

Automatismos y generación de pensamiento

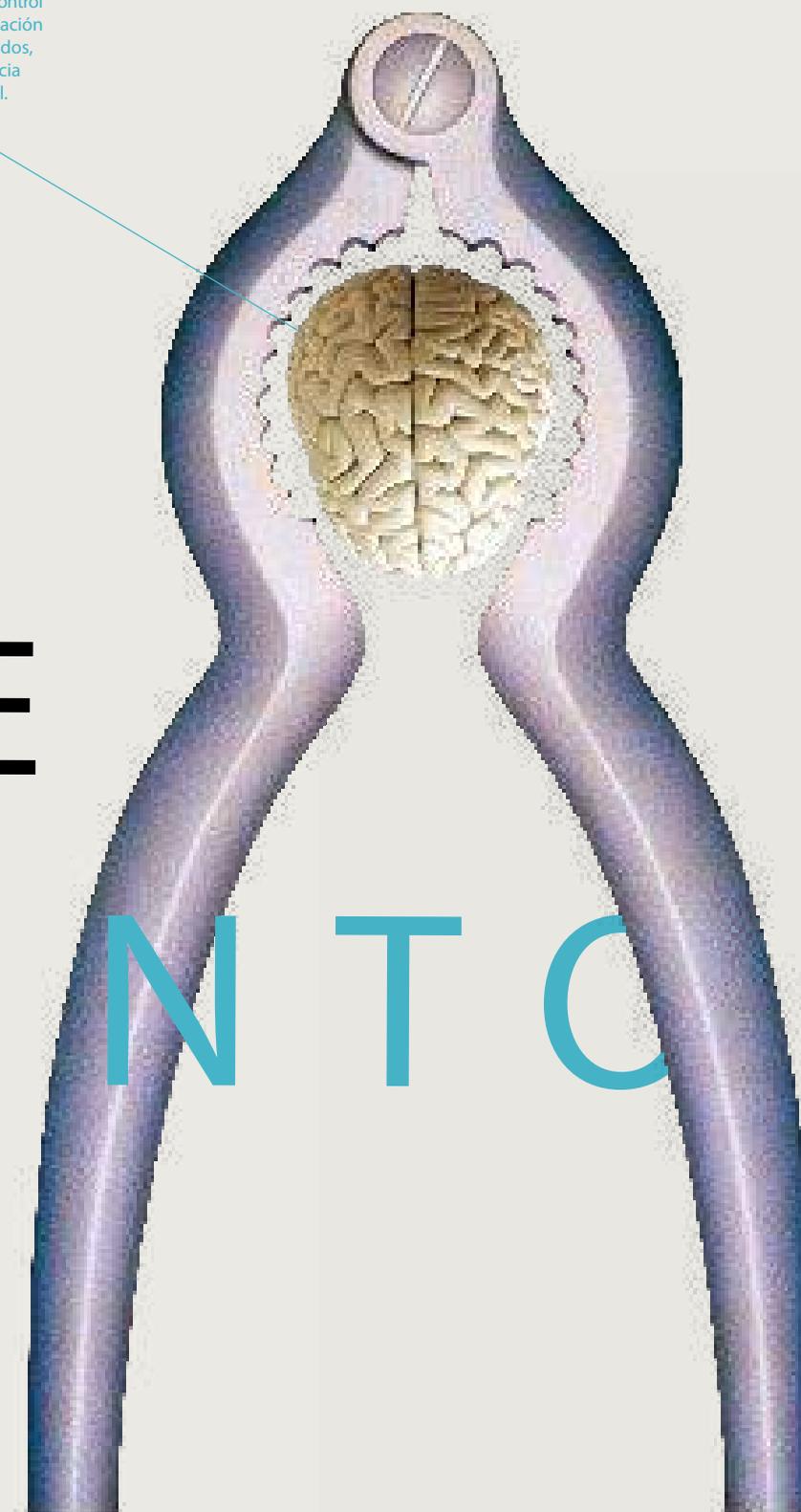


EVA MOYA
Directora corporativa de
Inteligencia en Mnemo

LIBERTAD DE PENSAMIENTO E INTELIGENCIA ARTIFICIAL

Los nuevos avances en inteligencia artificial han sorprendido por su capacidad de generar contenido que hasta ahora era propio de la creatividad humana. Entramos en una nueva era que obliga a una profunda reflexión sobre si los algoritmos de la IA pueden ayudarnos a avanzar como sociedad. Hay que analizar la posibilidad de que puedan ser utilizados de forma perversa para así favorecer las tácticas de manipulación social. Es el momento de revisar qué acciones deben ser contempladas para evitarlo.

Palabras clave:
posverdad, control social, generación de contenidos, inteligencia artificial.



Automatismos and thought generation
FREEDOM OF THOUGHT AND ARTIFICIAL INTELLIGENCE

New advances in artificial intelligence have surprised us with their ability to generate content that until now was the domain of human creativity. We are entering a new era that requires a deep reflection on whether AI algorithms can not only help us to advance as a society. We must analyze the possibility that they can be used in a perverse way to favor post-truth social manipulation tactics. It is time to review what actions should be contemplated to avoid it.

Keywords: post-truth, social control, content generation, artificial intelligence.

Vivimos una época histórica compleja, donde la interacción entre los descubrimientos tecnológicos y los nuevos comportamientos sociales nunca han sido tan relevantes para la configuración del futuro de la humanidad. En este sentido, el empuje recibido por los avances en inteligencia artificial (IA), las tecnologías cuánticas¹ y el 5G/6G² están facilitando la posibilidad de delegar tareas y funciones sobre los automatismos, así como mejorar los procesos de las operaciones. Los automatismos más conocidos están relacionados con la fabricación o sistemas industriales, algunos de los cuales utilizamos en nuestras casas, como por ejemplo los robots de limpieza, que estudian nuestros patrones de uso de la casa; o a través del consumo de servicios, como en el caso de las aplicaciones de servicios de transporte que pueden identificar las mejores rutas³. Sin embargo, también hay otras opciones que aparecen en el horizonte que pueden mejorar nuestra calidad de vida, como aquellas orientadas a la salud⁴, al cuidado de los dependientes⁵, a la mejora del entorno y medioambiente⁶, etcétera.

A los usos más tradicionales, sin embargo, se unen también otros de

gran impacto en el desarrollo de las relaciones humanas como son aquellos relacionados con la comunicación y el lenguaje. Por ejemplo, se reflexiona ya cómo la IA puede ayudar en los procesos de aprendizaje⁷. Pero también, recientemente la humanidad se ha sorprendido ante la manifestación más cercana al pensamiento creativo automatizado a través de una serie de chatbots⁸ capaces de dialogar sobre casi cualquier tema, como el ChatGPT⁹ del proyecto OpenIA. Este bot tiene la particularidad de que está disponible para ser consumido por cualquier persona de forma gratuita, lo que le está permitiendo ampliar la base de conocimiento a gran velocidad. Sin embargo, no solo existen aquellas inteligencias centradas en texto, sino también tenemos disponibles otras inteligencias multimedia centradas en imágenes, como la primera versión gratuita de DALL-E, del proyecto Craiyon¹⁰, que ha evolucionado con OpenIA en una segunda versión comercial que está siendo utilizada por grandes compañías para generar contenidos visuales¹¹. Y, por supuesto, también se pueden generar contenidos visuales a través de vídeo como es el caso del generador de vídeos

animados llamado Raw Shorts¹² o vídeos reales como Synthesia¹³.

Ahora bien, toca preguntarse hasta qué punto la creatividad del bot puede ir más allá de una mera reconstrucción de contenidos basados en la información de la que dispone la inteligencia artificial, o esta será capaz de construir contenidos creativos por sí misma. En este sentido hay más sorpresas, como por ejemplo la IA denominada Text-Davinci-003, que ha sido capaz de generar un poema basado en la teoría de la relatividad de Albert Einstein¹⁴, o Botto¹⁵, quien ya vende cuadros por más de un millón de dólares.

IA en espacios humanos

Es importante destacar que hoy el aprendizaje es dirigido. Esto significa que la IA necesita una serie de inputs de información para poder funcionar y conforme se interactúa con ella, la IA va generando el contenido. Ahora bien, la IA no está preparada para evaluar si el contenido sobre el que aprende está verificado o incumple los valores que consideramos universales¹⁶. En esta línea, son todos los movimien-

tos de la UE para conseguir “la ética de los algoritmos”¹⁷.

Ante este panorama se abren varios escenarios que deben ser contemplados, no solo a través de las numerosas oportunidades que nos pueden brindar, sino también a través de los diferentes riesgos a los que exponemos a la humanidad, que no solo pueden ser tecnológicos, sino que también pueden afectar a nuestra forma de interpretar la realidad.

Toca recordar el punto de partida. Estamos ante una sociedad

Las tecnologías
y la manipulación
del pensamiento
ya vienen operando
juntas desde
hace tiempo

1 Sending quantum information from Alice to Charlie: <https://cordis.europa.eu/article/id/442356-sending-quantum-information-from-alice-to-charlie/es>

2 6G-IA: <https://6g-ia.eu/>

3 Engineering More Reliable Transportation with Machine Learning and AI at Uber: <https://www.uber.com/en-ES/blog/machine-learning/>

4 INFORME C: Inteligencia artificial y salud: <https://oficinac.es/informes-c/ia>

5 GeoFamily, un proyecto que utiliza la inteligencia artificial para el cuidado de personas mayores: <https://www.ubu.es/noticias/geofamily-un-proyecto-que-utiliza-la-inteligencia-artificial-para-el-cuidado-de-personas-mayores-ganador-de-startup-weekend-burgos>

6 IA como clave para la preservación del medio ambiente y de la industria forestal: <https://ticnegocios.camaravalencia.com/servicios/tendencias/inteligencia-artificial-como-clave-para-la-preservacion-del-medio-ambiente-y-de-la-industria-forestal/>

7 43 Examples of Artificial Intelligence in Education: <https://onlinedegrees.sandiego.edu/artificial-intelligence-education/>

8 El sustantivo *chatbot*, empleado para denominar a los programas informáticos con los que se puede mantener una conversación, es un neologismo válido en español: <https://www.fundeu.es/recomendacion/chatbot-neologismo-valido>

9 Introducing ChatGPT: <https://openai.com/blog/chatgpt>

10 Hugging Face: <https://huggingface.co/spaces/dalle-mini/dalle-mini>

11 DALL-E 2: <https://openai.com/product/dall-e-2>

12 Transforma tu Texto en Vídeos Animados en Minutos: <https://www.rawshorts.com/creacion-de-videos-automatizada>

13 1 AI video creation platform. Create professional videos in 15 minutes: <https://www.synthesia.io/>

14 Una inteligencia artificial que recita poemas y escribe rimas: <https://www.infobae.com/america/tecnologia/2022/12/12/una-inteligencia-artificial-que-recita-poemas-y-escribe-rimas/>

15 Botto is a decentralized autonomous artist: <https://www.botto.com/>

16 Valores Universales: <https://unsdg.un.org/es/2030-agenda/valores-universales>

17 Marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html

totalmente perfilada por las grandes plataformas a las que regalamos nuestros datos y que les permiten identificar nuestros patrones de consumo y comportamiento. Gracias al big data de buscadores, redes sociales, plataformas de comercio electrónico y demás, hemos ido permitiendo que se desarrolle una identidad digital sobre nosotros con unas claras líneas que definen lo que nos motiva o lo que nos disgusta. Tanto es así que hoy en día es difícil leer en Internet algo que no sea de nuestro agrado, dado que los algoritmos nos encierran en todo lo que encaja en nuestras preferencias, mostrándonos alternativas para leer, ver, escuchar o participar en aquellas experiencias que sean similares a lo que nos gusta. Este efecto ha sido bautizado por el ciberactivista Eli Pariser como “el filtro burbuja”.

El marcador somático

Si añadimos al “filtro burbuja” todo lo aprendido ya desde que Antonio Damasio identificara su “marcador somático” en la obra *El error de Descartes* (1994), quien refiere la relevancia de las emociones en todas las decisiones, nos encontramos ante una palanca accionable, que puede llegar a facilitar la manipulación del pensamiento humano sobre decisiones más relevantes para la humanidad que el simple hecho de incitar a una compra sobre otra. Una de las perversiones en el centro de la confluencia tecnológica que venimos argumentando, es la ya muy conocida posverdad. Esa técnica de comunicación, donde el generador del contenido solo muestra una parte de la verdad y

la magnífica, de tal manera que parece la única verdad. Esa parte de la realidad, normalmente, suele coincidir con algunas creencias del grupo objetivo sobre el que se actúa, con la intención de activar el sesgo de confirmación¹⁸ para suministrar un anclaje mental de supuestos argumentos objetivos.

El Brexit es, sin duda, uno de los ejemplos más claros y recientes del impacto de la posverdad en la política europea. La campaña a favor de la salida del Reino Unido de la Unión Europea utilizó reconocidas técnicas de posverdad como por ejemplo, el hecho de centrarse solo en la cantidad de dinero que los ingleses estaban aportando a la UE. La manipulación consistía en la convicción de que todo ese dinero iría a parar directamente a las arcas de un Reino Unido libre que podría utilizar en su sistema de sanitario¹⁹. Boris Johnson utilizó la cifra bruta, sin los desembolsos posteriores que la UE devolvía al país, representando más de la mitad de ese dinero. Como Johnson sabía que una de las principales quejas ciudadanas venía por el sistema de salud, considerado deficiente por gran parte de la ciudadanía británica, simplemente diseñó una estrategia de inferencia argumental que los ciudadanos no querrían refutar.

Llegados a este punto, donde las tecnologías y la manipulación del pensamiento ya vienen operando juntas desde hace tiempo, se nos plantean nuevos escenarios de riesgo digital sobre los que debemos actuar. Vamos hacia una IA hiperconectada, gracias a una velocidad de procesamiento de datos cada vez mayor; con capacidad de aprender del big data, que le facilita interpretar cómo somos y cómo interac- ▶▶▶



Los algoritmos nos encierran
en todo lo que encaja en
nuestras preferencias

¹⁸ Sesgo de confirmación: https://es.wikipedia.org/wiki/Sesgo_de_confirmacion

¹⁹ Boris Johnson: we will still claw back £350m a week after Brexit: <https://www.theguardian.com/politics/2017/sep/15/boris-johnson-we-will-claw-back-350m-a-week-post-brex-it-after-all>

- 20 Tay, la robot racista y xenófoba de Microsoft: https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb
- 21 Pause Giant AI Experiments: An Open Letter: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 22 'It's an especially bad time': <https://edition.cnn.com/2023/04/06/tech/tech-layoffs-platform-safety/index.html>
- 23 Ética de la inteligencia artificial: <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>
- 24 Estrategia Nacional de Inteligencia Artificial: <https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx>

Los saltos tecnológicos son necesarios para evolucionar como humanidad. Frenarlos es alejar las oportunidades que generan

tuamos entre nosotros; y capaz de identificar todas las claves de nuestra manera de comunicarnos para dotar a los contenidos generados de un aspecto de veracidad. Es verdad que la capacidad del razonamiento inductivo no está al alcance de las IA, tal y como justifica Larson en su Mito de la inteligencia artificial (2022), pero la toma de decisiones basada en argumentos sesgados que van moldeando la mente con las técnicas de posverdad, tampoco se basan en la aplicación del pensamiento crítico por parte de los ciudadanos.

Bajo esta premisa, y en términos generales, hay dos escenarios vinculados a las IA que podrían magnificar y ampliar los alcances de las tácticas de posverdad, manipulación del pensamiento y control social.

En el primer escenario, estos algoritmos podrían mejorar las técnicas actuales de posverdad y ser utilizadas por grupos de poder con la intención de distorsionar la realidad y crear tendencias de opinión enfocadas a sus propios intereses, ya sean de tipo político, social o económico. En el segundo escenario, la propia IA podría aprender estas tácticas de manipulación como parte de su aprendizaje mientras observa y analiza las relaciones sociales. Algo muy similar a lo que ocurrió con Tay, el chatbot de IA que Microsoft tuvo que retirar de la red tras aprender conductas xenófobas²⁰.

Evidentemente, los saltos tecnológicos son necesarios para evolucionar como humanidad, al menos en algunos aspectos, por lo que frenar el desarrollo también es frenar todas las oportunidades que se abren a su paso. Sin embargo, a falta de unas líneas claras de defensa, esta es una propuesta que parte de la comunidad digital y grandes pensadores han realizado recientemente con la idea de ralentizar el primer impacto de la aplicación a gran escala a través de una carta abierta denominada Pause Giant AI Experiments²¹. Sin embargo, “pau-

sar” no es la solución definitiva, ni tampoco asegura una reflexión profunda que permita desarrollar mecanismos de control hacia el futuro.

Proteger a la humanidad

Hay dos puntos de intervención humana que están muy claros para conseguir una protección de los derechos y libertades reales.

Por un lado, está el hecho de seguir profundizando en tomar consciencia del impacto social de los desarrollos de código de las IA, al margen de solo el beneficio económico. Por el otro, empieza a ser muy relevante desarrollar sistemas de identificación temprana que señalen con rapidez cuándo una IA se ha desvirtuado o está actuando en una dirección incompatible con los valores deseados.

La cuestión ética sobre el desarrollo de los algoritmos lleva tiempo en las casas de las grandes plataformas que salieron al mercado en busca de expertos en la materia. Sin embargo, paradójicamente, estamos en un momento en el que estas grandes plataformas están despidiendo a estos expertos debido a la crisis en el sector tecnológico²². Así pues, se están adelgazando las estructuras de aquellos perfiles que más se necesitan para generar un entorno digital seguro y social, primando la visión

del negocio entendido desde su perspectiva más dura.

Respecto a la segunda opción, parece que queda algo lejos, pues de momento solo existen aproximaciones filosóficas y semillas de ideas normativas, como la de la Unión Europea y su ética de los algoritmos ya mencionada; o los avances reflexivos por parte de la UNESCO en su versión de la Ética de la inteligencia artificial²³. En España, por ejemplo, tenemos ya nuestra Estrategia Nacional de Inteligencia Artificial (ENIA)²⁴, que incluye los aspectos relacionados con el impacto social. Sin embargo, si se quiere explotar esta vía, conviene avanzar más rápido en el diseño de frameworks que puedan ser aplicados sobre la realidad que se nos presenta, determinando aquellos actores que están obligados a intervenir y levantando un sistema de indicadores sólido que facilite el alertamiento temprano para volver los entornos digitales más seguros y respetuosos con la libertad de pensamiento y derechos fundamentales.

Bibliografía

- Coeckelbergh, M. (2020): *AI Ethics*. The MIT Press Essential Knowledge Series. Cambridge, MA, The MIT Press.
- Damasio, A. (2022): *El error de Descartes. La emoción, la razón y el cerebro humano*. Barcelona, Booket.
- Larson, E. (2022): *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*. Barcelona, Shackleton Books.
- Pariser, E. (2017): *El filtro burbuja: cómo la web decide lo que leemos y lo que pensamos*. Barcelona, Taurus.
- Russell, S., Norving, P. (2021): *Artificial Intelligence: A modern Approach, Global Edition*. 4.ª edición. Pearson-Prentice Hall.