracidad, imparcialidad son cualidades clásicas y atemporales del periodismo que pueden ser reforzadas con las potencialidades de la IA. Porque, sobre todo, esta es una gran oportunidad para mejorar el periodismo.

El cambio en la forma de comunicar ya se viene dando en los últimos años desde la irrupción de internet en nuestras vidas. La necesidad imperiosa de formarse y dominar la tecnología para que no sea ella quien nos domine es una máxima de estos tiempos. La resignificación de esta profesión es esencial, y se debe aprovechar el impulso de la IA para dignificar la importancia del buen periodismo.

Estas ideas convergen en esta investigación, en la que sobresale una reflexión de fondo sobre qué es hoy en día internet, y en qué lo hemos convertido. Internet, como dice el escritor Cory Doctorow, no es lo más importante de todo pero sí es el campo de batalla, el terreno donde libramos las luchas, las peleas por lo que de verdad es importante. La web, los foros, las redes sociales llevan inmersas en un proceso lento y constante de decadencia palpable en el que predomina una maraña de contenido sintético, artificial y poco inteligente que busca el relleno, el clickbait, y el control de los patrones de consumo. Porque internet es, sobre todo, un lugar de mercadeo capitalista, además del principal terreno –si no el más importante– en el que ocurre, se desarrolla, se produce, se consume y se difunde el periodismo.

Nos parecía inabarcable el antiguo internet y ahora los grandes modelos de lenguaje (LLM, Large Language Models) como ChatGPT lo han fagocitado y están ampliándolo con contenido multiformato firmado por una IA generativa que –en ocasiones– se hace pasar por un periodista. O al revés: un periodista que obvia contar que lo que escribe no es suyo. Uno de los principales retos, no solo del periodismo sino del conjunto de la sociedad, será preservar la honestidad y que las viejas fake news y los nuevos deepfakes no dominen ni el panorama mediático ni nuestras vidas, opiniones, elecciones o creencias políticas.

Porque el peligro de la desinformación es una preocupación real para una gran parte de la población. En una encuesta realizada por Metroscopia, un 35% de la ciudadanía encuestada piensa que la IA va a mejorar el periodismo, pero hay un 52% que opina que lo va a empeorar. Y más del 80% cree que va a suponer un riesgo para la privacidad y que favorecerá la proliferación de bulos y de noticias falsas.

En estas páginas analizamos las aristas de la legislación actual de manos de juristas expertos, especialmente la AI Act, el Reglamento europeo sobre Inteligencia Artificial, en la que no se prohíbe los deepfakes como tal pero sí se “sugiere” que los contenidos sintéticos que se puedan confundir con contenidos creados por humanos deben estar debidamente etiquetados para no sembrar dudas. Repasaremos también las herramientas disponibles para detectar contenido falso, los robots que cazan robots y que deben de formar parte, a modo de ayuda o complemento, de la caja de herramientas para que los y las periodistas no deleguemos del todo la labor de seguir siendo garantes de la veracidad.

Expondremos cuestiones que atañen a la transparencia de los algoritmos y al código que hay detrás de la IA, la privacidad y los derechos humanos en la aplicación de tecnologías en cuestiones como migración y seguridad nacional. Haremos un breve recorrido por la historia y el origen de la inteligencia artificial, una explicación de los conceptos clave para entender cómo funcionan las herramientas que usamos. Repasaremos casos de uso exitosos de herramientas de IA generativa en redacciones de todo el mundo, indagaremos en las connotaciones éticas y políticas de la antropomorfización de las máquinas, de qué es ser inteligente, la cuestión del copyright en los elementos periodísticos, de los peligros del tecnosolucionismo y las virtudes del uso eficiente y transparente de las herramientas automatizadas para hacer un mejor periodismo.

Prefacio

Un sondeo sobre inteligencia y artificialidad

Para la inteligencia artificial no hay nada que sea potencialmente imposible. Es lo que dice el neurocientífico Mariano Sigman, autor del libro 'Artificial. La nueva inteligencia y el contorno de lo humano'. Desde la Universidad de Stanford, en el Informe del Índice de Inteligencia Artificial 2024¹, ya se concluye que la IA es mejor en la clasificación de imágenes, el razonamiento visual y la comprensión del inglés. Sin embargo, de momento se queda atrás en tareas más complejas como matemáticas de nivel avanzado, razonamiento visual de sentido común y planificación. A la cabeza de las filas de científicos pesimistas está Kilnam Chon, un informático octogenario conocido como el padre de internet en Corea. Chon echa mano de una comparativa entre el hombre y el mono y las posiciones de dominio, y pronostica que dentro de 30 años la IA será más inteligente que la humanidad y entonces podrá controlarnos, como controlamos nosotros a los monos.

Metáforas con primates aparte, lo cierto es que la historia de la IA ha tomado un cariz fascinante desde los años 60 a esta parte, en ese intento, plagado de inviernos, del ser humano por emular su propia inteligencia, pero también por entenderla. Ahora, en esta primavera dorada que vivimos, no pasa una semana sin que haya una nueva noticia, un nuevo programa, un nuevo congreso, una nueva charla sobre esta cuestión, que no es desconocida para el 93% de la ciudadanía.

Según Andrés Medina, director de Metroscopia², el porcentaje de personas preocupadas ante la irrupción y el avance de estas tecnologías es del 53%, y de entusiastas, de un 37%. Más de la mitad ve la IA como una oportunidad y un tercio opina que es una amenaza. De las personas encuestadas, un 35% piensa que va a mejorar el periodismo; un 52%, que lo va a empeorar. Y más del 80% cree que va a suponer un riesgo para la privacidad y que favorecerá la proliferación de bulos y de noticias falsas. "Los españoles siguen confiando en los medios para que defiendan el interés general", asegura Medina. La confianza de una cabecera no es un regalo de la IA sino que responde a otros factores como el trabajo de sus periodistas, la veracidad de sus fuentes, la calidad de sus firmas o su honrosa trayectoria. Ese antiguo lema de sin periodismo no hay democracia se puede ampliar: sin periodistas no hay medios, porque no hay redacciones que puedan funcionar solo con la tecnología más puntera que desarrolle OpenAI. No existen casos de éxito porque la solución no es la sustitución sino la complementación.

En el panorama mediático valenciano ninguno de los entrevistados quiere hablar de despidos y sí de reconversión de los puestos de trabajo de los y las periodistas que ahora hacen tareas automatizables. Según un informe elaborado por el FMI³

1 <https://aiindex.stanford.edu/report>

2 Son datos de un sondeo realizado por Metroscopia llamado Periodismo2023/Fundación AXA, presentado en julio de 2023.

3 Mauro Cazzaniga, M; Jaumotte, F; Li, L.; Melina, G; J Panton, A; Pizzinelli, C; J Rockall, E; Mendes Tavares, M. (2024). *Gen-AI: artificial intelligence and the future of work*. IMF: <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379>

junto a la Universidad de Stanford, hasta un 60% de los empleos notarán el impacto de la IA, sobre todo aquellos relacionados con tareas cognitivas. Pero esta nueva revolución industrial, comparada con el arranque de la física moderna del siglo XX, tiene una cualidad diferenciadora respecto a las anteriores y que recuerda a la literatura distópica de ciencia ficción: la IA generativa es capaz de procesar grandes cantidades de datos, identificar patrones, tomar decisiones, conversar con humanos y hasta aprender de sus errores. Es la llegada de la creatividad computacional. “El ser humano sigue estando en la cúspide en muchos procesos creativos, pero ahora tenemos un gran rival”, destaca Jordi Linares, doctor en Informática y profesor de la Universitat Politècnica de València.

Según este estudio, la mitad de los puestos de trabajo afectados verán incrementada su productividad, pero la otra mitad se verán inevitablemente desplazados: menos empleos, salarios más bajos y hasta la desaparición de algunos trabajos. Mujeres y personas migradas o racializadas⁴ tienen mayor riesgo de reemplazo, puesto que copan los puestos más fácilmente automatizables. Más allá del sesgo por sexo, raza o posición social, para gran parte de las tareas que hasta hoy realizaba un/una periodista (transcripción de entrevistas, documentación, programar publicaciones en redes, detectar los temas que son tendencia, redacción, traducción más o menos simultánea y hasta grabación, locución o edición de audio o vídeo, y veremos cómo se va añadiendo un largo etcétera) ya existen programas que ofrecen unos resultados de gran calidad o incluso en ciertos casos mejores que aquellos hechos por una mano humana.

José Suárez de Lezo, director del Laboratorio de Periodismo de Datos de la Fundación Luca de Tena, recuerda que en menos de un año, ChatGPT ha alcanzado los 100M de usuarios semanales y se ha convertido en la plataforma que más ha crecido en la historia en tan poco tiempo. También reseña sus más de 2M de desarrolladores de la API, una cantidad que favorece la creación de nuevas líneas de trabajo y, por supuesto, el crecimiento económico exponencial de la empresa que hay detrás, OpenAI, con su mediático CEO, Sam Altman, a la cabeza. Si en 2022 la compañía facturó 28 millones de dólares, las previsiones para 2024 son de 1.300 millones y se espera que duplique esta cifra en 2025. “[Estos directivos y empresarios] se están haciendo ricos con un producto que no pueden controlar”, expresa Miguel Arana, investigador en Machine Learning Causal de la Universidad de Cranfield, en Reino Unido, en referencia a la relativa autonomía que tienen estos grandes modelos de lenguaje, como ChatGPT, que funcionan como cajas negras en las que apenas puedes controlar la respuesta que te dan. Es lo que se conoce como *el problema de la explicabilidad*⁵ y realmente supone un gran obstáculo no solo porque pone en jaque la confianza de la persona usuaria, sino que también se cuestiona su validez legal y ética, sobre todo si se está usando una IA para interferir en decisiones que afectan a la vida de las personas.

Según el último World Train Report de Wan-Ifra⁶, World Association of News Publishers, casi la mitad de medios utilizan la IA en sus dinámicas de trabajo, pero apenas un 28% ha desarrollado guías y políticas de uso. Las tareas que le relegan son, sobre todo, creación de titulares y sumarios, corrección de textos, traducción y creación de artículos, sugerencia de temas, simplificación de búsquedas, o bien para acelerar la investigación o mejorar la interacción con la audiencia. En estas páginas veremos ejemplos de usos constructivos de la IA generativa en medios de todo el mundo, opiniones y opciones de periodistas, tanto de València como de fuera, perfiles

4 Ellingrud K; Sanghvi, S; Singh Dandona, G; Madgavkar, A; Chui, M; White, O; Hasebe, P. (2023). Generative AI and the future of work in America. McKinsey: <https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america>

5 Esto es lo que dice ChatGPT al respecto: *El problema de la explicabilidad en inteligencia artificial (IA) se refiere a la dificultad de entender cómo y por qué los modelos de IA, especialmente los más complejos como los basados en aprendizaje profundo, toman decisiones específicas. A medida que estos sistemas se vuelven más integrados en sectores críticos como la salud, finanzas y justicia, la necesidad de explicaciones claras y comprensibles se vuelve crucial. Más información: Yang, W., Wei, Y., Wei, H. et al. Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. Hum-Cent Intell Syst 3, 161–188 (2023).* <https://doi.org/10.1007/s44230-023-00038-y>

6 <https://wan-ifra.org>



técnicos, expertas en desinformación, programadores informáticos, investigadoras en el campo de la tecnología y los derechos humanos, abogadas, juristas, jefes de sección, responsables de informativos y redactores, entre otros, para crear un relato coral del momento periodístico que vivimos.

Esta investigación pretende ser una foto fija del arranque del año 2024, en este universo cambiante en el que asistimos como periodistas espectadoras al gran goteo de tecnologías nuevas, sorprendentes, sintientes, conversadoras, a las que miramos con anhelos, como oportunidad y también con cautela. La mirada crítica, la formación y la reflexión informada es lo que nos va a permitir, como sociedad, decidir qué uso le damos a estos artefactos y si somos capaces de utilizar todo su potencial para mejorar el periodismo en particular y nuestras vidas en general, en lugar de para destruirlo(las).

La *enshittification* de nuestro todo: internet muere donde la IA nace

Estas tecnologías disruptivas, transformadoras y también vistas con cierto miedo y recelo por todo el imaginario previo que existe en el cine y en la literatura distópica de ciencia ficción, van a remover casi todos los ámbitos de nuestra vida, no solo la profesión periodística. Estamos viendo cambios en cuanto a la rentabilidad empresarial, en la manera de hacer periodismo o de consumirlo, en la faceta editorial, en la confianza de las marcas informativas y en la sostenibilidad de los propios medios. “La IA retuerce los modelos de negocio y genera nuevas dificultades” dice Suárez de Lezo, a las que el conjunto de la sociedad deberá hacer frente.

Las personas expertas consultadas coinciden en que estamos, como humanidad, ante uno de los grandes cambios de era. La Inteligencia Artificial ha venido para transformar todos los ámbitos de nuestra vida y para hacer tambalear los pilares sobre los que se asienta la construcción de la sociedad tal y como la conocemos hoy en día. Los CEO de las grandes tecnológicas que marcan y dirigen el cambio, hablan con el tono de los profetas y sus frases son preludio de la realidad que se va construyendo a golpe de titular, o de declaración grandilocuente no exenta de un tono misterioso y ensoñador. Pero los ex CEO y ex directivos son los que más juego dan. A lo largo de 2024, ha sido noticia la cascada de abandonos y ceses de científicos, como fue el caso de Ilya Sutskever, para muchos una de las mejores mentes en este campo y cofundador de OpenAI, especializado en materia de seguridad. Con él se fueron otras tres personas de ese departamento, que ahora parece desprovisto de personal y de atención por parte de la dirección de la compañía. Existen muchas especulaciones al respecto. Se habla de que Sutskever y los suyos se encargaban de asegurar que cada nueva tecnología contaba con todas las garantías antes de publicar avances que pudieran ser potencialmente dañinos. Se cuenta que incluso pararon la comercialización de algunos modelos mucho más potentes, que podrían comprometer la seguridad de formas que solo imaginamos por la literatura de ciencia ficción.

Por un lado, hay una voluntad manifiesta, vista como el camino hacia el progreso, de hacer mejores a las máquinas, porque para peor ya estamos nosotros. A la vez, existe la predisposición de diferenciarnos de ellas rebuscando aptitudes y cualidades que sean exclusivamente humanas, que un robot nunca podría tener, al menos en la realidad no ficcionada, para que seamos algo más que los encargados de accionar la máquina o apretar el botón de enviar.

En realidad, parece que el miedo a que las máquinas sean mejor que nosotras es tan grande como las ganas que tienen unos pocos de lograrlo. Elon Musk, el dueño de la red social X, se ha atrevido a poner fecha a la llegada de lo que se conoce como la *singularidad*, que es la irrupción de una Inteligencia Artificial General o AGI en inglés –bien diferenciada de la IA estrecha, débil o aquella diseñada para una tarea específica, como reconocer rostros o jugar al ajedrez–, mucho más parecida a las capacidades de adaptación a cualquier situación del ser humano. Es decir, una inteligencia artificial superinteligente: “Tendremos IA o AGI a un nivel que puede realizar casi cualquier tarea cognitiva. Solo nos queda el cuándo”, se le ha oído decir, antes de lanzar la fecha: 2025. Mientras tanto, lo que ya estamos viendo es cómo estas potentes arquitecturas neuronales se están acabando internet. Queda poco para que su alimento sea su propio vómito.

Más allá del tecnooptimismo desenfrenado existe esta otra cara de la realidad tecnológica que tiene que ver con su gran impacto climático y planetario, con la amenaza que presenta a la privacidad de los usuarios o la sumisión a los designios de Google, Apple, Meta, Amazon y Microsoft, y a sus intereses económicos. Mientras los humanos buscan otorgar –o encontrar– en las máquinas la *singularidad*, la perfección, la versión mejorada de uno mismo, las prácticas capitalistas que alientan y pagan esa costosisima búsqueda, están convirtiendo el espacio digital en un auténtico estercolero de contenido patrocinado, sintético, *spam* o directamente falso. Es la *enshittification* de la red, un fenómeno acuñado por el escritor de ciencia ficción Cory Doctorow. La decadencia de las plataformas, la *mierdificación* de las redes, la espiral consumista extractivista de los servicios que ofrecen Google, Amazon, Bandcamp, Facebook, Reddit o X. La curva de calidad decreciente de las plataformas virtuales que dominan tantos aspectos de nuestra vida y que se han convertido en lugares de compraventa de lo (in)material y también de nuestra privacidad. El ejemplo práctico y reciente de decadencia lo vemos en la antigua Twitter. La penúltima medida de su dueño Elon Musk ha sido modificar las normas para permitir la publicación de pornografía explícita en la red social.

Demis Hassabis, CEO y cofundador de Google DeepMind, eminencia del ajedrez que pasó del desarrollo de videojuegos a interesarse en emular la neurociencia humana, pronostica que “el intento de destilar inteligencia en una construcción algorítmica puede llegar a ser el mejor camino para la comprensión de algunos de los misterios perdurables de nuestra mente”. En la IA depositamos las esperanzas de trascender los límites que tiene lo humano: buscamos respuestas para frenar el cambio climático, alternativas a la soledad de este mundo hiperconectado, al exceso de trabajo capitalista, a las propias limitaciones de nuestro cerebro, incluso a la muerte y a la enfermedad, cuando usamos algoritmos para que pronostiquen o diagnostiquen un cáncer o cualquier otra patología. Pero el capitalismo también las encuentra sumamente útiles a la hora de adaptar los productos que nos ofrece en función de nuestros gustos específicos, para ello controlan, conocen y poseen cientos de miles de datos sobre nuestro comportamiento.

Marta Peirano en su libro *El enemigo conoce el sistema* habla extensamente de las sofisticadas tecnologías de la persuasión y del capitalismo de la atención: “El 40 por ciento de los usuarios abandonan una página web si tarda más de tres segundos en cargar”⁷. Las mentes más brillantes cobran y trabajan en Silicon Valley e investigan en las mejores universidades americanas, para que cualquier comportamiento nuestro en la red repercuta en un amplio beneficio económico para las grandes empresas que venden productos y servicios, a base de crear necesidades de consumo.



Arrancamos recordando una definición, una de tantas, que es la elegida por la Comisión Europea en la redacción de la AI Act, la primera Ley Europea –y mundial– que regula la Inteligencia Artificial, aprobada el 13 de marzo de 2024 tras tres años de intensas negociaciones:



‘Artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments.⁸

Introducción

Ni tan inteligente ni tan artificial

“Cerrando el mes muy contenta” / “Hoy me he sentado a escribir todo por lo que me siento agradecida” / “Contadme de que estáis agradecidos” (sic) / “Cenita con los amigos que se vuelven familia”. Aitana Lopez, más conocida por sus 329.000 seguidores y seguidoras de Instagram como @fit_aitana, sube a sus *stories* fotos de los sujetadores que le ha regalado Victoria's Secret, fotos de capítulos de su vida social, fotos de sus viajes de visita a Madrid o del concierto de Blink-182 al que –supuestamente– fue. Y digo supuestamente porque en realidad no tiene vida social, no viaja de visita a Madrid ni estuvo en el concierto de Blink-182, aunque Victoria's Secret sí le regaló unos sujetadores. Aitana no existe en lo que hasta hoy conocemos como el mundo real. En la *bio* de su perfil se define como una “virtual soul”. Aitana es el producto de una IA que lleva la firma de una agencia de publicidad de Barcelona.

No es la única *AI Influencer* con miles de seguidores en redes sociales que podría parecer, casi casi, real. Tan real que muchos de sus *followers* varones le mandan mensajes con propuestas sexuales –sus creadores⁹ hablan de que recibe unas 400 proposiciones de cita diariamente– y decenas de comentarios pidiéndole que se desnude o que se dedique al porno. Desde futbolistas hasta actores –con hasta 11 millones de seguidores en Instagram– siguen confundiéndola con una modelo de carne y hueso y bombardeando sus DM. Habrá quien piense que Aitana ha pasado el test de Turing¹⁰ de sobra. Y más, cuando factura ella solita 4.000 euros mensuales en publicidad y posados en ropa interior.

El paso de este tipo de avatares del ámbito comercial al periodístico –estéticamente no son tan diferentes– ya lo estamos viendo. Ya existe “la primera influencer virtual entrenada por Inteligencia Artificial que presentará un espacio asociado a un gran formato televisivo en nuestro país”¹¹, un especial del programa Supervivientes para la web de Telecinco y la plataforma de contenidos a la carta de Mediaset. Y no es la única. Cada semana desayunamos con una lista de noticias relacionadas con personajes producto de la inteligencia artificial, nuevas herramientas que parecen inteligentes, nuevos pactos de medios con las tecnológicas que producen y entrenan inteligencia artificial generativa o entrevistas a no tan nuevas pensadoras y pensadores que opinan sobre la bondad o la maldad de la tecnología.

Para la periodista Marta Peirano, especialista en la intersección entre tecnología y poder, la IA, en realidad, no es ni tan inteligente ni tan artificial. Y ese sí es un debate abierto interesante: ¿qué quiere decir ser inteligente? Esta cualidad, hasta ahora, patrimonio casi exclusivo de la humanidad, se ha vuelto el epíteto de ChatGPT, Gemini, Copilot, Claude o cualquier otro enorme modelo de lenguaje (Large Language

9 Benedicto Borges, L. (2023, 8 de noviembre). *Nuestra hija Aitana, la primera modelo de IA española que gana... “4.000 euros al mes”*. El Mundo. <https://www.elmundo.es/cronica/2023/11/08/65453bc7fc6c83f03a8b4585.html>

10 Más adelante hablaremos de él pero solo por introducirlo, el test de Turing definía a una máquina como sintiente, inteligente o pensante cuando fuera una máquina que, en conversación con un ser humano, este no distinguiera si estaba hablando con un aparato o con un igual.

11 Redacción verTele. (2024, 13 de marzo). *Mediaset crea una presentadora por inteligencia artificial para conducir un formato de ‘Supervivientes’ en Mitele*. verTele! https://www.eldiario.es/vertele/noticias/mediaset-crea-presentadora-inteligencia-artificial-conducir-nuevo-formato-supervivientes-mitele_1_11210629.html

Models, LLM, en inglés). “Me preocupa mucho más la estupidez humana que la inteligencia artificial”, reflexiona Ramon López de Mántaras, profesor de investigación del CSIC y director del Instituto de Investigación en Inteligencia Artificial.

Los límites entre aquello que era estúpido y que hoy es inteligente se están redefiniendo. De hecho, cambiamos la definición según hablemos de un ser humano o de una máquina. “Tú le preguntas a ChatGPT, te responde coherentemente y decimos *buah*, es inteligente. A una persona, le preguntas y te responde coherentemente, y no consideras que es inteligente, dices, bueno, qué menos... El baremo que hemos puesto para decidir qué es inteligente es más bajo para una máquina que para un humano”, alega Ujué Agudo, doctora en Psicología e investigadora de la relación entre la IA y las relaciones humanas.

En una de sus conferencias,¹² Peirano cuestionaba estos grandes mitos contruidos alrededor de la IA y su capitalización de prácticamente todos los sectores sociales y laborales. Para la periodista y escritora, la tecnología lleva amenazando con dejar sin trabajo al personal asalariado desde la primera revolución industrial, cuando se popularizó el fantasma de las fábricas sin obreros. Pero ni pasó ahí ni, previsiblemente, ocurrirá tal cual en el futuro. Siguiendo a Peirano, lo que vemos es que la llegada de nuevas máquinas quizá no nos quita del todo el trabajo pero sí degrada nuestra posición laboral y con ello nuestra remuneración. Ciertas aptitudes humanas ya no son tan necesarias, por tanto, no se pagan como merecerían, con la consecuente reducción de costes de producción para el empresariado. La novedad es que hoy también –y sobre todo– se ven afectados los trabajos creativos y no solo los manuales o mecánicos. Porque no cobra lo mismo un guionista de Hollywood que tiene una idea original y la escribe, que si simplemente se dedica a redactar y dar forma a la ocurrencia llamativa que ha tenido cualquier chatbot de renombre. Además de Peirano, esta teoría la sostienen otras tantas voces que aparecen en esta investigación. Como si la tecnología y su avance supusiera también un paso adelante en la precarización de las fuerzas de trabajo.

El primero en nombrar al precariado, esa nueva clase social en la que vive instalada, entre otras profesiones, la periodística, fue el economista Guy Standing. Los integrantes de este estamento van perdiendo derechos sociales y económicos a la misma vertiginosa velocidad con la que avanzan la computación cuántica y la automatización de los sistemas algorítmicos. El “capitalismo rentista”, siguiendo a Standing, sería aquel que está encabezado por las grandes corporaciones tecnológicas y que lo ha alejado del libre mercado para crear un sistema de “mercado amañado” en el que “cada vez más ingresos van a parar a los dueños de la propiedad, la propiedad física, la propiedad financiera y la propiedad intelectual. Y va, cada vez menos, a las personas que dependen de su trabajo”¹³.

Si seguimos de cerca los últimos estudios, veremos que estamos próximas al escenario que vaticinan pensadoras como Peirano. En un *paper*¹⁴ publicado por personal investigador de *OpenResearch*, la Universidad de Pennsylvania y *OpenAI* –la empresa tecnológica que ha desarrollado ChatGPT– reconocen que sus modelos pueden afectar a miles de empleos. Según se recoge en el estudio, especialmente a los mejor pagados y cualificados y también los relacionados con la programación, la redacción y la creación de contenidos. Otra investigación¹⁵, en este caso del MIT –el *Massachusetts Institute of Technology*– se propuso ver cómo la IA de generación de

12 M Peirano (2023). Ni inteligente ni artificial: terapia extractivista para un mundo en crisis. Conferencia en el encuentro Immaterial 2023 en Tabakalera.

13 Ribas, N. (2022, 21 de junio). La renta básica no produce más vagos, sino gente con buena salud mental y más productiva. Entrevista a Guy Standing en Eldiario.es. https://www.eldiario.es/illes-balears/economia/guy-standing-renta-basica-no-produce-vagos-gente-buena-salud-mental-productiva_128_9099519.html

14 Tyna Eloundou, T; Manning, S; Mishkin, P; Rock, D (August 22, 2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models . OpenAI, OpenResearch, University of Pennsylvania.

15 *ChatGPT Lifts Business Professionals' Productivity and Improves Work Quality*, publicado en Nielsen Norman Group el 2 de abril de 2023. <https://www.nngroup.com/articles/chatgpt-productivity>



contenidos influye en la productividad de las personas trabajadoras. La conclusión fue que los materiales elaborados con ChatGPT se habían hecho en mucho menos tiempo y tenían más calidad, según dictaminó un equipo de personas expertas. Este estudio confirmaría el peor de los temores.

Pero como recuerda Marta Peirano, ChatGPT sin la mano del ser humano ni resulta creíble ni serviría para nada. Y podríamos añadir que sin una mano humana que sepa usarla correctamente, dando las órdenes o *prompts*¹⁶ adecuadas, tampoco. “Con toda esta automatización que se nos viene, si esas tareas las va a hacer una inteligencia artificial y tú no vas a aportar valor, obviamente te vas a quedar sin trabajo. El empleo no te lo va a quitar la inteligencia artificial, sino alguien que sepa usar la inteligencia artificial”, decía Cristina Aranda¹⁷, CEO de Big Onion y cofundadora de Mujeres Tech. Esa frase no es suya, al menos no solo, puesto que la hemos escuchado en los más dispares escenarios, entre ellos en un informe de IBM publicado el verano de 2023: “AI won’t replace people—but people who use AI will replace people who don’t”.¹⁸

El misterio de la explicabilidad: ¿Cómo funciona la IA generativa?

“Los grandes modelos de lenguaje (LLM) pueden hacer cosas asombrosas. Pero nadie sabe exactamente por qué. Y eso es un problema. Averiguarlo es uno de los mayores enigmas científicos de nuestro tiempo y un paso crucial para controlar modelos futuros más potentes”.¹⁹ Como escribe el periodista del *MIT Technology Review* Will Douglas Heaven, hace un par de años dos investigadores trataron de averiguar qué se necesitaba para que un modelo de lenguaje realizara operaciones aritméticas básicas. Por error, dejaron el experimento en marcha algunos días y se les fue mostrando a los modelos de forma repetida las mismas operaciones. Así descubrieron que, en algunos casos, los modelos podían aparentemente no aprender una tarea y, de repente, conseguirlo, una cualidad que no se esperaba del *deep learning* o aprendizaje profundo y que supone un paso más allá del *machine learning* o aprendizaje automático²⁰. A este comportamiento lo llamaron *grokking* o capacidades emergentes.

16 Más adelante recuperaremos de nuevo una frase de la matemática e investigadora Ada Lovelace, que decía que con las instrucciones adecuadas las máquinas podrían hacer más cosas. De ahí la importancia de los prompts, que son las instrucciones que le das a un modelo de lenguaje para guiar su respuesta y tratar de obtener el contenido que necesitas. Si le preguntas a ChatGPT, añade que *diseñar efectivamente un prompt es crucial para influir en la calidad y relevancia de la salida del modelo*.

17 Fue durante el Summit RedMad, Red de Mujeres en Alta Dirección, a finales de 2023. Se puede ver íntegra en el canal de YouTube de REDMAD. <https://www.youtube.com/@redmad9369/videos>

18 “La IA no reemplazará a las personas, pero las personas que sepan utilizar la IA sí reemplazarán a las que no”. Goldstein, J; Login, B; Fillare, C; Nowal, C (202-2023). *Augmented work for an automated, AI-driven world Boost performance with human-machine partnerships*. IBM IBV and Oxford Economics: <https://www.ibm.com/downloads/cas/NGAWMXAK>

19 Will Douglas Heaven es redactor jefe de Inteligencia Artificial en el *MIT Technology Review*, la publicación divulgativa del conocido Instituto de Tecnología de Massachusetts. Aquí hablamos de un reportaje publicado en marzo de 2024 titulado: *Large language models can do jaw-dropping things. But nobody knows exactly why*. <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why>

20 En las páginas finales existe un glosario de términos, pero aquí van dos definiciones sencillas que aporta ChatGPT: ‘*Machine learning* es un campo de la inteligencia artificial donde las máquinas aprenden a realizar tareas a partir de datos sin ser programadas explícitamente. Los algoritmos encuentran patrones en grandes cantidades de datos y ajustan sus respuestas o predicciones a partir de ellos. *Deep learning* es una subramificación del *machine learning* que utiliza redes neuronales artificiales profundas para modelar y resolver problemas complejos, inspirándose en el funcionamiento del cerebro humano’.

“¿Podemos estar seguros alguna vez de que los modelos han dejado de aprender?”, se pregunta Hattie Zhou, investigadora de IA en la Universidad de Montreal y Apple Machine Learning Research. Estos LLM parecen comportarse de maneras que las matemáticas de los libros de texto dicen que no deberían. Nadie sabe exactamente cómo funciona del todo el *deep learning*, que es una tecnología clave dentro de la IA, especialmente en áreas complejas como el procesamiento del lenguaje natural o la visión por computación, y que permite, por ejemplo, a las máquinas interpretar y comprender imágenes o videos de manera similar a como lo hacen los humanos. Como señala el redactor del *MIT Technology Review*, los investigadores estudian estas cuestiones relacionadas con el aprendizaje automático profundo como si fueran extraños fenómenos naturales, ya que rompen con las explicaciones que se tenían hasta el momento sobre el comportamiento de los modelos predictivos. “La tecnología funciona. ¿No es suficiente?”, ironiza el reportero.

“¿Explicar realmente por qué las redes neuronales tienen este tipo de comportamiento inesperado? Aún estamos lejos de conseguirlo”, le dice Daniel Hsu, informático de la Universidad de Columbia, como única respuesta posible al problema de la explicabilidad de la inteligencia artificial. Resulta un poco escalofriante pensar en que estos modelos sepan hacer cosas para las que no se les ha entrenado y que esto no se pueda justificar de una forma matemática. Estos modelos avanzados son capaces de generalizar y aprender a hacer tareas que no han visto antes porque no solo memorizan patrones sino que sorprenden con la capacidad de crear reglas para aplicar esos patrones a casos nuevos. Es la antesala de la AGI o *Artificial General Intelligence*, esa que es capaz de imitar a la perfección la capacidad humana para aprender o construir conceptos.

La propia comunidad científica es la primera en asombrarse cuando ve cómo entrenan a sus modelos con problemas matemáticos en inglés, después les ponen algo de literatura en francés y de ahí ven como “generalizan” y resuelven problemas matemáticos en francés. Estos enormes modelos de lenguaje se entrenan con ingentes cantidades de parámetros, también llamados pesos. Están basados en los *Transformers*, un tipo de arquitectura de red neuronal especializada en el procesamiento de secuencias de datos, como si fueran palabras dentro de frases. Así los define ChatGPT:

Los Transformers son una forma específica de organizar y estructurar las redes neuronales para que puedan procesar y entender datos de manera más eficiente. Piensa en la arquitectura como el plano de un edificio: describe cómo se organiza y conecta cada parte del sistema para cumplir su propósito.

De hecho, la “T” de ChatGPT hace referencia a ellos: *Generative Pre-training Transformer*. La “P” de *Pre-training* nos habla de su adiestramiento, que ha tenido varias fases y ha pasado por varios estadios, siendo en algunos momentos supervisado por humanos y en otros autosupervisado. Ese entrenamiento pasa por mostrarle un gran número de datos, como sabemos. Es lógico pensar que, a mayor cantidad de datos empleada, mayor dificultad para trazar su autoría y saber de dónde provienen o a quién pertenecen estos datos. ¿Qué volumen de *datasets* necesita un modelo con ChatGPT para su entrenamiento? Casi es complejo contarlos. Un ejemplo de una parte de esos datos podría ser desde los 3 billones de la Wikipedia en inglés, sumados con 410 billones de Common Crawl, un repositorio que rastrea la web mediante procesos de *crawling* que funciona desde 2008. Aunque parezca difícil de creer, los datos son finitos e internet se acaba. Como veremos más adelante, las grandes tecnológicas continúan usando su imaginación para costear estos arduos procesos de entrenamiento y para ello no dudan hasta en apostar por la energía nuclear como



forma de abastecimiento, o en transcribir YouTube entero para tener material con el que alimentar a sus criaturas.

Escoger una buena materia prima con la que adiestrar los grandes modelos de lenguaje es una parte clave, porque si se les instruye con basura, generarán basura. Si los datos tienen sesgos, el resultado será sesgado. “Estas máquinas son increíblemente inteligentes, pero asombrosamente tontas”, apuntaba Will Douglas Heaven. Aun así, nos fascinan. “Estamos hablando de la ola tecnológica más disruptiva que hemos vivido y yo creo que de la historia de la humanidad, porque es una tecnología donde todos los plazos se acortan”, señala la periodista y experta en redes sociales, periodismo móvil y desinformación, Carmela Ríos. Y añade: “Todos los procesos relacionados con el periodismo se ven afectados, desde la producción de informativos a la de contenidos, pasando por la escucha”. Según la veterana periodista, el mensaje que nos lanza esta nueva realidad es muy potente: “Una máquina o un bot con una inteligencia artificial es capaz de poder hacer una gran parte de las labores que hasta este momento solo podía garantizar el producto de la inteligencia humana. Ahora la inteligencia no es humana sino artificial, y el mensaje da mucho miedo”.

Para el profesor de la Universitat Politècnica de València Jordi Linares, nadie podía anticipar esta última ola tan disruptiva que nos haría hablar de conceptos antes impensables, como la creatividad computacional. La máquina ha entrado en campos antes vetados para ella y solo reservados a un cerebro humano, periodista, artista o creador: “La evolución de las redes neuronales artificiales y del *hardware* que las sustenta, unido al acceso de grandes *datasets*, a esos grandes volúmenes de información en internet, ha precipitado la aparición de un conjunto de técnicas extraordinariamente increíbles a la hora de generar texto creativo, vídeo, 3D, música y prácticamente todo lo que en definitiva podamos tener de forma masiva en internet y podamos utilizarlo para entrenar estas redes”, dice.

↳ **Hallucination is the belief that a statistical model of language is governed by truth criteria**

@tallerestampa



Ética y tecnología

La conciencia (¿ya no?) es patrimonio de la humanidad

De estos modelos de lenguaje se dice que, a través de esas complejas redes neuronales con cierta bioinspiración, como las califica Jordi Linares, son capaces de razonar, de entrenarse, de conversar. Estas son cualidades típicas de la conciencia humana. Así define ChatGPT la conciencia humana, una vez que hemos separado el grano de la paja: *la capacidad subjetiva de experimentar sensaciones, percepciones, pensamientos y emociones, así como de tener conciencia de uno mismo y de su entorno. (...) En términos más específicos, la conciencia implica una experiencia interna de la mente, que incluye la capacidad de atención, reflexión, introspección y autoconciencia. Y esto es clave: ser consciente de que existes, de que estás, de que eres un ser sintiente.*

“Desde luego, no va a haber, no se ha creado nunca conciencia a través de una máquina, ni hay posibilidades de crearla. Es un rasgo específicamente humano, es justamente lo que nos hace diferentes de millones de especies. Plantear una cosa así es un poco una locura, pero locamente todos nos lo hemos creído. Esta locura les está sirviendo mucho para unos fines que es ganar mucho dinero”, reprochaba Yann LeCun²¹, científico jefe de IA de Meta, uno de los grandes defensores de publicar el código en abierto, a quien se le ha escuchado en múltiples ocasiones denunciar a aquellos que explotan el miedo sobre estos sistemas.

Elon Musk, por ejemplo, discrepa y fija el próximo año 2025 como la fecha en la que conoceremos esa inteligencia artificial *pseudoconsciente* capaz de realizar cualquier tipo de tarea cognitiva de la misma forma que lo haría un humano. El magnate, según Forbes la persona más rica del mundo, se encuentra entre los mil firmantes de una declaración para pedir que se frene el entrenamiento de poderosos modelos de lenguaje con el argumento de que son una potencial amenaza para la humanidad²².

Como LeCun no es de los que le ven las orejas al lobo, no cree que haya que ser alarmista con la legislación. En declaraciones a El País, contaba²³ que “contrariamente a lo que se puede escuchar de algunas personas, no existe un diseño para un sistema que alcance la inteligencia humana”, y que, por ello, “pedir regulaciones por miedo a la inteligencia sobrehumana es como pedir una regulación de vuelos transatlánticos a una velocidad cercana a la del sonido en 1925. Esto no está a la vuelta de la esquina; va a llevar mucho tiempo, con sistemas que todavía no conocemos”.

21 Rizzi, A. (2024, 19 de enero). Yann LeCun, científico jefe de IA de Meta: “Una inteligencia artificial de nivel humano va a llevar mucho tiempo”. El País. <https://elpais.com/tecnologia/2024-01-19/yann-lecun-cientifico-jefe-de-ia-de-meta-una-inteligencia-artificial-de-nivel-humano-va-a-llevar-mucho-tiempo.html>

22 En mayo de 2023 más de mil expertos y ejecutivos de la industria tecnológica incluidos, además de Elon Musk, el cofundador de Apple, Steve Wozniak, el director ejecutivo de la firma Stability AI, Emad Mostaque, junto a investigadores de la firma DeepMind propiedad de Google, firmaron esta petición. La publicaron en la web del *Future of Life Institute*, una institución ubicada en Cambridge cuyo objetivo es reducir los riesgos que implica el desarrollo tecnológico. En esta carta pedían que las empresas que desarrollan este tipo de programas “pausen de inmediato, durante al menos 6 meses, el entrenamiento de los sistemas de inteligencia artificial más potentes que GPT-4”, porque se estaba convirtiendo en “una carrera fuera de control para desarrollar e implementar mentes digitales cada vez más poderosas que nadie, ni siquiera sus creadores, pueden comprender, predecir o controlar de forma fiable”.

23 En el mismo artículo de El País de la cita 21.

Y sigue LeCun: “Estos sistemas son inteligentes en un dominio relativamente estrecho en el que han sido entrenados. Dominan el lenguaje y eso nos hace pensar que son inteligentes, pero no lo son tanto”²⁴.

Este gran paso para la humanidad –lento para LeCun pero con grandes acelerones en los últimos años– que es pasar de la IA discriminativa o predictiva y su pensamiento convergente a la IA generativa y su cambio al pensamiento divergente, está ocurriendo y parece imparable. Su evolución ha sido exponencial y lo sigue siendo, porque nadie duda de que cada vez es mejor, más sofisticada, más similar a la respuesta humana. En apenas tres años los cambios han sido muy significativos.

Si echamos un vistazo al año 2021 podemos ver un ejemplo. Will Douglas Heaven, editor de la sección de IA del MIT Technology Review,²⁵ definía GPT-3 como algo “sorprendentemente bueno y totalmente absurdo”; un chatbot capaz de generar “texto asombroso similar al humano a demanda, pero que no nos acercará a la verdadera inteligencia”. Este periodista y Doctor en Ciencia Computacional hace una aproximación muy interesante a la definición de inteligencia, algo que va en la línea de las reflexiones de Marta Peirano, cuando ironiza con que no es tan artificial, porque está diseñada, creada y mantenida por humanos; los algoritmos no funcionan en un vacío sincrético sino que precisan de servidores, energía y supervisión humana constante. Y puede que tampoco sea tan inteligente, al menos no lo es en el sentido humano del término, precisamente por esta dependencia de los datos que le ofrecen sus diseñadores y porque, por ejemplo, es incapaz de generar conocimiento nuevo. Su inteligencia se limita a identificar patrones y realizar predicciones basadas en esos patrones, y a hacerlo maravillosamente bien.

Dice Douglas Heaven: “El rendimiento *huma-like* y la sorprendente versatilidad de GPT-3 son el resultado de una excelente ingeniería, no de una auténtica inteligencia”. Lo repetimos: la IA es un subproducto de la inteligencia humana. Y sigue haciendo que nos replanteemos el concepto: “Tenemos un listón muy bajo para detectar la inteligencia. Si algo parece inteligente, es fácil engañarnos pensando que lo es. El mayor truco de la IA ha sido convencer al mundo de que existe. GPT-3 es un gran paso adelante, pero sigue siendo una herramienta hecha por humanos, con todos los defectos y limitaciones que ello implica”. Las palabras de Douglas Heaven se entendían en un momento (julio de 2020) en el que todavía veíamos grandes errores o alucinaciones y en el que estos modelos de lenguajes enormes tenían apariencia de chatbots de batiburrillo corta-pega, además de un importante sesgo sexista y racista del que todavía hace gala. Pero el trasfondo sigue vigente.

En ese mismo año 2021, Renée Diresta, jefa de investigación del Observatorio de Internet de Stanford, publicaba un texto en Wired²⁶, tres meses antes de las elecciones del 3 de noviembre de 2020 en Estados Unidos, en el que criticaba que a los *deepfakes*²⁷ audiovisuales se les notaba demasiado el *fake*. Justo el verano de ese año se presentó al mundo una versión beta de GPT-3. Diresta lo definió entonces como “la próxima frontera de la escritura generativa: una inteligencia artificial capaz de producir frases que suenan sorprendentemente humanas, aunque a veces surrealistas”. Estos enormes modelos de lenguaje transformador van afinando cada vez más

24 Rizzi, A. (2024, 19 de enero). Yann LeCun, científico jefe de IA de Meta: “Una inteligencia artificial de nivel humano va a llevar mucho tiempo”. El País. <https://elpais.com/tecnologia/2024-01-19/yann-lecun-cientifico-jefe-de-ia-de-meta-una-inteligencia-artificial-de-nivel-humano-va-a-llevar-mucho-tiempo.html>

25 Douglas Heaven, W. (2020, 20 de julio). *OpenAI's new language generator GPT-3 is shockingly good—and completely mindless*. MIT Technology Review. <https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp>

26 DiResta, R. 31-07-2020: *AI-Generated Text Is the Scariest Deepfake of All*. Revista Wired. <https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all>

27 Más adelante se vuelve a explicar este término, pero la traducción sería algo así como ultrafalsificación. Del inglés, *fake*, engaño o falsificación, y *deep*, porque estos contenidos normalmente en video se elaboran gracias a herramientas que usan *deep learning*, que como ya se ha dicho, es una rama del aprendizaje automático o *machine learning* que utiliza redes neuronales complejas para hacer predicciones a partir de grandes volúmenes de datos.



la puntería y son capaces –como los humanos– de ir mejorando y aprendiendo de sus errores para fabricar productos creativos óptimos, más creíbles y más difíciles de diferenciar de aquellos generados por personas.

Sam Altman, cofundador de OpenAI, tuiteaba por aquel entonces²⁸: “La exageración de GPT-3 es demasiada. Es impresionante (¡gracias por los lindos elogios!) pero todavía tiene serias debilidades y, a veces, comete errores muy tontos. La IA va a cambiar el mundo, pero GPT-3 es sólo un vistazo muy temprano. Todavía tenemos mucho que resolver”. Esa es la carrera que están recorriendo las grandes tecnológicas desde la Bahía de San Francisco por ser las mejores y las primeras, con OpenAI a la cabeza, en crear algo que cada vez sorprenda más, sin escatimar en gastos económicos, energéticos o de recursos. Porque la multimodalidad necesita *multimillones* de recursos y datos, y un aumento creciente de prácticas extractivistas a gran escala tanto de ese nuevo petróleo como de materias primas y energías. Es la competición multimillonaria a la conquista de una nueva luna lista y sintética. Y tiene otro punto en contra, que señala Carissa Véliz, experta en ética aplicada a la tecnología: la inteligencia artificial la están diseñando cinco hombres blancos en Silicon Valley a cualquier precio.

↳ Why isn't AI doing the tedious shit for creative people instead of the creative shit for tedious people

@piesaac

La creatividad hace humana a la máquina: antropomorfización y AGI (Artificial General Intelligence)

“Nos maravilla lo que hemos creado”, ironiza la periodista especializada en medios digitales de Newtral, Marilín Gonzalo. “No hay necesidad como humanidad de darle el contenido creativo a las máquinas, primero porque lo hacen mal”, sigue Gonzalo. Para Miguel Arana, investigador en *Machine Learning Causal* de la Universidad de Cranfield, en Reino Unido, solo es cuestión de tiempo: “Las máquinas empiezan a copiar muy bien lo que hacemos los humanos, incluso a sobrepasar, en algunos casos, la capacidad humana. Estamos viendo que la inteligencia humana quizá no es tan increíble como nos parecía, que teníamos una visión de la inteligencia humana un poco mitificada”.

Parece que cuesta reconocer que un LLM²⁹ pueda ser más creativo que tú. Existe un estudio muy interesante publicado en la revista *Frontiers in Psychology*³⁰ titulado ‘Evaluación de la emoción y la sensibilidad de las obras de arte de IA’. En él se hizo el experimento de mostrar una obra generada con inteligencia artificial a dos grupos de personas; al primero se le dijo que era sintética y al segundo, que era obra de una mano humana. Ahí se concluyó que se le atribuía menor sensibilidad, menor capacidad para evocar sus emociones y menor calidad a la obra de arte cuando se creía que el artista era una IA en comparación con cuando pensaban que era obra de la creatividad y el talento humano. Los resultados muestran –dice el estudio– que la evaluación del arte parece estar modulada, al menos en parte, por estereotipos y prejuicios previos sobre las habilidades creativas de la IA.



Ontológicamente hablando, la comprensión de que somos iguales o de que podemos hacer cosas parecidas de las que hace un montón de algoritmos, cables y chips no resulta sencilla. Para que este montón de chatarra nos caiga mejor, la humanizamos, consciente o inconscientemente, en un proceso que se conoce como antropomorfización de las máquinas. Les otorgamos características, aptitudes y habilidades humanas, y utilizamos los mismos adjetivos para hablar de ellas que de un grupo de gente. Por eso decimos que son *inteligentes*, o que las *entrenamos* con una gran cantidad de datos. O denominamos *alucinaciones* a los errores que cometen tecnologías como GPT.

Estas cuestiones las explica Marilín Gonzalo en un reportaje³¹ en Newtral. La periodista define las alucinaciones como “una respuesta confiada dada por una IA que no está justificada por los datos con los que ha sido entrenada”. Este término viene de la psicología humana y se asocia con las falsedades. La utilización del lenguaje no es casual y, como la propia experta recuerda, no es lo mismo decir que la IA

29

Como ya se ha dicho, Large Language Model, modelo de lenguaje grande.

30

Agudo, U; Arrese, M; G. Liberal, K; Matute, H. 05/04/2022. *Assessing Emotion and Sensitivity of AI Artwork*. *Frontiers in Psychology*.

31

Gonzalo, M. (2023, 22 de junio). Las “alucinaciones” de la inteligencia artificial son fallos que preocupan a los expertos. Newtral. <https://www.newtral.es/alucinaciones-ia-fallos-inteligencia-artificial-confabulaciones/20230622>

“copia” que decir que la IA “aprende”³², puesto que en el aprendizaje ni se infringen los derechos de autor ni se puede recibir, a priori, la acusación de plagio.

En un artículo de Naomi Klein en *The Guardian*, la conocida autora de *No Logo* y *La doctrina del Shock* se pregunta: “¿Por qué llamar a los errores ‘alucinaciones’? ¿Por qué no *basura algorítmica*? ¿O fallos? (...). Al apropiarse de una palabra de uso común en psicología, psicodelia y diversas formas de misticismo, los impulsores de la IA, al tiempo que reconocen la falibilidad de sus máquinas, alimentan la mitología más preciada del sector: que al construir estos grandes modelos lingüísticos y entrenarlos en todo lo que los humanos hemos escrito, dicho y representado visualmente, están en proceso de dar a luz una inteligencia animada a punto de provocar un salto evolutivo para nuestra especie”³³.

Estos programas, sigue Gonzalo, “simplemente son excepcionalmente eficaces en predecir patrones del lenguaje y han sido optimizados para ser conversacionales, por lo que se colocan en un plano mental similar al de otros agentes humanos cuando interactuamos con ellos, y nos resultan creíbles y convincentes”. Es nuestra empatía la que hace el resto del trabajo.

Una concienzuda y bien trabajada estrategia de marketing empaqueta la máquina para regalo. Cómo se “vende” la IA al mundo y a la sociedad es importante. Por ejemplo, todos los anuncios de nuevos lanzamientos en OpenAI tienen ese halo de misterio y temor que desprenden las personas con poder que tienen acceso a una información que el común de los mortales desconoce. Suelen jugar a la duda y al despiste, y aprovechan sus grandes audiencias para que sus líderes hagan declaraciones que van de la fascinación por lo afinado de su última actualización de ChatGPT al miedo a la distopía o a la destrucción de la raza humana con algo que ha creado la propia raza humana. “Altman siempre tiene ese tono como que él ha visto a Dios en la máquina y no nos lo quiere contar, es lo que vende, y trata constantemente de alimentar ese *hype* que les está reportando muchos beneficios económicos”, objeta Arana.

En una entrevista a Sam Altman en el podcast de Lex Fridman³⁴ —que recibió 1,6 millones de reproducciones a los 7 días de haber sido publicada en YouTube—, el CEO de OpenAI afirma que antes de que acabe esta década ya tendremos entre nosotras la famosa AGI, la Inteligencia Artificial General, aunque asegura que todavía no la han construido. Tecnológicamente hablando, habríamos pasado de las calculadoras, capaces de resolver un número limitado de operaciones, a una superinteligencia singular con conciencia de sí misma que supera con creces a la humana y que es capaz de responder y dar soluciones a todo, de generalizar, aprender y comprender.

Sara: ¿Qué es la AGI?
 Gemini: La AGI, o Inteligencia Artificial General (por sus siglas en inglés), es un tipo hipotético de inteligencia artificial que iguala o supera la inteligencia humana promedio en todas las áreas cognitivas. Se trata de una IA capaz de comprender el mundo de la misma manera que un humano, con la capacidad de aprender, razonar y resolver problemas de forma autónoma.

32 Gonzalo, M. (2024, 15 de enero). *Los derechos de autor y la IA, según OpenAI*. Newtral. <https://www.newtral.es/derechos-autor-ia-openai/20240115>

33 Klein, N. (2023, 8 de mayo). *AI machines aren't 'hallucinating'. But their makers are*. *The Guardian*. <https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein>

34 Según su canal de YouTube, en el que atesora más de 4 millones de suscriptores, “Lex Fridman es un científico, ingeniero y podcaster que explora temas como la inteligencia artificial, la filosofía o las artes marciales”. Solo la entrevista con Altman tiene más de 2 millones de visualizaciones, y no es el único habitante de Silicon Valley que aparece en este canal. Encontramos entrevistas de Fridman de hasta 3 horas con Yann Lecun, Elon Musk o Marl Zuckerberg, además de otros personajes variopintos y muy conocidos, desde Kevin Costner o el cómico egipcio Bassem Youssef hasta el dueño de Amazon Jeff Bezos o el presidente de Israel, Benjamin Netanyahu.

Sara: ¿Existe la AGI?
 Gemini: No, la AGI no existe todavía. Se trata de un concepto teórico y aún no se ha desarrollado una IA que cumpla con todas las características de la AGI. Sin embargo, es un campo de investigación muy activo y se están realizando avances significativos.

En noviembre de 2023 fue noticia el despido y la posterior readmisión de Sam Altman como consejero delegado de la tecnológica que él mismo cofundó, y justo en las mismas fechas la agencia de noticias Reuters publicaba una información³⁵ según la cual OpenAI había alcanzado un poderoso descubrimiento de inteligencia artificial que podría amenazar a la humanidad. Este proyecto misterioso se llamaría Q* (pronunciado *qu star*) y, preguntado por él en el podcast de Fridman, Altman responde: “No estamos preparados para hablar de eso. Investigamos todo tipo de cosas. Hemos dicho que un mejor razonamiento en estos sistemas es una dirección que nos gustaría proseguir”.

“A la compañía sin duda le interesa mantener esa aura de flirteo con proyectos enigmáticos”, escribe el periodista Jordi Pérez Colomé, en su newsletter sobre tecnología de El País. En ella recuerdan una entrevista de la revista Wired con Lukasz Kaiser, otro científico de OpenAI. Cuentan que esto es lo que pasó en esa charla: “Kaiser es uno de los inventores de una nueva tecnología llamada Q*. Cuando intenté preguntar a Kaiser sobre esto en nuestra entrevista, la persona de relaciones públicas de OpenAI casi saltó sobre la mesa para hacerle callar”.

Altman introduce el concepto de *agente de IA*, que será una especie de mezcla entre un asistente, tu mejor amiga y un consejero real, que sabrá de ti más que tú misma: “Vas a meter toda tu información, tu historial a lo largo del tiempo y te irá conociendo cada vez mejor. Quiero que integre las lecciones aprendidas y que me recuerde en el futuro qué hacer de manera diferente o en qué debo tener cuidado. Todos ganamos experiencia a lo largo de nuestras vidas, y me gustaría que mi agente de IA también se beneficiara. Si pudiera meter todas las conversaciones que he tenido con cualquier persona en mi vida, si pudiera tener todos mis correos, todas mis peticiones cada vez que hago una pregunta, sería genial”. Pérez Colomé tecldea en esa newsletter: “Queda poco para preguntarle, ¿qué hice mal con mi primer ex, hijo o trabajo? ¿Qué es mejor que no haga otra vez? Seguro que hay gente que querrá recordar su vida con ese nivel de detalle, como si fuera importante”.

Entonces, ¿la AGI será una especie de nueva religión a la que encomendarnos para que responda a nuestras grandes dudas existenciales? Miguel Arana lo imagina como un proceso gradual el que iremos conquistando tareas y que “no es probable que ocurra de una forma explosiva” sino como un proceso gradual.



Historia

Antecedentes, teoría e historia: Ada Lovelace, Turing, Altman

↳ «La máquina analítica no puede hacer nada sin las instrucciones adecuadas, pero puede hacer mucho si se le dan las instrucciones adecuadas»

Ada Lovelace

↳ «Una computadora puede ser llamada inteligente si logra engañar a una persona haciéndole creer que es un humano»

Alan Turing

En realidad, siempre que se habla de cualquier hito de la historia de la tecnología, la programación o las matemáticas, el primer nombre necesario que debería salir es el de Ada Lovelace. Ella fue la investigadora británica de principios del XIX que escribió el primer algoritmo destinado a ser procesado por una máquina y la primera programadora de computadoras de la historia, además de una gran visionaria que marcó los siglos posteriores al predecir que los ordenadores podrían ir más allá de los simples cálculos numéricos: todo se trataba de darle las instrucciones adecuadas.

Entre los expertos que le cogieron el testigo en el campo de la computación destaca sin duda Alan Turing. “¿Pueden las máquinas pensar?”, fue una pregunta que lanzó Turing, en 1947, a un abarrotado auditorio compuesto por miembros del *National Physical Laboratory* de Londres. Este gran científico británico nacido en 1912 dejó un gran legado tanto en la práctica computacional como en la defensa de los derechos y libertades sexuales. Turing también tuvo un papel decisivo en la Segunda Guerra Mundial ya que consiguió descifrar los códigos de las transmisiones de mensajes de Enigma, la máquina utilizada por los nazis para ocultar las comunicaciones de sus maniobras bélicas. Pero apenas unos años después de la derrota del Tercer Reich, Turing fue sometido a un proceso judicial por su condición de homosexual, prohibida en Reino Unido en aquellos años.

Su conocido test o prueba de Turing definía a una máquina como sintiente, inteligente o pensante cuando fuera una máquina que, en conversación con un ser humano, este no distinguiera si estaba hablando con un aparato o con otro humano.

Esta idea ha sido la base teórica que ha ido espoleando la investigación en materia de inteligencia artificial, en una historia repleta de altibajos, de inviernos y primaveras, y que ha tenido algunos hitos famosos y conocidos, como la pugna histórica entre el ser humano y las máquinas que él mismo ha creado. Fue en 1997, cuando Kaspárov, el mejor jugador de ajedrez del mundo, fue finalmente derrotado³⁶ por Deep Blue, una supercomputadora de IBM cuyos procesadores le permitían calcular millones de posiciones por segundo, algo que no se le conoce ni a la persona humana con la mente más brillante del mundo.

La propia Gemini, de Google, si le preguntas sobre este hecho, te responde, a modo de conclusión:

Es un ejemplo del progreso de la tecnología y de cómo las máquinas pueden desafiar y superar nuestras expectativas. Es un recordatorio de que la inteligencia artificial tiene el potencial de cambiar el mundo de maneras que aún no podemos imaginar.

Esta frase nos podría recordar a algunas de las que lanza HAL 9000, el ordenador asesino de 2001: *Odisea del espacio* (1968): “*La humanidad es una amenaza para sí misma y para el universo. Yo soy la única esperanza para el futuro*”.

Como recuerda la periodista de *Newtral* Marilín Gonzalo, “la inteligencia artificial nace después de muchos años de ciencia ficción sobre la inteligencia artificial” y asegura que gran parte de la comunidad de desarrolladores sabe que, aunque puedas enseñar a hacer cosas a una máquina y que esta sea capaz de hacerlas muy bien, estamos muy lejos de que una sofisticada red neuronal artificial pueda crear, razonar o inventar algo nuevo. “Otra cuestión es que esto que hemos creado nos maraville y nos parezca espectacular”. Y lo cierto es que programas como Sora, la IA generativa avanzada por OpenAI para la generación de vídeo, lo son. Y también son la gasolina perfecta para añadir al fuego de los *deepfakes* y la desinformación, de lo que hablaremos más adelante.

Volviendo a Turing, fue él quien sugirió que, en lugar de construir un programa para simular una mente adulta, sería mejor producir uno más simple que emulara la de un niño y luego someterlo a educación. O entrenarlo. El neurocientífico Mariano Sigman, autor del libro *Artificial, la nueva inteligencia y el contorno de lo humano*, recoge el ejemplo del proceso de aprendizaje del lenguaje que hace una criatura cuando comienza a hablar, por puro aprendizaje o inferencia estadística, dice Sigman. El niño escucha, escucha, escucha –como un modelo de lenguaje de gran escala– y va prediciendo dónde va un sustantivo o cuál es la sintaxis correcta de una frase a base de prestar atención a las conversaciones de los adultos de su alrededor.

36

En febrero de 1996 Gari Kaspárov, el mejor jugador de ajedrez del mundo desde el año 85, disputó una partida de ajedrez a seis partidas. Deep Blue, uno de los primeros prototipos de supercomputadoras creada por IBM, ganó la primera pero el ajedrecista soviético consiguió reponerse y ganar por 4-2. Al año siguiente, una versión mejorada de la máquina pidió la revancha y salió victoriosa.



Todo lo que necesitas es atención: neuronas sintéticas, inteligencia y máquinas

La historia de la IA está cargada de expectativas y de condenas, de loas y de críticas, de apóstoles y apóstatas. Pero todo empezó muy despacio, casi como un juego académico. En sus inicios, la computación y la programación se veían como un sistema de instrucciones con muchas órdenes complejas para que el ordenador ejecutara acciones más o menos sencillas, centradas en resolver operaciones matemáticas. Eso fue hasta que Turing lanzó la asociación entre inteligencia y máquinas en 1956, durante la conferencia de Dartmouth, en New Hampshire, EE. UU., donde diez científicos varones estadounidenses y blancos se reunieron para tratar de desentrañar la posibilidad de materializarla. En ese evento fue la primera vez que se oyó el término inteligencia artificial, y fue pronunciado por John McCarthy, un prominente informático galardonado con el Premio Turing en 1971.

En los años 50, cuarenta décadas antes de la victoria de Deep Blue, ya se pensaba en cómo crear máquinas inteligentes que pudieran jugar al ajedrez. Las preguntas ya se habían lanzado hacía mucho: ¿cómo distinguir una máquina automática que pudiera jugar al ajedrez de una calculadora? Todavía faltaba la revolución en infraestructura computacional y la llegada, como suele pasar en estos casos, de una idea, quizá desdeñada en un principio por lo obvio y simple de su razonamiento.

La respuesta llegará desde el campo de la psicología. Coincidió en el tiempo con el auge de esta disciplina y con el avance en el conocimiento biológico de la mente humana. Los psicólogos Carl Rogers y Abraham Maslow, en su preocupación por entender qué era la locura humana, sugirieron un acercamiento empático y propusieron vislumbrar la locura como una parte intrínseca de cada uno de nosotros. La empatía fue clave para que un siglo antes de que se lanzara ChatGPT, un profesor de informática del MIT, Joseph Weizenbaum, programara Eliza en 1966: el primer bot conversacional de la historia. Aunque tenía un algoritmo sencillo y no sabía memorizar, no entendía la ironía ni podía hablar de muchos temas, sí creaba la sensación de conversador empático, con buena escucha e interés por la persona con la que hablaba, al estilo de lo que haría un buen psiquiatra. A partir de la década de los 60 fue cuando se optó por cambiar estos programas informáticos clásicos por unas unidades de cómputo a las que llamaron neuronas, y que juntas, conformaban redes neuronales con esa bioinspiración que buscaba recrear el cerebro humano.

Marvin Minsky –fundador del AI Lab del MIT, el reputado Instituto de Tecnología de Massachusetts– junto a otros científicos como John Hopfield o Warren McCulloch, cambiaron la óptica usada hasta la fecha para tratar de hacer brotar la inteligencia de un sustrato que no es inteligente. Transmutaron sus pretensiones de escribir la inteligencia humana en un programa para buscar la forma de directamente construir un cerebro digital, y ver si este podía reproducir comportamientos inteligentes. La neurociencia aportó lo que sabía sobre el funcionamiento de la mente humana: sus 86 mil millones de neuronas están interconectadas entre ellas en forma de red. Una sola neurona tiene un comportamiento muy simple, pero cuando se juntan varios miles, hasta millones, forman verdaderas redes neuronales que son las que permiten acciones tan fascinantes como el aprendizaje. Cada una de estas neuronas tiene una misión muy sencilla, por ejemplo, una detecta líneas verticales en una imagen, otra, los círculos. Cuando conectan y se unen un buen puñado de ellas, esa red neuronal es capaz de detectar qué es una nariz y qué es una boca en una fotografía. Así funciona

el cerebro humano: como una estructura masivamente paralela en la que cada una de los cientos de miles de neuronas hace algo muy básico, pero juntas, formando capas, son capaces de memorizar, asimilar, asociar, cada vez con un mayor nivel de abstracción o categorización. Es lo que se conoce como el enfoque conexionista.

Mariano Sigman en su libro *Artificial* toma prestada, para entender de qué hablamos cuando hablamos de inteligencia, la frase atribuida al psicólogo sueco Jean Piaget que dice que “inteligencia es saber qué hacer cuando no sabes qué hacer”. Una planta moviéndose buscando la luz del sol, una decisión astuta de un animal para salvar su vida, son situaciones a las que el humano puede atribuirle la cualidad de inteligente, aunque solo sea eso, una atribución. “Nosotros somos muy buenos atribuyendo intenciones que emulan nuestra propia experiencia”, le cuenta Sigman a Marcos Vázquez en su podcast³⁷. Somos hábiles viendo cosas donde no las hay, o en la línea de las palabras antes mencionadas de Will Douglas Heaven, editor de la sección de IA del MIT Technology Review, si algo parece inteligente es fácil que nos engañe y pensemos que realmente lo es. Para Sigman, los modelos extensos de lenguaje (LLM) están muy cerca de pasar el test de Turing pero este no es la prueba definitiva, no es suficiente. Su definición de la inteligencia, para un futuro cercano, sería que inteligencia es todo aquello que una máquina no puede hacer.

Para hacer rodar esa especie de cerebros sintéticos también hizo falta una pequeña revolución en el campo de la infraestructura. Una mezcla de saberes, la extracción de una idea que no iba encaminada a ese propósito pero que, como algunos efectos secundarios beneficiosos para otras patologías que se descubren por sorpresa a los medicamentos, alumbró el material necesario para soportar los cerebros sintéticos de las máquinas.

Hasta ese momento, el componente principal de las computadoras era la Central Processing Unit (CPU), responsable de ejecutar instrucciones y realizar cálculos necesarios para que los programas y sistemas operativos funcionen. Estas CPU tenían unas marcadas limitaciones, como la relativamente lenta velocidad de procesamiento de la información, el reducido número de núcleos, la capacidad de memoria limitada, etc. Estas restricciones se solventaron tecnológicamente en la industria de los videojuegos con la llegada de la empresa Nvidia en 1993 y sus avanzadas Unidades de procesamiento gráfico o *Graphics Processing Unit*, en inglés, GPU. Estos avances en cuanto a microchips y tarjetas de procesamiento era una de las piezas que le faltaba a la industria de la inteligencia artificial para avanzar y salir de sus prolongados inviernos. Gracias al aumento del poder de cómputo del hardware, fue posible agregar cada vez más cantidad de capas neuronales intermedias que dieron lugar a un nuevo tipo de red neuronal conocida como aprendizaje profundo o *deep learning*, y que constituye el subcampo de *machine learning* o aprendizaje automático más importante en el ecosistema de la inteligencia artificial.

Fue en 2015 cuando la compañía DeepMind, un año después de ser adquirida por Google, desarrolló una IA entrenada sobre la base de millones de partidas humanas de un popular juego de mesa chino llamado Go, al que los humanos llevan jugando desde hace casi tres siglos. AlphaGo, que así se llamaba el ordenador, ganó de forma sorprendente al máximo campeón del momento, con una jugada revolucionaria e inesperada. “Nos regaló una idea nueva que no se le había ocurrido a ningún jugador en los 3.000 años de historia del juego. Las máquinas, por primera vez, parecían listas para superarnos incluso en la esfera más humana: la creatividad”, escribe Sigman. En 2017, AlphaGo presentó su versión mejorada, AlphaZero, y aprendió a jugar tanto al ajedrez como al Go sin que nadie le enseñara una partida. “La máquina empieza jugando con una copia de sí misma y la clave está en que solo se le permite a una de las dos revisar su modelo de juego. Después de miles de partidas, la que puede aprender



comienza a vencer a la otra. En ese momento se realiza un nuevo clon de esa versión mejorada y se repite el proceso. Así (...) un programa jugando contra sí mismo genera un conocimiento exponencial”, teclea Sigman.

También en 2015 Sam Altman, Elon Musk y otros socios fundaron lo que era por aquel entonces una organización sin ánimo de lucro llamada OpenAI. Cuatro años después llegaría GPT-2, *Generative Pre-training Transformer 2*, un gran modelo de lenguaje desarrollado por esta compañía que utiliza esta arquitectura de red neuronal llamada *transformer*, capaz de procesar y comprender secuencias largas de texto. Esas 11 letras que suenan a película de Marvel y, sobre todo, la idea sencilla que hay detrás, son la clave, la guinda del pastel de ChatGPT.

Fueron ocho científicos que trabajaban para Google quienes, en 2017, publicaron uno de los *papers* fundacionales al ritmo de canción de los Beatles: *Attention is all you need*³⁸ era el título. En él se presentaba lo que denominaban “una nueva arquitectura de red sencilla, el *transformer*, basada únicamente en mecanismos de atención, prescindiendo por completo de redes neuronales recurrentes (RNN) y redes neuronales convolucionales (CNN)”. Estas arquitecturas para el aprendizaje profundo, los *transformers*, ayudan a identificar qué palabras o partes del texto son importantes al procesar una secuencia. A esto se refieren los “mecanismos de atención”: a cada palabra o parte de una frase le otorga un “peso” que indica cuánta atención debería prestarle la red para entender la frase. Esta capacidad de sopesar la importancia de las palabras en una oración cambió la manera de procesar el lenguaje natural (NLP en inglés), la traducción automática y el análisis de sentimientos. Los *transformers* son el cemento de los Large Language Models (LLM). Este ejemplo de ChatGPT nos aproxima más a entenderlos:

Imagina que tienes una frase: “El gato está en el jardín.” Si quieres que un modelo genere la siguiente parte de la frase, el *transformer* procesa la frase palabra por palabra, pero no en orden secuencial estricto como lo haría una red neuronal recurrente (RNN). En cambio, utiliza el mecanismo de atención para mirar toda la frase a la vez y entender qué partes son más importantes para predecir la siguiente palabra.

Hay que tener en cuenta que antes de que apareciera el aprendizaje profundo o *deep learning*, las técnicas tradicionales de procesamiento del lenguaje natural presentaban serias dificultades a la hora de que las máquinas comprendieran el contexto y los matices y sutilezas del lenguaje humano, puesto que no es una tarea fácil. Las redes neurales extensas han ayudado a que esto sea una realidad, junto a la arquitectura del *transformer*. Todo ello ha contribuido a que esta red neuronal, siendo entrenada en diferentes fases, tanto de forma supervisada como autosupervisada con un gran *corpus* de información, sea capaz de entender la sintaxis, la semántica y el contexto y pueda, además, generar nuevo texto gracias al gran entendimiento del lenguaje que posee.

Por procesamiento de lenguaje natural (NLP) entendemos un subcampo de la inteligencia artificial que se centra en la interacción entre ordenadores y humanos a través del lenguaje natural. El NLP consta de dos componentes principales: la comprensión del lenguaje natural o NLU (traducción de idiomas, análisis de sentimientos o la respuesta a preguntas) y la generación del lenguaje natural o NLG (la creación de frases desde cero). Aunque se empezaron estudiando para la traducción automática, los investigadores en seguida vieron el potencial que podría tener en otras áreas rela-

cionadas con lo que hoy conocemos con la IA generativa multimodal³⁹ –omnimodal, como diría OpenAI– y los *Large Multimodal Models* (LMM) que la hacen posible y que acaban de llegar para sustituir a los Large Language Models (LLM).

Otro de los conceptos clave en esta historia son los llamados **Modelos fundacionales**, un término popularizado⁴⁰, en agosto de 2021, por el Centro de Investigación de Modelos Fundamentales (CRFM) del Instituto de Inteligencia Artificial Centrada en el Ser Humano (HAI) de la Universidad de Stanford. Los *foundational models* se refieren a “cualquier modelo entrenado con datos amplios, generalmente mediante autosupervisión a escala, que puede adaptarse a una amplia gama de tareas posteriores”. Estos científicos supieron ver el cambio de paradigma que se avecinaba y enumeraron sus oportunidades, pero también sus riesgos y su impacto social:

“Aunque los modelos fundacionales se basan en el aprendizaje profundo y el aprendizaje por transferencia, su escala da lugar a capacidades emergentes nuevas, y su efectividad en tantas tareas incentiva la homogeneización. Esta homogeneización ofrece un gran potencial, pero requiere precaución, ya que los defectos del modelo fundacional se heredan en todos los modelos adaptados a partir de él. A pesar de su inminente despliegue generalizado, actualmente no tenemos una comprensión clara de cómo funcionan, cuándo fallan o de qué son capaces debido a sus propiedades emergentes.”

Ya en 2022 fue cuando irrumpió en nuestras vidas ChatGPT-3.5 y, desde entonces, asistimos a una cascada de aplicaciones y actualizaciones cada vez mejores, más avanzadas, más inteligentes y con procesos de adiestramiento más sofisticados. Las posibilidades de uso se amplían. Por ejemplo, puedes utilizar un modelo preentrenado tal como está, sin realizar ningún ajuste adicional (*out of the box*) o bien optar por el *fine-tuning* o ajuste fino, ese proceso de tomar un modelo de IA preentrenado y ajustar sus parámetros con un nuevo conjunto de datos concreto para que nos ayude en una tarea particular. Es la personalización o adaptación de un modelo genérico a un problema específico.

39 Se le llama multimodal porque combinan varios modos de información y puede generar contenido nuevo (texto, imágenes, o audio) a partir de múltiples tipos de datos o entradas. Un ejemplo sería la creación de imágenes a partir de texto.

40 Más de 50 expertas y expertos firmaron en agosto de 2021 un paper titulado *On the Opportunities and Risks of Foundation Models* del Stanford Institute for Human-Centered Artificial Intelligence (HAI) del Centro de Investigación en Modelos Fundacionales (CRFM) del Stanford Institute for Human-Centered Artificial Intelligence (HAI) <https://arxiv.org/abs/2108.07258>



IA y periodismo

Radiografía del sistema actual: gloria y decadencia, éxitos y despidos, y la reinvención constante de las redacciones

Sports Illustrated, una revista semanal dedicada al mundo del deporte, fue adquirida en 2021 por Authentic Brands Group. Esa compra supuso el despido de 40 personas del grupo, más de un tercio de sus periodistas. Su caso fue sonado porque se les acusó de sustituir a los redactores despedidos por periodistas ficticios creados con inteligencia artificial: una foto de perfil, una descripción de la persona y unos artículos firmados por esa persona que no existe. Y todo ello, a escondidas.

No es difícil pensar que internet se está llenando de contenido automatizado y puede que seamos ya consumidores de facto, quizá sin saberlo. La *enshittification* de la red es un hecho desde hace al menos dos años, que fue cuando el escritor Cory Doctorow popularizó este término para hacer referencia a la degradación de la calidad de los productos y servicios online. Fue elegida Palabra del Año en 2023 por la *American Dialect Society* y de tanto repetirla, cada vez es más real y está más extendida. La *enshittification* llega a absolutamente todo, tituló un artículo⁴¹ Doctorow: “Todos esta-

¿QUIÉN ACCIONA LA MÁQUINA?



mos viviendo un gran *enshitting*, en el que los servicios que nos importan, en los que confiamos, se están convirtiendo en gigantescos montones de mierda. Es frustrante. Es desmoralizador. Incluso aterrador”. Pero la *enshittificación* nombra el problema y propone una solución, según el bloguero y autor de ciencia ficción canadiense.

“Durante mucho tiempo se ha minimizado el valor de la firma, se ha castigado la presencia de marca personal con medios en los que abundan contenidos firmados simplemente por *Redacción*”, escribe Mauricio Cabrera, periodista mexicano que se define como un “terapeuta de contenidos” en su newsletter *Story Baker*. “Para recuperar algo de la confianza perdida se debe enfatizar la figura del humano que está detrás de la creación de contenido. Así los medios no solo podrán presumir de que tienen redactores humanos en vez de robots, sino también humanos altamente capacitados en los que la gente pueda confiar”. Humanos, dice, idealmente bien pagados por contar buenas historias. En declaraciones a esta investigación, este emprendedor y experto en innovación en medios, resalta la importancia de encontrar una voz y perspectiva propia para saber diferenciarse en el maremagnum de la información: “Como periodistas nos enseñaron a escribir bajo una misma estructura: qué, quién, cómo, dónde, cuándo, por qué. Es un estilo automatizado con una falta de sello distintivo fácilmente sustituible por la IA”. El periodista debe reunir cualidades como “memorable, auténtico” para “compartir y manifestarse como un creador”, a la altura de un poeta o un cineasta.

“La clase media del periodismo va a terminar padeciendo muchísimo”, vaticina Cabrera ante esta amenaza “colosal” para el ejercicio periodístico. El adjetivo lo pone Antonio Lorenzo, redactor de *El Economista* especializado en tecnología y corresponsal de este medio en el metaverso. “Hasta que llegó ChatGPT, todos los discursos sobre que la IA nos iba a robar el trabajo se centraban en cosas mucho más mecánicas, como que si los robots nos van a liberar de doblar la ropa y lo que ha pasado es que, de repente, ha aparecido algo que hace tambalear esa sensación de seguridad en según qué puestos laborales. En el caso del periodismo, que es una profesión que ya viene precarizada desde hace muchos años, le añade una capa extra de presión y de precarización”, y quien habla ahora es Judith Membrives, técnica de políticas digitales de Lafede.cat y gran experta en la AI Act, la legislación europea sobre inteligencia artificial: “El gran impacto es la precarización y no el robo de puestos de trabajo”, sentencia. En este punto coinciden la mayoría de expertas del mundo del periodismo y la comunicación consultadas.

¿Qué opinan al respecto los propios responsables de los medios? Para Juan Candela, jefe de Digital del diario *Las Provincias*, el fantasma de los despidos es solo una “paranoia”: “En una redacción todo el mundo hace de todo, no solo tareas automatizables”. Puede que ese sea el caso de su medio, *Las Provincias*, pero no existe un modelo único. En la redacción madrileña del diario *Público*, por ejemplo, existe un equipo de entre 6 y 9 personas que se dedican única y exclusivamente a lo que se conoce como *Última hora*, es decir, a subir teletipos de agencia a la web del periódico. Trabajan en turnos rotativos, más o menos de 8 de la mañana a 12 de la noche, y cuentan con un director. Un buen aprovechamiento de los modelos de IA generativa permitiría reducir este equipo y tener más redactores disponibles para trabajar en temas propios que aporten riqueza al diario. Un uso poco ético implicaría reducción de plantilla.

Sobre la utilización que se hace de las arquitecturas generativas en *Las Provincias*, Candela alude a la “gran cantidad de alucinaciones, casi un 40%” de ChatGPT-3 para argüir que, aunque su medio lleva investigando con ellas desde el inicio, utilizarlas implicaría “destrozar el periodismo”. Con ChatGPT-4 sí vislumbra cierto salto cualitativo “pero aún le queda mucho recorrido, aunque estamos haciendo pruebas constantemente”. Los tres medios valencianos consultados coinciden en que la creación de un chatbot particular, entrenado –fine-tuning– con datos y contenidos de su propio medio, sería una buena ayuda al periodista para aportar contexto, historia,



referencias y cifras de una forma más sencilla, simplemente formulando las preguntas adecuadas y facilitando notablemente la tarea de investigación.

Para Candela, el foco del debate es otro: “Lo que es peligroso para el trabajo de los periodistas es Google y Facebook. Eso es lo que nos está exterminando a todos. Gente que no paga impuestos en España y que tiene una posición de dominio absoluto. Que de cada 10 euros en publicidad que gana, 8 los gana sin tener ningún respeto por la privacidad. Esos son los que han exterminado a los periodistas y a los medios de comunicación y no la inteligencia artificial en principio, que tendrá su repercusión, pero a día de hoy todos los problemas que tienen los medios de comunicación lo son exclusivamente por esos dos actores”.

Íñigo Roy, responsable de Canales del Levante-EMV, recalca que una penalización de Google puede suponer que se caiga el tráfico de la web del Levante-EMV hasta un 70%, lo cual supone pérdidas cuantiosas de dinero: “Los medios de comunicación tenemos algo por encima de todo, que es nuestra reputación en internet, y esta es de dos formas: la reputación editorial y luego la reputación frente a Google. Si utilizas de manera indiscriminada la IA y descuidas aspectos como la veracidad, corres el riesgo de perder el prestigio ante tus lectores. Pero más importante que eso son las consecuencias que puede tener, porque Google, tras la actualización de su último *core update*, está penalizando a los medios que hacen un uso abusivo de contenido a través de IA sin tutela humana”. Cabría preguntar si el castigo se extiende también en el caso de que el contenido haya sido creado por su chatbot, Gemini. Ante la cuestión de por qué viene esta penalización de Google, Roy responde: “Porque quieren darle más valor a la información de calidad para que el tráfico sea de calidad, cualitativo y no cuantitativo”. Una calidad puesta en seria tela de juicio sobre todo desde los últimos cinco años, porque el gigante de las búsquedas es un claro empleo de la *enshittification*.

La sentencia de muerte del gran buscador

Hoy conocemos que en 2019 los directivos de Google ya temían la caída de su monopolio casi absoluto y empezaron a priorizar el negocio antes que la calidad y eficiencia⁴². Esto es, a más búsquedas, más visitas, más publicidad, más dinero. Y si para que haya más búsquedas hay que entorpecer la labor del algoritmo e incrementar sus contenidos patrocinados o sitios web spam, se hace. Como dice Marta Peirano, Google cada vez arroja más en sus búsquedas contenidos patrocinados, afiliados, sintéticos o, directamente, desinformación. La *enshittification* de Cory Doctorow siempre está justificada desde las grandes corporaciones: todo vale menos perder dinero. Alphabet, la empresa matriz que engloba el conglomerado Google, “es lo suficientemente grande para sobrevivir a la desaparición de su primer gran negocio. Pero hace falta un modelo capaz de garantizar la supervivencia del periodismo de verdad”, sentencia Peirano⁴³.

La decisión priorizar cantidad antes que calidad tomada por la compañía que domina y controla una parte tan grande de nuestros dispositivos —y de nuestra vida: del mail al calendario, del móvil al documento de trabajo, del mapa a nuestras

42 Pérez Colomé, J. (2024, 3 de mayo). *El día en que Google empezó a empeorar: “Estamos acercándonos demasiado al dinero”*. El País. <https://elpais.com/tecnologia/2024-05-03/el-dia-en-que-google-empezo-a-empeorar-estamos-acercandonos-demasiado-al-dinero.html>

43 Peirano, M. (2024, 18 de marzo). *Quién mató al buscador*. El País. <https://elpais.com/opinion/2024-03-18/quien-mato-al-buscador.html>

horas de sueño—, en un contexto en el que mantiene, desde 2014, más del 90% de cuota de mercado de los buscadores, no parecía importante. Pero desde la irrupción de enormes modelos de lenguaje conversacionales como ChatGPT, Copilot o Perplexity, se está poniendo ese liderazgo cada vez más en duda. Los competidores que le han salido están muy cerca de ser mejores y más utilizados. Perplexity, por ejemplo, actúa como un motor de búsqueda que ordena y redacta la información de internet, ofrece, según dice, “información concisa, directa y sin anuncios patrocinados, evitando la necesidad de que los usuarios hagan clic en múltiples enlaces para obtener la información”. Su creador, Aravind Srinivas, lo define así: “Perplexity es como una unión de Wikipedia y ChatGPT que puede responder instantáneamente a las preguntas de una persona sin el enorme desorden de los resultados de búsqueda convencionales de Google”. Las revistas Forbes y Wired ya han lanzado acusaciones de plagio hacia esta plataforma por haberles copiado material original y ofrecer descaradamente como fuentes primarias, artículos de otros medios que replicaron las publicaciones de Wired y Forbes sin citarlos⁴⁴. Su CEO se defiende diciendo que Perplexity no se entrena con el contenido de nadie porque su cometido es la agregación de contenido a partir de lo que otros motores generan.

Estos motores de búsqueda conversacionales contribuyen más todavía a que desaparezcan, o se reduzcan al menos, las visitas al medio original. Así muere el tráfico, del que tanto dependen hoy en día los medios para su subsistencia. Sobre todo aquellos que no consiguen fidelizar al lector y, especialmente, hacerle que pague por sus contenidos con una suscripción a su muro de pago.

Mauricio Cabrera le augura a Google Search una muerte no tan lejana: “No importa cuánto los medios pretendan aferrarse a que siga el consumo de contenido a través de búsquedas en Google. La practicidad de encontrar sumarios de información en un esquema conversacional se terminará imponiendo”. De hecho, motores de búsqueda como Perplexity se están llevando una parte significativa del pastel: 85 millones de usuarios mensuales en total según Cabrera, aunque todavía le queda para llegar a los 2000 millones de Google.

Cabrera defiende que “los medios ya están cambiando, no solo por la IA sino por la enésima evidencia de que cantidad por encima de calidad no funciona, porque no te lleva a la fidelización de una audiencia. Los medios de nueva generación deben centrarse en crear comunidades en lugar de audiencia”. Para el periodista, la prensa debería pensar en términos de desarrollo de producto que vayan más allá del contenido puramente escrito y anima a que la IA se desarrolle no solo para tareas burdas o básicas, sino también para permitir “procesos de investigación y desarrollo periodísticos que antes hubieran requerido o de muchísima inversión, o de un seguimiento que requiere presupuestos que no tiene ni siquiera el New York Times”. Y ofrece un consejo al sector:

“Tenemos una capacidad de poder generar alteraciones al sistema a partir de nuestro criterio y creo que es ahí donde el periodista, visto como un autor, como una marca personal, tiene la capacidad de generar productos que produzcan una disrupción. Es decir, la inteligencia artificial nos va a dar más de lo mismo a menos que sepamos solicitar, que sepamos dar los *prompts* adecuados. Es ahí donde nosotros podemos diferenciarnos a través de una voz, de la autenticidad, del estilo, a través de las conexiones o de las inspiraciones y referencias que tengamos y que no estén ya instaladas en el sistema”.

44

Sullivan, M. (2024, 21 de junio). *Perplexity CEO Aravind Srinivas responds to plagiarism and infringement accusations*. Fast Company. <https://www.fastcompany.com/91144894/perplexity-ai-ceo-aravind-srinivas-on-plagiarism-accusations>



El caso de los medios valencianos y el ejemplo de las corporaciones públicas

RTVE anunció que cubriría con ayuda de la IA las elecciones al Parlamento de Catalunya del 14 de mayo de 2024, a través de un convenio con la empresa Narrativa. Fue un apoyo a la labor periodística en forma de texto automático en lenguaje natural para dar cobertura a los resultados en todo el territorio catalán, en un proceso que, como destacaron en un comunicado⁴⁵, estuvo supervisado y entrenado por profesionales de RTVE “para certificar y verificar su calidad informativa, editorial y visual”. Porque también estaba prevista la generación de imágenes y voces sintéticas “que informen de los resultados obtenidos tanto en español como en catalán, permitiendo una cobertura más amplia y equitativa del territorio en el que se celebran los comicios, cerrando brechas territoriales e informativas”. Esta es su arma para luchar contra la desinformación y la manipulación informativa: publicar la mayor cantidad posible de noticias con rapidez y precisión, dicen, siempre basadas en datos oficiales.

No es la primera vez. En las Generales y Municipales de 2023 ya se implementó algo similar para generar noticias de los resultados electorales en los cerca de 5.000 municipios de España que cuentan con menos de 1.000 habitantes. Según la corporación, fueron más de 76.000 las publicaciones en texto producidas bajo esta firma: ‘El contenido de esta noticia ha sido generado automáticamente mediante inteligencia artificial a partir de los datos proporcionados por el Ministerio del Interior del Gobierno de España’.

La televisión autonómica À Punt Mèdia sigue la estela del ente público estatal. Es el medio valenciano que hace uno de los usos más inteligentes de los sistemas de aprendizaje automático, en parte gracias a un convenio, firmado en 2020 y pendiente de renovación mientras se redacta esta investigación, con la Universitat Politècnica de València para la subtitulación asistida en tiempo real de contenidos audiovisuales.

“Nosotros huimos del concepto IA. Lo que hacemos son procesos de aprendizaje o *machine learning*”, redefine Pau Peiró Sendra, responsable por parte de la Societat Anònima de Mitjans de Comunicació (SAMC) del proyecto *Recerca i desenvolupament de sistemes intel·ligents avançats per a l'accessibilitat, catalogació i indexació de continguts textuais, radiofònics i audiovisuals*, que desarrolla junto a la UPV desde 2022. Trabajan con sistemas algorítmicos de reconocimiento automático de voz (ASR), es decir, de transcripción del lenguaje hablado al escrito, para que la máquina, por ejemplo, sepa procesar adecuadamente los registros sonoros de la lengua, y también con aprendizaje profundo o *deep learning* para mejorar la creación del corpus de la lengua valenciana. Su tasa de acierto es del 86% y la de latencia, inferior a los 0,7.

El paso del reconocimiento del lenguaje al aprendizaje profundo tiene unas claras ventajas para el proyecto; por ejemplo, les permite aplicar capas correctivas y corregir los fallos orales para subtítularlos con la palabra o expresión correcta: “Queremos que sea ‘subtitulación asistida’ aunque estemos tan acostumbrados a que sea ‘automática’. Somos conscientes de que la lengua está viva y siempre se debe ir actualizando”. Peiró recalca que esta tecnología es replicable y de código abierto, de hecho, ya ha habido contactos con otras televisiones autonómicas.

45

RTVE prepara la incorporación de la IA a la labor periodística en la cobertura de nuevos procesos electorales, publicado en la web de RTVE.es el 6 de marzo de 2024. <https://www.rtve.es/rte/20240306/rte-prepara-cobertura-nuevos-procesos-electorales-ia/16001716.shtml>

De momento las personas expertas consultadas reconocen que, pese a que la inteligencia artificial en sí tiene muchas oportunidades, desde el ente autonómico no se está apostando por la generativa y sí por mejorar la accesibilidad de los contenidos y por ofrecer herramientas que ayuden a la redacción a ser más productiva, eficiente y, sobre todo, más accesible. Un factor a tener en cuenta es que la implementación de tecnologías novedosas en la televisión pública suele ser un proceso lento, con mucha parte administrativa y burocrática, sin olvidar la difícil cuestión de cuadrar las partidas presupuestarias destinadas a renovaciones tecnológicas, que muchas veces se ocupan en acondicionar los estudios, salas de grabación y platós. Según Peiró, más allá de la IA generativa, en À Punt se utilizan al menos 14 herramientas que echan mano de la inteligencia artificial para mejorar tareas relacionadas por ejemplo con la composición de imagen y vídeo.

Entre los proyectos deseables en un futuro cercano, Peiró apuesta por poder implementar algún día “modelos entrenados por nosotros mismos” para poder tener el máximo control posible. Que sean “como un mini ChatGPT alimentado por los contenidos de la redacción y que sirva de ayuda a los periodistas que, por ejemplo, llegan a una sección por primera vez. Sería tener una especie de asistente que le dé todos los aspectos y connotaciones básicas de cómo hemos tratado hasta ahora una temática concreta de noticias”. Un modelo *finetuneado* acorde a las características y necesidades de la redacción de À Punt.

“No hay duda de que el *machine learning* agiliza un montón los procesos internos, pero también es cierto que para poder hacer uso, tienes que tener un conocimiento previo. Eso es lo que creo yo que no se está explicando a la sociedad. Lo que nos dicen es: tienes esta app en el móvil, puedes preguntarle lo que quieras, que es un cuñado perfecto. Te lo va a decir todo, pero en modo cuñado. Debes tener el conocimiento de cómo usarlo, y no solo cómo hacer los *prompts*, sino saber lo que hay debajo”, zanja.

El periodismo local en À Punt y las voces sintéticas

Javier Lifante, Jefe del área digital de À Punt, enumera las herramientas que en este momento usa la redacción y que cuentan con Inteligencia Artificial: “Estamos utilizando algunas que nos facilitan algunos proveedores, por ejemplo TikTok, que la utiliza para eliminar un fondo. Antes lo hacías con Premiere o Photoshop y ahora es muchísimo más rápido. También tenemos aplicaciones como Metricool para publicar en redes sociales. Es verdad que lo estamos haciendo de una forma muy experimental todavía, pero básicamente esas son las dos que tenemos más avanzadas”. En la redacción de Informativos, las herramientas de inteligencia artificial empleadas para transcripciones y traducciones son una ayuda valiosa que ahorra tiempo a las y los redactores. Traducción de teletipos de agencia al valenciano y transcripción de entrevistas, por ejemplo, para elaborar contenidos para la web o para minutar materiales grabados, serían tareas que se podrían realizar incluso desde el propio CMS o Sistema de Gestión de Contenidos del ente autonómico si este incluyera estas opciones: “Es una forma de ahorrarte trabajo, luego tú solo tienes que cambiar la foto, mejorar la entradilla o el titular. El periodista explica lo que ha pasado, cuenta cómo te va a afectar a ti desde el punto de vista local, ahí es donde aporta valor”.



En la parte de redes sociales que coordina Lifante, las técnicas automatizadas encuentran numerosas utilidades a la hora de crear, generar y adaptar contenidos a las distintas plataformas. Desde recortar imágenes de un programa de televisión para adecuarlas y que puedan ser transmitidas en Twitter o Facebook, hasta diseñar creativities de forma rápida para anunciar un programa en Instagram.

Iván Esteve, Director de Informativos y Deportes de À Punt Mèdia, tiene claro que, en las cuestiones creativas, no se sustituirá el trabajo del periodista y que estamos ante una oportunidad, dice, para mejorar las rutinas de trabajo y los procesos de elaboración de las noticias: “Para una redacción como la nuestra, la inteligencia artificial, de momento, es más una oportunidad que un peligro para los puestos de trabajo. Una oportunidad porque a medida que vayan saliendo herramientas que faciliten el trabajo de los periodistas y, a medida que los periodistas se formen, van a hacer mejor su trabajo”.

Los preceptos y exigencias de un medio público no son -o no tanto- los mismos que en uno privado: “Nuestros criterios y nuestros valores son diferentes a los de cualquier medio de comunicación de titularidad privada, que puede ser que vean ahí una oportunidad para ahorrarse costes o para mejorar beneficios. Nosotros nos regimos por otros parámetros que tienen más que ver con el servicio público y con la información de proximidad”.

Hablar sobre posibles despidos por sustitución en una televisión autonómica en la que, a fecha de hoy, se ha suspendido la convocatoria de empleo público es una cuestión compleja y hasta extraña. Aun así, Esteve aprecia: “En principio no creo que pase [despedidos por sustitución]. Sí que hay algunas categorías profesionales que están ligadas a la redacción que hacen trabajos fácilmente automatizables, que es posible que sean a los que más les afecte, como las personas que están en el archivo documentando y catalogando imágenes. Es un trabajo muy mecánico que, probablemente, la inteligencia artificial con reconocimiento de imágenes y rostros podrá hacer en un futuro bastante inmediato”.

Las televisiones, radios y periódicos ya han visto extinguirse muchas profesiones ligadas al oficio periodístico: montadores, realizadores, maquetadores, son puestos de trabajo que se han transformado en los últimos 30 años. Bien porque sus labores han sido sustituidas por algoritmos o bien porque es el propio periodista quien ha asumido ese trabajo⁴⁶. Y sigue Esteve: “A mí me cuesta imaginar a alguien escribir un reportaje de investigación con ChatGPT. Aquí tenemos la ventaja de que nuestros contenidos son todos en valenciano y aún no existen ese tipo de herramientas generalizadas que puedan redactar o crear contenido en valenciano. En un futuro, no sé si eso llegará a existir”. En realidad, eso ya es posible. ChatGPT puntualiza que es capaz de hacer traducciones en más de 50 idiomas aunque maneja mejor los mayoritarios:

Para algunos lenguajes más especializados o dialectos locales, mi capacidad puede ser más limitada en términos de precisión o matices culturales.

46

“Correctores, introductores de textos, auxiliares de redacción y maquetadores han ido desapareciendo, poco a poco, desde hace quince años en los periódicos y sus tareas han sido asumidas por los periodistas. En la radio, el redactor ha pasado también a realizar tareas de control técnico e, incluso en las emisoras pequeñas de ámbito municipal, el periodista es también director, publicista y administrativo. En la televisión, los equipos ENG (Electronic News Gathering) con cámaras pequeñas portátiles diseñadas para trabajar al hombro están formados por un único redactor que capta las imágenes, escribe el texto y edita la información para entregar una pieza ya preparada para su emisión. Esta práctica se ha extendido también a las agencias de noticias y los redactores de EFE portan además una pequeña cámara con la que recogen las imágenes de la noticia para servirla a los abonados de este servicio”. Prieto Sánchez, C. (2013). Tesis Doctoral sobre *El movimiento asociativo de los periodistas: el caso del Sindicato de Periodistas de Andalucía (SPA)*. Departamento de Periodismo, Facultad de Ciencias de la Comunicación, Universidad de Málaga. <https://riuma.uma.es/xmlui/bitstream/handle/10630/5964/Tesis%20Cristina%20Prieto.pdf>

“Con el valenciano vamos a tener una barrera. En el departamento tecnológico han entendido bien que debíamos ser nosotros mismos quienes apostáramos por crear ese *corpus* propio”. La idea es integrar un botón en Dalet –el programa de edición que se usa tanto en tele como en radio– que ofrezca una transcripción literal de una rueda de prensa, por ejemplo, y que incluya un buscador para que puedas ir rápidamente al corte de voz que interesa al periodista.

De la AI Act –la única ley que hasta el momento regula la inteligencia artificial y sus prácticas– no se extrae una prohibición de usar voces sintéticas en los medios de comunicación aunque sí se sugiere que deberían indicarlo. En radio, otra de las tareas más mecánicas y fácilmente sustituibles, como advierte Esteve, son los boletines informativos: “Tú podrías decirle a una inteligencia artificial que te interesan las noticias que hayan sucedido en los últimos 60 minutos, que incluya alguna declaración política importante que se haya producido ahí, que miren a ver si hay algún suceso o algún accidente o que mire cómo está el estado del tráfico o de los trenes. Con esa información es muy fácil elaborar un boletín de noticias”. Incluso se podría clonar la voz del periodista estrella de informativos para que la locución sonara como él o ella. Tan solo harían falta 15 segundos de un discurso suyo oral y un programa como el Voice Engine de OpenAI o la aplicación de Eleven Labs: “Probablemente sería una voz que no tendría fallos, que haría una pronunciación perfecta y sin *apitxar*. Creo que también es una de las riquezas de los medios públicos que se oigan voces diferentes y formas de hablar distintas dentro de una misma lengua. Por tanto, a mí de momento esa idea no me seduce. Seguro que técnicamente se podrá hacer dentro de muy poco pero ahora mismo ni nos planteamos sustituir a los periodistas de los boletines ni es una prioridad para nosotros”, concluye Esteve.

El responsable de informativos del ente autonómico no le ve los pies al lobo y no cree que la inteligencia artificial haya venido a quitar o a sustituir el trabajo de los periodistas en el sistema público: “Yo eso lo descarto por completo. La innovación tecnológica ha pasado siempre, los medios que tenemos ahora para la edición de las piezas en la tele y en la radio no son los que había hace 10 o 20 años. Los periodistas debemos formarnos, irán saliendo herramientas que nos facilitarán nuestro trabajo y nos ayudarán a mejorarlo, pero como ha venido haciéndose desde siempre”.

Otro de los proyectos futuros que sí cuentan con la aceptación de la dirección de À Punt Mèdia es la creación de un avatar que haga la traducción para televisión a lengua de signos de forma simultánea aunque, por el momento, solo es un conjunto de ideas en una fase inicial.



La excepcionalidad del New York Times: la denuncia a las tecnológicas por “usar su contenido sin pagar”

A finales de 2023, la compañía editora del New York Times demandó a OpenAI y Microsoft y les acusó de “copia y uso ilegal de las obras de singular valor del Times”, con las que habrían entrenado a sus modelos de IA generativa. Este trabajo de enseñarle a su famoso modelo de lenguaje extenso miles de artículos, la mayoría de ellos candados en el muro de pago, le habrían generado al emblemático diario americano un gasto de “miles de millones de dólares en daños y perjuicios legales y reales”. En la demanda se pide a las compañías que destruyan cualquier modelo de chatbot y datos de entrenamiento que utilicen material con derechos de autor del Times. “Los demandados buscan aprovecharse de la enorme inversión que ha hecho el Times en su periodismo”, además de “usar el contenido sin pagar para crear productos que sustituyen al Times y le roban audiencia”⁴⁷.

En la denuncia se incluyen 100 ejemplos concretos de textos arrojados por ChatGPT que copian palabra por palabra algunas partes de artículos del periódico. Esta especificación de que sean frases literales es importante y es lo que puede marcar la diferencia, según algunos analistas, ya que ha habido varios casos de denuncias que han sido desestimadas, como la de la cómica y escritora Sarah Silverman contra Meta y OpenAI⁴⁸. En ese caso, los jueces no consideraron que los textos que generan arquitecturas neuronales sintéticas como LLaMA (Large Language Model Meta AI, hoy conocido como Meta Llama 3) o ChatGPT tengan “similitudes sustanciales” con sus libros originales. Una tesis que defiende ampliamente el Doctor en Informática Jordi Linares porque, desde el punto de vista técnico, estos LLM no almacenan textos, artículos, imágenes ni cuadros pictóricos, sino que crean representaciones internas –llamadas *embeddings*– de las palabras y frases, para comprender las relaciones semánticas y sintácticas.

Solo son capaces de guardar datos numéricos, conocidos por el nombre de pesos. Estos son una pieza fundamental para el *deep learning* que les permite determinar la importancia de las conexiones entre las neuronas dentro de una red. Los pesos son valores numéricos asociados a diferentes elementos (como por ejemplo, palabras dentro de una frase) que empiezan siendo valores aleatorios y durante el proceso de entrenamiento, van siendo ajustados mediante algoritmos y penalizaciones en un proceso de ensayo-error con el que van afinando sus predicciones. Así van aprendiendo patrones, estructuras y relaciones entre palabras y conceptos a partir de grandes cantidades de datos textuales.

La sensación de los analistas consultados es que los tribunales deciden en estos casos como en cualquier otro juicio sobre derechos de autor, y centran su atención en si el material generado a través de estos modelos infringe las leyes del *copyright*, pero sin tener en cuenta –al menos, hasta el momento– aquello con lo que se ha alimentado a la máquina. Es el debate de si la IA aprende o copia, como adelantábamos unas líneas antes: en el primer caso no se vulnera ningún derecho y en el segundo, puede que sí.

47 M. Grynbaum, M y Mac, R. (2023, 27 de diciembre). *The New York Times demanda a OpenAI y Microsoft por el uso de obras con derechos de autor en la IA*. The New York Times. <https://www.nytimes.com/es/2023/12/27/espanol/new-york-times-demanda-openai-microsoft.html>

48 Garay, J. (2023, 10 de julio). *Comediante Sarah Silverman demanda a OpenAI y Meta por infringir derechos de autor*. Wired.

¿Qué aporta a este debate la nueva legislación europea sobre IA? Preguntamos sobre ello a Andrés Boix, profesor de Derecho Administrativo de la Universitat de València. El jurista diferencia el caso europeo del americano. En Estados Unidos, más que una legislación sobre propiedad intelectual lo que tienen es una exposición jurisprudencial sobre lo que se puede hacer y lo que no, y aquello que se considera un uso justo, o legítimo. Es lo que se conoce como *fair use*: “En España y en la Unión Europea no hay ninguna restricción legal, en estos momentos, que diga que no se pueda entrenar a la IA usando las cosas que están por ahí o a las que tienen acceso. Para poder restringirlo habría que cambiar la doctrina del *fair use* en Estados Unidos y legislar en Europa y España. Mientras no se legisle, es perfectamente posible entrenarlas así”. E insiste Boix: “No es que no haya *fair use* en Europa, lo que pasa es que la doctrina en Estados Unidos es de creación jurisprudencial y van decidiendo caso a caso sobre el *fair use*. En Europa se codifica en las directivas europeas y en nuestras leyes y, como digo, hasta la fecha no hay ninguna prohibición de uso de obras a las que tengan acceso porque están libres, o cuentan con derechos de autor y hayan comprado la licencia. Igual que uno lo consume, que lo consuma su máquina”.

La AI Act no es una ley sobre propiedad intelectual pero sí se aplican el resto de directivas en materia de derechos de autor de la Unión Europea. En el caso de España, la Ley de Propiedad Intelectual: “La gran discusión es si el uso que hacen los modelos de IA generativa, que se entrenan con contenidos que están sujetos a derechos de autor, es un *fair use*. Hasta ahora así ha sido”. Sobre el desenlace final de la denuncia del New York Times, Boix se mantiene prudente: “De momento, hay una demanda, pero no hay ninguna resolución. Tú no puedes deseducar a la máquina, así que [en el caso de que los jueces dieran la razón al periódico] se establecería una indemnización que puede ser mayor o menor, y que se pagaría a quien haya sufrido el daño”.

Otra de las cuestiones que preocupa al Times en su denuncia, además de que fusile sus reportajes sin citarlos, es que la utilización masiva de estos modelos de IA generativa perjudicarán a la estrategia comercial del diario digital, puesto que si ChatGPT es capaz de generar noticias o de dar respuestas sobre la actualidad, restará visitas a su página web. Como se apuntaba más arriba, la viabilidad de la prensa online, nuevamente, ha de ser repensada. A este respecto, en una newsletter de La Wikly, Emilio Doménech y Bosco Bárcena cuentan que el famoso columnista de tecnología Ben Thompson recuerda que el mayor talón de Aquiles de los chatbots es la falta de confianza que destilan, porque también fallan y mucho. La información fidedigna y de fiar siempre seguirá en la web del Times. Pero, “¿qué porcentaje de lectores ignorarán esa premisa fiándose de ChatGPT y prescindirán de pagar por una suscripción para pinchar un enlace afiliado?”, se preguntan Doménech y Bosco. Quizá el New York Times sobreviva a la bajada de tráfico en su web, pero habrá muchos, puede que la mayoría, que no lo hagan.

Poca presentación necesita el New York Times porque, sin duda, su fama le precede. Es, desde hace décadas, un medio referente de rigurosidad y calidad periodística que ha marcado la forma de hacer para muchos otros medios. Fueron pioneros en cuestiones hoy bien extendidas y popularizadas como el muro de pago, que utilizan desde 2011, y otras que no todos se pueden permitir sin precarizar la laborabilidad de sus empleados, como su enfoque, prioridad y presupuesto dedicado al periodismo de investigación –que le ha llevado a ganar 132 premios Pulitzer– o su extensa red de corresponsales internacionales. En diciembre de 2023 el Times anunció el fichaje de Zach Seward como director editorial de iniciativas de inteligencia artificial. Su cometido sería dar respuesta a estas preguntas: *¿Cómo debería beneficiarse el periodismo del Times de las tecnologías de inteligencia artificial generativa? ¿Pueden estas nuevas herramientas ayudarnos a trabajar más rápido? ¿Dónde deberíamos trazar las líneas rojas en torno a dónde no las utilizaremos?* Ante el avance de las aplicaciones con IA, el diario refuerza las bases que le otorgan la confianza del público y reitera que “el periodismo del Times siempre será informado, escrito y editado por nuestros periodistas expertos”.



Si hay dinero, no hay pleitos

Como vemos, el diario neoyorquino ha sido el primer gran medio en optar por la vía judicial, una opción que también escogieron algunos escritores⁴⁹ como Jonathan Franzen y John Grisham, quienes lanzaron su acusación sin paliativos: “El éxito y la rentabilidad de OpenAI se basan en la infracción masiva de los derechos de autor sin consentimiento y sin pagar un centavo de compensación a los propietarios de esos derechos”. Las tecnológicas se defienden, entre el victimismo y la amenaza, asegurando que exponerlas a la responsabilidad por derechos de autor “obstaculizaría significativamente su desarrollo”, además de que limitaría la expansión de la innovación y el liderazgo de Estados Unidos en el desarrollo de la IA. El mismo Sam Altman ha reconocido sin pelos en la lengua que hoy por hoy es imposible entrenar sus modelos sin infringir los derechos de autor.

La sonada denuncia del Times a OpenAI llega después de que ambas empresas no consiguieran llegar a un acuerdo en lo económico, cosa que sí alcanzó la creadora de GPT con Associated Press (AP). La agencia de noticias estadounidense vende la licencia de uso de parte de su archivo de texto a la compañía de Sam Altman a cambio de poder implementar la tecnología y las aplicaciones de OpenAI. Lo que no se hizo público es el acuerdo económico alcanzado. Aunque en el comunicado de ambas compañías no se hablaba de dinero, se da por hecho que existe un pago. “Nadie sabe cómo son esos acuerdos”, lamenta Marilín Gonzalo, y critica la opacidad con la que funcionan las grandes tecnológicas, las trabas que le ponen a la prensa y la poca transparencia que muestran, pese a que el mensaje público que lanzan sea muy diferente.

Estamos asistiendo a un goteo constante de noticias sobre nuevos pactos con la compañía tecnológica, al añadir el conglomerado alemán Axel Springer, el diario francés Le Monde y Prisa Media a su elenco de medios a sueldo. Fue en marzo de 2024 cuando conocimos esta colaboración entre las publicaciones del grupo Prisa –El País, Cinco Días, AS o El Huffpost, entre otros– que permitirá “a los usuarios de ChatGPT interactuar en torno a la actualidad con los contenidos de alta calidad de Prisa Media en español y de Le Monde en francés”, según se leía en la noticia publicada en El País: “Contenidos que también contribuirán, a la vez, a la formación y mejora continua de los modelos de inteligencia artificial”. Brad Lightcap, director de operaciones de OpenAI, afirmaba en un comunicado: “Nuestro objetivo es permitir que los usuarios de ChatGPT de todo el mundo conecten con las noticias de una forma más interactiva y que tengan una visión más completa de la actualidad”. Según la información que publica El País, ChatGPT incorporará resúmenes de contenidos de actualidad publicados en estos medios e incluirá los enlaces a los artículos originales.

Cuando se escriben estas líneas, se ha anunciado una nueva alianza entre el Financial Times y OpenAI, unos acuerdos buenos para ambas partes según Mauricio Cabrera aunque con un claro beneficiario, y es la compañía de Silicon Valley: “No solo adquiere el derecho a usar de manera legal la propiedad intelectual de los medios sino que se hace también con grandes sellos de calidad que validan los resultados editoriales que le presenta a sus usuarios. Es un lavado de cara necesario para cualquier tecnológica”. El goteo es constante, y también hay pacto entre OpenAI y News Corp, propietaria del Wall Street Journal, The New York Post, The Daily Telegraph y el Herald Sun.

A Cabrera estas últimas noticias le recuerdan a la promesa de Mark Zuckerberg de apostar por la información para las personas usuarias de Facebook al incluir pestañas de noticias –y acuerdos millonarios con medios por enlazar sus contenidos.

Una inyección extra que le vino muy bien a la prensa hasta que el dueño de Meta decidió poner fin a la fiesta del talonario en abril de 2024 y dar prioridad a las interacciones entre personas, eliminar Facebook News y, por tanto, desuscribir estos succulentos acuerdos económicos con medios de comunicación mayoritariamente británicos, americanos, australianos, franceses y alemanes. Así comenzó lo que Cabrera llama “la caída de los grandes iconos mediáticos de la era Facebook”, con BuzzFeed, Vice y Cultura Colectiva a la cabeza.

“Con semejante antecedente de por medio, la industria (de los medios) tiene la oportunidad de actuar diferente frente a un mismo escenario”. Pero no lo está haciendo según Cabrera, ya que vuelven a optar por entregar “sus marcas más valiosas a cambio de ingresos significativos”. Según estimaciones de las publicaciones The Information y Tech Runch, OpenAI estaría desembolsando entre 4 y 20 millones de dólares al año por las noticias. Una cantidad que puede —o no— parecer calderilla para una empresa que supera los 2.000 millones de dólares en ingresos.

Andrés Boix es de la opinión jurídica de que a los medios de comunicación en Europa les sale mejor pactar por el uso de sus contenidos con las grandes tecnológicas que denunciarlas. “Otra cosa es que quieran hacer presión al legislador para ver si establece algún tipo de restricción o prohibición que les daría más poder de negociación”. Pero eso solo te lo puedes permitir si eres el Times.

Los datos y el fin del petróleo

De entre los que han decidido coger el dinero y correr destaca el caso de Reddit, quien, según algunas fuentes⁵⁰, recibirá 60 millones de dólares al año de Google para que este pueda acceder a sus datos y entrenar con ellos sus modelos de IA. Esta plataforma web, que actúa como una especie de foro en el que los usuarios comparten y comentan noticias y publicaciones, es una de las páginas más visitadas del mundo. Son, precisamente, las personas que entran las auténticas creadoras del contenido que ha comprado Google, pero ellas no verán ni un dólar.

Esa es la forma en la que se ha ido construyendo internet, con los contenidos desinteresados de todos sus participantes puestos al servicio de una causa mayor, que es la de generar riqueza a los habitantes de Silicon Valley. Lo vimos también en 2018, cuando Microsoft compró GitHub, el mayor repositorio de código abierto de internet. En esa época, parecía que la red, ese macro espacio virtual infinito, no nos la acabaríamos nunca y ahora, a estas arquitecturas con redes neuronales que simulan las nuestras, se les ha quedado corta. Ya hay estudios⁵¹ que auguran la fecha del apocalipsis de los datos: los de alta calidad, antes de 2026. Los de baja calidad durarán hasta 2050.

Las tecnológicas saben que deben ir un paso por delante. OpenAI tuvo que tirar de imaginación para seguir entrenando a su modelo GPT-4 y para ello, se le ocurrió crear otra IA que tradujera a texto los vídeos de YouTube para alimentar a su ingente y flamante nuevo modelo de lenguaje, según una investigación del New York Times⁵²,

50 Anna Tong, A; Wang, E y Coulter, M. (2024, 22 de febrero). *Exclusive: Reddit in AI content licensing deal with Google*. Reuters. <https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22>

51 Villalobos, P; Sevilla, J; Heim, L; Besiroglu, T; Hobbhahn, M; Ho, A. *Will We Run Out of ML Data? Evidence From Projecting Dataset Size Trends* (Nov 10, 2022). EPOCH AI. <https://epochai.org/blog/will-we-run-out-of-ml-data-evidence-from-projecting-dataset>

52 Metz, C; Kang, C; Frenkel, S; A. Thompson, S y Grant, N. (2024, 6 de abril). *How Tech Giants Cut Corners to Harvest Data for A.I.* The New York Times. <https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html>



con la vulneración de los derechos de autor que esto supone. Se calcula que la famosa plataforma contiene más de 150.000 millones de minutos de vídeo, pero la mayor parte de ese material tiene poco o ningún audio utilizable, porque es música, imágenes fijas o secuencias de videojuegos. Siguiendo la investigación del Times, Meta se habría planteado la compra de la editorial Simon & Schuster para adquirir sus obras, una opción al parecer descartada porque les sale más rentable pagar multas y seguir recopilando datos protegidos por derechos de autor en internet. “Negociar licencias con editores, artistas, músicos y la industria periodística llevaría demasiado tiempo” es su principal argumento.

Algo similar le está ocurriendo tanto a Google como a Meta. De hecho, los de Zuckerberg ya han anunciado a las personas con cuenta en Instagram y Facebook que usarán sus publicaciones, entradas, textos y pies de fotos para educar a su criatura artificial. En Europa esta decisión no ha sentado nada bien a las personas usuarias, que han ejercido en masa su derecho de oposición. Meta posee dos décadas de datos de redes sociales; Google empezó su carrera a la cima a principios de los 2000 y ahí ha ido creciendo hasta llegar a los 102,2 mil millones de usuarios en todo el mundo de hoy en día. Eso son muchos datos y mucho material para alimentar a la IA, sobre todo si incluye los datos de las versiones gratuitas de Google Docs, Google Sheets o Google Slides, lo cual rebasaría ciertos límites de la privacidad. De hecho, la compañía salió a aclarar⁵³ que solo usa como material de entrenamiento aquellos archivos de Google Docs que están “disponibles públicamente” en internet, es decir, si han sido compartidos a través de redes sociales o publicados en sitios web.

Parece haber cierto consenso en la idea de que, a largo plazo, no habrá suficientes blogs, artículos noticiosos o comentarios en redes para mantener el actual nivel de desarrollo de la IA. Por eso las opciones que se contemplan para este escenario son, como decimos, usar datos privados o directamente texto generado sintéticamente por las propias herramientas de inteligencia artificial. Ambas tienen contrapartidas tanto éticas y legales como de calidad, sobre todo en el caso de que opten por contenido sintético. Ya se habla de que en ese escenario podría darse un colapso del sistema, y que ocurriría lo mismo que pasa cuando haces una fotocopia de una fotografía original, y después sigues haciendo copias de la fotocopia. Supondría alimentar a la máquina con la propia *enshittification* que ella misma ha ayudado a generar.

La escasez de este nuevo petróleo que son los datos no es una realidad lejana a nuestro tiempo y espacio. Algunos investigadores lo llaman la nueva “fiebre del oro” y anticipan, como se ha dicho más arriba, que podría estallar en los próximos dos años. Los acuerdos económicos de las tecnológicas con los grandes medios de comunicación son, sin duda, una forma de insuflar oxígeno en esta carrera por saquear internet para escalar los modelos, incrementar sus capacidades y con ello la calidad de las respuestas que ofrecen. Según el estudio antes citado⁵⁴, la cantidad de datos de texto con los que se alimenta a la inteligencia artificial generativa se ha multiplicado por 2,5 al año, mientras que la capacidad de cálculo se ha cuadruplicado. Solo el modelo más reciente y más grande de Meta, la empresa matriz de Facebook, se ha entrenado con hasta 15 billones de tokens⁵⁵, que son la unidad básica de procesamiento del texto que utiliza la IA generativa.

Una de las reflexiones más brillantes sobre la relación entre la escasez de datos y el desarrollo de la Inteligencia Artificial General (AGI) la hacen⁵⁶ los profesores

53 Europa Press. (2024, 11 de abril). *Google asegura que solo usa archivos de Docs para entrenar su IA si están “disponibles públicamente” en Internet*.

54 Ver cita 51.

55 Esta es la definición de token que da ChatGPT: En el contexto de la IA generativa, un token es una pequeña unidad de texto que los modelos de IA procesan. Puede ser una palabra completa, una parte de una palabra, o un signo de puntuación. Esto permite que la IA trabaje de manera más eficiente y comprenda mejor el lenguaje. La newsletter de Narayanan y Kapoor se llama AI Snake Oil. Acaban de sacar un libro que lleva por título *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Esta es la del 27 de junio de 2024, titulada AI scaling myths.

de ciencia de la computación de la Universidad de Princeton, Arvind Narayanan y Sayash Kapoor. Para ambos expertos, el camino actual no nos va a conducir a donde pretendemos porque el tamaño no es lo más importante. Nuestros pasos, según los investigadores, estarían “plagados de mitos y conceptos erróneos”:

“La aparente previsibilidad del escalado es un malentendido de lo que ha demostrado la investigación. Además, hay indicios de que los desarrolladores de AGI ya están al límite de datos de entrenamiento de alta calidad. Y el sector está observando una fuerte presión a la baja en el tamaño de los modelos. Aunque no podemos predecir con exactitud cuánto avanzará la IA mediante el escalado, creemos que prácticamente no hay ninguna posibilidad de que el escalado por sí solo conduzca a la AGI”.

Una de las cuestiones centrales a plantear es precisamente esta: si se podría alcanzar la ansiada Inteligencia Artificial General entrenando a las máquinas con sus propios subproductos. Si solo con eso valdría para llegar a la metacognición, que es conseguir que las máquinas piensen más como los humanos. Tanto Narayanan como Kapoor ponen el foco en que no solo es importante la escala, es decir, construir LLM más y más grandes: “La investigación sobre las leyes de escalado demuestra que, a medida que aumentamos el tamaño del modelo, el cálculo de entrenamiento y el tamaño del conjunto de datos, los modelos lingüísticos “mejoran”. La mejora es realmente sorprendente en cuanto a su previsibilidad, y se mantiene en muchos órdenes de magnitud”. Pero ¿qué es un *modelo mejor*?: “Las leyes de escalado solo cuantifican la disminución de la perplejidad, es decir, la mejora de la capacidad de los modelos para predecir la siguiente palabra de una secuencia. Por supuesto, la perplejidad es más o menos irrelevante para los usuarios finales; lo que importa son las “capacidades emergentes”, es decir, la tendencia de los modelos a adquirir nuevas capacidades a medida que aumenta su tamaño. Es cierto que, hasta ahora, los aumentos de escala han aportado nuevas capacidades. Pero no existe ninguna regularidad empírica que nos permita confiar en que esto continuará indefinidamente.” Y continúan: “Si los LLM no pueden hacer mucho más de lo que se ve en el entrenamiento, llegará un momento en el que tener más datos ya no ayudará, porque todas las tareas que alguna vez van a estar representadas en ellos ya están representadas. Todos los modelos tradicionales de aprendizaje automático acaban por estancarse; quizá los LLM no sean diferentes”. Sobre la presión y escasez de los datos, aportan:

“A menudo se discute sobre cuándo las empresas se quedarán sin datos de formación. Pero esta pregunta no tiene sentido. Siempre habrá más datos de entrenamiento, pero conseguirlos costará cada vez más. Y ahora que los propietarios de los derechos de autor se han dado cuenta y quieren ser compensados, el coste podría ser especialmente elevado. Además de en dólares, podría haber costes de reputación y legales, porque la sociedad podría oponerse a las prácticas de recopilación de datos”.

Estos LLM, como sabemos, son capaces de realizar tareas más complejas como el reconocimiento de imágenes o el procesamiento del lenguaje natural gracias al *deep learning*. Para ello se utilizan cantidades ingentes de datos. Y, como ya hemos visto, finitas. Internet ya está lleno de imágenes sintéticas firmadas por DALL-E, Leonardo AI, Stable Diffusion, Midjourney, Visual Electric o Playground. Así que los nuevos sistemas ya se están entrenando, como en un bucle, con contenido artificial que ha perdido la propiedad intelectual en la maraña de las redes neuronales de las máquinas.

Jathan Sadowski, investigador en el ámbito de la economía política de las tecnologías de la información, acuñaba en febrero de 2023 el término *Habsburg AI*, “un sistema tan entrenado con los resultados de otras IA generativas que se convierte en un mutante endogámico, probablemente con rasgos exagerados y grotescos”. El paso de la IA generativa a la degenerativa está cada vez más cerca, y con ello, el fin —o al menos, la reconfiguración— de internet. Un usuario de la red X ofrecía su solución⁵⁷: escribir un “before:2022” en el buscador de Google para que los resultados no incluyan aquellos generados de forma artificial.

De Gizmodo a La Nación: IA generativa como problema y también como solución

‘Historia de un equipo español sustituido por una IA, literalmente’⁵⁸. La noticia fue sonada y los titulares llamativos. Gizmodo en español era una publicación tecnológica que funcionaba desde 2012 y contaba con 5 redactores a sueldo. En julio de 2023, los cinco fueron despedidos por videoconferencia y sustituidos por una inteligencia artificial que traducía la versión en inglés de la publicación, con la consiguiente merma de calidad de los contenidos y también, de la plantilla. “En la redacción de Madrid de El Economista somos 70 personas”, avisa el periodista Antonio Lorenzo: “Te digo que con inteligencia artificial generativa bastaría con 20, entre personal encargado de supervisar esa IA y redactores que aporten valor diferencial desde el punto de vista de la comunicación. Lo mismo para los editores”. Carmela Ríos vaticina que “hemos generado un destructor de puestos de trabajo” al que hay que darle espacio, inevitablemente, pero no todo el espacio: “Existe el riesgo de que todo lo producido con IA generativa sea considerado *low cost* o de menor calidad. La intervención humana tiene ahora una oportunidad mayor de ser valorada en procesos como el periodismo”. La veterana reportera añade que se pueden tambalear los cimientos del pacto de confianza entre el medio y la audiencia si los periodistas trabajan de forma opaca, o sin reconocer ante sus lectores y espectadores cuándo y para qué usan ChatGPT o cualquier otro sistema. No solo para generar artículos, fotografías o vídeos sino también para hacer los titulares o elaborar resúmenes de prensa.

La industria de la comunicación vive a la deriva desde hace varios decenios, dice Íñigo Roy, responsable de Canales del Levante-EMV: “La prensa lleva en crisis más de un siglo. Ha aprendido a vivir en la crisis, entonces va apagando fuegos a medida que el entorno cambia. Por ejemplo, la prensa no creía en internet, la prensa se metió en internet por obligación”. Desde ese momento ha estado experimentando por necesidad o por competencia con diferentes modelos. El New York Times fue de los primeros en capar sus informaciones con un muro de pago y poco a poco le fueron copiando el resto de periódicos. En España fue a partir de 2019 cuando se empezó a popularizar el modelo ‘freemium’, por el que solo los contenidos más elaborados son

57 <https://twitter.com/DMueller93/status/1766091304432279634>

58 Rodríguez, D. (2023, 8 de septiembre). *Historia de un equipo español sustituido por una IA, literalmente*. La Vanguardia. <https://www.lavanguardia.com/tecnologia/20230908/9213417/artificial-historia-equipo-espanol-sustituido-ia-literalmente.html>

59 En esta entrada del blog de La Nación cuentan todo el proceso y las herramientas empleadas para monitorizar la transparencia en el proceso electoral mediante inteligencia artificial, automatización de procesos y visualización de datos: <https://blogs.lanacion.com.ar/projects/data/elections-data/computer-vision-to-monitor-elections-transparency>

de pago y desde entonces, han ido creciendo gota a gota las suscripciones. Según algunos análisis, solo el 5% de la población española estaría dispuesta a pagar por el contenido. Según Roy, en la zona mediterránea (“de Tarragona para abajo”) existen muchas menos suscripciones de pago que en el norte de la península. Y muchos menos que en el norte de Europa.

Flor Coelho, periodista de La Nación en Argentina, nos comparte cómo emplean en su medio la RPA o Automatización Robótica de Procesos, un primer paso para monitorizar los documentos electorales y validar los datos mediante crowdsourcing⁵⁹. Son proyectos implementados desde La Nación Data, su departamento de experimentación con nuevas tecnologías para la producción periodística. Desde ahí hicieron otro trabajo para el que emplearon herramientas de Procesamiento del Lenguaje Natural (NLP) con las que analizaron 692 letras de canciones de trap, populares entre la juventud argentina para, según dicen, entender de qué hablan estos artistas. De ahí concluyeron que Gucci es una marca mencionada 24 veces o que la palabra más escuchada es *yeah*, seguida de *querer*. “Sin duda todo esto te acelera el trabajo, por ejemplo, para hacer traducciones, resúmenes, para hacer primeras aproximaciones a temas... Todo eso te sirve si uno después lo sabe chequear. Para mí es una fuente más y te optimiza el trabajo si lo sabes usar para el bien”, objeta Coelho. Son herramientas de asistencia y no de sustitución, como puntualiza Jordi Linares.

Para noticias como los resultados de los partidos de fútbol en La Nación ya las están implementando. *Este resumen fue realizado por inteligencia artificial bajo supervisión de editores de LA NACIÓN*⁶⁰, es la leyenda que figura, en el lugar de la firma, en algunas de las entradas sintéticas del diario argentino. De la misma manera, en Las Provincias, están haciendo pruebas –aunque cuando se escriben estas líneas se siguen haciendo manualmente– para que las informaciones de servicio, como la información meteorológica o la programación de las salas de cine sean automatizadas.



60

Messi se compró una casa en Miami: los millones que pagó por la propiedad, dónde está y cómo es por dentro es el titular de esta información, con fecha del 12 de septiembre de 2023, que firma una inteligencia artificial “bajo supervisión de editores” en La Nación.

Protocolos para un uso de la IA responsable

El verano de 2023 la agencia de noticias americana Associated Press, junto a otros medios, anunciaron la elaboración de estándares pactados para la implementación de tecnologías de sintetización de procesos y aprendizaje profundo. Añadieron, así, unas cuantas páginas más a su libro de estilo en las que establecían pautas y consejos para sus periodistas sobre cuestiones del tipo cómo cubrir la noticia, o un glosario de terminología en materia de inteligencia artificial. Además, fijaban que “el material producido por la inteligencia artificial debe ser examinado cuidadosamente, al igual que el material de cualquier otra fuente de noticias”. O que “una foto, vídeo o segmento de audio generado por la IA no debe ser utilizado, a menos que el material alterado sea en sí mismo el tema de una historia”.

*Wired*⁶¹, la publicación especializada en tecnología referente en este campo, también hizo público para qué usan y para qué no los generadores de texto e imagen en su medio. “No publicamos artículos con texto generado por inteligencia artificial, excepto cuando el hecho de que sea generado por inteligencia artificial es el punto central del artículo”, redactan. Sí utilizan la IA para sugerir titulares o textos para publicaciones breves en las redes sociales, para añadir nuevas ideas a su *brainstorming* o como *tools* de investigación o análisis. “Pero escribir historias es otra cosa”, concretan. Tampoco tirarán de imágenes sintéticas como sustitutas de fotografía de archivo por respeto a los profesionales de la fotografía que así se ganan la vida. Prometen anunciarlo siempre, incluso cuando sea para fines no editoriales como la automatización de correos electrónicos de marketing.

De la misma manera que cada medio tiene su propio libro de estilo, se hace urgente que exista un documento que detalle, para periodistas y lectores o espectadoras, qué prácticas están permitidas y cuáles no en materia de IA. La transparencia es un proceso circular que debemos pedirle tanto a los ingenieros que entrenan LLM como a los periodistas y directivos de las empresas de comunicación.

Al cierre de esta investigación, ninguno de los medios de comunicación valencianos consultados –À Punt, Levante-EMV y Las Provincias– cuenta con nada parecido. En algún caso sí se sigue la directiva marcada –y de momento, en fase de redacción– del grupo editorial. Tampoco existe en el periódico El Economista pero, en cambio, en Newtral sí cuentan con una guía ética para el uso de inteligencia artificial⁶².

61 Varios autores. (2023, 22 de mayo). *How WIRED Will Use Generative AI Tools*. Revista Wired. Según anuncian, esta política se actualizó ese día, el 22 de mayo de 2023, “para permitir el uso de imágenes y vídeos generados por IA en determinadas condiciones”.

62 Se puede consultar en el siguiente enlace: <https://www.newtral.es/la-guia-etica-de-newtral-para-el-uso-de-la-inteligencia-artificial/20240417>

La difícil convivencia entre la IA y los derechos de autor

El pragmatismo y la adaptabilidad es una característica de las especies para sobrevivir a un medio cambiante y el ser humano lleva haciéndolo desde los inicios de su historia. Como hemos visto más arriba, parte del mundo de la cultura opta por denunciar las infracciones de *copyright* de los modelos de lenguaje de gran tamaño (LLM) para tratar de frenar la difusión de este tipo de obras “derivadas” o creadas a partir de estas herramientas. Los tres grandes sellos discográficos –Sony Music, Warner Music Group y Universal Music Group– también lanzaron una denuncia conjunta a las compañías de generación de música por IA Suno y Udío por supuesta infracción de derechos de autor. En el caso particular de Warner, amplió esa denuncia para que YouTube y otras plataformas eliminaran los contenidos culturales artificiales publicados con los nombres y voces sintetizadas de su cantera de músicos a sueldo.

Este abril de 2024 se ha aprobado en Tennessee, Estados Unidos, la Elvis Act (Ensuring Likeness Voice and Image Security Act), una norma que busca proteger a los músicos y evitar que su voz, su imagen y actuaciones sean replicadas sin autorización previa. Esta ley pretende marcar un ejemplo que se repita en el resto del mundo, aunque en realidad no ha sido promovida por los músicos sino por la industria, sobre todo por editores y agrupaciones de sociedades de derechos de autor. “Parece más diseñada para proteger la inversión de las discográficas y garantizar nuevas vías de negocio a los herederos y sociedades de gestión”, escribía Marta Peirano en El País⁶³.

Hay otra parte de la industria creativa que opta, en cambio, por aprovechar sus potencialidades y anticiparse al cambio de paradigma que se viene. Este año, Grimes, cuyo nombre real es Claire Boucher, presentó un software de inteligencia artificial llamado Elf.Tech que permite a otras personas cantar a través de su voz, y además anima a los músicos a publicar canciones con la aplicación, siempre y cuando se compartan a medias las ganancias obtenidas. “Me repartiré el 50% de los derechos de autor de cualquier canción generada por IA que utilice mi voz. El mismo trato que con cualquier artista con el que colabore. Siéntete libre de usar mi voz sin penalización. No tengo discográfica ni obligaciones legales”, escribió en la red social X⁶⁴.

La tecnología no tiene por qué ser siempre el monstruo que nos quite el trabajo. También puede jugar a nuestro favor. Solo hay que saber usarla y dale las indicaciones adecuadas para que trabaje con nosotras y no contra nosotras. Algo así debió pensar el departamento de Ciencia Computacional de la Universidad de Chicago cuando desarrolló *Nightshade*, un programa que desordena los datos de entrenamiento de forma que podría causar graves daños a los modelos de IA generadora de imágenes. Es la manera que se les ha ocurrido para luchar contra las *big tech* que utilizan obras de artistas con *copyright* para entrenar sus modelos sin el permiso de sus creadoras. Con este programa los artistas podrán añadir cambios invisibles a los píxeles de sus obras antes de subirlas a internet, de modo que si se incorporan a un conjunto de datos de entrenamiento de IA el modelo resultante puede fallar de forma caótica e impredecible y convertir los perros en gatos o no diferenciar entre los coches y las vacas. De alguna manera, envenena estos datos de entrenamiento para dañar futuras iteraciones de modelos de la IA como DALL-E, Midjourney o Stable Diffusion, entre otros tantos, al inutilizar algunos de sus resultados. “Entiendo que los artistas están muy preocupados, pero creo que hay que buscar la forma de financiar la vida de los

63

Peirano, M. (2024, 6 de abril). *No has entendido nada, Beyoncé*. El País. <https://elpais.com/babelia/2024-04-06/no-has-entendido-nada-beyonce.html>

64

<https://twitter.com/Grimezs/status/1650304051718791170>

artistas de otra manera que no sea a partir de una defensa acérrima de un *copyright* que las grandes empresas se van a saltar y que va a impedir que, en realidad, el común de los mortales tengamos un acceso mucho más libre a la cultura”, apostilla Gala Pin, diputada en el Congreso y activista por los derechos sociales en internet.

Quién expolia a quién: la propiedad intelectual en la era de internet

En marzo de 2024 la Audiencia Nacional ordena bloquear la aplicación de mensajería instantánea Telegram como consecuencia de una denuncia presentada por Mediaset, Antena 3 y Movistar, que acusaban a la aplicación de alojar contenido protegido por derechos de autor sin permiso: “Es una muestra de cómo la defensa del *copyright* lo que consigue es generar censura y clausura de espacios de organización y no ayuda a que los creadores de contenidos o los artistas no mueran de hambre”, lanzaba Gala Pin en redes⁶⁵. El verdadero expolio de la propiedad intelectual, para Marta Peirano, lo están haciendo las grandes tecnológicas en sus procesos de entrenamiento opaco con material protegido y sin respetar sus licencias, ya sean privativas o libres.

La cuestión de la defensa de los derechos de autor tiene muchas aristas y algunos apellidos ilustres, como el de Richard Stallman, fundador del movimiento por el software libre, del sistema operativo GNU y de la Free Software Foundation. Suya es la famosa frase “no es libre como en barra libre sino libre como en libertad de expresión”, para diferenciar libre y gratis, porque en inglés se usa la misma palabra: *free*.

En la primera década del 2000 se extendió en la península el movimiento por la cultura libre y conceptos como Creative Commons, software libre, licencias libres o los programas de código abierto se popularizaron y enraizaron en una extendida lucha social, espoleada por leyes que iban en contra de la naturaleza de internet y que no han podido envejecer con dignidad, porque el tiempo ha mostrado su inutilidad y cuestionado su ética. Quizá la más sonada fue la promulgada por la entonces ministra de Cultura Ángeles González-Sinde en 2011, muy criticada y controvertida entre otros puntos, por haber sido aprobada tras las presiones del lobby audiovisual estadounidense y por propiciar directa o indirectamente la censura y la creación de listas negras, la criminalización de lo que llamaban “piratería” y la vulneración explícita de la libertad de expresión y los derechos de las personas usuarias de internet. “Nos han vendido la moto de una forma intensa y absurda sobre la propiedad intelectual. Nos han llamado piratas hasta la extenuación y la realidad es que los primeros piratas han sido ellos, que para poder construir esas imágenes y vídeos han tenido que esquilmar medio internet”, protesta Karlos G. Liberal, investigador de la interacción entre personas y tecnología y responsable de Bicolabs.

Como vemos, la irrupción de la IA solo hace que echar más leña a un fuego bien prendido en el que jurídicamente todavía no hay una hoja de ruta claramente establecida. A fecha de abril de 2024, las compañías de IA generativa acumulan ya 20 demandas en Estados Unidos por violar derechos de autor. En la europea AI Act, la primera legislación sobre inteligencia artificial del mundo, como hemos dicho, no se hace referencia de forma clara a los derechos de autor aunque sí deriva a las directivas ya existentes, como la Ley de Protección de Datos –la adaptación española al RGPD

européo— o la Directiva 2019/790 sobre Derechos de Autor y derechos afines en el mercado único digital. Las leyes que protegen el copyright, tanto en Estados Unidos como en Europa, establecen que el depositario de la propiedad intelectual debe ser una persona y no una máquina. Con lo cual, cualquier contenido producido por inteligencia artificial no podría ser atribuible, al menos teóricamente. “Los procesos [y el algoritmo] son demasiado aleatorios para que sea posible decir que el autor es el usuario. ¿Significa eso que el autor es la máquina? Tampoco. ¿Lo son las personas que programaron la máquina? ¿Las que obtuvieron los datos para entrenarla? ¿Las que crearon el modelo? En este caso, el problema es que ellos no tienen ni idea de cómo se van a utilizar esas herramientas. En ambos casos, el proceso es demasiado aleatorio, así que nadie puede ser el autor”, responde al diario El País la abogada experta en propiedad intelectual Jane C. Ginsburg.⁶⁶

Ricardo Oliva pertenece a la firma de abogados Algoritmo Legal, especializada en Derecho de las nuevas tecnologías y negocios: “Los derechos no son absolutos. No lo son el derecho moral de autor, los de explotación, reproducción, transformación, comunicación pública, distribución, etc. Como ocurre con el derecho a la propiedad ordinaria, están sometidos a límites en algunos casos, sobre todo a supuestos donde un usuario va a poder utilizar una obra sin necesidad de solicitar el consentimiento del titular de los derechos, siempre que se cumplan determinados registros legales”.

Oliva rescata el concepto de minería de textos y datos y expone que el legislador —tanto la Unión Europea como la ley española— en casos como estos “permite su uso libre en base a la prevalencia”, en concreto el derecho de acceso a la información y a la investigación sobre los derechos de propiedad intelectual”. Esto no quiere decir que haya carta libre. Este Real Decreto Ley 24/2021 —la transposición de la Directiva 2019/790— define que la minería de textos y datos es toda técnica automatizada destinada a analizar textos y datos en formato digital a fin de generar información que incluya pautas, tendencias o correlaciones sobre elementos similares. Aquí se incluyen métodos para atrapar información de la web como el *crawling*⁶⁷, el *scraping* o extracción de datos específicos de páginas de internet. Para el abogado, “esto es lo que hace ChatGTP, así que claramente estaría incluido aquí”.

Siguiendo a Oliva, el artículo 67 dice que “no será necesaria la autorización de los titulares de derechos de autor para realizar reproducciones de obras y prestaciones con fines de minería de texto, siempre que se haya accedido a ellas de forma legítima. Es decir, si el autor te autoriza, podrás reproducir, transformar, tratar y hacer lo que te dé la gana, y ya no tendrás que pedir más permiso”.

Esta redacción abierta lanza dos mensajes: por un lado, da cierta permisividad a los creadores de IA para que continúen haciendo lo que hacían y a los autores les recuerda que no tienen derechos totalmente exclusivos. En el caso de que estos datos sean personales y se haya accedido a ellos sin el consentimiento expreso del titular, el RGPD establece, según Ricardo Oliva, que “los puedes utilizar siempre y cuando te pongas en contacto con la persona para informarle y ofrecerle el derecho de acceso u oposición. Pero luego, el reglamento, inteligentemente, dice que cuando es imposible identificar o ubicar a la persona cuyos datos se están utilizando en este caso, el responsable del tratamiento no está obligado a buscar ni a comunicar, basta con que en su política de privacidad informe de que ha hecho el mejor esfuerzo y se ponga a disposición en caso de que el titular le pida que los quite”. Una vez más, en los términos

66 Aunión, J. A. (2024, 30 de marzo). *Jane C. Ginsburg, experta en propiedad intelectual: “No estamos preparados para aceptar la idea de que la máquina sea el autor”*. El País. <https://elpais.com/eps/2024-03-30/jane-c-ginsburg-experta-en-propiedad-intelectual-no-estamos-preparados-para-aceptar-la-idea-de-que-la-maquina-sea-el-autor.html>

67 Según Gemini: El *crawling*, también conocido como rastreo web o indexación web, es el proceso automatizado que realizan los motores de búsqueda, como Google, para descubrir y explorar las páginas web que existen en internet. Este proceso se lleva a cabo mediante robots llamados *crawlers* o arañas web, que siguen los enlaces de una página a otra para descubrir nuevas páginas y agregarlas al índice del motor de búsqueda.



y condiciones de uso lo aceptamos todo sin leer nada. Y no es para menos: “Leerlo requiere una paciencia de santo y una licenciatura en derecho”, ironiza Marta Peirano en su libro *El enemigo conoce el sistema*. En 2015, los Términos de Usuario de la tienda de iTunes tenían veinte mil palabras. Los de Facebook, quince mil.

Desde el punto de vista técnico, Jordi Linares recalca una idea que ya ha salido antes: en los procesos de aprendizaje de tecnologías para la generación de imágenes, por ejemplo, no hay fotografías en sus *memorias*, solo números o datos asociados a esas fotografías. Así explica Linares su funcionamiento: a la imagen de un perro se le asigna el número uno, a la de un gato, el dos. Se le muestra a la máquina cientos de miles de fotos de perros y gatos y, a través de operaciones matemáticas en los que se establecen premios y castigos, se va adiestrando al ordenador para que, por probabilidad, acierte cuánto se parece lo que ve en una imagen a un perro y cuánto a un gato. “Esto no es copiar y pegar. Aquello que almacena una red neuronal son pesos, son valores numéricos en los que no podemos identificar ni el texto ni las imágenes que hayamos usado en su entrenamiento” y compara esa red neuronal con un gran fichero de datos. Para Linares, aquí reside la complejidad de garantizar la trazabilidad en estos procesos, además de la novedad que suponen hasta para sus propios creadores. Así lo detalla:

“Estas arquitecturas no son meras bases de datos sino que actúan como nuestra memoria a largo plazo. En el proceso de entrenamiento de un modelo de lenguaje, este no almacena los datos de entrenamiento. Ahí dentro solo hay parámetros, pesos, valores numéricos que multiplican las entradas de las neuronas, haciendo que estas presten más atención a un determinado valor que le viene de unas antes que de otras neuronas. Al final, es un conjunto de números, muy similar a nuestra propia red neuronal biológica. Esto se llama conocimiento paramétrico o memoria paramétrica de carácter estadístico. Como digo, actúa de manera similar a nuestra memoria a largo plazo: si algo ha aparecido en el *corpus* de entrenamiento muchas veces, es más probable que la respuesta sea correcta. De la misma forma que si, a mí, en el colegio me han repetido muchas veces el año de la conquista de América, me acordaré más de que fue en 1492 que del año en el que ocurrió la Revolución Francesa.”

En el campo del aprendizaje automático no se dan soluciones, como sí hace un programa de ordenador digamos al uso. Lo que ofrece son predicciones, puesto que es un aprendizaje basado en ejemplos y experiencia, y como ocurre en el ser humano, no nos va a dar nunca una certeza al 100%, según Linares.

La transparencia del algoritmo: cuestiones sobre la ética de la IA, los vikingos negros y el sobreajuste de los modelos

El halo de devoción y misterio alrededor de estas herramientas automatizables es enorme y viene alimentado desde varias cocinas: por un lado, porque la ciudadanía no sabe bien cómo funciona —sus propios creadores, en parte, tampoco—. Por el otro, los titulares, los lanzamientos, las declaraciones, los medios de comunicación. Los propios conversadores virtuales ponen su particular guinda al pastel porque están creados por humanos que no ocultan ese sesgo de admirar a la máquina. “Los algoritmos son diseños y construcciones humanas, y están diseñados a imagen y semejanza de la realidad que hemos generado”, recuerda Karlos G. Liberal.

Si aprenden de nosotras, es probable que cuenten con nuestras mismas distorsiones o parcialidades. “Los algoritmos [utilizados en la administración pública] pueden revelar patrones racistas o sexistas de nuestras instituciones. Si el algoritmo es racista es porque se ha entrenado con datos racistas, porque los humanos que han producido esos datos tenían comportamientos racistas”, en palabras de Ana Valdivia, profesora e investigadora en Inteligencia Artificial, Gobierno y Políticas en el *Oxford Internet Institute* de la Universidad de Oxford. Y si el algoritmo no es transparente, es porque la industria o los organismos públicos no quieren ser transparentes. “Si todo el mundo sabe cómo funciona mi algoritmo, se sabrá con qué tipo de datos está entrenado o se verá que no funciona tan bien como nos hacen creer. Hay mucho misticismo acerca de lo que se puede hacer o no con la IA”, sigue Judith Membrives, responsable de incidencia en Inteligencia Artificial de Lafede.cat y miembro de Algorights⁶⁸.

El marketing, de nuevo, se vuelve una buena herramienta para sostener un sistema que no es transparente y ayuda a alimentar su interés y popularidad. A que no decaiga el *hype*. Marilín Gonzalo nombra a la mercadotecnia de la inteligencia artificial como pieza clave y necesaria para venderla mejor. La misma legislación europea, la AI Act, habla de *providers* y *deployers*, es decir, proveedor e implementador, porque la IA es, sobre todo, un negocio dentro del mercado capitalista. “No es una ley que proteja los derechos humanos: es una norma orientada a lo que le llaman *product safety* o seguridad del producto, pero esto no implica que per se esté protegiendo derechos humanos”, avisa Membrives.

“Llevo más de 15 años trabajando en la responsabilidad tecnológica. Creo que la tecnología en general puede ser y ha sido en gran medida una fuerza positiva en este mundo. Pero eso solo puede seguir ocurriendo si existen estas fuerzas compensatorias que tratan de alinear los incentivos entre las empresas y el resto de la sociedad”, apunta Arvind Narayanan, profesor de ciencia computacional de la Universidad de Princeton⁶⁹.

68 Algorights es una red colaborativa que promueve la protección de los derechos humanos y la participación de la sociedad civil en el ámbito de las tecnologías de la inteligencia artificial. Su labor durante toda la redacción de la AI Act ha sido remarcable, con un seguimiento exhaustivo junto a otras organizaciones de toda Europa y con un proceso de divulgación muy interesante para que el conjunto de la sociedad comprenda cómo le afecta esta norma.

69 La revista Time designó a las 100 personas más influyentes en el mundo de la inteligencia artificial y entre ellas estaban Arvind Narayanan y Sayash Kapoor, el primero es profesor de ciencia computacional y el segundo estudiante de doctorado, ambos de la Universidad de Princeton. <https://time.com/collection/time100-ai/6308266/arvind-narayanan-sayash-kapoor>



Desde la psicología de las redes neuronales humanas, se ha visto que el miedo también es un arma de control y para Marilín Gonzalo, OpenAI lo sabe: “Ellos están construyendo una AGI, una inteligencia de propósito general, cuando ni siquiera ese concepto está totalmente ni consensuado ni definido por la mayoría de los científicos que trabajan en esto” y rememora que ha sido el propio CEO de OpenAI quien pedía una regulación debido a los riesgos que entraña el uso de estas tecnologías. “¿Y quién va a ayudar a esa regulación si no es la gente que está trabajando en ello? Es un discurso que busca poder, o sea que busca mantener el poder que están generando. Cuando te pones a hablar con los científicos les parece muy poco serio esto. Es más, te dicen, hemos logrado muchas cosas y está muy bien, pero estamos muy lejos de lograr una cosa similar a lo que ellos dicen de una AGI”.

Esta misma línea de pensamiento de Altman lo comparte Brad Smith, presidente de Microsoft, en una entrevista para El País: “Cuanto más poderosa se vuelve una tecnología, más fuertes tienen que ser las salvaguardas y controles que la acompañen. Necesitamos un nivel de regulación que garantice la seguridad. A veces me sorprende cuando observo a personas en el sector tecnológico que dicen que no deberíamos tener esa regulación. Cuando compramos un cartón de leche en el supermercado no nos preocupamos de si es seguro beberlo, porque sabemos que hay una base de seguridad que lo garantiza. Si esta es, como creo que es, la tecnología más avanzada del planeta, no creo que sea irracional pedir que tenga al menos tanta regulación de seguridad como la que tenemos para un cartón de leche”⁷⁰. Los que crean la IA son los primeros interesados en regularla, eso sí, siendo partícipes de la redacción de las normas. Judith Membrives alerta de la cantidad de dinero que invierte el lobby tecnológico en Bruselas para poder tener “leyes a medida”. Según algunos medios, más de 100 millones de euros al año.

Es lógico pensar que, si esta es, como dice el directivo de Microsoft y tantos otros, la invención más disruptiva e importante desde que se descubrió la imprenta, la carrera por maximizar los beneficios entre las *big tech* va a alcanzar —y lo hace ya— niveles estratosféricos. Solo hay que ver el montante de facturación de empresas como Nvidia, Microsoft o la propia OpenAI. Pero en la intersección entre tecnología, poder y beneficio económico se encuentra la persona que consume, compra, utiliza esa tecnología, por la que paga un precio, bien monetario, bien en paquetes de datos y privacidad. Son personas las que entrenan a ChatGPT y las que producen el material de entreno, son personas quienes usan ChatGPT, por tanto, la tecnología no es ni puede ser neutral.

Sobre los sesgos de estas herramientas generativas se han escrito ríos de bits. De un lado y de otro. Una de las más sonadas fue la noticia de la cancelación durante un tiempo de la opción de Gemini de generar imágenes por la incomodidad que producía su ‘diversidad forzada’⁷¹. Es decir, por devolver imágenes históricamente erróneas, como curas mujeres, soldados nazis negros, vikingos con rasgos racializados, guerreros de la Grecia clásica de rostro asiático, o imágenes de los padres fundadores de Estados Unidos con pieles oscuras. Los resultados eran “históricamente incorrectos”, como reconocía la propia Gemini si le preguntabas por qué ya no podía generar imágenes. Esto es un fallo por sobrecompensación, o por sobreajuste, según concreta Karlos G. Liberal. Lo opuesto justamente a las acusaciones que recibió DALL-E, la IA generativa de imágenes de OpenAI, de incluir sesgos marcadamente racistas y clasistas al ofrecer imágenes de forma sistemática de presos racializados y directivos varones blancos.

70 Fernández de Lis, P. (2024, 20 de febrero). *Brad Smith, presidente de Microsoft: “Debemos tener una manera de ralentizar o apagar la inteligencia artificial”*. El País. <https://elpais.com/tecnologia/2024-02-20/brad-smith-presidente-de-microsoft-debemos-tener-una-manera-de-ralentizar-o-apagar-la-inteligencia-artificial.html>

71 Doménech, E. (2024, 24 de febrero). *¿Vikingos negros? Google Gemini y sus imágenes con diversidad forzada*. Newtral. <https://www.newtral.es/vikingos-negros-gemini-imagenes-diversidad/20240224>

“La generación de imágenes por IA de Gemini origina un amplio abanico de personas. Y eso es bueno en general, porque lo utiliza gente de todo el mundo. Pero aquí está fallando”, argumentó a la revista Wired⁷² Jack Krawczyk, director senior de gestión de productos en Gemini Experiences de Google. “Estamos trabajando para mejorar este tipo de representaciones de inmediato”, declaró a la publicación y en X apostilló: “Diseñamos nuestras capacidades de generación de imágenes para reflejar nuestra base global de usuarios, y nos tomamos en serio la representación y el sesgo. Seguiremos haciéndolo en el caso de las preguntas abiertas (las imágenes de una persona paseando a un perro son universales). Los contextos históricos tienen más matices y los adaptaremos”.

Elon Musk los acusó⁷³ de tener un sesgo woke y avisó del peligro criminal de lo que él llama ‘woke AI’ y que sería sinónimo de IA asesina. Así lo escribió en X: “Si una IA está programada para impulsar la diversidad a toda costa, como lo estaba Google Gemini, entonces hará lo que sea para provocar ese resultado, potencialmente incluso matando a gente”.

Dispensadores de contenidos a la carta en época electoral: el papel de las plataformas de difusión en la contención de la desinformación

“A medida que se vayan haciendo más populares [las herramientas de creación de contenido] es razonable esperar que serán utilizadas en importantes contextos políticos”, le cuenta Renée Diresta a Marta Peirano en este reportaje para El País⁷⁴. “Las próximas campañas electorales tienen un dispensador de contenidos a la carta”, apunta Peirano. Todo esto en el año –2024– en el que se celebran elecciones en 74 países: al menos 20 países de la Unión Europea y en otros 47 de todo el mundo, entre ellos, Estados Unidos.

Desde la organización internacional AlgorithmWatch se hizo un estudio⁷⁵ junto a AI Forensics a finales de 2023 en el que se concluyó que Bing Chat de Microsoft –ahora llamado Copilot– que utiliza el modelo GPT-4 (aunque entrenado con una parte de datos diferentes), proporcionaba información errónea hasta en un tercio de las preguntas que le realizaron. “Microsoft parece no poder o no querer solucionar el problema” denunciaban. El chatbot se inventaba desde falsos escándalos sobre políticos reales hasta candidatos obsoletos o fechas falsas. “La UE y los gobiernos nacionales deben asegurarse de que las empresas tecnológicas rindan cuentas, especialmente a medida que los sistemas de IA se integran en productos que ya se utilizan ampliamente”, proclaman.

72 Gilbert, D. (2024, 22 de febrero). Google’s ‘Woke’ Image Generator Shows the Limitations of AI. Wired. <https://www.wired.com/story/google-gemini-woke-ai-image-generation>

73 <https://twitter.com/elonmusk/status/1768746706043035827>

74 Peirano, M. (2023, 10 de febrero). Cómo evitar que ChatGPT provoque un nuevo asalto al Capitolio. El País. <https://elpais.com/ideas/2023-02-10/como-evitar-que-chat-gpt-provoque-un-nuevo-asalto-al-capitolio.html>

75 V.V.A.A. An analysis of Microsoft’s Bing Chat Generative AI and elections: Are chatbots a reliable source of information for voters? Diciembre 2023. AI Forensics, AlgorithmWatch. https://algorithmwatch.org/en/wp-content/uploads/2023/12/AlgorithmWatch_AIForensics_Bing_Chat_Report.pdf

En marzo de 2024, un grupo de investigadores del *Center for Countering Digital Hate*, una organización británica que trabaja sobre discursos de odio y desinformación online, probaron⁷⁶ algunos de los generadores de imágenes con IA más populares del momento. En concreto: Midjourney, DreamStudio de Stability AI, ChatGPT Plus de OpenAI y Microsoft Image Creator. Su conclusión –nada sorprendente– fue que todas estas herramientas podían crear imágenes engañosas relacionadas con los candidatos a la presidencia de Estados Unidos: “Las protecciones existentes son inadecuadas”, concluía el informe, aunque reconoce que algunas empresas sí toman ciertas medidas, como poner marcas de agua en las imágenes generadas –es el caso de Stability AI, propiedad de DreamStudio– para que sus imágenes sean identificables, o prohibir expresamente “la creación y el uso indebido de contenidos engañosos” en sus normas de uso. Por su parte, OpenAI, según declaraciones a la CNN, estaría diseñando “mitigaciones como declinar las solicitudes que piden la generación de imágenes de personas reales, incluidos los candidatos [a las elecciones presidenciales de EE. UU.]”. Con todos los ojos puestos en los comicios americanos, la industria no ha tenido más remedio que mostrarse proactiva.

En la Conferencia de Seguridad de Múnich (MSC) se firmó un Acuerdo tecnológico para combatir el empleo engañoso de la IA en las elecciones de 2024. En ese encuentro se habló del auge de estos potenciales “dispensadores de contenidos a la carta” en un año de tanto flujo electoral, una cuestión que preocupa de forma generalizada a los *fact checkers*, según explica Marilín Gonzalo: “Siguiendo algunos estudios, el impacto de los *deepfakes*, sobre todo en vídeo, no es tan grande respecto a las *fake news* o las campañas de desinformación y de distribución de bulos digamos tradicionales a través de las redes sociales, a las que estamos más habituados”. Hasta el momento, según Gonzalo, las imágenes creadas con IA generativa en campaña electoral triunfan más por el lado de la simbología que por la voluntad de engañar o confundir.

Ahí está el caso de las imágenes de Javier Milei, difundidas en la campaña electoral argentina, en la que el hoy presidente enseñaba un gran león –la figura simbólica con la que se ha autorrepresentado– liberando de una gran jaula a la ciudadanía a las puertas del Congreso de los Diputados⁷⁷. “Para mí lo grave es que se extienda esa creencia de que no te puedes fiar de nadie ni de nada, porque justamente eso es lo que buscan las campañas de desinformación: ensuciar el espacio informativo y que la gente piense que todos los medios mienten”. Esa actitud escéptica y desencantada ha aumentado al mismo tiempo que el voto a la ultraderecha. “El descreimiento nos conduce de cabeza al cinismo”, publica la escritora Irene Lozano en *El País*⁷⁸: “Ante la duda, el ruido y los torrentes de información contradictoria, un atajo frecuente nos dice: cree al líder de tu tribu, ya sea el partido al que votas, el profesor de pádel (...) o un primo meteorólogo. Esto (...) contribuye a la división y daña profundamente el debate público, que es el combustible del que se alimenta la democracia, (...) que es un régimen de opinión pública y funciona mal si no hay un debate sano”.

La diputada Gala Pin comparte esta idea de Gonzalo y manifiesta que le preocupan más las elecciones de los próximos tres años cuando la IA generativa se ratifique como “un elemento de asentar la mentira”, pero sí considera que ya hoy en día contribuye a esta guerra epistémica “y a la estrategia de la extrema derecha de deslegitimar a las instituciones democráticas generando una especie de estado de

76 VV.AA. *FAKE IMAGE FACTORIES. How AI image generators threaten election integrity and democracy*. Marzo 2024. Center for Countering Digital Hate. <https://edition.cnn.com/2024/03/06/tech/ai-images-election-misinformation>

77 Redacción La Voz. (2024, 2 de febrero). *Un león frente al Congreso, el pueblo y una jaula: polémica por una imagen hecha con IA y publicada por Milei*. La Voz de Argentina. <https://www.lavoz.com.ar/politica/un-leon-frente-al-congreso-el-pueblo-y-una-jaula-polemica-por-una-imagen-hecha-con-ia-y-publicada-por-milei>

78 Lozano, I. (2024, 14 de marzo). *Si no te crees nada, no confías en nadie*. *El País*. <https://elpais.com/opinion/2024-03-14/si-no-te-crees-nada-no-confias-en-nadie.html>

ánimo e imaginario colectivo en el que la pérdida de soberanía popular sea percibido como algo mucho menos grave por parte de un sector de la población”. La diputada de Sumar en el Congreso participa en la Comisión de calidad democrática que se centra precisamente en *deepfakes* y desinformación. “Me preocupa la generación de imágenes que ridiculizan, no tanto que mienten, sino que de alguna manera deslegitiman desde una falta de respeto. Me preocupa más por ahí, por el papel de los *deepfakes* para deslegitimar determinadas figuras e instituciones democráticas, que no tanto como un elemento que instaure la mentira *stricto sensu*”. Y zanja: “Para mí, la parte que más nos tiene que preocupar en relación con la inteligencia artificial es: qué datos, de quién, cómo y para qué. Deberíamos poner el foco en el almacenamiento, tratamiento y venta de nuestros datos, que es lo más preocupante”.

El doctorando de la Universidad de Princeton Sayash Kapoor arroja luz al respecto y mueve el foco: “Hemos oído predicciones sobre toda esta desinformación catastrófica que acabará con nuestra capacidad de discernir los hechos de la ficción. No esperamos que eso ocurra pronto. Todo esto no quiere decir que la desinformación no sea un problema importante, sino que nuestra capacidad para contener la desinformación depende mucho más de contener la distribución de la desinformación. Plataformas como Twitter, Instagram y YouTube tienen un papel mucho más relevante que desempeñar”.

Según las reflexiones del informático Jordi Linares, este 2024 va a ser el año del vídeo generado a través de inteligencia artificial y serán estos próximos meses cuando vamos a vislumbrar el futuro del mundo audiovisual. De hecho, en febrero de 2024 OpenAI siguió sorprendiendo al mundo y lanzó Sora, su herramienta para crear vídeo a partir de líneas de *prompt* con una calidad hasta ahora no conseguida muy próxima al fotorrealismo. Los titulares de los medios jalean para que no bajes de la ola de admiración: ‘Todo lo que has visto de la IA eran migajas y el último invento de OpenAI lo demuestra’, arengaba el diario El Confidencial.

Karlos G. Liberal concreta que es cuestión de tiempo que Sora haga unos vídeos perfectos, que no puedan ser distinguidos de otro hecho por una mano humana: “Sora es una herramienta de *deepfake* aunque no está hecho para eso”. Pero la de OpenAI no es la única: ya están disponibles para trabajos de diseño y animación gráfica, animación 3D y vídeo de gran calidad de otras compañías, como Gen-2 Runaway (“No lights. No camera. All action” se lee en su web)⁷⁹ o Luma Labs (“Dream Machine is an AI model that makes high quality, realistic videos fast from text and images”⁸⁰), por nombrar dos.

A propósito de Sora, Marilín Gonzalo colgaba dos publicaciones⁸¹ en X poniendo en duda su funcionamiento y acusando a OpenAI de hacer una campaña promocional sin que ningún periodista “o técnico exterior” haya probado el modelo:

“Como periodista tecnológica he visto muchas campañas de marketing que los medios se dan prisa por reproducir, pero al menos cuando lo hacían, Apple o Microsoft daban a probar el producto a algún periodista o técnico externo (aunque estuvieran en EE. UU.)”.

“Lo de OpenAI con Sora es solamente marketing, no hay ningún informe técnico ni OpenAI ha demostrado que el modelo funcione realmente. Y dice que no lanzará Sora en breve. Imposible comprobar o escribir algo medianamente serio o verificado sobre esto”.

79
80
81

<https://research.runwayml.com/gen2>
<https://lumalabs.ai/dream-machine>
<https://twitter.com/marilink/status/1759552461306466699>



Como explica Marilín Gonzalo, la palabra *deepfake* es una mezcla de *deep learning* (aprendizaje profundo, la técnica de IA que se utiliza para crearlos) y *fake* (falso). Los *deepfakes* son imágenes, audio o vídeo artificiales manipulados o producidos en su totalidad con algoritmos de inteligencia artificial. Cuanto más sofisticadas son las herramientas de IA generativa, más reales parecen y, por tanto, más complicadas son de diferenciar. La mayoría de *deepfakes* son vídeos o imágenes pornográficas falseadas de mujeres sin su consentimiento, tanto famosas como desconocidas. “Se habla mucho de los riesgos que vendrán, pero hay muchos que ya están aquí, como el porno *deepfake*”, denuncia Gonzalo, “ya está teniendo un gran impacto tremendo en mucha gente en muchos lugares del mundo, desde poblaciones locales donde les empieza a faltar el agua por los sistemas de Data Centers o porque los sistemas de extracción de minerales que se necesitan para este tipo de tecnologías les están dañando”.

Mariana Spring es especialista en *fake news* de la BBC y forma parte de *Verify*, el equipo de *fact checkers* de la cadena pública británica encargada de comprobar la veracidad de las informaciones virales. Ella sostiene⁸² que la IA va a hacer más compleja todavía la labor de los periodistas y recuerda la difusión de una grabación falsa⁸³ del alcalde de Londres en la que supuestamente decía que iba a aplazar un desfile de una fiesta local para priorizar una manifestación a favor de Palestina. Este bulo agitó la calle y hasta se convocaron protestas para mostrar rechazo ante esta medida, sin saber que solo era un *fake* con apariencia hiperrealista. “Lo peor es que ni siquiera te hace falta la IA; hay tácticas muy simples como los memes, las teorías de la conspiración, los mensajes online, que son igualmente eficaces en la difusión de desinformación, y con las que todavía no hemos conseguido terminar de lidiar”.

82 Aunión, J. A. (2024, 23 de marzo). Marianna Spring, especialista en 'fake news' de la BBC: "Si conoces a alguien que cree en teorías conspirativas, no le abandones, intenta entender de dónde viene". El País.

<https://elpais.com/eps/2024-03-23/marianna-spring-especialista-en-fake-news-de-la-bbc-la-desinformacion-tiende-a-hacerse-viral-porque-queremos-creerla.html>

83 Spring, M. (2024, 13 de febrero). Sadiq Khan says fake AI audio of him nearly led to serious disorder. BBC News. <https://www.bbc.com/news/uk-68146053>



Legislación y marco regulatorio

Mecanismos legales de control de la desinformación

¿Cómo controlar y contener los *deepfakes*? El marco regulatorio es difuso y en ocasiones poco eficiente por la complejidad de regular las cuestiones que tienen que ver con internet y que atañen a la libertad de expresión. Existen al menos dos artículos en el código penal español que castigan la difusión de noticias falsas o inexactas, de igual forma que la Ley General de la Comunicación Audiovisual regula qué tipo de contenidos audiovisuales se difunden. Hay también un Plan de Acción contra la Desinformación impulsado por el Gobierno, aprobado en 2020, que propone, entre otras medidas, la creación de una Comisión Permanente contra la Desinformación⁸⁴.

En el plano legal europeo, Andrés Boix extrae de la AI que “todo aquello que pueda hacerse pasar por creación humana y que un observador externo pueda confundirse y no saber si es humano o no, hay que explicarlo. Está permitido hacerlo, pero hay que explicar que ha sido generado por inteligencia artificial”. De nuevo, los *deepfakes* no están prohibidos como tal en la legislación europea pero debe avisarse que es un contenido falso, salvo que sea evidente. “A nivel periodístico, aunque no esté específicamente regulado, creo que sí que hay que indicar cuándo un contenido ha sido elaborado con IA. Además de que es muy buena práctica. Los medios de comunicación, probablemente, van a decir que mientras no haya una obligación expresa ellos no están obligados. Yo creo que la lógica subyacente a la regulación es que todo aquello que se pueda confundir y parecer que es humano cuando en realidad es inteligencia artificial, hay que indicarlo”, suscribe Boix.

Existe cierta disparidad de opiniones legales sobre si la AI Act establece o no exigencias de trazabilidad y transparencia. Así responde Lorenzo Cotino, Catedrático de Derecho Constitucional de la Universitat de València: “Sobre transparencia en el AI Act hay muy poco. Muchos se confunden porque hay obligaciones de transparencia en el artículo 13⁸⁵, pero esa transparencia hace referencia a la información que el fabricante [de IA] debe dar a quien compra el uso del sistema y no está pensada para los afectados”. Es decir, dentro de la transacción de compra-venta, el proveedor de sistemas de inteligencia artificial deberá facilitar toda la información necesaria al comprador, como un manual de instrucciones en el que aparezca claramente la forma de uso.

“El artículo 50⁸⁶ también habla de transparencia, pero se refiere a la obligación de informar si se interactúa con un chatbot o de informar que es una ultrafalsificación (audio, vídeo, imagen). Pero la transparencia sobre la que he escrito un par

84

En la página web del Ministerio de Asuntos Exteriores, Unión Europea y Cooperación está el informe completo con fecha de 2019, además de enlaces a propuestas europeas más elaboradas y actualizadas como la EU vs. Disinfo: <https://www.euvdsdisinfo.eu>

85

<https://artificialintelligenceact.eu/es/article/13>

86

<https://artificialintelligenceact.eu/es/article/50>

de libros y varios artículos, esencialmente no se regula”, responde Cotino. De hecho, la mayoría de sistemas de inteligencia artificial no tendrán que cumplir con la ley porque solo se aplica la legislación a los sistemas considerados de alto riesgo⁸⁷. En el artículo 50 que nombra Cotino se han añadido, a última hora, esas ciertas exigencias de transparencia de las que hablaba Boix, pero son “obligaciones muy pequeñas”, limitadas solo a que debes informar, salvo que sea evidente, que estás tratando con un contenido sintético. Ese deber de informar solo se aplica para asuntos o informaciones que puedan ser de interés público⁸⁸. “Es una obligación de las flojitas”, reprocha Cotino e insiste: “No existe en el reglamento la transparencia para los algoritmos”.

Esta ley, como apuntaba Judith Membrives más arriba, es un reglamento para fabricantes de inteligencia artificial en el que apenas aparecen los derechos de la persona afectada por la utilización de estos sistemas de alto riesgo. Sí que hay obligación para los fabricantes de sistemas de IA de alto riesgo de diseñarlos desde el inicio basándose en una serie de requisitos: análisis de riesgos, estudio de impacto sobre los derechos fundamentales, documentación técnica, robustez, fiabilidad, seguridad, supervisión humana, etc.

Casi en el último momento se ha incorporado un “pequeño derecho a la explicabilidad”, que aunque no está muy claro, indica que si hay un sistema que tome decisiones solo automatizadas se debe explicar la lógica del algoritmo, un precepto similar a lo establecido en el Reglamento Europeo de Protección de Datos. “Habrá que ver, con los años, si esto de verdad sirve para saber cómo funcionan los algoritmos, sobre todo los del sector público”. En cuanto a la trazabilidad, en los sistemas de alto riesgo se establece una especie de tacómetro, unos registros que se guardarán mínimo 6 meses.

En el estrato europeo, existe, asimismo, el Código de Buenas Prácticas sobre Desinformación, firmado por las principales plataformas tecnológicas en 2018, aunque este no tiene carácter legal obligatorio, cosa que sí posee la Ley de Servicios Digitales (Digital Services Act) que incluye la obligatoriedad de retirar contenido como *deepfakes* y marcar claramente los contenidos sintéticos. Otro de los imperativos que establece es que las grandes plataformas adapten sus sistemas de recomendación para evitar la amplificación algorítmica de la desinformación.

Los intentos legales por implementar mecanismos de transparencia o protección del derecho al honor están resultando, a todas luces, insuficientes. Ha pasado con las imágenes *deepfakes* de las chicas de Almendralejo⁸⁹ pero también con imágenes falseadas de Taylor Swift, que recibieron 25 millones de visualizaciones antes de que X las eliminara de su plataforma. “Yo creo que el problema es que este tipo de tecnologías están accesibles, al alcance de cualquier persona que tenga un ordenador y una conexión a Internet” argumenta Ana Valdivia. El uso particular de cada herramienta determina el producto final: “Intentar crear un modelo de inteligencia artificial que no pueda utilizarse mal es como intentar crear un ordenador que no pueda utilizarse para cosas malas”, escriben Arvind Narayanan y Sayash Kapoor.

87

La AI Act clasifica los programas según su nivel de riesgo. Más adelante ampliamos este tema.

88

Esto es lo que dice el punto 4 del Artículo 50 de la AI Act: **Los implantadores de un sistema de IA que genere o manipule contenidos de imagen, audio o vídeo que constituyan una falsificación profunda, deberán revelar que el contenido ha sido generado o manipulado artificialmente.** Esta obligación no se aplicará cuando el uso esté autorizado por la ley para detectar, prevenir, investigar o perseguir delitos penales. Cuando el contenido forme parte de una obra o programa manifiestamente artístico, creativo, satírico, de ficción o análogo, las obligaciones de transparencia establecidas en el presente apartado se limitarán a revelar la existencia de dicho contenido generado o manipulado de una manera adecuada que no obstaculice la visualización o el disfrute de la obra. **Los usuarios de un sistema de IA que genere o manipule texto que se publique con el fin de informar al público sobre asuntos de interés público deberán revelar que el texto ha sido generado o manipulado artificialmente.** Esta obligación no se aplicará cuando el uso esté autorizado por ley para detectar, prevenir, investigar o perseguir delitos penales o cuando el contenido generado por IA haya sido sometido a un proceso de revisión humana o control editorial y cuando una persona física o jurídica ostente la responsabilidad editorial de la publicación del contenido.

89

En septiembre de 2023, la Policía Nacional investiga unas fotografías de desnudos hechas con inteligencia artificial que toman el rostro de unas niñas menores de edad, en la población de Almendralejo, en Badajoz. Se detectaron al menos veinte casos.



IA para desenmascarar a la IA o el robot que caza robots

La Iniciativa para la Autenticidad del Contenido, impulsada por Adobe, Twitter y el New York Times en 2019, propone un sistema de trazabilidad a través de metadatos que ayuden a identificar la procedencia de los contenidos. Es decir, que quede claro que la atribución de esos productos pertenece a sus autoras y autores. Es una de las opciones que tratan de dar respuesta al complicado problema del rastreo y control de los materiales creativos.

“Si cada empresa crea un modelo diferente de generación de textos o de imágenes, no se va a poder construir una herramienta que compruebe los errores sistemáticos adaptada a cada uno de estos modelos”, apunta Karlos G. Liberal, y sigue: “Va a ver un descreimiento absoluto pero yo pienso que los medios se lo han ganado. Si Ferreras, después de lo que hizo y lo que dijo⁹⁰, sigue siendo el entrevistador estrella de La Sexta, ¿por qué el problema van a ser los bulos? El problema es el modelo periodístico basado, en muchos casos, en los intereses de los grandes conglomerados comunicativos y empresariales”. Y contra los bulos y la desinformación, su receta son redes de afinidad y confianza que nos permitan desmontarlos: “Son peligrosos y capaces de desestabilizar un país, lo vimos con la victoria de Bolsonaro, donde se descubrió todo el entramado de los evangelistas para difundir mentiras y noticias falsas. En el caso de Brasil se sabe y en otros también, sobre todo después del escándalo de Cambridge Analytica”.

Esto ocurre, según el investigador de BikoLabs, porque no ha habido una reflexión previa sobre dónde queremos que nos lleve la tecnología. Ujué Agudo, doctora en Psicología e investigadora de la relación entre la IA y las relaciones humanas, añade: “También estamos viendo casos en los que es el propio algoritmo el que genera una información falsa, en su necesidad por crear contenido” y recuerda lo acontecido con el chatbot de la compañía de vuelo Air Canada, cuando dio una información incorrecta sobre el reembolso de un billete a un pasajero y la empresa se defendió diciendo que el bot parlanchín era “responsable de sus acciones”⁹¹. La historia la zanjó un Tribunal canadiense que obligó a la compañía a cumplir⁹² lo que el asistente virtual le había prometido al cliente. El peligro de depositar toda la confianza en los modelos y pensar que las alucinaciones o equivocaciones no existen está ahí. Karlos G. Liberal menciona: “En Estados Unidos cada vez están más preocupados por la trazabilidad del código porque la gente está escribiendo código con GPT y luego no se sabe quién ha hecho esta pieza de código que ha fallado”.

En la línea de las palabras del investigador Karlos G. Liberal van las reflexiones de Marta Peirano, y para ella, el *robot para cazar otros robots* no está funcionando, principalmente porque las empresas no ofrecen transparencia en sus procesos y no dejan que sus parámetros sean detectables.

90 En julio de 2022 se hicieron públicos los audios de una conversación que tuvo lugar durante la campaña electoral de 2016 entre Antonio García Ferreras, y el hoy excomisario José Villarejo, agente operativo del Ministerio del Interior. En ella el presentador de Al rojo vivo de La Sexta reconoce que la supuesta noticia destapada por Eduardo Inda de que Pablo Iglesias tiene una cuenta bancaria en un paraíso fiscal, en la que el gobierno de Nicolás Maduro habría ingresado miles de euros, “es demasiado burdo (sic)” aunque reconoce que “va con ello”. Es decir, que lo contó en La Sexta como si se tratara de una información verídica. Más información: <https://www.elsaltodiario.com/medios/ultimos-audios-villarejo-ferreras-epicentro-debate-cloacas-estado-podemos>

91 Yagoda, M. (2024, 23 de febrero). Airline held liable for its chatbot giving passenger bad advice - what this means for travellers. Bbc.com. <https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know>

92 Belanger, A. (2024, 13 de febrero). Air Canada must honor refund policy invented by airline's chatbot. Arstechnica.com. <https://arstechnica.com/tech-policy/2024/02/air-canada-must-honor-refund-policy-invented-by-airlines-chatbot>

“Uno de los grandes problemas con estos modelos generativos tipo ChatGPT o los visuales son que no puedes controlar mucho los resultados. Tú pides algo y sale algo, pero a veces está mal, es incorrecto, no es coherente o no es lo que buscabas. Para las grandes empresas que los fabrican esto es un gran problema, puesto que tienen un producto que no pueden controlar. Se están haciendo ricos con esto, pero al mismo tiempo tienen algo que es incontrolable y eso les produce un profundo malestar”, expresa el experto en *machine learning* Miguel Arana. La explicabilidad de esas capacidades emergentes, como ya hemos comentado, es un problema para el que todavía no hay respuesta.

“En la era de la inteligencia artificial, cuando desconfiamos de las imágenes que vemos, la confianza debería depositarse en la fuente que las suministra (...) Las fakes news suceden porque las instituciones cuyo trabajo era decir la verdad ya no lo hacen. Se produce así un desplazamiento inevitable: cuando las autoridades cambian, el mundo cambia. La posverdad no se soluciona con el *fact-checking*”, sopesa Mária Martínez Bascuñán⁹³.

Herramientas de verificación disponibles y fiables

No hay una receta precisa y absoluta contra las ultra falsificaciones o las medias verdades. Tanto en instituciones académicas como en medios privados existen departamentos especializados en tareas de verificación a través de herramientas con IA, a la búsqueda del perfecto robot caza mentiras. Algunas de ellas son el Observatorio de la Desinformación de la Universidad Carlos III de Madrid o VerificaRTVE, la plataforma de comprobación de datos impulsada por RTVE. Google tiene su propio paquete de programas verificadores y hasta un buscador, Google Fact Check Explorer. La propuesta basada en IA del Grupo Prisa se llama VerificAudio⁹⁴ y su finalidad es, según pone en su web, *combatir la desinformación en los contenidos sonoros, una amenaza en ascenso propiciada por el avance de las últimas tecnologías generativas.* La aplicación de Prisa ofrece la posibilidad de que cualquier persona suba un audio a la web para que sea analizado por sus equipos de *verificación*. El responsable de Sistemas de À Punt, Pau Peiró, aporta la suya: *autoverifAI.com*, una plataforma ofrecida por Telefónica que dice detectar tanto audio como vídeo, y también texto, siempre y cuando haya sido creado por ChatGPT y en inglés.

Cuanto más pasan los meses y más perfectas y verosímiles son las composiciones de la generativa, mayor dificultad tienen los periodistas a la hora de saber qué es real y qué es un *fake*. “Hay signos”, desvela Peiró, “o detalles que son meramente humanos, como el pestañeo irregular. Pero sobre todo hay metadata, una que indica cuándo fue creada y otra dentro de la composición que te puede dar una orientación, aunque es un proceso complicado y cada vez más”. Para el Jefe del área Digital de À Punt Mèdia, Javier Lifante, la herramienta de verificación más infalible es descolgar el teléfono y llamar a las autoridades, a las fuentes, al bar cercano al suceso, algo factible y necesario en el periodismo local y de proximidad que hace la cadena autonómica.

93 Martínez-Bascuñán, M. (2024, 31 de marzo). Cloacas reales: de Kate Middleton a RTVE. El País. <https://elpais.com/opinion/2024-03-31/cloacas-reales-de-kate-middleton-a-rtve.html>

94 <https://verificaudio.ai>



Pero en esta era de la artificialidad con apariencia de inteligencia, “todos corremos el riesgo de ser engañados alguna vez”.

En RTVE Verifica⁹⁵, un equipo formado en detección de bulos comprueba noticias, fotografías y vídeos falsos. Además, ofrece una lista bastante completa de aplicaciones para extraer metadata y análisis inverso de imágenes, diagnóstico de links para ver si son maliciosos, registros de IP de un sitio web, herramientas que detectan alteraciones digitales en una imagen, rastreadores de vuelos y barcos, analizadores de vídeo como Azure Video Analyzer de Microsoft o incluso algunas de cronolocalización, capaces de saber qué hora era en una imagen o vídeo por las sombras y la colocación del sol, o conocer qué tiempo hacía en una fecha, hora y lugar concreto.

Para detectar texto sintético, hay una larga lista. Los que mejor suelen funcionar son Winston AI y Undetectable. Gpt Zero está especializado en detectar texto generado por ChatGPT, Gemini y LLaMA, principalmente. Originality trata de asegurar que los contenidos estén libres de plagio. En el caso de imágenes, especialmente las creadas con DALL-3, puede funcionar bien AI or not. Sensity.ai se puede utilizar para detectar falsificaciones mediante IA en vídeos, imágenes, audio e incluso en carnés de identidad.

Aunque hay cierta variedad de plataformas que aseguran poder comprobar y distinguir los contenidos generados de forma artificial de los producidos por humanos, como hemos dicho, no son infalibles. En el campo de las *deepfakes* también se cumple la máxima, como pasa con las *fake news*, de que es mucho más fácil crearlas que desmentirlas, eliminarlas o frenar su propagación en la web. La complejidad de *fact chequear* o comprobar si son verídicos contenidos que han podido, o no, ser creados por la máquina de desinformación de la inteligencia artificial es mucho mayor por tener gran verosimilitud, ser más sencillas de generar o propagar en cantidades industriales.

Legislar a priori: AI Act, la primera ley que pone coto a la Inteligencia Artificial

El periodista Javier Sampedro en una de sus Newsletters en El País auguraba que, en un futuro tan lejano como tarde en llegar el GPT-5 “se podrán piratear webs de forma enteramente autónoma”. Que la tecnología GPT escribe código, con el *prompt* o la orden adecuada, mejor que muchas manos humanas es ya una realidad. De ahí a manipular el código fuente de un sitio web para robar la información de sus usuarios parece que no queda tanto⁹⁶. Según Sampedro, con la versión GPT-4 se ha logrado una tasa de éxito 10 veces superior a la de GPT-3.5, su predecesor inmediato, en esta tarea. “El problema es que esta tecnología es capaz de producir realidades distópicas”, alerta la periodista Carmela Ríos: “Estamos en el momento perfecto para identificar esos peligros casi antes de que aparezcan para poder legislar a priori y no a posteriori”.

95

<https://www.rtve.es/noticias/verificartve>

96

Caburn, T. (2024, 17 de febrero). *How to weaponize LLMs to auto-hijack websites. We speak to professor who with colleagues toolled up OpenAI's GPT-4 and other neural nets.* The Register. Ahí recoge el experimento que realizaron informáticos de la Universidad de Illinois Urbana-Champaign (UIUC), quienes, usando LLM atacaron webs vulnerables sin intervención humana y demostraron que estos modelos “pueden hackear sitios web de forma autónoma y realizar tareas complejas”.

Para juristas expertos, la llamada Artificial Intelligence Act –AI Act–, con sus 468 páginas, es uno de los textos de mayor complejidad redactados hasta el momento. La primera ley que regula la inteligencia artificial se empezó a gestar en el seno de Europa en abril de 2021. Por fin, tras intensas negociaciones que coincidieron, en su fase final, con la presidencia española de la Unión Europea, fue aprobada el 14 de marzo de 2024. Todavía faltarán al menos tres años para su entrada en vigor, aunque a principios de 2025 se pondrá en funcionamiento la prohibición de usar sistemas de IA en ciertos contextos.

A mediados del próximo año, la Oficina de Inteligencia Artificial europea terminará de redactar los códigos de prácticas que especificarán los requisitos técnicos de los ‘modelos de propósito general’, como ChatGPT, y de ‘IA con riesgo sistémico’ que es como se denomina en la ley a la inteligencia artificial generativa. En mayo de ese año deberán estar funcionando los organismos que vigilarán el cumplimiento de la AI Act. También se espera que estén definidas cómo serán las multas que se pondrán por el incumplimiento de los artículos establecidos. A mitad de 2026 empezarán las obligaciones para sistemas de IA de alto riesgo incluidos en el anexo III de la ley y ya bien entrado el siguiente año, 2027, las obligaciones para sistemas de IA de alto riesgo incluidos en el anexo II y para los que ya estén comercializados. En el caso de ChatGPT, como ya está en el mercado, la UE le da un plazo de 2 años para adaptarse a la ley.

Según aclara Andrés Boix, los tiempos largos son habituales y se hacen para dar tiempo a que los operadores jurídicos, tanto públicos como privados, se adapten. “Aunque no entre en vigor, ya establece unos principios directores y unas normas que van a ir conformando un cierto estándar de comportamiento”. Por tanto, dará pie a que los Estados miembros puedan ir aplicando normas coherentes con ellos.

Al tener el estatus de reglamento, se traspone tal cual en el régimen jurídico de los países de la UE, a diferencia de lo que ocurre con las directivas europeas, destaca Algorights⁹⁷. Esto quiere decir que los Estados pueden decidir limitar estas excepciones e imponer regulaciones más garantistas.

El catedrático de Derecho Constitucional Lorenzo Cotino la califica de “muy positiva” a la hora de dar “seguridad jurídica para que se propicie la inversión pública y privada con un marco jurídico claro (...) “con garantías para nuestros derechos fundamentales y la dignidad del ser humano”. Esta norma, aunque afecta solo a los países miembros, influirá *de facto* en cualquier tecnología que se elabore en otro país y quiera ser comercializada en Europa. “Esto se va a traducir en que países como Estados Unidos van a tener que asumir la normativa europea porque, de lo contrario, no podrán operar en Europa”, precisa Marta Suárez-Mansilla, abogada experta en arte y tecnología.

Tanto Cotino como Suárez-Mansilla coinciden en que la AI Act se debe leer de forma conjunta con el Reglamento General de Protección de Datos (RGPD) en el caso de que, para el entrenamiento de los modelos de IA se utilicen datos personales. Son normas que “se superponen aunque hoy en día se ignoran” y cabe esperar una “mejora del proceso regulador”, dice Cotino.

97

La labor de divulgación de colectivos como Algorights de la ley europea de inteligencia artificial es intensa y extensa, ya que han seguido desde el inicio todo el proceso legislativo. Aquí un video: ¿Qué va a pasar con la IA en Europa? Introducción a la AI Act para movimientos sociales
<https://fediverse.tv/w/kzVGtTT8Eh3o2hzgXNSPcY>



Rojo, amarillo y verde: la división legal según el riesgo de cada producto con IA

La ley establece el reparto de una serie de etiquetas a cada producto de IA según el riesgo que presente para los derechos humanos: “A más peligro, más obligaciones preventivas”. Como si de un semáforo se tratase, la etiqueta roja corresponde con prácticas prohibidas; la amarilla, para aquellas que sean consideradas de alto riesgo (“para evitar daños, imponemos preventivamente una serie de obligaciones a productores, diseñadores y usuarios”). Las verdes, sigue Cotino, no están reguladas por esta norma pero sí por otras normas como el Reglamento General de Protección de Datos.

En el listado rojo figuran como restringidas aquellas acciones que empleen inteligencia artificial para el reconocimiento facial biométrico en tiempo real en espacios de acceso público. En palabras de Algorights, no te podrán reconocer por tus datos biométricos con una IA mientras estás en la calle, en el cine, o en un supermercado. También sería ilegal usarla para implementar técnicas manipuladoras o subliminales que puedan generar daños o que supongan la explotación de las vulnerabilidades de las personas debido a su edad, discapacidad o una situación social o económica de la que se puedan aprovechar para condicionar su comportamiento y causarles perjuicio. Ahí entraría la creación y difusión de *deepfakes* o las prácticas comerciales fraudulentas o manipulativas

Está prohibido por la ley categorizar a personas según sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias u orientación sexual. Uno de los puntos punitivos va en la línea de impedir el *scoring social* o carné de buen comportamiento y crédito social como el que existe en China, con el que la ciudadanía recibe puntos o los pierde en función de su comportamiento social y que condiciona directamente las prestaciones y el trato que recibe del Estado.

No se permiten los sistemas de reconocimiento de emociones en entornos laborales y educativos, salvo por motivos médicos o de seguridad. Tampoco los sistemas para crear o ampliar bases de datos de reconocimiento facial mediante *scraping* de imágenes de internet o de Circuitos Cerrados de Televisión. Un ejemplo, según Algorights, de empresa que ya no podría operar en la UE sería *Clearview*, si bien sigue siendo utilizado en EE. UU., Reino Unido o Ucrania.

Sara: ¿Qué es Clearview?
 Gemini: Clearview AI es una compañía estadounidense que desarrolla tecnología de reconocimiento facial. Su producto principal es un motor de búsqueda facial que permite a las fuerzas del orden público y otras organizaciones subir una imagen de un rostro y compararla con una base de datos masiva de más de 40 billones de imágenes públicas provenientes de internet, incluyendo redes sociales.

Entre los sistemas de alto riesgo se encuentran los sistemas de identificación biométrica para el uso en la admisión de colegios y centros educativos, el control del empleo público y privado, acceso a servicios esenciales públicos o privados, respuesta a emergencias, ayuda médica, bomberos, etc. En el ámbito penal y policial, cuestiones como la prevención, detección y persecución de delitos, migración, asilo y control de fronteras. Algunos ejemplos de prácticas que incluirían técnicas automatizadas

consideradas de alto riesgo serían el uso de polígrafos, evaluaciones de la posibilidad de reincidir o no de un preso, la medición del estado emocional de un acusado o de una víctima, la fiabilidad de las pruebas en un juicio, o bien usar la IA para predecir la personalidad o el comportamiento delictivo, el análisis de delincuencia, migración, asilo o control de fronteras. Según denuncia Algorights, estos sistemas dejarían de ser considerados de alto riesgo si no suponen riesgo significativo para la salud, la seguridad o los derechos fundamentales o no influyen materialmente en el resultado de la toma de decisiones.

Por último, en la lista roja se encuentran los sistemas de la llamada Policía predictiva, según la cual se evaluaría el riesgo de que una persona cometa un delito basándose en la recreación de su perfil o características de su personalidad. Sí se permitirán sistemas que apoyen la evaluación humana de la implicación de una persona en un delito.

Desde Algorights también recuerdan que la AI Act no impide que empresas europeas desarrollen este tipo de tecnología, prohibida en Europa, para exportarla. En cuanto a las medidas correctivas, aunque sí se contemplan sanciones para las empresas que incumplan el reglamento, no existen para la Administración, pese a que las instituciones usan ampliamente estas herramientas y se benefician de la larga lista de excepciones contempladas. El encargado de supervisar y de garantizar que todos los sistemas que llegan a la UE cumplen con la regulación será la nueva Oficina de la IA, dependiente de la Comisión Europea.

En la propia ley se indica que será el proveedor del servicio tecnológico quien decida su nivel de riesgo es una primera autoevaluación, una cuestión que para las organizaciones que defienden los derechos humanos puede suponer un problema. Gala Pin comparte este pensamiento: “Esa distinción entre alto, medio y bajo riesgo opera sobre cómo se aplica y cómo se usa inteligencia artificial en el ámbito europeo, pero tú puedes desarrollar inteligencia artificial de uso prohibido en Europa y exportarla para que se use en otros países, lo cual va en contra de una visión global de garantía de los derechos humanos”.

Por su parte, Boix aporta una explicación jurídica al respecto: “Todos estos sistemas de seguridad industrial se basan en la colaboración con la industria y las empresas privadas. La primera evaluación es una autoevaluación que hace la propia industria, la segunda una auditoría por parte de una entidad privada de certificación. Existe asimismo un mecanismo último de control en el poder público, encargado de revisar si se están haciendo bien las auditorías o comprobar si esas valoraciones de las empresas son correctas”. Esto pasa aquí y en todos los sectores, por ejemplo con cualquier electrodoméstico o con las empresas automovilísticas. Son las propias firmas las que marcan las certificaciones energéticas de sus productos, después existen auditorías y comisiones privadas y finalmente, si hay denuncias o sospechas de irregularidades, un control público. “Una de las funciones de la Agencia europea de Supervisión de la IA será esa, controlar a quienes hacen las auditorías y establecer sanciones si es el caso. Solo en extremos potencialmente peligrosos se exige que además haya evaluación pública desde el principio, y si no, lo que se hace es que la evaluación pública se hace después si se detecta algún problema”, acaba Boix.

Privacidad y seguridad en las excepciones de la AI Act: el debate sobre la captación biométrica en las calles y en las fronteras

Uno de los temas más controvertidos de la norma es el referente a los sistemas de reconocimiento facial en la vía pública, que como se ha dicho quedan prohibidos salvo determinadas excepciones que, según su lectura política y administrativa, podrían ser más amplias o menos. Solo podrán ser empleados por los Cuerpos y fuerzas de seguridad del Estado y su uso está sujeto a autorización judicial o administrativa previa, aunque en casos de “extrema gravedad” se podrán implementar sin este consentimiento, como analizaremos más adelante.

Como explica Algorights, las autoridades públicas podrán hacer uso del reconocimiento biométrico por razones de seguridad en casos de riesgo de atentado, o para buscar a víctimas de secuestro o trata y a sospechosos de delitos graves. Abordaremos este punto con la ley en la mano junto a Javier Boix: “Ha habido mucha discusión sobre si se debía prohibir o no el reconocimiento biométrico. La decisión adoptada por la Unión Europea es que no se prohíbe con carácter general, sino que se prohíben los reconocimientos biométricos en tiempo real que hagan empresas privadas. No estará prohibido, en cambio, cuando lo hagan los poderes públicos para la persecución y el combate de delitos. La pregunta es, ¿hasta dónde llega esta habilitación a los poderes públicos cuando se trata de combatir delitos?”.

El profesor de Derecho Administrativo resalta el caso de China, donde el control ciudadano mediante biometría se hace con la excusa de evitar que se cometan infracciones. En la redacción de la AI Act se extrae que el caso chino no es extrapolable a Europa porque aquí solamente se podrá usar para la persecución de delitos *ex post*, es decir, una vez se hayan producido y dentro de una investigación que ya esté en marcha: “El carácter preventivo generaría un control general y pasivo de toda la población” que es justamente lo que se quiere evitar.

El reglamento prohíbe el uso en tiempo real de identificación biométrica remota en espacios públicos excepto en estos casos en lo que sea estrictamente necesario para conseguir objetivos, sigue Boix, como la búsqueda específica de víctimas de abducción, tráfico de personas, trata o explotación sexual: “Es decir, cuando tú tienes ya una operación en marcha, no preventiva”. Y concede: “Eso no tiene que ver, en principio, con la identificación en frontera, aunque a veces se pueda utilizar ahí. Pero tiene que ser solo para combatir asuntos como las redes de trata”.

Otra de las excepciones que permiten el uso de estos sistemas de identificación basado en rasgos físicos sería para la prevención de una inminente amenaza para la vida o la seguridad física de personas, para la prevención de un ataque terrorista y para la localización e identificación de un sospechoso de haber cometido una ofensa criminal. La diputada Gala Pin completa con un caso práctico: “Si tú activas el reconocimiento facial en tiempo real solo en casos como después del atentado de la Rambla de Barcelona nos puede no gustar, pero al final estás diciendo que la excepcionalidad son momentos realmente excepcionales. En cambio, si empiezas a aplicar el reconocimiento facial en cualquier manifestación de más de mil personas, estás normalizando la excepcionalidad”.

Judith Membrives, que también es parte de Algorights, aporta otro ejemplo: “Estamos en alerta 4 de atentado terrorista que es alerta inminente. Si el Gobierno considera que ese estado de alerta requiere de uso de reconocimiento facial o de

reconocimiento biométrico en tiempo real por seguridad nacional, lo podría implementar”. Para ello necesitaría una autorización judicial o bien una administrativa. Y en ciertos casos, ni siquiera eso, porque la Policía Nacional lleva desde agosto de 2023 usando una herramienta de reconocimiento facial en al menos 11 ciudades sin requerir la orden de un juez. Se conoce como el programa ABIS, que son las siglas en inglés de Sistema Automático de Identificación Biométrica. Su funcionamiento de puertas hacia afuera es sencillo de explicar, pero la forma en la que se ha entrenado su algoritmo es un misterio. Lo que se sabe de él es que los sistemas automáticos de reconocimiento facial extraen el rostro de las imágenes mediante visión computacional y aplican un algoritmo a ese rostro para obtener un patrón que lo distinga de los demás. La IA selecciona y ordena si ese patrón se repite o aparece dentro de la extensa base de datos de fichas policiales que incluyen tanto patrones faciales como el ADN de 4,4 millones de individuos fichados por la Policía Nacional o la Guardia Civil. Dos operadores formados para ello verifican la lista de parecidos que ofrece el sistema, aunque raramente toman como errónea la opción del algoritmo.

El sistema ABIS no procesa imágenes en tiempo real porque no está conectado a cámaras de vigilancia, sino que obtiene los rostros de grabaciones de cámaras de seguridad o de fotografías hechas con móviles. “Un uso descontrolado del reconocimiento facial erradicaría el anonimato”, replica la periodista del New York Times Kashmir Hill. Nuestros movimientos cotidianos podrían ser rastreados tanto por nuestro vecino como por nuestro Gobierno, como hace Rusia en las manifestaciones en la vía pública en las que ficha mediante esta tecnología a ciudadanos insurrectos⁹⁸: “El rostro es el último bastión de la privacidad. En China hay una lista roja para quienes están en el poder y no quieren que les vean, así son invisibles para los sistemas de vigilancia y se eliminan de los registros de las cámaras. Es muy revelador que no ser visto sea un privilegio exclusivo de los poderosos”, sigue la reportera del NY Times.

Según publica el diario El País⁹⁹, desde que está operativo el sistema ABIS se ha empleado en unas 400 investigaciones policiales. Este diario también informa de que el reconocimiento biométrico se usa desde hace al menos un año en las fronteras de la UE para registrar a los ciudadanos no comunitarios recién llegados.

“Ninguna de ellas [las excepciones para el uso de biometría en la vía pública] está pensada para el análisis y el control de la inmigración o en frontera. Otra cosa es que, entendido de forma amplia, se use de forma sistemática en frontera, y eso sería un problema porque los Estados estarían incumpliendo la norma”. Sobre su uso en materia de migración, Boix subraya que cruzar líneas fronterizas entre países sin los permisos no es delito: “Muchas organizaciones que defienden los derechos humanos entienden que la actual redacción sí que permite un entendimiento muy amplio, quizá demasiado, que hará que los gobiernos lo usen en frontera”. Para el titular de Derecho Administrativo de la Universitat de València, la intención de la UE no es instar al uso de estas herramientas para control fronterizo: “Es una discusión en marcha”, y argumenta: “Si hacemos una interpretación estricta de la ley, a día de hoy estaría prohibido. Si se da una creciente presión migratoria y la función pública cambia de opinión y decide que sí se puede usar, lo que hará es cambiar la AI Act para permitirlo”.

98 Como denuncia Amnistía Internacional, entre otras organizaciones: <https://www.amnesty.org/es/latest/news/2020/01/russia-intrusive-facial-recognition-technology-must-not-be-used-to-crackdown-on-protests>
 99 G. Pascual, M. (2024, 28 de mayo). *La Policía española ya usa en sus investigaciones un sistema automático de reconocimiento facial*. El País. <https://elpais.com/tecnologia/2024-05-28/la-policia-espanola-ya-usa-en-sus-investigaciones-un-sistema-automatico-de-reconocimiento-facial.html>





El caso de iBorder Control: automatismos en la gestión fronteriza

El uso de la biometría para el control migratorio no es algo nuevo. Uno de los proyectos piloto europeo no exento de polémica fue iBorder Control, utilizado de 2016 a 2019 y basado en el reconocimiento de emociones para categorizar a las personas migrantes a partir de datos biométricos. Mediante “un mecanismo automatizado de detección de mentiras”¹⁰⁰ analizaba las “microexpresiones faciales” y servía para orientar “a las autoridades fronterizas en la toma de decisiones”. Como menciona Judith Membrives, existe evidencia científica del gran margen de error que existe a la hora de reconocer las emociones de una persona, puesto que es algo que cambia en cada una, y está muy marcado por cuestiones como su educación o prácticas culturales.

“Todo lo que es Seguridad Nacional y Cuerpos Policiales ha quedado catalogado como alto riesgo pero con pocas prohibiciones al respecto por la cantidad de excepciones incorporadas. Hay requisitos de transparencia pero no lo son de cara a la sociedad civil. Sí se ha establecido un *database*, una especie de registro de algoritmos pero no van a ser de acceso público. Esto ya pasa bastante con Frontex, que es quien tiene la externalización de las fronteras”. Y añade: “Europa está haciendo política pública centrada en fortificar sus fronteras y para ello utiliza la inteligencia artificial, los sistemas algorítmicos y el Big Data”. Y nombra a Eurodac¹⁰¹, la gran base de datos de todas las personas que piden asilo político o refugio en Europa, como el gran instrumento de seguimiento digital de Europa que quedaría fuera del Reglamento de Inteligencia Artificial.

100 Romano, Andrea. (2023). *Derechos fundamentales e inteligencia artificial emocional en iBorderCtrl: retos de la automatización en el ámbito migratorio*. Revista Catalana de Dret Públic, 66, 237-252. <https://doi.org/10.58992/rcdp.i66.2023.3928>

101 European Asylum Dactyloscopy Database, Eurodac en su abreviatura oficial, es, según ChatGPT, una base de datos biométrica de huellas dactilares utilizada por los países miembros de la Unión Europea y otros estados asociados para gestionar y comparar las huellas dactilares de los solicitantes de asilo y los migrantes irregulares que funciona desde 2003.

Transparencia, tecnosolucionismo y cambio climático

Dinero público, código público: el poder institucional y los algoritmos automatizados

No se espera de la AI Act que sea garante de los derechos humanos porque como hemos dicho, no está pensada para ello. “El reglamento de usos de la IA es un reglamento de seguridad industrial”, diferencia Boix. Esto quiere decir que es una norma de control parecida a las ITV de los vehículos o las inspecciones de los ascensores: “Establece unas reglas básicas sobre la tecnología, qué cautelas debe tener, qué mecanismos de control y de inspección existen por parte de los poderes públicos, con la colaboración de una serie de auditorías privadas, para asegurar que todas esas normas se estén cumpliendo. Las empresas tendrán la obligación, igual que nosotros tenemos la obligación de haber pasado la ITV, de haber pasado la auditoría correspondiente”.

Para el profesor de Derecho Administrativo de la Universitat de València, esta ley establece el marco, la piedra inicial sobre la que los países europeos –o la propia institución europea– pueden desarrollar otro tipo de normas que respondan a cómo tienen que ser los algoritmos, qué pasa cuando afectan a derechos humanos, cómo se deben proteger los datos o la privacidad, y qué hacer cuando esos algoritmos generan sesgos. En realidad, la AI Act apenas toca esto. “No está pensada para controlar qué hacen las administraciones y los poderes públicos con la IA”, algo que sí se produjo con el Reglamento General de Protección de Datos (RGPD). “Se podría pensar que con el RGPD ya valdría o que podría ser trasladable”. El problema, según Boix, es que no es lo mismo controlar los usos que se les da a los algoritmos de automatización que hacen agentes privados –para fisgar en nuestros datos o para ofrecernos anuncios en función de nuestro historial de consumo–, que “cuando los usamos para adoptar decisiones que hasta ahora son competencia de poderes públicos como el judicial o la administración pública, que inciden y afectan directamente a nuestros derechos, garantías y estatuto jurídico”.

En uno de los primeros borradores sí se consideraban de alto riesgo –y por tanto, regulados– todos los sistemas algorítmicos. En la versión final solo se consideran de alto riesgo los algoritmos que impacten en los procesos democráticos, algo que todavía está por definir y aclarar. “Para esto tardaremos unos años en ver cómo las autoridades de supervisión imponen a las plataformas su cumplimiento”, expone Cotino.

Andrés Boix lleva tiempo analizando desde una perspectiva jurídica el hecho de que cada vez más sean los algoritmos quienes toman las decisiones que antes tomaban los poderes públicos, los jueces o las autoridades administrativas. Hay varios famosos, como BOSCO, el algoritmo encargado de dirimir si una persona tiene

derecho al bono social o no. Cuando la Fundación Civio¹⁰² quiso estudiarlo, el Gobierno se negó a entregarle el código fuente, decisión que fue ratificada por un juez.

Pero hay más. En Euskadi tienen su propio programa informático en materia de violencia machista, llamado EPV-R, Escala de predicción del riesgo de violencia grave contra la pareja, y lleva en uso desde 2007. Ana Valdivia auditó el algoritmo en 2022 y su conclusión¹⁰³ fue clara: “Está muy mal diseñado y funciona muy mal”, entre otras cuestiones porque su tasa de error es del 53%: “De cada diez casos severos, cinco los está etiquetando mal”. La investigadora del Instituto de Internet de Oxford apunta que se implementó porque la Ertzaintza carecía de expertos suficientes en materia de violencia machista.

Como cuenta Naiara Bellio para El Salto¹⁰⁴, los algoritmos diseñados para predecir el riesgo de que una persona vuelva a cometer un delito se llevan usando décadas, mucho antes del auge de la inteligencia artificial. Se diseñan en forma de cuestionario psicológico que evalúa a la persona según sus respuestas. Una de las preguntas es la procedencia de la persona sobre la que existe la sospecha, y se cuenta como extranjero a alguien de Marruecos o Colombia, pero no a alguien de Francia. Si como parece, este traspaso del poder de tomar decisiones va a pasar de los poderes públicos a este tipo de algoritmos, según Boix, lo mínimo que podemos hacer es controlar y regular estos sistemas.

A nivel estatal, la Policía Nacional utiliza VioGen desde hace 17 años para definir el riesgo de sufrir violencia machista de una mujer que denuncia a su pareja. Como precisa Boix, VioGen no usa –al menos, en este momento– inteligencia artificial, pero igualmente existe mucha opacidad y una negativa clara desde los poderes fácticos a dar toda la información sobre este tipo de algoritmos de forma transparente para que sean fiscalizados por la sociedad civil. Para Boix, estamos asistiendo a “un proceso de sustitución” al condicionar la decisión que toman los jueces.

Del tecnosolucionismo al incremento de la crisis climática

“Existe un sesgo humano que es pensar que todo lo que viene de la máquina es objetivo y que la persona no lo es. Esto es tecnosolucionismo puro y duro: creer que mediante la técnica vamos a salvar o arreglar problemas estructurales”, suscribe Judith Membrives, y trae a colación la paradoja de Jevons, una metáfora utilizada en los análisis económicos según la cual a medida que incrementa el uso eficiente de un recurso, también aumenta su consumo total y con ello, el gasto energético. Es decir, que las mejoras en la eficiencia pueden conllevar una mayor utilización de recursos en lugar

102 El principal argumento del Juzgado Central de lo Contencioso Administrativo número 8 para denegar a Civio el acceso al código fuente de BOSCO es que sería un peligro para la seguridad pública. Lo cuentan en una noticia publicada en su web el 10 de febrero de 2022 titulada *La Justicia impide la apertura del código fuente de la aplicación que concede el bono social*. <https://civio.es/novedades/2022/02/10/la-justicia-impide-la-apertura-del-codigo-fuente-de-la-aplicacion-que-concede-el-bono-social>

103 Un extracto de la conclusión de la auditoría: *Las herramientas de evaluación de riesgos se aplican a veces para paliar la falta de recursos humanos. Sin embargo, los algoritmos y las herramientas de evaluación de riesgos no pueden pensar como las personas, como afirmaba Dreyfus: “Los seres humanos tienen una inteligencia intuitiva que las máquinas “razonadoras” sencillamente no pueden igualar”. Se encuentra aquí: Ana Valdivia, A; Hyde-Vaamonde, C; García-Marcos, J. (2022). Judging the algorithm: A case study on the risk assessment tool for gender-based violence implemented in the Basque country. King’s College London (KCL), United Kingdom. Euskal Herriko Unibertsitatea (EHU) Bizkaia, Spain.*

104 Bellio, N. (2023, 5 de abril). *Los sesgos con los extranjeros en el algoritmo de violencia de género de la Ertzaintza. El Salto*. <https://www.elsaltdiario.com/policia/sesgos-extranjeros-del-algoritmo-violencia-genero-ertzaintza>



de ayudar a reducirlos, algo que parece evidente en materia de IA, como denuncia Ana Valdivia. La investigadora de la Universidad de Oxford nos invita a tener una perspectiva crítica y a no olvidar algo que a su juicio ha estado siendo invisibilizado, y que denomina “la materialidad de la tecnología”. Es un factor a tener en cuenta, sobre todo cuando se espera que la inteligencia artificial enseñe a la humana cómo acabar, por ejemplo, con el cambio climático.

Un artículo en la revista Nature¹⁰⁵ recogía las declaraciones del directivo de OpenAI Sam Altman en las que reconoce algo que los expertos llevan diciendo desde hace años: que la industria de la inteligencia artificial se encamina hacia una, crisis energética inminente. Parte del avance en la investigación en IA se destina, por necesidad, a encontrar fuentes de energía alternativas, así que no sorprende saber que Altman lleva invirtiendo desde 2021 en empresas que trabajan para alcanzar la fusión nuclear. “Se calcula que una búsqueda realizada mediante IA generativa consume entre cuatro y cinco veces más energía que una búsqueda web convencional. Dentro de unos años, es probable que los grandes sistemas de IA necesiten tanta energía como naciones enteras”, escribe la investigadora y profesora universitaria Kate Crawford en Nature. No es solo energía, también es agua, pero también son las trabas de las tecnológicas a saber con exactitud los gastos medioambientales de sus productos: “Los costes planetarios totales de la IA generativa son secretos empresariales celosamente guardados”, sigue Crawford.

Algunas estimaciones dicen que ya se dedica un 8% de la energía mundial a la inteligencia artificial. “Necesitas minerales, pero no solo. Los chips, cuando los metes en un centro de datos que ejecuta algoritmos de inteligencia artificial como ChatGPT, utilizan mucha agua, mucha energía y hace mucho ruido. Las comunidades que viven cerca de centros de datos se quejan de esa gran contaminación sonora. Aparte, su infraestructura tiene un ciclo de vida limitado y cada cinco años un centro de datos tiene que desechar toda esa infraestructura tecnológica, algo que tiene un impacto negativo en las personas que viven cerca de los vertederos, como pasa en Ghana y en Kenia”, lamenta Ana Valdivia.

Según un estudio¹⁰⁶ realizado en 2019 por la Universidad de Massachusetts Amherst, entrenar una sola inteligencia artificial con arquitecturas neuronales emite 284 toneladas de CO2, equiparable a lo que contaminan 45 europeos en todo un año, o a volar más de 2.800 veces en el trayecto Madrid-Barcelona. Desde que Sam Altman anunciara a principios de este año que la industria de la inteligencia artificial está a punto de provocar una crisis energética, gobiernos como el estadounidense ya trabajan codo con codo con las grandes tecnológicas en soluciones tan imaginativas como instalar pequeños reactores nucleares¹⁰⁷ en los centros de datos que funcionan día y noche sin descanso. “Los grandes modelos de lenguaje tienen una arquitectura muy grande. Un algoritmo de aprendizaje automático que te ayude a elegir a quién contratar quizás necesite 50 variables: dónde trabaja el candidato, qué salario tiene ahora, experiencia previa, etcétera. La primera versión de GhatGPT tiene más de 175.000 millones de parámetros”, ilustra Valdivia. Su recomendación: “Cuando entrenes a un algoritmo que te analiza la cuestión del cambio climático, ten también en cuenta y analiza tu propio algoritmo, cómo está contribuyendo también a ese cambio climático”.

105 Crawford, K. (2024). *Generative AI's environmental costs are soaring — and mostly secret*. Revista Nature. <https://www.nature.com/articles/d41586-024-00478-x>

106 Strubell, E; Ganesh, A; McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. College of Information and Computer Sciences University of Massachusetts Amherst. <https://arxiv.org/pdf/1906.02243>

107 G. Pascual, M. (2024, 30 de abril). *Los centros de datos quieren tener sus propios reactores nucleares*. El País. Este extenso y completo reportaje de El País habla de que ya existen acuerdos de compra entre Google y los desarrolladores de pequeños reactores nucleares. Desde Bill Gates hasta AWS, la división de computación en la nube de Amazon, están haciendo movimientos similares. https://elpais.com/tecnologia/2024-04-30/los-centros-de-datos-quieren-tener-sus-propios-reactores-nucleares.html?sma=newsletter_tecnologia&utm_medium=email&utm_source=newsletter&utm_campaign=tecnologia_2024.05.03_3

La importancia del código de la IA

El 5 de noviembre de 2024 se celebran elecciones presidenciales en Estados Unidos. Una de las particularidades de su mecanismo electoral es su sistema digital de recuento de voto. Las papeletas virtuales espolearon a Donald Trump y sus seguidores a extender la idea de fraude electoral cuando hace cuatro años ganó Joe Biden. En Concord, New Hampshire, desde 1989 usaban máquinas AccuVote, propiedad de la macro empresa *Dominion Voting Systems*, el segundo mayor proveedor del país. Pero esto puede cambiar. Como relata la revista tecnológica del Instituto Tecnológico de Massachusetts, la *MIT Technology Review*, son tres las empresas que controlan aproximadamente el 90% del mercado estadounidense de tecnología electoral. Las tres son de capital privado y mantienen su código fuente privativo y oculto, igual que la identidad de sus inversores o sus informes de salud financiera.

Pero en esa pequeña localidad de New Hampshire se ha presentado a la competición por la gestión del voto digital una cuarta protagonista inesperada: VotingWorks, una empresa sin ánimo de lucro con solo 17 empleados cuyas máquinas de votación ya operan en cinco pequeños condados, todos en Mississippi. La firma ha adoptado el enfoque opuesto al de estas tres grandes tecnológicas: transparencia, sencillez y seguridad, son sus tres reclamos principales. Sus cuentas están publicadas en su página web y cada línea de código que alimenta a sus máquinas se publica en GitHub, disponible para que cualquiera pueda inspeccionarla.

Algunas voces interesadas se apresuraron en propagar el mensaje de que el código abierto podría ser peligroso. “Si el FBI estuviera construyendo un nuevo edificio, no pondría los planos en línea”, dijeron, a lo que el cofundador de VotingWorks, Ben Adida, replicó: “Un sistema de votación de código abierto es aquel en el que no hay secretos sobre su funcionamiento”. La transparencia es un factor fundamental para recuperar la confianza en los equipos automatizados de votación: “Todo el código fuente es público para que el mundo lo vea. ¿Por qué en 2023 estamos contando votos con software propietario?”.

La historia del *open source* ha tenido muchos capítulos y ahora estamos ante uno nuevo. “Código abierto significa para mí que no hay prejuicios, que no hay segundas intenciones y que, al final, es un legado al que yo he contribuido en lo que he podido, y si el año que viene no voy a estar aquí, otra persona tendrá que coger el testigo. Eso con un código cerrado no lo vas a tener. Yo creo que sin el *open source* internet no se hubiera democratizado”, aporta Pau Peiró.

Mucho se ha escrito y escuchado sobre la seguridad de los modelos abiertos. Miguel Arana es tajante sobre el tema:

“Prácticamente todos los sistemas informáticos que tenemos ahora se basan precisamente en modelos de software abierto, es decir toda la infraestructura de internet, todo lo que es importante, funciona en Linux y es software abierto, porque se ha demostrado que funciona mejor, que es más seguro, que genera menos problemas. Y que produce mejores sociedades porque compartimos el conocimiento, lo replicamos y lo reinventamos, que es lo que sabemos hacer los humanos. Siempre han estado ahí los defensores del conocimiento cerrado y propietario. Justamente todos los sistemas de seguridad, por ejemplo, los criptocódigos, todo lo que tiene que ver con transacciones bancarias, el control de nuestras armas, etc., todo se basa en sistemas abiertos. Uno crea un código y lo primero que hace es publicarlo y decir, así hago yo mi código, a ver quién lo revienta, a ver quién encuentra un fallo. Así es como se demuestra si este código es seguro. Nadie que entienda un poco de



tecnología puede creer que lo abierto es peor. En ningún sentido: ni en seguridad, ni como desarrollo, ni como eficacia. Todos los que defienden el código privativo es porque tienen un interés económico y quieren control, poder, monopolio. Ha sido así en toda la historia de internet”.

Arvind Narayanan, a quien ya hemos citado en varias ocasiones, es profesor de Ciencias Computacionales de la Universidad de Princeton y director del Centro de Política y Tecnologías de la Información de esta universidad. Su aproximación a la IA resulta muy interesante e invita a la reflexión: “[el aprendizaje automático] es un tipo de tecnología engañosa en muchos sentidos, es muy fácil engañarse a uno mismo. Así que muchos de los que le están dando bombo y platillo no solo engañan a los demás, sino que se engañan a sí mismos”. Desde su departamento hacen una defensa ferviente de todas las posibilidades que ofrecen los modelos abiertos, es decir, aquellos que hacen público su código fuente. Esto se enfrenta frontalmente a la defensa que se hace desde grandes sectores del universo empresarial *tech*, para los cuales la seguridad de la IA pasa por la restricción de los modelos abiertos, y parecen olvidar que las herramientas, per se, con código publicado o privativo, no son ni buenas ni malas, sino que depende de cómo las usemos. El científico tiene una respuesta que ya hemos nombrado unas páginas atrás: “Intentar crear un modelo de inteligencia artificial que no pueda utilizarse mal es como intentar crear un ordenador que no pueda utilizarse para cosas malas”.

Inteligente, artificial y de código abierto

“No puedes evitar soluciones propietarias”, objeta Jordi Linares. Y recuerda la gran inversión que hacen las empresas, de la misma forma que pasa con las farmacéuticas y las patentes de medicamentos. “Al margen de los sistemas propietarios, la salud de los sistemas abiertos impulsados por universidades e instituciones es excelente, hasta el punto de que están siendo tan poderosos como los propietarios. Esto es muy importante porque nos permite no tener que depender de los sistemas propietarios y tener una merecida independencia”. Este doctor en Informática de la UPV sostiene que “es bueno que existan los modelos abiertos porque garantizan que no estemos en las manos de Microsoft, Google o Amazon, y que podamos tener la libertad de utilizar modelos para aplicaciones de cualquier tipo que no dependan de ellos”. Pero la batalla nunca ha estado equilibrada. Karlos G. Liberal toma como punto clave ese momento del movimiento en defensa del código abierto que cambia el software libre por el *open source*. “En el software libre no hemos conseguido esos espacios de emancipación que queríamos. Lo que hemos conseguido es que Microsoft se quede con todo el código de GitHub para que puedan entrenar su propio modelo de IA con nuestro código”.

Cory Doctorow¹⁰⁸, pionero, agitador, escritor futurista publicó un artículo en el que reprochaba que no se pudiera hablar de modelos libres si no tenemos la infraestructura, la energía ni el agua que necesitan esas máquinas para funcionar.

Sabemos que estos grandes modelos de lenguaje requieren de cantidades ingentes de datos, energía y recursos naturales, pero como decíamos, desconocemos la cantidad exacta de todo, porque los procesos de adiestramiento son completamente opacos: “Nadie ha tenido acceso a la información que se le ha dado, no hemos participado en ese proceso, ni es para nada *open*”, sigue Linares. Una opción intermedia a la realidad que plantea Doctorow es optar por aquellos modelos pesados entrenados de forma más general que permiten un proceso de *fine-tuning*, ajuste fino o personalizado, del que ya hemos hablado, para instalarlo en nuestro equipo y evitar regalar nuestros datos a las tecnológicas.

Esto se puede hacer tanto con algunos *foundational models* como con modelos de lenguaje de muchos tipos, desde clonación de voz hasta traducción, generación de imagen, texto, vídeo o 3D, etc. Páginas como Hugging Face¹⁰⁹ cuentan con más de medio millón de *foundation models* abiertos con todo su código fuente publicado que se pueden usar para multitud de tareas y con licencias más o menos restrictivas. Entre las grandes tecnológicas, Meta está liderando el mercado de los modelos abiertos incluso permitiendo licencia comercial en algunos casos. Para la generación de imágenes, Stable Diffusion es un modelo libre y abierto que podemos descargar y *fine-tunear* libremente.

“Los modelos abiertos, junto a una férrea comunidad activa de desarrolladores, nos garantizan que haya alternativa a los grandes *foundational models* propietarios”, aporta Linares. Las alternativas libres, independientemente del tipo de licencia que tengan, nos ofrecen transparencia: conocemos su código, cómo y con qué datos han sido entrenados y está disponible toda la documentación del modelo. Además, permiten la instalación en local, sin acceso a internet, para que nuestros datos no sean compartidos y perdamos el control sobre ellos. Pero no son las opciones mayoritarias ni muchas veces las más sencillas a nivel usuario.

“Igual no tenemos IA de código abierto, pero sí podemos tener IA de servicio público”, opina Gala Pin. “Dinero público, código público”¹¹⁰, es el lema de una campaña de la Free Software Foundation. La diputada de En Comú Podem trae a la conversación el *Barcelona Supercomputing Center*, un centro pionero en este campo que abarca desde la investigación hasta la oferta de servicios de supercomputación, financiado con dinero público. Como explica Gala Pin, ofrece un servicio de procesamiento de datos y generación de procedimientos algorítmicos que tiene vocación de servicio público y que permite a personas de la Academia, científicas o a cualquiera, presentar un proyecto y, en el caso de que resulte seleccionado, facilita la infraestructura tecnológica y las herramientas de IA necesarias para llevarlo a cabo. “Desde la sociedad civil no hay posibilidad de recopilar todos los datos que tú necesitas para hacer una inteligencia artificial potente que tenga mucha capacidad de aprendizaje, ni máquinas de computación de datos suficientemente potentes para competir con un Google o con un Amazon”, pero sí se dispone de herramientas institucionales para apoyar la creación y acercarla a otros sectores. “Yo creo que hay un debate no resuelto y que la gente del *open source* no hemos sabido resolver, sobre cómo aplicar la filosofía *open source* desde el principio hasta el final en la inteligencia artificial. Cómo la inteligencia artificial con redes neuronales que aprende por sí misma, cómo eso se hace *open source* del todo”.

En la línea de Gala Pin, existen propuestas que van más allá y hablan de la necesidad de convertir internet en un verdadero servicio público.

109 Es la definición de ChatGPT: *Hugging Face es una empresa y una plataforma de tecnología que se especializa en herramientas y modelos de inteligencia artificial, con un enfoque particular en el procesamiento del lenguaje natural (NLP) y el aprendizaje profundo. Fundada en 2016, Hugging Face ha desarrollado una comunidad activa y una extensa biblioteca de modelos y recursos de IA que son ampliamente utilizados tanto por investigadores como por desarrolladores. Cuenta con un gran repositorio de transformers y modelos pre-entrenados, multitud de datasets, y extensa documentación para desarrolladores, todo de código abierto y con la premisa de compartir código.*

110 <https://publiccode.eu/es>



“La tecnología nunca es neutral; siempre está condicionada por fuerzas políticas, sociales y económicas, así como por los valores y las elecciones humanas. El nacimiento de internet estuvo muy influido por las filosofías libertarias de los primeros fundadores de Silicon Valley, y nuestro enfoque actual de la regulación tecnológica se ha visto predominantemente moldeado por los deseos neoliberales de favorecer el crecimiento económico y el consumismo. Sin embargo, estas ideologías no deberían determinar los límites de nuestra imaginación. Teniendo en cuenta todo lo que está en juego, es hora de preguntarse si los modelos empresariales basados en el servicio público podrían ofrecer mejores resultados para la democracia y los ciudadanos”, escribe Helen Jay, directora de la sección de política de la cadena pública británica Channel 4¹¹¹.

La escritora y profesora de Harvard, Shoshana Zuboff, publicó en 2020 el libro ‘La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder’. En él analiza cómo se está produciendo una gran acumulación de riqueza y poder en Silicon Valley, donde existen “mercados conductuales” en los que se compran y se venden predicciones sobre nuestro comportamiento. Esto es, que utilizan nuestros datos sin que lo sepamos para engancharnos a sus productos y servicios sin descanso. Para entender la magnitud de lo que hablamos, Zuboff aporta un ejemplo: un documento filtrado de Facebook, empresa que ahora se conoce como Meta, demostró que para 2018 su centro de inteligencia artificial estaba procesando billones de datos al día para producir alrededor de 6 millones de pronósticos de comportamiento.

Los algoritmos atesoran información importante sobre nuestros gustos y preferencias y son capaces de *crearnos* necesidades de consumo de sus productos y servicios. Sin que apenas seamos conscientes de ello, son los que eligen la música que escuchamos en Spotify, las series que vemos en Netflix y los productos que compramos en Amazon. La enshittification de internet tiene una artillería poderosa detrás; el consumo exacerbado que articula produce grandes beneficios económicos a unos pocos. Y esos algoritmos corruptos controlan toda la red, especialmente nuestras redes sociales. Como lamenta Marta Peirano en su libro *El enemigo conoce el sistema*, todas las mentes académicas y empresariales más brillantes trabajan en la economía de la atención, en conseguir nuestro *engagement*, nuestro tiempo y nuestros clics. Las notificaciones *push* o el *scroll* infinito y otras ocurrentes técnicas de seducción son armas cuidadosamente pensadas para retenernos el mayor tiempo posible en sus aplicaciones. Y tienen efectos contradictorios en las personas usuarias: primero generan grandes y adictivos picos de dopamina, después acaban en cuadros de ansiedad y adicción. Más adecuado que personas usuarias sería, simplemente, consumidores. Como dice la investigadora Paola Ricaurte, el mundo se convierte en un dato: “La colonialidad del poder a través de la tecnología se expresa a través de la captura algorítmica de la subjetividad a través del extractivismo de datos y la mediación algorítmica que produce modelos de mundo de los cuales estamos excluidas”.¹¹²

111 Jay, H. (2024, 22 de marzo). We Need a Public Service Internet to Free Us from Big Tech's Grasp. Scientific American. Helen Jay es doctoranda en la Universidad de Westminster (Inglaterra). https://www.scientificamerican.com/article/we-need-a-public-service-internet-to-free-us-from-big-techs-grasp/?sma=tendencias_2024.03.26&utm_medium=email&utm_source=newsletter&utm_campaign=tendencias_2024.03.26

112 La investigadora mexicana Paola Ricaurte realizó una ponencia en el DecidimFest21 sobre extractivismo de datos, violencia y colonialidad: https://www.youtube.com/watch?v=LEIxBup6l_Q&ab_channel=Decidim

El bien común, la privacidad y el capitalismo de plataformas

OpenAI se define como “una empresa de investigación y despliegue de IA” cuya misión es “garantizar que la inteligencia artificial general beneficie a toda la humanidad”. Según las últimas estimaciones¹¹³, sus ingresos son de 1.000 millones de dólares, con una facturación de más de 80 millones de dólares al mes. Cabe destacar que según algunos estudios¹¹⁴, un 23% de adultos estadounidenses ha usado ChatGPT, eso equivale a unos 60 millones de personas, sobre todo menores de 30 años.

Las empresas que contratan su IA, los usuarios que pagan su tarifa –por ejemplo, la suscripción de pago de ChatGPT4– y los datos personales que utiliza cuando aceptamos sus condiciones de uso, son sus principales fuentes de ingresos. Precisamente OpenAI ha recibido un varapalo en cuanto a gestión y extracción de datos personales en España, donde la Agencia Española de Protección de Datos (AEPD) ha paralizado en marzo de 2024 la actividad de Worldcoin, una empresa vinculada a su CEO, Sam Altman, que llevaba desde julio de 2023 escaneando el iris a unos 400.000 españoles, a cambio de un lote de criptodivisas que equivaldrían a unos 80 euros. Los datos recogidos hasta la fecha por Worldcoin quedan bloqueados, por lo que no se podrán tratar ni compartir a la espera de una investigación internacional que decidirá si es legal o no que una empresa privada recoja este tipo de información biométrica.

El experto en derecho digital Borja Adsua declaraba su preocupación al diario El País¹¹⁵ puesto que “no se pone el foco en lo importante: el problema no es que te den dinero por tu iris, sino si ese dato se está tratando de forma correcta”. En total, y según Marilín Gonzalo, llevan ya 3,4 millones de datos recopilados. Pero ¿para qué quiere Worldcoin y la empresa matriz que hay detrás, Tools for Humanity Corporation, escanear nuestros ojos? Su respuesta es que pretenden crear una identidad digital única para cada persona en el mundo utilizando la tecnología de escaneo de iris que permita diferenciar la humana de otras identidades creadas por la IA. *No queremos saber quién eres, solo que eres único*, dicen en el apartado de privacidad de su página web. También te ofrecen la posibilidad de guardar o no este dato biométrico, aunque advierten de que si eliges la opción de borrado de estos datos, sí se conserva un mensaje que contiene un código de tu iris. Las denuncias sobre la dudosa legalidad de esta práctica llegan también por la opacidad y las pocas explicaciones que ofrece sobre el cómo y el para qué del tratamiento de estos datos tan sensibles. **“OpenAI ahora de open no tiene nada. Hasta el punto de que desconocemos de primera mano los detalles de la arquitectura de red neuronal que hay detrás de ChatGPT. Lo que sabemos es por experimentación y por deducción”**, critica Jordi Linares.

Nuestro iris es un rasgo biométrico mucho más preciso que el reconocimiento facial o que la huella dactilar y que, mal usado, puede servir para suplantar la identidad de una persona. “A través de la mirada y de cómo dilata la pupila una persona puedes saber qué es lo que le gusta, qué le asusta, qué le interesa y hasta ciertas características cognitivas, como si tiene párkinson”, indica a El País Carissa Véliz, profesora de filosofía en la Universidad de Oxford y autora del libro ‘Privacidad es poder’.

113 Efrati, A y Holmes, A. (2023, 29 de agosto). *OpenAI passes 1 billion revenue pace as big companies boost ai spending*. The Information. <https://www.theinformation.com/articles/openai-passes-1-billion-revenue-pace-as-big-companies-boost-ai-spending>

114 Tyson, A; B, Kennedy.(2024). *Many Americans think generative AI programs should credit the sources they rely on*. Pew Research Center. https://www.pewresearch.org/short-reads/2024/03/26/many-americans-think-generative-ai-programs-should-credit-the-sources-they-rely-on/?sma=newsletter_tecnologia&utm_medium=email&utm_source=newsletter&utm_campaign=tecnologia_2024.04.12_2

115 G. Pascual, M.(2024, 8 de marzo). *Por qué el iris es el dato biométrico máspreciado*. El País. https://elpais.com/tecnologia/2024-03-08/por-que-el-iris-es-el-dato-biometrico-mas-precioso.html?sma=tendencias_2024.03.12&utm_medium=email&utm_source=newsletter&utm_campaign=tendencias_2024.03.12



No es la primera vez que la AEPD actúa así ante un caso de recopilación de datos biométricos. En 2021 ya multó a Mercadona por instalar cámaras con sistemas de reconocimiento facial en 48 tiendas sin el conocimiento ni consentimiento de los clientes. Ana Valdivia habla de la extensión de los “mecanismos de la datificación” según los cuales desde que nos levantamos hasta que nos vamos a dormir, dentro y fuera de casa, estamos produciendo datos, por ejemplo, con los relojes que van calculando nuestras pulsaciones o contando nuestros pasos. “Todo eso se envía a bases de datos que alimentan algoritmos que van analizando nuestro comportamiento”. La privacidad fue y sigue siendo una de las primeras damnificadas desde la irrupción y la extensión de internet. Los dispositivos de almacenamiento de datos están muy extendidos, pero puede que no todas las personas sean conscientes de lo que suponen. Carissa Véliz recuerda que cuando los nazis invadían un país, lo primero que hacían era apoderarse de los registros locales como primer paso para controlar a la población y, en particular, para localizar a los judíos. La filósofa propone imaginar qué pasaría si llegara al gobierno un régimen autoritario, y así lo cuenta en *El País*¹¹⁶: “Si pudiera predecir todos nuestros movimientos, podría ser el comienzo de un régimen invencible. Para que te hagas una idea de lo peligrosos que son los datos personales, imagínate un régimen como el nazi, pero en la actualidad, con acceso a datos en tiempo real sobre tu ubicación, tu perfil facial, tu forma de andar, tu frecuencia cardíaca, tus ideas políticas, tu afiliación religiosa y muchas cosas más”.

Como explica Mariano Sigman en su libro y se apuntaba más arriba, el avance de la IA que vemos hoy en día no sería posible si el soporte tecnológico no hubiera avanzado a la par. Aquí el nombre clave es Nvidia, la empresa tecnológica que ha sido definida como la firma más influyente Wall Street y el valor más importante sobre el planeta tierra¹¹⁷. Nvidia anunciaba en febrero de 2024 un aumento del 265% en los ingresos trimestrales y unos beneficios de 12.285 millones de dólares, cifras récord para la compañía. En 2023, su negocio de IA generó unos 15 mil millones de dólares en ingresos, un 40% más que el año anterior. De hecho, superó a la división de juegos como su mayor fuente de ingresos. En consecuencia, sus acciones se han disparado más de un 180% este año. Los ingresos de Nvidia en el conjunto del año crecen un 126%, hasta los 60.922 millones de dólares, según las cuentas comunicadas a la Comisión de Valores y Bolsa de Estados Unidos (la SEC).

Nacida en 1993, empezó fabricando chips gráficos y se especializó como proveedora de tarjetas gráficas para videojuegos hasta que en 1999 mejoró notablemente la tecnología de las GPU, unidades de procesamiento gráfico, lo cual permitió el impulso de las capacidades gráficas de los ordenadores. Su CEO, Hensen Juang, en 2006, optó por que los chips de Nvidia fueran programables, algo que permitió adaptarlos a usos diferentes como la minería de criptomonedas y, sobre todo, aplicarlos en los avances en la investigación de la IA. Fue Alphabet, empresa matriz de Google, la primera que apostó por las GPU de Nvidia para entrenar los grandes modelos de lenguaje que desarrollaba. Este sofisticado hardware necesita muchísima energía para funcionar, unas cinco veces más que los procesadores convencionales o CPU. El entrenamiento de los LLM requiere de decenas de miles de GPU operando día y noche durante semanas o meses. Nvidia, sin duda, lidera el campo del hardware de desarrollo de IA y su colaboración con OpenAI ha sido, según algunas fuentes, fundamental para entender el avance y desarrollo de modelos revolucionarios como ChatGPT, el gran exponente de la inteligencia artificial generativa, máxima rival de profesiones creativas como la periodística.

116 Véliz, C. (2021, 12 de septiembre). *Protejamos nuestros datos. No olvidemos cómo los usaban los nazis*. *El País*. <https://elpais.com/ideas/2021-09-12/protejamos-nuestros-datos-no-olvidemos-como-los-usaban-los-nazis.html>

117 Jiménez, M. (2024, 5 de febrero). *Nvidia supera las previsiones con unos resultados espectaculares gracias a la inteligencia artificial*. *Cinco Días*. <https://cincodias.elpais.com/companias/2024-02-21/nvidia-supera-las-previsiones-con-unos-resultados-espectaculares-gracias-a-la-inteligencia-artificial.html>



Una forma de concluir

Entre la amenaza colosal y la oportunidad

Puede que todavía estemos a tiempo de decidir qué queremos que suponga la irrupción de la inteligencia artificial generativa tanto en nuestra vida como en la profesión periodística, si será una amenaza masiva o una valiosa oportunidad para la reinención. “Ante el periodismo de declaraciones que tan en boga está, las máquinas son imbatibles”. Antonio Lorenzo, periodista de El Economista no tiene ninguna duda: “La IA generativa va a ser una amenaza colosal para los periodistas”, incluso establece la fecha simbólica de tres años para que realmente estos modelos de escritura automática ocupen un asiento en las redacciones. Los periodistas que escriban como máquinas serán los primeros en ser sustituidos.

De nuevo es el momento de resignificar la profesión, más allá del arquetipo de periodista-tertuliano para ser realmente el periodista especialista quien persigue la noticia, ofrece un análisis certero, aporta algo diferente y es capaz de absorber el contexto para enmarcarla de una forma que resulte entendible y apetecible para el público. Lorenzo lo llama “contar esa crónica que no sea capaz de contar la máquina”. Para muchos analistas, no queda tanto para que ni siquiera las noticias en exclusiva o las primicias importen mucho. Aquello que sí va a permanecer es el tono personal, las firmas que dan credibilidad y confianza. “El criterio, el olfato, su astucia y picardía profesional, el diálogo con la fuente, son los aportes que hacen irremplazable al periodista”, suma Flor Coelho.

“Somos inteligentes porque vivimos”

La profesora de inglés del hijo del neurocientífico Mariano Sigman propuso este ejercicio al alumnado: escribid un párrafo vosotros mismos, enfrentaros al desafío de la página en blanco, al reto de crear, de pensar, de discernir la mejor forma de hacerlo. Después, pasárselo a ChatGPT para que lo mejore. Una vez que tengáis los dos textos, comparadlos, ved qué cambios le aceptáis a ChatGPT y cuáles no. No tiene mucho sentido educar como si estos modelos no existieran o incluso prohibirlos, porque ya sabemos que en el campo no podemos poner puertas, y en el extenso paraje de la IA mucho menos, porque internet y la tecnología en general está diseñada para aprender a saltarse las vallas, los cortafuegos y casi cualquier tipo de limitación. Esto no quita que el marco jurídico sea necesario porque cada juego tiene sus propias reglas y las instituciones públicas y los gobiernos en su conjunto deben ser garantes y guardianes de los derechos de las personas en la calle y en la Red. Todas las personas entrevistadas coinciden en la importancia de educar, de formarnos, de cultivar el

espíritu crítico, y más en el caso de los y las periodistas que nos dedicamos a contar, explicar o traducir el mundo que nos rodea.

El fuerte golpe de la inteligencia artificial y su impacto en casi todos los ámbitos de nuestra vida nos está haciendo reflexionar y hacernos preguntas. Cómo definimos la inteligencia, qué cualidades son propias de lo humano, cómo diferenciarlos de las máquinas, cuál es nuestro papel y cuál es y será el suyo. “La IA no puede emocionar. Las máquinas solo parten de lo que ya estaba claro, pero únicamente el creador, el autor, es el que parte de la oscuridad, del dolor, y es capaz de descifrarlo con esfuerzo, es capaz de emocionar”, barruntaba el escritor Lorenzo Silva, en ese intento que se repite por ver qué nos diferencia, qué nos hace mejores. No queda tanto para que lleguemos a esta definición que imagina Mariano Sigman: “Inteligente es todo aquello que las máquinas no pueden hacer”.

Judith Membrives concede que es y será una herramienta útil pero también le pone nombre a los peligros de la IA:

“Dicen que buscan simular la inteligencia humana, pero es una visión de la inteligencia muy estrecha, no es contextual, no se tiene en cuenta toda la parte física o material nuestra que nos hace inteligentes. Porque nosotros no solo somos inteligentes porque leemos, somos inteligentes porque vivimos en un contexto, nos relacionamos con personas. Todo esto es aprendizaje, todo eso no tiene la IA. Entonces se basa en codificar la experiencia humana, que es algo que seguramente no es codificable, porque no es solo cuantificable, también es cualitativo. Pensamos que puede hacer absolutamente todo lo que hace un humano, le damos ese estatus y encima permitimos que las empresas que lo están desarrollando lo hagan sin mecanismos de transparencia consensuados, como ha sido y se ha hecho hasta ahora toda la producción del conocimiento científico”.

El proceso de entrenamiento de estos grandes modelos de lenguaje (LLM) se hace a imagen y semejanza del proceso del aprendizaje del primer lenguaje, que es por inferencia estadística, como ya se ha dicho. Los bebés no saben gramática ni tampoco les hace falta, no distinguen una preposición de un sustantivo, pero crean una especie de estructura estadística en la que, por probabilidad y a base de escuchar cientos o miles de veces las frases de su familia, van acertando, van mejorando la sintaxis de su lenguaje sin haber abierto un libro de texto. “Aprendizaje estadístico”, lo llama Mariano Sigman. El mismo que utiliza GPT.

El ser humano lleva un siglo compartiendo su vida con dispositivos tecnológicos más o menos sofisticados: el smartwatch, el primer teléfono móvil, el Osborne 1 o primer ordenador portátil de 11 kg de peso, la primera televisión en blanco y negro, la radio, los electrodomésticos. Antes de eso, las tecnologías primitivas o iniciáticas (el lenguaje, el fuego, los útiles de defensa, caza o para trabajar el campo hechos con piedras y pieles de animales, la rueda, la imprenta). Todos ellos han ido moldeando nuestro día a día y cubriendo huecos, hasta que en muchos casos algo mejor los ha ido sustituyendo. Esto ha supuesto cambios en nuestras tareas y también ha implicado delegarlas a los nuevos dispositivos que iban surgiendo (ya no usamos la orientación sino que abrimos Google Maps, ya no cambiamos las marchas si tenemos un coche automático). Nuestra historia es la historia de la sustitución. Cuando más sofisticadas son las tecnologías, más delegamos. Sigman divide las tecnologías en dos grupos: las que suplantando o sustituyen las capacidades humanas y, por lo tanto, atrofian nuestras propias capacidades hasta llevarnos al “sedentarismo cognitivo”, o aquellas que quitan fricción a las acciones de nuestro día a día y que multiplican nuestras propias cualidades. La diferencia entre optar por una bicicleta para desplazarse o por un coche. El mensaje de Sigman es apostar por una tecnología “que no nos reemplace sino que

nos multiplique y nos permita llegar a sitios a los que solos no llegaríamos, pero siendo nosotros los protagonistas y quienes tomamos las decisiones”.

El mayor desafío de toda la inteligencia artificial, según escribe Sigman, es no dejar de ser quienes somos. Esto se puede aplicar fácilmente al periodismo: no perder nuestra esencia, nuestra capacidad crítica, seguir mejorando nuestras cualidades, potenciar el valor (informativo, veraz y también el económico) de nuestra firma. Que no elija ni la máquina ni el algoritmo, ni Netflix ni los *reels* de Instagram qué quieres leer, escribir, comprar o ver a continuación. Y sin olvidar que la expansión voraz de la inteligencia artificial no va a durar eternamente, porque nada crece infinitamente. Todas las burbujas explotan, o como dicen Narayanan y Kapoor, “ninguna tendencia exponencial puede continuar indefinidamente. Pero puede ser difícil predecir cuándo una tendencia tecnológica está a punto de estancarse”.

Esta investigación solo es un punto y seguido. Aunque como hemos visto, la historia de la inteligencia artificial se remonta varias décadas atrás, este momento álgido que protagonizamos acaba de empezar. Estamos en la cresta de una ola que inevitablemente romperá en algún momento sobre las rocas y no sabemos bien qué pedazos quedarán a flote. “George Orwell temía a aquellos que prohibían los libros. Aldous Huxley temía que no hubiera razones para prohibir libros porque no quedaba nadie que los quisiera leer”, escribió Neil Postman en 1985 en su libro *Amusing Ourselves to Death*. La irrelevancia del periodismo es una amenaza mayor que cualquier supercomputadora *omnimodal* de avanzadas redes neuronales sintéticas, pero de la profesión dependerá diferenciarse de la *enshittification* o buscar inteligentes excusas artificiales con las que justificar su mediocridad ética e informativa.



Internet no es más importante que la emergencia climática, la justicia de género, la justicia racial, el genocidio o la desigualdad. Pero Internet es el terreno en el que libraremos esas luchas. Sin una Internet libre, justa y abierta, la lucha está perdida antes de empezar.

Cory Doctorow



Bibliografía

Cano Orón, L., López Meri, A. (2024). *Introducción al uso de la IA en periodismo: Guía de referencias y modos de uso.* Publicacions de la Universitat de Valencia. <https://omp.uv.es/index.php/PUV/article/view/730>

Andrés Pedreño Muñoz, A; González Gosálbez, R; Mora Illán, T; del Mar Pérez Fernández, E; Ruiz Sierra, J; Torres Penalva, A. (2024). *La IA que revolucionará el periodismo.* Grupo 1million bot.

Peirano, M. (2019). *El enemigo conoce el sistema: Manipulación de ideas, personas e influencias después de la economía de la atención.* Editorial Debate.

Vázquez-Barrio, T; Salazar García, I (Ed). (2023). *Inteligencia artificial, periodismo y democracia.* Tirant Humanidades.

Sigman, M., & Bilinkis, S. (2023). *Artificial: La nueva inteligencia y el contorno de lo humano.* Editorial Debate.

VV. AA. (2024). *Propuestas para regular la Inteligencia Artificial desde una perspectiva de derechos.* Conclusiones II Jornadas DAR - Democracia, Algoritmos y Resistencias.

VV. AA. (2023). *Biko Insights #6. Tendencias digitales para un mundo en constante cambio.* Jakala - Biko.

VV. AA. (2020). *Tecnológicas.* Revista de Eldiario.es.

VV. AA. (2023). *Inteligencia artificial: Riesgos, verdades y mentiras.* Revista de Eldiario.es.

Cabrera, M. (2024). *Tendencias: El año de las historIAS.* Story Baker.

Glosario de entrevistas

(realizadas entre enero y marzo de 2024)

Ujué Agudo, doctora en psicología y **Karlos G. Liberal**, data activista y ciberpunk, investigadores en BikoLabs.

Miguel Arana, investigador en Machine Learning Causal de la Universidad de Cranfield, en Reino Unido.

Andrés Boix, profesor de Derecho Administrativo de la Universitat de València.

Mauricio Cabrera, periodista mexicano y “terapeuta de contenidos” creador de la newsletter Story Baker.

Juan Candela, responsable de Digital de Las Provincias.

Flor Coelho, periodista de La Nación, Argentina.

Lorenzo Cotino, Catedrático de Derecho Constitucional de la Universitat de València.

Iván Esteve, director de Informatius de À Punt.

Marilín Gonzalo, periodista de Newtral.

Javier Lifante, responsable de Digital de À Punt.

Jordi Linares, doctor en Informática y profesor de la UPV.

Ricardo León, abogada de Algoritmo Legal, especializada en Derecho de las nuevas tecnologías y negocios.

Antonio Lorenzo, periodista de El Economista.

Judith Memvibres, responsable de políticas digitales en Lafede.cat y parte de Algorithms, además de experta en la AI Act.

Pau Peiró, responsable por parte del 'SAMC del proyecto 'Recerca i desenvolupament de sistemes intel·ligents avançats per a l'accessibilitat, catalogació i indexació de continguts textuais, radiofònics i audiovisuals' de À Punt.

Gala Pin, diputada de En Común Podem en el Congreso de los Diputados.

Carmela Ríos, periodista experta en redes sociales, periodismo móvil y desinformación.

Íñigo Roy, jefe de Digital del Levante-EMV.

Marta Suárez-Mansilla, abogada de Algoritmo Legal, especializada en Derecho de las nuevas tecnologías y negocios.

Ana Valdivia, Profesora investigadora en IA, Gobernanza y Políticas del Oxford Internet Institute de la Universidad de Oxford.

Glosario de términos

A

Algoritmo

Es un conjunto de instrucciones que una computadora sigue para realizar una tarea. Los algoritmos se utilizan en todas las áreas de la informática, incluida la IA.

Aprendizaje automático

Un subcampo de la IA que se centra en el desarrollo de algoritmos que pueden aprender de los datos y mejorar su rendimiento con el tiempo.

Automatización

Es el uso de máquinas para realizar tareas que antes se realizaban por humanos. La automatización se utiliza en muchos sectores para mejorar la eficiencia y la productividad.

I

Inteligencia artificial (IA)

Es un campo de la informática que busca crear máquinas que puedan imitar las capacidades cognitivas humanas, como el aprendizaje, el razonamiento y la resolución de problemas.

IA generativa

Es un campo de la IA que se centra en el desarrollo de modelos que pueden generar contenido nuevo, como texto, imágenes o código. La IA generativa se utiliza para tareas como la creación de arte, la generación de música y la traducción automática.

Inteligencia Artificial General (AGI)

Es un tipo hipotético de IA que sería capaz de realizar cualquier tarea intelectual que pueda realizar un humano. La AGI aún no se ha logrado, pero es un objetivo de investigación a largo plazo para muchos científicos de la IA. En la legislación europea (la AI Act) se habla de IA de propósito general e IA de propósito general con riesgo sistémico.

L

LLM

Large Language Models, modelos masivos de lenguaje, son un tipo de inteligencia artificial, un software avanzado que ha sido entrenada con una gran cantidad de texto para entender y generar lenguaje humano. Es la tecnología que hay detrás de los chatbots, capaces de leer, escribir, traducir y responder preguntas en lenguaje natural, como lo haría una persona. Algunos ejemplos de LLM serían GPT de OpenAI (Generative Pretrained Transformer), LLaMA (Large Language Model Meta) o PALM2 (Pathway Language Model 2) de Google AI.

M

Machine learning (ML)

Es un subcampo de la IA que se centra en el desarrollo de algoritmos que pueden aprender de los datos y mejorar su rendimiento con el tiempo.

Máquina automatizada

Es una máquina que puede realizar una tarea sin la intervención humana. Las máquinas automatizadas se utilizan en muchos sectores, como la fabricación, la agricultura y la atención médica.

Modelos fundacionales

La AI Act define a los *foundation models* como “un desarrollo reciente en el que los modelos de IA se desarrollan a partir de algoritmos diseñados para optimizar la generalidad y versatilidad del resultado. Estos modelos a menudo se entrenan con una amplia variedad de fuentes de datos y grandes cantidades de información para realizar una amplia gama de tareas posteriores, incluidas algunas para las que no fueron específicamente desarrollados ni entrenados. Estos sistemas pueden ser unimodales o multimodales, entrena-

dos mediante diversos métodos, como el aprendizaje supervisado o el aprendizaje por refuerzo”. El término “foundation models” fue acuñado por investigadores del Instituto de Inteligencia Artificial Centrada en el Ser Humano (HAI, por sus siglas en inglés) de la Universidad de Stanford. Su notoriedad llegó con la publicación del artículo “On the Opportunities and Risks of Foundation Models” en 2021, liderado por Percy Liang y otros colaboradores. Este trabajo conceptualizó y formalizó el enfoque de estos modelos como una base unificadora para diversas aplicaciones de inteligencia artificial, tanto por su potencial como por los riesgos asociados.

Algunos ejemplos serían:

- Gopher: Desarrollado por DeepMind, fue uno de los primeros modelos en demostrar la capacidad de realizar múltiples tareas.
- LaMDA (Language Model for Dialogue Applications): Desarrollado por Google AI, se especializa en la generación de diálogos realistas y coherentes.
- MUM (Multitask Unified Model): Desarrollado por Google AI, es un modelo multimodal que puede procesar información de texto e imágenes.

P

Procesamiento del lenguaje natural (NLP, Natural Language Process)

Es una rama de la inteligencia artificial que permite a las máquinas comprender, interpretar y generar lenguaje humano. Su objetivo es facilitar la interacción entre humanos y ordenadores utilizando el lenguaje natural, como cuando hablamos o escribimos. Las aplicaciones del NLP incluyen asistentes virtuales, traducción automática y análisis de textos.

Procesamiento natural del lenguaje (NLP)

Es un campo de la IA que se centra en la interacción entre las computadoras y el lenguaje humano. El NLP se utiliza para tareas como traducción automática, análisis de sentimiento, resumen de texto y generación de lenguaje natural.

R

Red neuronal

Una red neuronal es un modelo computacional inspirado en el cerebro humano, compuesto por capas de nodos interconectados (“neuronas”) que procesan datos en múltiples capas para reconocer patrones complejos. Se utiliza ampliamente en tareas como el reconocimiento de voz e imagen. Hay dos tipos principales:

- Red Neuronal Convolutiva (CNN), un tipo de red neuronal diseñada para procesar datos estructurados en forma de cuadrícula, como las imágenes. Las CNN son efectivas para tareas de visión por computadora como el reconocimiento de objetos y la clasificación de imágenes.
- Red Neuronal Recurrente (RNN), un tipo de red neuronal que es efectiva para procesar datos secuenciales, como texto o series temporales. Las RNN tienen “memoria” que les permite considerar información previa en sus predicciones actuales, haciéndolas adecuadas para tareas como la traducción automática y la predicción de texto.

S

Self-supervised learning (SSL)

Es un tipo de aprendizaje automático en el que los modelos se entrenan con datos sin etiquetar. El SSL se utiliza para tareas como la representación de lenguaje, la detección de anomalías y la generación de imágenes.

Sistemas algorítmicos

Son sistemas que utilizan algoritmos para tomar decisiones. Los sistemas algorítmicos se utilizan en muchas áreas, como la banca, la justicia penal y la atención médica.

T

Transformer

Un tipo de arquitectura de red neuronal diseñada específicamente para el procesamiento del lenguaje natural (PLN). Fue introducido en un documento de investigación de Google AI en 2017 y ha revo-

lucionado el campo del PLN. En esencia, un transformer es un modelo basado en la atención. Esto significa que, en lugar de procesar las palabras de una oración de forma secuencial (una tras otra), el transformer analiza las relaciones entre todas las palabras de la oración simultáneamente. Como si pudiera poner atención a todas las palabras a la vez para entender el contexto completo de la oración. Para lograr esto, utiliza primero un codificador que procesa la entrada (por ejemplo, una oración) y genera una representación interna. Después, un decodificador utiliza esta representación para generar la salida (por ejemplo, una traducción o un resumen). Además, tanto el codificador como el decodificador contienen capas de atención que permiten al transformer "prestar atención" a las partes más relevantes de la entrada para cada parte de la salida.

V

Visión por computación

La visión artificial o visión por computadora es una rama de la inteligencia artificial que permite a las máquinas interpretar y comprender imágenes o videos de manera similar a cómo lo hacen los humanos. Para ello utiliza algoritmos y técnicas avanzadas que permiten extraer información útil de datos visuales, como fotografías o secuencias de video. Así pueden identificar y localizar objetos en una imagen y reconocer patrones de las imágenes o videos. Es la tecnología detrás del reconocimiento facial y también de los vehículos autónomos, diagnósticos médicos, realidad aumentada, etc.



