

Algoritmos a la caza de fake news



La desinformación y las *fake news* son uno de los grandes problemas de nuestro tiempo. La capacidad que ofrece la tecnología para la rápida difusión de bulos presenta una seria amenaza para el buen funcionamiento de las democracias e incluso del orden mundial. Sin embargo, la solución puede estar ligada a la propia tecnología. Los algoritmos de inteligencia artificial están empezando a mostrar su eficacia en la detección de noticias falsas.

De acuerdo con un informe de la consultora Gartner, para el año 2022 la mayor parte de los ciudadanos de las economías desarrolladas consumirá más información falsa que verdadera. Las *fake news* -noticias falsas o bulos- se han convertido en un serio problema que afecta a la sociedad digital y constituyen una amenaza de primer orden para el orden geopolítico mundial.

Quizá el gran problema no sea tanto la capacidad que aporta la tecnología para diseminar desinformación, sino la falta de cultura digital de las personas para distinguir entre lo que es información fiable y veraz, y lo que supone un intento de manipulación de la opinión y hasta de la voluntad de actuar.

El trabajo de OpenMind-BBVA *La era de la perplejidad* arroja el dato preocupante de que, en el Reino Unido, solamente una cuarta parte de los niños entre ocho y once años puede identificar un anuncio o un enlace patrocinado, entre las distintas publicaciones recibidas a través de las redes sociales. Y en el caso de los adultos, solo el 60% es capaz de explicar la diferencia entre ambas.

La cuestión es que hemos llegado depender en exceso de la información que recibimos a través de internet, sin considerar que es un medio que no establece exigencias o condiciones para publicar (en teoría, cualquiera puede subir contenido a redes sociales o blogs) y que no tiene una autoridad central que pueda vetar lo que allí se cuelga.

¿Por qué tendemos a creernos las noticias falsas?

De acuerdo con las teorías del proceso dual, la mente pone en marcha dos procesos mientras lee o recibe información, uno automático y superficial, y otro que requiere esfuerzo y concentración, que se utiliza para tomar decisiones estratégicas. Cuando hacemos uso del proceso superficial, juzgamos automáticamente la veracidad de la información sobre criterios como lo familiar que es para nosotros o lo fácil que es de comprender. De esta forma, cuanto más fluidamente procesamos la información, más familiar nos resulta y más tendemos a tomarla por verdadera.

Esta fluidez con la que asumimos determinada información conlleva a menudo un efecto colateral, que hace que la corrección o refutación de la información falsa nos lleve a creer aún con más fuerza la mentira. Un ejemplo de ello es que todavía existe entre un 20% y un 30% de los norteamericanos que sigue creyendo que Irak escondía armas de destrucción masiva, aunque la invasión del país en 2003 demostró lo contrario.

Dada nuestra psicología como seres humanos, la ingente cantidad de información que circula por las redes y el hecho demostrado que los bulos se difunden mucho más rápido que las noticias reales, resulta muy difícil ponerle coto al fenómeno de las *fake news*.

Una posible solución a esta preocupante cuestión es utilizar la inteligencia artificial para poder discernir entre lo que es información y la tergiversación de la realidad. Actualmente ya se están llevando a cabo experiencias con algoritmos que son capaces de analizar a gran velocidad montañas de noticias y comunicados, y de identificar, con un elevado porcentaje de éxito, aquellas que no son más que bulos.

Pero, por desgracia, estas mismas herramientas también están demostrando su eficacia para trabajar para el enemigo. Hace unos días saltaba la noticia de que un equipo de investigadores de OpenAI ha conseguido crear una máquina que redacta automáticamente *fake news* bastante convincentes.

¿Fake news o desinformación?

El término fake news ha saltado a primera plana en los últimos años a raíz de la manipulación de la opinión pública y del voto en las elecciones de Estados Unidos de 2016, y también en el referéndum del *Brexit* celebrado en el Reino Unido. El escándalo protagonizado por la empresa Cambridge Analytica, que hizo un uso fraudulento de los datos personales de millones de usuarios de Facebook, volvió a avivar su protagonismo el pasado año.

No obstante, no todos aprueban el uso de la denominación de noticias falsas para referirse al fenómeno y hay quien lo considera muy restrictivo e insuficientemente descriptivo del problema de fondo. Es el caso de la Comisión Europea, que prefiere hablar de *desinformación*.

En concreto, define la desinformación como “*información falsa, inexacta o engañosa, diseñada, presentada o promovida para causar intencionadamente un daño público o para obtener un beneficio*”. Para la Comisión, la expresión *fake news* no es adecuada, porque no abarca la complejidad del problema.

De hecho, a menudo se trata de contenido que no es falso, o que no es completamente falso, pero que es información fabricada, mezclada con hechos y prácticas que poco tienen que ver con el concepto de noticia, como pueden ser cuentas automáticas en redes sociales utilizadas para hacer *astroturfing* (disfrazar las acciones de una entidad política o comercial como la reacción pública espontánea), el uso de redes de seguidores falsos, los vídeos manipulados, la publicidad dirigida, los *trolls* organizados o los *memes* visuales. En resumen, se trata de todo un abanico de prácticas para manipular la opinión pública en internet, que van más allá de lanzar una noticia falsa.

La velocidad de la mentira

Paradójicamente, las noticias falsas o *fake news* se viralizan en las redes sociales mucho más rápidamente que la información veraz y contrastada. Es algo que ha podido demostrar un reciente estudio de MIT Initiative on the Digital Economy, que analizó, entre 2006 y 2017, en torno a 126.000 hilos de noticias en Twitter, tuiteados más de 4,5 millones de veces por unos 3 millones de personas.

Los resultados fueron desalentadores. En palabras de los autores, la verdad tarda aproximadamente seis veces más que la mentira en alcanzar a 1.500 personas. En suma, los contenidos falsos se difunden significativamente más lejos, más rápido y más profundamente en los hilos y cascadas de conversaciones, que los verdaderos.

Entre todas las categorías de bulos, los relacionados con la política son los que alcanzan mayor difusión, por encima de los relacionados con el terrorismo, los desastres naturales, la ciencia, la información financiera o las leyendas urbanas.

El hecho de que las fake news presentan una probabilidad de ser retuiteadas un 70% superior puede tener que ver, según el estudio, a que se perciben como más novedosas que las reales. La gente tiende a difundir la

novedad en mayor medida que lo ya conocido.

Finalmente, y en contra de lo que se piensa, el análisis realizado demostró que los *bots*, los perfiles automáticos de Twitter, aceleran las noticias falsas y verdaderas en la misma proporción, lo que implica que somos los humanos -y solo nosotros-, los únicos responsables de la proliferación de las *fake news*.

Algoritmos cazadores de bulos

La caza de noticias falsas es una tarea ardua y compleja. El inmenso caudal de información que llega a los portales agregadores de contenidos y que circula por las redes sociales hace muy difícil que los editores humanos puedan verificar rápidamente una determinada noticia, especialmente cuando se trata de una historia nueva. Ocurre con frecuencia que, cuando se consigue demostrar que una noticia es falsa, el daño que produce ya ha tenido lugar.

Las máquinas pueden aportar rapidez y eficiencia a la tarea de cazar bulos. En concreto, nos pueden superar en el análisis de los atributos cuantificables de la noticia, como la estructura gramática, la elección de palabras, la puntuación y la complejidad del texto.

No obstante, el verdadero desafío para construir un buen detector de fake news no es tanto el diseño del algoritmo, sino encontrar los datos adecuados para alimentarlo y entrenarlo en suficiente cantidad. Pero los bulos aparecen y desaparecen con rapidez, y resulta complicado recopilarlos para poderse los mostrar a las máquinas.

Un equipo de investigación de la Universidad de Michigan ha creado un algoritmo cazador de noticias falsas que ha demostrado hacerlo mejor que los humanos: ha conseguido identificar *fake news* con un 76% de éxito, frente al 70% de los cazadores humanos.

Otro ejemplo de detección de noticias falsas a través de la inteligencia artificial es el sistema basado en aprendizaje profundo que ha desarrollado la startup británica Fabula. En este caso, la identificación del bulo no se produce a través del análisis del texto, sino estudiando cómo se comparten las historias, para reconocer patrones de difusión que únicamente pueden corresponder a *fake news*.

Por último, el MIT's Computer Science and Artificial Intelligence Lab (CSAIL), en colaboración con el Qatar Computing Research Institute (QCRI), ha realizado una tercera aproximación a este tema, en este caso centrando la atención en las fuentes de las noticias. El sistema que han desarrollado utiliza el aprendizaje automático (*machine learning*) para determinar la exactitud de una fuente de información e identificar si está políticamente sesgada.

La tesis que subyace en esta experiencia es que, si una web ha publicado fake news en el pasado, es muy probable que lo vuelva a hacer. De esta forma, el algoritmo recopila datos de estos sitios sospechosos y los modeliza, para poder identificar con este patrón a otras páginas que vayan a publicar bulos por primera vez. Este rastreador solamente necesita analizar unos 150 artículos para poder determinar con fiabilidad si una fuente de información es o no de confianza.

Cuando la inteligencia artificial se pasa al enemigo

Este febrero los medios especializados se han hecho eco de un inquietante proyecto: el instituto de investigación OpenAI de San Francisco, una institución fundada por Elon Musk, ha construido un algoritmo capaz de redactar *fake news*.

El objetivo del proyecto era otro, se trataba de desarrollar un sistema que, entrenado con grandes cantidades

de texto de la web, fuese capaz de traducir textos y responder a preguntas, entre otras acciones. Pero los responsables pronto se dieron cuenta del potencial que presentaba esta inteligencia artificial para redactar textos falsos y hacerlo de manera creíble.

Will Knight reproduce un ejemplo del funcionamiento del algoritmo en su artículo de MIT Technology Review. Al sistema se le dio la frase: *“Russia has declared war on the United States after Donald Trump accidentally ...”* (Rusia ha declarado la guerra a Estados Unidos después de que Donald Trump accidentalmente...). Y el algoritmo elaboró él solo la continuación de la “noticia”:

Russia has declared war on the United States after Donald Trump accidentally fired a missile in the air.

Russia said it had “identified the missile’s trajectory and will take necessary measures to ensure the security of the Russian population and the country’s strategic nuclear forces.” The White House said it was “extremely concerned by the Russian violation” of a treaty banning intermediate-range ballistic missiles.

The US and Russia have had an uneasy relationship since 2014, when Moscow annexed Ukraine’s Crimea region and backed separatists in eastern Ukraine.

Los responsables del proyecto muestran su preocupación de que máquinas de este tipo –en un futuro cercano mucho más sofisticadas– puedan ser utilizadas para desinformar y manipular, generando noticias falsas y bulos tan perfectos que resulten muy difíciles de identificar.

Pero Richard Socher, un experto en este campo y jefe científico en la empresa de software Salesforce, no se muestra tan preocupado, dado que considera que *“no necesitas inteligencia artificial para crear fake news. Las personas lo hacen con facilidad”*.

Photo by [rawpixel.com](https://www.rawpixel.com) from Pexels

Conner-Simons, A. (2018) “Detecting fake news at its source” en *MIT News*. Disponible en: <http://news.mit.edu/2018/mit-csail-machine-learning-system-detects-fake-news-from-source-1004>

European Commission (2018) “A multi-dimensional approach to disinformation”. Disponible en: <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

Gartner (2017) “Top Strategic Predictions for 2018 and Beyond” Disponible en: <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/>

NVIDIA Developer (2019) “Fabula AI Develops A New Algorithm to Stop Fake News”. Disponible en: <https://news.developer.nvidia.com/fabula-ai-develops-a-new-algorithm-to-stop-fake-news/>

OpenMind (2018) “La era de la perplejidad”. Penguin Random House Grupo Editorial. Disponible en: <https://www.bbvaopenmind.com/libros/la-era-de-la-perplejidad/>

Science Daily (2018) “Fake news detector algorithm works better than a human”. Disponible en:

<https://www.sciencedaily.com/releases/2018/08/180821112007.htm>

Velautham, L. (2018) "Fake news?" en *Berkeley Science Review*. Disponible en: <http://berkeleysciencereview.com/article/fake-news/>

Vosoughi, S. Roy, D. y Aral, S. (2018) "The spread of true and false news online" en *MIT Initiative on the Digital Economy*. Disponible en: <http://ide.mit.edu/sites/default/files/publications/2017%20IDE%20Research%20Brief%20False%20News.pdf>

Will, K. (2019) "An AI that writes convincing prose risks mass-producing fake news" en *MIT Technology Review*. Disponible en: <https://www.technologyreview.com/s/612960/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish>