

Andrés Ortega

Cómo la nueva Inteligencia Artificial puede manipularte como votante

El País, 17 de septiembre de 2023.

La IA generativa, unida a la ciencia de redes, puede conseguir que los usuarios cambien de opinión casi de forma inconsciente.

No se la llamaba entonces inteligencia artificial (IA), pero la manipulación que puso en marcha la empresa [Cambridge Analytica](#), esencialmente a través de la combinación de redes sociales y datos personales por medio de algoritmos, de los votantes en el referéndum sobre el Brexit —que triunfó— y en las elecciones de 2016 en EE UU, que ganó (no en votos populares sino electorales) Donald Trump, cambiaron un equilibrio interno en las democracias. Esa capacidad de influir —no de decidir— en las preferencias de los votantes con mensajes individualizados adaptados a los gustos de cada uno es solo un primer paso comparado con lo que entraña la simbiosis que está al llegar entre la IA generativa basada en lenguaje natural (como ChatGPT o Bard, entre las principales), con unas redes sociales mucho más ágiles que Facebook (como TikTok) y otros futuros ecosistemas como el Metaverso. ¿Lo veremos ya en 2024 en las diversas elecciones europeas (incluidas a la Eurocámara) o en las cruciales de EE UU? En principio, esta simbiosis va contra la democracia, pero la nueva IA (aceptemos el nombre, aunque no sea propiamente “inteligencia”) aporta también nuevas posibilidades de participación directa de los ciudadanos en las decisiones colectivas.

“La IA generativa, basada en lenguajes naturales, y la ciencia de redes pueden converger en una tormenta perfecta, para dar lugar a nuevas formas de manipulación a nivel societal”, señala Manuel Cebrián, uno de los investigadores más importantes en este campo a lo largo de dos décadas en diversos centros punteros del planeta, ahora en la Universidad Carlos III de Madrid. La capacidad de esta tecnología para adaptarse, aprender y crear, junto con la ciencia de redes que aporta comprensión detallada de la propagación de la información y el cambio social, podría utilizarse para socavar la seguridad, la privacidad y la autonomía a una escala sin precedentes, señala. Lo que va mucho más allá de la política. “Para combatir estos desafíos”, estima Cebrián, “es fundamental desarrollar y adoptar nuevos enfoques”.

En contra

[Esta simbiosis va a generar más capacidad de desinformación.](#) “La IA generativa no solo inundará internet con más mentiras, sino que también puede crear desinformación convincente dirigida a grupos o incluso a individuos”, señala un reciente análisis en *Wired*. El uso de las redes sociales (entonces embrionarias) en las campañas electorales, antes de esta nueva IA, lo empezó Barack Obama en 2008. Le permitió no solo hacer llegar sus mensajes a una parte importante de la población, sino también escuchar esa “conversación pública” que no sabían recoger las encuestas que, en general, le daban como perdedor en su primer y exitoso intento. Algo similar, con las redes sociales ya plenamente desplegadas, ocurrió en las últimas elecciones en España del pasado 23 de julio. Todos los partidos con posibles hacen análisis de redes e intervienen, pero, sorprendentemente, no así los medios de comunicación. La consultora Llorente & Cuenca sí realizó este tipo de estudios. Y así como en la campaña (en un sentido amplio) para los comicios locales y autonómicos del pasado 28 de mayo, el PP logró imponer sus temas en la conversación pública (al final,

contra la *ley del sí es sí*, y los votos comprados), no ocurrió lo mismo en las generales del 23-J, en las que, en la última semana de campaña, dicha conversación fue mucho más emborronada y favorable a los temas del PSOE, que recuperó mucho terreno. Pero había que saber escuchar.

Esto antes de que intervenga la citada simbiosis entre la IA generativa y la ciencia de redes que va a permitir una desinformación personalizada. Si las redes sociales clásicas usan ya la IA para atraer la atención, para que el usuario pase más tiempo en ellas, le saquen más información, y le metan más publicidad, la nueva combinación, con las posibilidades de la IA generativa puede llevar a un objetivo mucho más peligroso: que el usuario cambie de opinión (política, cambiando de voto; de consumo u otras formas) casi de forma inconsciente, con mensajes diseñados para cada cual que cambian de forma casi instantánea según las circunstancias a lo largo de una campaña, adaptándose a cada individuo de forma muy rápida, gracias a lo que se llama *reinforcement learning*, aprendizaje por refuerzo.

Un estudio sobre Facebook de [Sandra González-Bailón](#) y otros autores, publicado en *Science*, concluye que “la derecha está más expuesta a bulos intencionales que la izquierda”. En todo caso, esto dará mucho poder a los que estén detrás, desde un candidato, su oponente o un boicoteador —hay muchos y de varios tipos—, a empresas o a potencias extranjeras con intereses en los resultados.

Falta poco más de un año para las elecciones estadounidenses, y menos de nueve para las europeas. El uso de la IA generativa en las campañas estará en sus comienzos. Pero son numerosos los expertos que alertan contra sus efectos, no solo con enormes cantidades de desinformación y noticias falsas, con capacidades técnicas muy avanzadas (por ejemplo, en vídeos *fake*), que rápidamente se van amoldando a las preferencias de sus destinatarios, e inundando todo el ecosistema de falsedades difícilmente diferenciables de las veracidades, especialmente cuando los medios de comunicación tienen menos capacidades profesionales. Esta herramienta permite que personas con pocos conocimientos técnicos puedan generar esos contenidos, y [reduce la capacidad de los receptores de diferenciar lo que viene de otra persona de lo que viene de una máquina o programa](#). Un reciente experimento en la Universidad de Cornell envió cartas a siete millares de legisladores en EE UU, la mitad generadas por IA y la otra mitad por humanos. El número de respuestas de los legisladores a unas y otras fueron muy similares.

Algunos expertos consideran que hay una aceleración y democratización de la propaganda. Archon Fung y Lawrence Lessig, en un análisis sobre cómo la IA podría tomar el control de las elecciones y socavar la democracia, citan cómo, preguntado en este sentido por el senador Josh Hawley, el propio director ejecutivo de OpenAI (ChatGPT), Sam Altman, respondió en una audiencia parlamentaria que algunas personas podrían utilizar estos modelos lingüísticos para manipular, persuadir y entablar interacciones individuales con los votantes.

En los últimos meses —[ChatGPT va a cumplir un año](#) de su lanzamiento público—, empresas como OpenAI, que la creó, o Google con Bard, han lanzado sus productos al mercado de masas y así han surgido herramientas de IA fáciles de utilizar capaces de crear imágenes fotorrealistas, imitar el sonido de la voz y escribir textos humanos de forma bastante sencilla.

No todos los expertos, sin embargo, coinciden en que la manipulación de las redes genera más polarización. Algunos estudios, con muestras amplias, como el de Andrew Guess y

otros, también publicado en *Science*, concluyen que las personas no informaron de cambios en sus comportamientos políticos después de que se cambiaran lo que recibían por Facebook o Instagram. Pero el estudio es anterior a la IA generativa. La polarización está ya en las sociedades, pero se puede alimentar con estas tecnologías, aunque la televisión — en la que está por ver cómo entra esta herramienta, un proceso que ya ha empezado— es más determinante.

Hay también otra peligrosa simbiosis en marcha, la de la IA y la neurociencia. Es un campo que investigan tanto los servicios de inteligencia de países como EE.UU., como, especialmente, China, y también numerosas grandes empresas que quieren influir a través de la manipulación de nuestras mentes en nuestras decisiones comerciales. Como señala Nita Farahany en su libro *The Battle for Your Brain* (La batalla por tu cerebro), empresas de *marketing* descodifican nuestro cerebro para vendernos productos e incluso adaptar elementos de consumo (como las películas o las series) a los deseos de nuestra mente. Y para controlar. En esta ciencia, China parece más avanzada, y algunos estudios, como el del Centro para una Nueva Seguridad en EE UU, ven un peligro en la difusión global de lo que llama el “autoritarismo digital”, aunque EE UU también controla, como puso de manifiesto el exanalista de la CIA Edward Snowden con sus filtraciones. Son numerosos los gobiernos con medios para llevar a cabo investigaciones en este terreno, o comprar la tecnología necesaria, y aplicarlas. El desarrollo de la IA empodera a los gobiernos, a las empresas con medios y a los individuos. Por estas razones, el neurocientífico español Rafael Yuste ha promovido un nuevo tipo de derechos, que llama “neuroderechos” —véase al respecto la Declaración de Valencia a la que se quiere dar mayor alcance— porque “los datos cerebrales pueden ser uno de los pocos baluartes que quedan contra el compromiso total de la privacidad en la vida moderna”. Aunque está por demostrarse que nuevos derechos o declaraciones impidan este tipo de abusos en el mundo si no van acompañados de técnicas específicas para impedirlos que habrá que desarrollar.

Entre los previsibles efectos nocivos de la IA generativa en modelos amplios de lenguaje natural está que, como lo definen Markus Anderljung y Paul Scharre, se “democratice la violencia”, la capacidad de hacer daño. Ya no solo por parte de *hackers* o codificadores experimentados, sino por todo tipo de gente. La generación de desinformación va a resultar mucho más fácil de conseguir, y la diferencia entre la mentira y la verdad, mucho más difícil de discernir. El concepto de “posverdad” para la era que vivimos se puede quedar corto. Y la fabricación de armas biológicas y químicas por parte de individuos y grupos con escasos conocimientos de ellas también resultará facilitada. ¿Hacia un terrorismo generativo? La IA puede generar catástrofes, y no solo impedir las que es para lo que se suponía que nació.

Incluso en vez de unir, la IA puede separar, justamente por la uniformidad que fomenta. En países básicamente democráticos como India, hay nuevos *chatbots* que fomentan la intolerancia religiosa (de los hindúes contra los musulmanes) y el extremismo.

A favor

La IA aporta grandes ventajas, por ejemplo, en avances en medicina o en el estudio del cambio climático. [No todo con esta nueva tecnología tiene por qué ser negativo para la democracia](#). También brinda nuevas posibilidades a los ciudadanos, para empezar, forzando una nueva transparencia —además de un mayor control— en las administraciones o en las

empresas privadas que sustentan estas nuevas tecnologías (a menudo con apoyo del sector público).

Puede permitir nuevos tipos de participación ciudadana en, por ejemplo, los usos de los fondos públicos. Es algo que se está empezando a estudiar. Así, desde el European Institute of Science in Management, un centro de investigación sin ánimo de lucro con sede en Barcelona, se está intentando poner en marcha con un equipo internacional de politólogos, físicos teóricos, matemáticos, biólogos, neurocientíficos, y economistas de Israel, USA y Europa un programa para enlazar la matemática de la física cuántica con un nuevo tipo de gobernanza democrática multinivel para la que la nueva IA será esencial, con unos primeros ensayos en ciudades europeas. Son proyectos de investigación que permitirían una participación más profunda de la ciudadanía no sólo en la elección de sus representantes, sino también de las prioridades presupuestarias a adoptar en ayuntamientos.

¿Regular es la solución?

[Crecen las demandas, incluso por los propios impulsores de la nueva tecnología, de regular el uso de la IA generativa en las elecciones democráticas](#), aunque hay dudas sobre la efectividad de tales regulaciones. OpenAI se ha comprometido a monitorear el uso de su ChatGPT para tratar de detectar operaciones de influencia política.

Algunos expertos, como los citados Anderljung y Scharre, hablan de la necesidad de “alineación de la IA”, es decir, de asegurar que los objetivos de un sistema de IA coinciden con las intenciones de sus usuarios y los valores de la sociedad. Se podrían prohibir este tipo de programas. Pero estos autores reconocen que “hasta ahora, nadie ha descubierto cómo controlar de forma fiable el comportamiento de la IA”, y menos cuando se prioriza (no es el caso de China) la libertad de expresión, que habrá que reforzar con la libertad de pensamiento, si nos controlan este último como George Orwell entrevió.

La UE lo está intentado con la nueva Ley de IA, la primera del mundo de su tipo, que está pendiente de un acuerdo final entre las instituciones, y cuya efectividad está por ver. El proyecto establece cuatro niveles de riesgo para los sistemas de IA, cada uno con su propio conjunto de requisitos de transparencia, trazabilidad, no discriminación y respeto por los derechos humanos. Los sistemas de IA de alto riesgo, como los que se usan en la toma de decisiones automatizadas con implicaciones legales o significativas para las personas, deben ser sometidos a una evaluación de impacto de la IA antes de su despliegue.

Sobre todo, hay que educar a los ciudadanos, desde los más jóvenes hasta los más mayores, en lo que implican estas tecnologías. Ya casi todo el mundo sabe que lo gratuito, para el usuario, tiene un precio, el de los datos cedidos a cambio, y ahora la capacidad de manipulación de nuestras propias mentes. Mucha IA generativa es gratuita. La alimentamos con nuestras preguntas, imágenes o escritos. Lo gratuito se acaba pagando. De otra forma.

[La Eurocámara aprueba negociar la primera ley sobre inteligencia artificial del mundo](#)

Andrés Ortega es escritor y periodista e investiga sobre el impacto de la IA en las sociedades y en la geopolítica.