

Confluencias (invisibles) entre la inteligencia artificial y la desinformación

La capacidad de ciertos algoritmos para transformar la realidad y la adaptación -casi obsesiva- del comportamiento humano a sus requerimientos profundizan y complejizan la relación entre las nuevas aplicaciones de inteligencia artificial y la producción y propagación de contenidos desinformativos.

En 2016, dos investigadores de la Universidad de Shanghai publicaron un polémico estudio en el que afirmaban haber construido un algoritmo capaz de predecir la conducta criminal a partir de la forma de la cara de los individuos. Para ello, entrenaron al algoritmo con miles de fotografías de personas de origen chino. Las imágenes de los sujetos no delincuentes procedían de los servicios *online* de búsqueda de empleo, mientras que las fotografías de los convictos fueron extraídas de los registros policiales y procedían de documentos oficiales de identidad. De acuerdo con este trabajo, existirían rasgos faciales que caracterizan a los delincuentes frente a aquellos que no lo son. El algoritmo detectó diferencias entre los rostros de ambos tipos de personas en aspectos como la distancia entre los vértices interiores de los ojos, el ángulo entre la nariz y el labio superior y la curvatura del labio superior.

Años más tarde, los profesores de la Universidad de Washington Carl Bergstrom y Jevin West demostraron que ese algoritmo fallaba con mucha frecuencia y que el origen de sus errores se encontraba en la validez de los datos con que fue entrenado. En las fotografías de los sujetos no delincuentes utilizadas para entrenar al sistema aparecían individuos sonrientes porque esas imágenes iban a estar incluidas en el curriculum de estas personas. Por el contrario, las fotos de los delincuentes mostraban rostros totalmente serios porque tales imágenes se tomaron en contextos muy distintos: servirían para estar identificados en registros policiales. Lo que hizo el algoritmo fue detectar patrones de las caras en ambos tipos de personas y vincularlos (erróneamente) con características reales del rostro. El algoritmo detector de conductas delincuentes resultó ser, por tanto, un fiasco.

Una sociedad sometida a una sobrerrepresentación de determinados relatos es una sociedad desinformada

Este ejemplo ilustra una de las grandes preocupaciones de la creciente utilización de algoritmos en las sociedades avanzadas. Estos ingenios matemáticos son alimentados utilizando grandes cantidades de datos, por lo que la calidad y ausencia de sesgos de esta información resulta fundamental para su correcto funcionamiento. Los instrumentos de inteligencia artificial generativa, basados en potentes algoritmos nutridos con información masiva procedente de Internet, están sujetos al principio anteriormente formulado: si se entrenan con datos erróneos pueden resultar problemáticos en un contexto donde cada vez mayor número de personas confían en este tipo de aplicaciones. Si un sistema de creación automática de textos como ChatGPT utiliza como referencia información alojada en la Web, no es de extrañar que sus textos puedan contener datos falsos, dado el inmenso volumen de información errónea, inventada, inexacta y sesgada

producida y alojada en los circuitos digitales.

El espacio de convergencia entre la desinformación y la inteligencia artificial está repleto de luces y sombras. Por un lado, las tecnologías generativas de texto, audio e imágenes son capaces de multiplicar la escala y velocidad de los contenidos falsos. Por otro, las mismas tecnologías algorítmicas ya están siendo aplicadas por las entidades de verificación (*fact-checkers*) para combatir la desinformación (Carrilho Santos, 2023). Sin embargo, además de esta relación visible entre algoritmos y posverdad, existe una confluencia menos evidente entre la inteligencia artificial y la representación falseada de la realidad.

Negociación tácita

Un aspecto clave es la actitud con la que el usuario interactúa con los algoritmos en los espacios digitales. Cuando un individuo es consciente de que sus acciones serán mediadas –y medidas– por un algoritmo, tiende a adaptar su comportamiento a lo que tal algoritmo requiere. Se configura, por tanto, una suerte de negociación tácita entre persona y máquina (Gabelas-Barroso, García-Marín y Aparici, 2023). El uso del *prompt* adecuado para crear justo la imagen sintética que queremos es una habilidad clave para el uso efectivo de este tipo de inteligencia artificial generativa. El sujeto actúa para que el algoritmo obtenga información útil y fácil de procesar a cambio de que este resuelva la necesidad que el individuo pretende solventar. Esta negociación se establece a diario cuando buscamos en Google: colocamos en el cajón de búsquedas las palabras que, bajo nuestra intuición, el algoritmo comprenderá mejor para que el buscador nos ofrezca los resultados pretendidos.

Este principio de negociación tácita produce efectos más allá de las búsquedas cotidianas que cualquier sujeto puede efectuar. Por ejemplo, cuando una industria como la periodística descubre que los algoritmos de los buscadores posicionan mejor las noticias cuando están redactadas y estructuradas de una determinada forma, gran parte de los profesionales de la información utilizarán las mismas estrategias para componer sus textos. El efecto es una preocupante estandarización de las formas compositivas de la información digital a fin de adaptarse a lo que les gusta a los buscadores, convertidos en poderosos *gatekeepers* de la información digital.

A la vez, cuando se detecta que determinados temas se posicionan de forma más visible, se tenderá a producir contenidos sobre tales temáticas y, a la vez, se evitarán las historias con menor capacidad para ser propagadas. Es decir, algunos relatos serán reproducidos masivamente y otros quedarán ocultos. Una sociedad sometida a una sobrerrepresentación de determinados relatos es una sociedad desinformada. La creciente presión por las métricas que inunda ciertas redacciones (un tipo de despotismo numérico), muchas veces acompañada por incentivos para aquellos periodistas cuyas piezas obtienen más clics, ahonda en esta problemática, que puede tener efectos transformadores a partir de la acción performativa de estas tecnologías algorítmicas.

Performatividad algorítmica

En *La condición posmoderna* (2006), el filósofo francés Jean-François Lyotard define los relatos performativos como aquellos que, además de enunciar características de objetos o situaciones, son capaces de transformar la realidad. Cuando un líder político enuncia una declaración de guerra contra un país enemigo, no solo produce un hecho informativo de máxima relevancia, sino que también está modificando el contexto de las relaciones entre ambas naciones. Cuando el rector de una universidad declara que el curso universitario queda inaugurado, su discurso no solo anuncia denotativamente la apertura del año académico, sino que provoca de facto que el curso comience. El relato performativo, por tanto, “tiene la particularidad de que su efecto sobre el referente coincide con su enunciación: la universidad queda abierta puesto que se la declara tal en esas condiciones” (Lyotard, 2006, p. 11).

La desinformación es un viejo problema que se ha reconfigurado en la época actual

Al algoritmo –y, por extensión, a los complejos sistemas de inteligencia artificial– les podemos atribuir tales capacidades performativas. Los algoritmos de los buscadores que operan *online* no solo ofrecen respuestas a nuestras búsquedas. No solo miden y jerarquizan el grado de coincidencia entre lo que buscamos y las páginas que encuentran *online*, sino que contribuyen a transformar la realidad que opera alrededor de tales búsquedas. No olvidemos que lo que hacen los buscadores es producir un listado de páginas y/o contenidos –es decir, un *ranking*– a propósito de un *input* ofrecido por el usuario. Cuando un buscador coloca en primera posición un determinado relato (o una empresa determinada), sus algoritmos están provocando que ese relato (o esa empresa) se convierta en el más visible (y tenga, por tanto, mayor popularidad y probabilidad de éxito) solo porque el buscador lo coloca como el más visible. Se produce, por tanto, una profecía autocumplida donde opera de nuevo la lógica dicotómica del posicionamiento y la ocultación de los contenidos: hay discursos que terminan muy visibilizados y otros que quedan silenciados en los circuitos digitales. No olvidemos que la capacidad de hacer que los contenidos sean fáciles de encontrar resulta esencial en unos entornos virtuales dominados por la saturación informativa y la lucha constante por el limitado tiempo de atención de los usuarios (Webster, 2014).

La desinformación es un viejo problema que se ha reconfigurado en la época actual como consecuencia de varios factores, entre ellos las tecnologías digitales. Aunque no podemos situar toda la responsabilidad en el efecto que los algoritmos tienen en la propagación de la información falsa ni en la consiguiente representación sesgada del mundo, es urgente analizar de forma crítica las consecuencias derivadas de la colonización del algoritmo en un creciente volumen de aplicaciones de nuestra cotidianeidad, así como sus efectos en la percepción de la realidad. No se trata de demonizar al algoritmo, ya que no deja de ser un instrumento cuyas implicaciones dependen del uso que hacemos de él. Tampoco debemos olvidar que estos sistemas presentan indudables ventajas, ya que facilitan infinidad de labores por su capacidad y velocidad en el procesamiento de ingentes cantidades de datos.

Sin embargo, resulta determinante que, tal como numerosos expertos en inteligencia artificial han recomendado, nos paremos a reflexionar sobre aspectos esenciales que repercuten en sus posibles efectos beneficiosos: cómo son producidos, cómo se pueden hacer transparentes para la sociedad, cómo pueden reconfigurar las nociones de autoría y propiedad intelectual, en qué medida impactan en la privacidad de los usuarios, qué efecto medioambiental tiene la base tecnológica sobre la que operan y cómo condicionan la representación social y cultural del mundo. Cómo pueden contribuir a mejorar la sociedad, hacerla más eficiente en el manejo de la información y, por consiguiente, producir una ciudadanía mejor formada –y menos desinformada– sobre las cuestiones relevantes que impactan en su vida. Es urgente una aproximación reflexiva y alejada de la espuma del titular apresurado para conocer mejor cómo producir y utilizar estas tecnologías algorítmicas. Para abrir sus cajas negras y entender su funcionamiento. Para evitar, en definitiva, confiar ciegamente en un prometedor sistema de detección de delincuentes cuando, en realidad, solo es un irrelevante detector de sonrisas.

Carrilho Santos, F.C.: «Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis» en *Journalism and Media*, 2023. Disponible en: <https://doi.org/10.3390/journalmedia4020043>

Gabelas-Barroso, J.A., García-Marín, D. y Aparici, R. (2023): *La invasión del algoritmo*. Barcelona, Gedisa.

Lyotard, J.F. (2006): *La condición posmoderna*. Madrid, Cátedra.

Webster, J.G. (2014): *The Marketplace of Attention. How Audiences Take Shape in a Digital Age*. Cambridge, The MIT Press.