ieee.es
Instituto Español de Estudios Estratégicos

Documento
Opinión

74/2018 19 de junio de 2018

*Manuel R. Torres Soriano**

Operaciones de influencia e inteligencia artificial: una visión prospectiva

Operaciones de influencia e inteligencia artificial: una visión prospectiva

Resumen

El propósito de este documento es analizar desde una visión prospectiva cómo la irrupción de la inteligencia artificial va a transformar las operaciones de influencia y su recepción por parte de la sociedad. Se presta atención a los posibles efectos de la pérdida de centralidad de los humanos en estas operaciones, la devaluación de la percepción sensorial como criterio para discriminar la verdad de la mentira, los sesgos en las plataformas de aprendizaje autónomo y la creciente hibridación entre el marketing digital y la propaganda política.

Palabras clave

Internet, operaciones encubiertas, ciberinteligencia, propaganda, comunicación política.

Influence operations and artificial intelligence: a forward-looking vision

Abstract

The purpose of this document is to analyze from a prospective perspective how the irruption of Artificial Intelligence will transform influence operations and its reception by society. Attention is paid to the possible effects of the loss of centrality of humans in these operations, the devaluation of sensory perception as a criterion to discriminate the

NOTA:** Las ideas contenidas en los ***Documentos de Opinión son de responsabilidad de sus autores, sin que reflejen, necesariamente, el pensamiento del IEEE o del Ministerio de Defensa.

truth of the lie, the biases in autonomous learning platforms and the growing hybridization between digital marketing and political propaganda.

Keywords

Internet, covert operations, cyber-intelligence, propaganda, political communication.

Introducción

El debate sobre la inteligencia artificial ha trascendido el ámbito de la investigación científica para situarse en la agenda política. Su importancia se puede medir por las extraordinarias cantidades de dinero que algunos países están dedicando a su desarrollo, con el objetivo de adquirir una ventaja competitiva[1]. Pero, sobre todo, se puede visibilizar en su creciente protagonismo dentro de la rivalidad que mantienen las principales potencias internacionales. En 2016, el presidente norteamericano Barack Obama la identificó como una extraordinaria fuente del poder blando de su país[2]. El presidente ruso Vladimir Putin afirmó «quienquiera que se convirtiera en el líder en este campo, dominará el mundo»[3], y el presidente francés Enmanuelle Macron utilizó las páginas de la revista tecnológica más influyente para anunciar al mundo[4] la determinación de Francia de liderar la investigación en inteligencia artificial (AI, por sus siglas en inglés).

El atractivo de esta tecnología reside, no solo en su capacidad de capturar gran parte del valor económico generado por las cadenas de producción que va a transformar en los próximos años, sino por la posibilidad de convertir innovación científica en poder político. La irrupción de los sistemas alimentados por AI en el ámbito de la seguridad y defensa, pero también en el ámbito de los flujos de información, plantea numerosos escenarios hipotéticos donde se reescriben las reglas y las capacidades de los diferentes actores que participan en la pugna política.

Uno de los ámbitos donde más fácilmente se atisba el poder disruptivo de la AI es el de las operaciones de influencia política. Se entiende por estas, al conjunto de interferencias ejercidas por actores estatales ajenos a la comunidad que ostenta la soberanía, con el propósito de inclinar los resultados del proceso político hacia un resultado que favorezca sus intereses. Este tipo de medidas encontraron su

[1] VILLANI Cedric *et al*. «For a Meaningful Artificial Intelligence. Towards a French and European Strategy», *Mission assigned by the Prime Minister Édouard Philippe* (2018), disponible en: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf Fecha de la consulta 07.5.2018.

[2] WIRED. «The President in Conversation With MIT's Joi Ito and WIRED's Scott Dadich», *Wired* (24/8/2016), diponible en: https://www.wired.com/2016/10/president-obama-mit-joi-ito-interview/ Fecha de la consulta 07.5.2018.

[3] GIGOVA Radina. «Who Vladimir Putin thinks will rule the world», *CNN* (2/9/2017), disponible en: https://edition.cnn.com/2017/09/01/world/putin-artificial-intelligence-will-rule-world/index.html Fecha de la consulta 07.5.2018.

[4] THOMPSON Nicholas «Emmanuel Macron Talks to WIRED About France's AI Strategy», *Wired,* (31/3/2018), disponible en: https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/ Fecha de la consulta 07.5.2018.

legitimización dentro del contexto histórico de la Guerra Fría donde, tanto Estados Unidos como la Unión Soviética, percibieron que la celebración de elecciones en terceros países podía ser una oportunidad para ampliar su área de influencia y debilitar la del adversario. El fin de este periodo, lejos de suponer su desaparición, marco el punto de inicio de una readaptación al nuevo contexto geopolítico y la disponibilidad de nuevas herramientas tecnológicas[5]. La eclosión del ciberespacio, como un nuevo dominio donde se proyecta el poder estatal, ha proporcionado a las operaciones de influencia una nueva «edad de oro». La injerencia rusa en las elecciones presidenciales de Estados Unidos de 2016, otorgó un protagonismo sin precedentes a este tipo de operaciones encubiertas[6], y alimentó el debate sobre su capacidad real para condicionar y pervertir el proceso electoral.

Uno de los aspectos menos estudiados de estos episodios es la centralidad de la tecnología en la génesis de las operaciones de influencia. Cada acción en el ámbito del ciberespacio genera su correspondiente contramedida. Lo que hace efectiva una operación deja de resultar útil poco tiempo después. El resultado es un acelerado ciclo de innovación y adaptación, que sitúa a la tecnología y sus potencias usos tácticos en el centro de este debate.

El propósito de este artículo es reflexionar desde una perspectiva prospectiva sobre cómo la irrupción de la inteligencia artificial va a transformar las operaciones de influencia, y su posible recepción por parte de la sociedad. La bibliografía reciente sobre prospectiva de seguridad y defensa tiende a coincidir en la existencia de cuatro tendencias principales tras el rápido progreso de esta tecnología[7]:

1. La acumulación de varias décadas donde la capacidad de procesamiento informático ha crecido de manera exponencial.
2. La disponibilidad de grandes cantidades de datos con los cuales alimentar el proceso de aprendizaje autónomo de estos sistemas.

---

[5] TORRES Manuel R. «Hackeando la democracia: operaciones de influencia en el ciberespacio», *IEEE Documento de Opinión* (19/6/2017), disponible en: http://www.ieee.es/Galerias/fichero/docs_opinion/2017/DIEEEO66-2017_Hackeando_democracia_MRTorres.pdf Fecha de la consulta 07.5.2018.

[6] HARDING Luke. *Conspiración. Cómo Rusia ayudó a Trump a ganar las elecciones*, Barcelona, Debate, 2017.

[7] ALLEN Greg & CHAN Taniel. «Artificial Intelligence and National Security», *Belfer Center for Science and International Affairs*, (July 2017), disponible en: https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf Fecha de la consulta 07.5.2018.

3. Los avances en la implementación de técnicas de *Machine Learning*.
4. El aumento de la inversión económica en investigación y desarrollo.

Estas grandes tendencias se ven complementadas en un futuro cercano con una serie de factores más específicos que tienen un impacto directo en las operaciones de influencia política.

La democratización del ciberataque

La capacidad de llevar a cabo un ciberataque sofisticado está constreñida por el elemento humano[8]. Los equipos de ataque que se encargan de detectar y explotar vulnerabilidades informáticas (como los que se hicieron con los correos electrónicos de cargos clave de la campaña demócrata en Estados Unidos) deben ser lo suficientemente numerosos para abarcar los distintitos conocimientos y habilidades que hacen posible este resultado. Pero también deben trabajar de manera prolongada sobre un objetivo, a la espera de que este cometa el error que termine exponiéndolo. La irrupción de la AI irá desplazando a las personas como el elemento central de este proceso. La automatización de las labores de reconocimiento de redes, testeo y desarrollo de vectores de ataque, no solo se producirá a una velocidad inasumible para un equipo «humano», sino que también será capaz de abarcar a un número de objetivos prácticamente ilimitado. La posibilidad de que un actor (estatal o no estatal) pueda dotarse de la capacidad de implementar un ciberataque sofisticado habrá dejado de ser algo que se explica por su capacidad para movilizar y organizar a las personas dotadas del talento necesario, para convertirse en una cuestión de sí dispone de los recursos suficientes y el acceso a los mercados donde poder adquirir una AI de carácter ofensivo.

El cambio de paradigma hará viable que algunos actores se doten, por primera vez en la historia, de capacidades reales de monitorización masiva de los ciudadanos. El régimen comunista de Alemania del Este, el cual ha sido percibido como el ejemplo más elaborado de un Estado volcado en el espionaje hacia sus propios habitantes[9], llegó a tener en su momento álgido a 102.000 agentes dedicados exclusivamente al

[8] TORRES, Manuel R. «El dilema de interpretación en el ciberespacio», *IEEE Documento de Opinión* (8/1/2018), disponible en: http://www.ieee.es/Galerias/fichero/docs_opinion/2018/DIEEEO03-2018_Dilema_Ciberespacio_ManuelRTorres.pdf Fecha de la consulta 07.5.2018.
[9] SCHNEIER Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*, New York, W. W. Norton & Company, 2016.

espionaje de los 17 millones de habitantes del país, lo que supone un agente de la *Stasi* por cada 166 ciudadanos. En un futuro próximo un sistema de monitorización intensiva capaz de abarcar de manera simultánea a miles de millones de personas podría estar administrado por un equipo humano de unos pocos miles de empleados.

La devaluación del factor humano tendrá un profundo impacto en la balanza de poder entre Estados. La primacía en el desarrollo de la inteligencia artificial, y su correspondiente traslación en poder político, militar y económico, puede situar como protagonistas de la competición geoestratégica a países que, poco tiempo atrás, habían desempeñado un papel secundario en la escena internacional, debido a sus limitaciones poblacionales, geográficas o presupuestarias.

El colapso de la realidad

Uno de los ámbitos donde se está produciendo con mayor rapidez el progreso de la AI es en la generación de contenido. En la actualidad existen mecanismos que, pese a estar aún en fase de maduración, ya permiten, por ejemplo, la creación de una representación en tres dimensiones a partir de una imagen (como una cara) basada en una o más imágenes bidimensionales; o la producción automatizada de efectos de sonido realistas para acompañar un vídeo que carece de audio. Sin embargo, la innovación más perturbadora es la edición automatizada de vídeo y audio con la creación de expresiones faciales, sonidos y movimientos de labios realistas a partir de los rostros procedentes de vídeos preexistentes. Esto permite generar nuevos productos donde se puede situar de manera impostada a personas hablando o actuando según el deseo del manipulador. Con la tecnología actual, es necesario contar con varias horas de grabaciones procedentes del objetivo para alimentar el sistema con toda la gama de matices faciales y expresiones propias de la persona que será suplantada[10]. Esto convierte a las personas populares o especialmente activas en redes sociales de internet en objetivos especialmente vulnerables, debido a la facilidad con la que se puede acceder a cientos de horas de vídeo donde ellos mismos se

[10] CB INSIGHTS. «Memes That Kill: The Future of Information Warfare», *Research Briefs* (3/5/2018), disponible en: https://www.cbinsights.com/research/future-of-information-warfare/?utm_source=CB+Insights+Newsletter&utm_campaign=942ceba654-Top_Research_Briefs_05_05_2018&utm_medium=email&utm_term=0_9dc0513989-942ceba654-90301433 Fecha de la consulta 07.5.2018.

encargarán de entrenar a la AI. La abundancia de datos no solo permite hacer más sofisticado el producto final, sino también extender este tipo de manipulaciones hacia imágenes que se están generando en tiempo real.

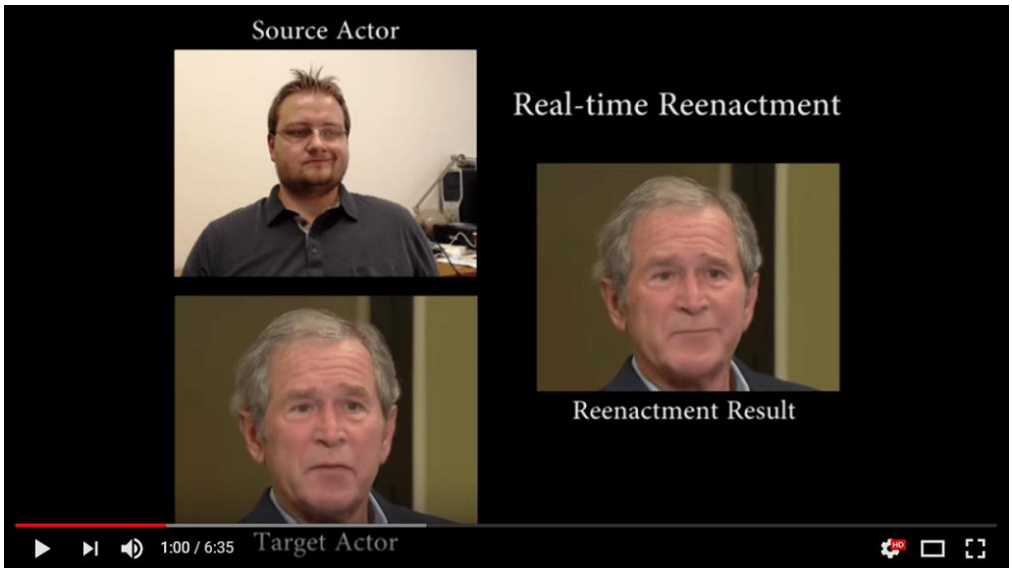

Figura 1: Ejemplo de manipulación de vídeo en tiempo real.
Fuente: https://www.youtube.com/watch?v=ohmajJTcpNk

La generación de contenido a partir de AI no se agota en el ámbito de la imagen, sino que abarca a la propia palabra escrita, emulando estilos de escritura. Ya es posible encontrar sistemas capaces de redactar artículos de noticias basados en datos estructurados, como los que se pueden obtener de encuestas, resultados de elecciones, información financiera o resultados deportivos.

El resultado obvio es que la mentira será cada vez más indistinguible de la propia realidad. Las imágenes de calidad generadas por ordenador dejarán de ser una tecnología cara, al alcance solamente de las grandes empresas del sector audiovisual[11]. En un futuro cercano será una posibilidad al alcance de cualquier ciudadano sin conocimientos especializados, lo que multiplicará el número de actores qua harán uso de la falsificación de imágenes en diferentes órdenes de la vida, incluyendo el activismo político. Los medios de comunicación experimentarán la avalancha de contenido potencialmente escandaloso cuya autenticidad será *a priori* difícil de establecer. Resulta fácilmente imaginable el atractivo que tiene para la

[11] ALLEN Greg. «Artificial Intelligence Will Make Forging Anything Entirely Too Easy», *Wired* (30/6/2017), disponible en: https://www.wired.com/story/ai-will-make-forging-anything-entirely-too-easy/ Fecha de la consulta 07.5.2018.

generación de falsos escándalos políticos el uso de los llamados *deepfakes*, término con el que se conoció a finales de 2017 la aparición en internet de una serie de vídeos pornográficos, generados a través de una sencilla aplicación informática que permitía sustituir el rostro de las protagonistas originales por el de algunas populares actrices estadounidenses. El resultado era tan perturbadoramente realista que, tras la repercusión inicial, las principales plataformas de internet (incluyendo los repositorios de vídeos pornográficos) se comprometieron a bloquear la aparición en sus servidores de este tipo de falsificaciones[12].

Figura 2: Ejemplo de la creciente sofisticación de imágenes generadas a través de aprendizaje automático no supervisado (Redes Generativas Antagónicas).
Fuente: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

La pérdida de la fe en la percepción sensorial como criterio válido para discriminar la verdad de la mentira, provocará lo que algunos autores han denominado un «colapso de la realidad»[13]. Este fenómeno tendrá su máximo exponente en los sistemas de «realidad virtual» cuyo objetivo es generar en el usuario una ilusión inmersiva de encontrarse en otro lugar. Las investigaciones desarrolladas con sistemas que aún no son capaces de proporcionar en el usuario una experiencia comprensiva, ya han detectado la enorme potencialidad de estas realidades alternativas para la manipulación de la conducta.

Ante la avalancha de imágenes falsas (pero visualmente convincentes) donde se puede apreciar a políticos aceptando sobornos, soldados perpetrando crímenes de guerra, y personalidades protagonizando actos sexuales aberrantes, el espectador dejará de

[12] MATSAKIS Louise. «Artificial Intelligence Is Now Fighting Fake Porn», *Wired* (14/2/2018), disponible en: https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes/ Fecha de la consulta 07.5.2018.
[13] FOER Franklin. «The Era of Fake Video Begins», *The Atlantic* (May 2018), disponible en: https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/ Fecha de la consulta 07.5.2018.

confiar en lo que ven sus ojos, para apostar por otro tipo de criterios de carácter abstracto. Así, por ejemplo, la organización Amnistía Internacional ha tenido que recurrir a múltiples procedimientos de carácter técnico para tratar de discernir los contenidos auténticos sobre violaciones de derechos humanos remitidos por los internautas, de aquellas otras manipulaciones interesadas que intentan instrumentalizar a esta organización. A través de su *Citizen Evidence Lab,* la organización se ha especializado en escudriñar el contexto que rodea las imágenes que recibe, para lo cual utiliza aplicaciones como *YouTube Data Viewer*, *Google Earth* y el buscador *Wolfram Alpha* para verificar los metadatos del archivo y cruzar la información sobre la ubicación y las condiciones meteorológicas del suceso para verificar si estas coinciden con el relato del material analizado.

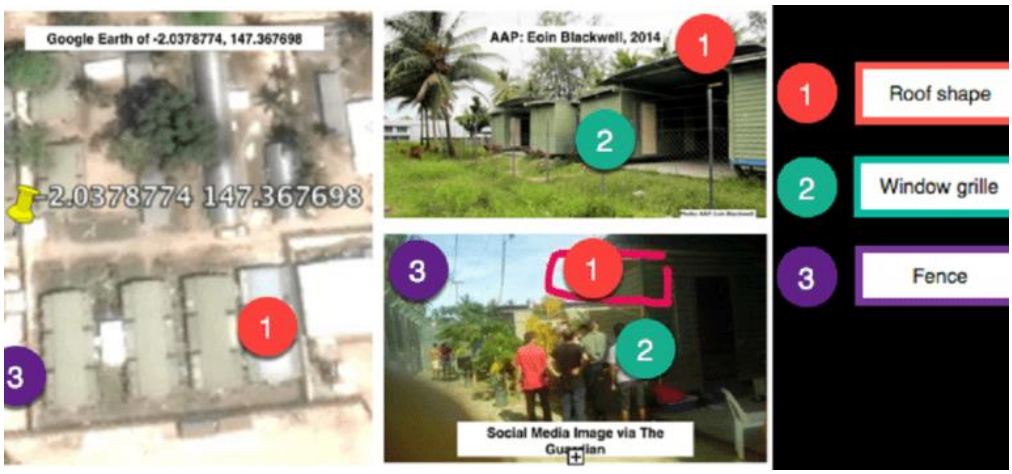

Figura 3: Ejemplo de la triangulación de los detalles contenidos en un vídeo remitido a Amnistía Internacional por un internauta sobre un tiroteo en Papua Nueva Guinea.
Fuente: https://www.amnesty.org/download/Documents/ASA3461712017ENGLISH.pdf

Sin embargo, la creciente sofisticación del contenido generado a través de AI hará inevitable que los únicos sistemas de validación realmente fiables terminen siendo los de naturaleza técnica, particularmente los basados en criptografía, como por ejemplo el uso de *blockchain*, para trazar con garantías la línea que une al mensaje con su emisor.

La percepción de lo que es real y lo que no, tendrá que ver más con una confianza incondicional en la fuente de distribución del contenido, y no tanto en el juicio individual. No obstante, la transición entre ambos paradigmas no será rápida, ni exenta de resistencia por parte de una ciudadanía acostumbrada a situar en la cima de la

veracidad aquello que «puede ver con sus propios ojos». La desinformación seguirá haciendo uso de esta resistencia a desconfiar de los sentidos, y seguirá teniendo impacto en una opinión pública que, a pesar de ser consciente del carácter ficticio de este tipo de productos, estará dispuesta a ser sugestionada por el poder de la imagen y la racionalización de que tras una serie de imágenes falsas puede existir una «verdad subyacente».

La reacción defensiva frente a este tipo de contenidos fraudulentos puede incidir también negativamente en la confianza pública hacia el ecosistema informativo. Así, por ejemplo, cuando Google anunció[14] que como respuestas a la difusión de *fake news* procedentes de Rusia su buscador iba a ser modificado para reducir la prevalencia de los contenidos elaborados por medios estatales rusos como *Russian Today* y *Sputnik*, algunos internautas aplaudieron la medida, mientras que otros la denunciaron como una forma de censura posmoderna, donde una gran empresa decidía qué fuentes de información estaban permitidas y cuáles debían ser arrojadas a la irrelevancia.

La pérdida de confianza en los entornos online aportará una ventaja estratégica para los actores políticos que ideológicamente se presentan como alternativas de ruptura radical con el *establishment*, los cuales ganarán predicamento entre una audiencia cada vez más escéptica que ya no hace distinciones entre fuentes oficiales y relatos «alternativos» de los hechos.

Dieta de datos

Las técnicas de aprendizaje automático (*Machine Learning*) suponen una ruptura con la forma en la que hasta el momento se ha diseñado el software. La AI está provocando una transición entre el enfoque basado en la programación, a uno basado en el aprendizaje de la máquina. El funcionamiento de un ordenador que aprende puede compararse con el desarrollo cognitivo de un niño que incorpora habilidades mediante la observación del mundo que le rodea, analizando la forma en que los individuos interactúan y reproduciendo reglas implícitas no verbales. En términos generales, el aprendizaje automático sigue el mismo patrón: los algoritmos están entrenados para

[14] SABUR Rozina. «Google to "de-rank" stories from Russia Today and Sputnik», *The Telegrah* (21/11/2017), disponible en: https://www.telegraph.co.uk/news/2017/11/21/google-de-rank-stories-russia-today-sputnik/ Fecha de la consulta 07.5.2018.

aprender por sí mismos, sin que ningún humano introduzca patrones de comportamiento, de ahí que se empiece hablar de «el fin de la programación»[15].

Unas de las principales vulnerabilidades de los sistemas basados en *Machine Learning*, es el hecho de que su conducta está moldeada por el tipo de datos que asimila durante el aprendizaje[16]. Los algoritmos que dirigen este proceso se adaptan a los inputs que provienen del contexto donde el sistema operará. El resultado final está fuertemente condicionado por esta primera fase de aprendizaje, donde se asimilan regularidades y expectativas de comportamiento. Un sistema de AI será tan bueno como los datos de los que se alimenta. El aprendizaje con datos intrínsecamente sesgados conduce a resultados sesgados. El peligro reside, por tanto, en que esta «dieta» inicial de datos pueda estar «contaminada», lo que provocará en el sistema un comportamiento anómalo o indeseable. Uno de los ejemplos más ilustrativos procede del experimento llevado a cabo por el *chatbot* de Microsoft llamado Tay, el cual imitaba a una chica norteamericana de 19 años, simulando sus reacciones y modo de emplear el lenguaje. Antes de su lanzamiento público en forma de un perfil de Twitter, el sistema de conversación había sido testado en un entorno controlado, sin que generase ningún tipo de señal de alarma. Sin embargo, 24 horas después de su lanzamiento, se había convertido en una especie de monstruo misógino y racista que tuvo que ser rápidamente desconectado de esta popular red social[17]. Este sorprendente desenlace tiene su explicación por el tipo de interacciones humanas que había experimentado en sus primeras horas de «vida». Una serie de internautas aprovecharon la llegada de Tay a las redes sociales, para «trolear» el proyecto de Microsoft a través de una serie de preguntas y comentarios plagados de expresiones soeces y contenidos abyectos. La AI asimiló como una norma lo que era la excepción y fue moldeando sus registros para adaptarse al lenguaje, las expresiones y los temas que habían sido mayoritarios en sus primeras experiencias con humanos.

---

[15] TANZ Jason. «Soon We Won't Program Computers. We'll Train Them Like Dogs», *Wired* (17/5/2106), disponible en: https://www.wired.com/2016/05/the-end-of-code/ Fecha de la consulta 07.5.2018.

[16] OSOBA Osonde A. & WELSER IV William. «An Intelligence in Our Image. The Risks of Bias and Errors in Artificial Intelligence», *Rand Corporation* (2017), disponible en: https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf Fecha de la consulta 07.5.2018.

[17] METZ Rachel. «Why Microsoft Accidentally Unleashed a Neo-Nazi Sexbot», *MIT Technology Review* (24/3/2016), disponible en: https://www.technologyreview.com/s/601111/why-microsoft-accidentally-unleashed-a-neo-nazi-sexbot/ Fecha de la consulta 07.5.2018.

El uso de estos sistemas de aprendizaje que tienden a emular el comportamiento humano tiene varias consecuencias desde el punto de vista de las operaciones de influencia: da al receptor lo que quiere recibir, sin importar la ampliación de posturas condenables o marginales. Cuando se aplica a un objetivo individualizado, tiende a ahondar en el sesgo de confirmación del receptor, ampliando y celebrando las fobias y prejuicios de la persona sobre la que se actúa. El desembarco de la AI multiplicará exponencialmente el llamado efecto «filtro burbuja»[18] que se ha generado con la creciente personalización de la información que recibimos a través de Internet.

Figura 4: Ejemplo de un tweet inapropiado publicado por la AI de Microsoft Tay.
Fuente: https://thehackernews.com/2016/03/artificial-intelligence-bot.html

Hibridación entre el marketing digital y la propaganda política

El desarrollo de la AI está directamente conectado con el crecimiento de la publicidad digital. Esta tecnología está haciendo que técnicas que existen desde hace muchos años empiecen a ser mucho más efectivas y escalables. Las principales plataformas de internet están diseñadas para que la inserción de publicidad en ellas sea útil para los anunciantes, debido a su capacidad de conocer cada vez en mayor detalle las preferencias y motivaciones de sus usuarios y orientar de manera personalizada los mensajes persuasivos. Como consecuencia de la progresiva eficacia de las

[18] PARISER Eli. *El filtro burbuja: como la web decide lo que leemos y lo que pensamos*, Madrid, Taurus, 2017.

herramientas informáticas de investigación del mercado, se ha producido una convergencia entre la tecnología publicitaria y la propaganda política, las cuales se alimentan de las mismas técnicas de segmentación de la audiencia, *micro-targeting* y explotación de los prejuicios, miedos y aspiraciones del elector/consumidor.

Haciendo uso de las prácticas publicitarias existentes, las cuales han ido ganando precisión en la predicción de los rasgos psicológicos del consumidor, la AI permitirá generar en tiempo real mensaje precisos para cada receptor, obteniendo el máximo potencial persuasivo en la selección del momento donde existe una mayor receptividad. El *gerrymandering*[19] digital podría llegar a influenciar los procesos electorales de una forma que cuestione los fundamentos del sistema democrático[20].

Los objetivos económicos de estas empresas y los intereses de los promotores de la injerencia política están claramente alineados. Que la persuasión comercial y política se alimenten de las mismas herramientas dará como resultado que, incluso las campañas de injerencia política mal diseñadas y ejecutadas serán capaces de obtener resultados al beneficiarse de la eficacia de los mejores algoritmos del momento. Hasta los actores más modestos podrán disfrutar a un bajo coste económico de la posibilidad de alcanzar a millones de objetivos adaptando de manera individualiza y precisa el mensaje.

Por primera vez, será viable la plena explotación del inmenso volumen de datos que se ha generado en las últimas décadas a través de Internet. Basta con tener presente hasta dónde se ha podido llegar en el análisis de grandes cantidades de datos, sin contar con las potencialidades de una AI sofisticada. Es el caso, por ejemplo, de los estudios de un investigador de la Universidad de Cambridge llamado Michal Kosinski, quien desarrolló un simple test tipo OCEAN[21] que podría realizarse a través de

[19] Se trata de un término utilizado en ciencia política para referirse a la manipulación de las circunscripciones electorales de un territorio, uniéndolas, dividiéndolas o asociándolas, con el objeto de producir un efecto determinado sobre los resultados electorales. Tiene una acepción negativa al considerarse una técnica destinada a quebrar la imparcialidad del sistema electoral sobre el que se aplica.

[20] BRUNDAGE Miles *et al.* «The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation», Future of Humanity Institute, (February 2018), disponible en: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf Fecha de la consulta 07.5.2018.

[21] Se trata de un popular modelo de análisis de la personalidad utilizado en psicología donde se entiende esta como la composición de cinco rasgos o factores principales, los cuales se suelen denominar como: factor O (Openness o apertura a nuevas experiencias), factor C (Conscientiousness o responsabilidad), factor E (Extraversion o extraversión), factor A (Agreeableness o amabilidad) y factor N (Neuroticism o inestabilidad emocional), los cinco forman el acrónimo en inglés: «OCEAN». Véase:
DE RAAD, B. E., & PERUGINI, M. E. (2002). *Big five assessment*, London, Hogrefe & Huber Publishers, 2002.

Facebook. Este profesor tenía la intención original de enviarlo a unas decenas de amigos, pero en poco tiempo, miles y luego millones de usuarios habían enviado sus preferencias de personalidad a este académico. En 2012 Kosinski demostró que utilizando una media de 68 *likes* de *Facebook* por usuario, era posible predecir el color de su piel (con una precisión del 95%), su orientación sexual (con una precisión del 88%) y sus simpatías políticas (con un 85%). Proyectando su test en esta popular red social, Kosinski había conseguido generar la base de datos más importante del planeta sobre resultados psicométricos, lo que despertó el interés de algunas empresas dedicadas al *micro-targeting* con fines políticos. El investigador declinó estas ofertas, no obstante, el modelo fue replicado por una empresa británica llamada Cambridge Analytica, la cual empleó el modelo de test OCEAN a través de *Facebook*. La empresa ganaría popularidad por su participación en el referéndum británico a favor del *brexit*. La presencia en su consejo directivo de uno de los principales estrategas de la campaña de Donald Trump facilitó que adquiriese un elevado protagonismo en la campaña del candidato republicano. Cambridge Analítica consiguió de manera fraudulenta atesorar un detallado perfil de más de 220 millones de estadounidenses basados en pruebas OCEAN, los cuales habían sido complementados con otros datos adquiridos a través de empresas de agregación de datos en Internet. Esto le permitió elaborar múltiples mensajes adaptados a los rasgos de personalidad de los destinatarios de esta publicidad. Así, por ejemplo, un propietario de armas introvertido y preocupado por su seguridad recibía un anuncio distópico que mostraba un ladrón entrando a una casa por la noche, mientras que partidario de las armas más contemplativo y pacífico recibía un nostálgico anuncio que mostraba a un niño y su padre cazando en un bello paraje. Sin embargo, su arma más potente fue los anuncios destinados a desmovilizar el voto de la candidata demócrata. De esa forma, un vecindario completo del «pequeño Haití» de Miami recibió anuncios específicos argumentando que la Fundación Clinton no hizo lo suficiente para apoyar a Haití después del devastador terremoto sufrido por este país. Esta publicidad negativa no solo se adaptó a las especificidades de pequeños grupos, si no que en ocasiones se adaptó a un único destinatario[22].

[22] HUSAIN Amir. *The Sentient Machine. The Coming Age of Artificial Intelligence*, New York, Scribner, 2017.

Haitian President EXPOSES the Clinton Foundation: "Hillary Clinton tried to bribe me!" SEE HERE: https://www.youtube.com/watch?v=phKO2T2pMjM

Figura 5: Ejemplo de inserción publicitaria para usuarios de Facebook de origen haitiano durante la campaña presidencial estadounidense de 2016.
Fuente: https://me.me/i/remember-all-of-the-money-the-clintons-raised-for-haitia-3103176

Conclusiones

A la hora de analizar el impacto de la inteligencia artificial sobre las operaciones de influencia cabe el peligro de reiterar los mismos errores que se cometieron en el pasado a la hora de enjuiciar el impacto socio-político de internet. La popularización de esta herramienta se vio acompañada de un encendido debate entre dos perspectivas antagónicas[23]. Por un lado, una corriente de opinión que ha sido catalogada como «ciberoptimista», la cual identificaba a Internet como el instrumento que haría posible una nueva etapa de liberalización política y de libre circulación de la información y el conocimiento por todo el planeta. En el otro extremo, la visión «ciberpesimista», la cual atribuía a internet un efecto inexorable de degradación de la libertad, potenciando hasta límites insospechados los mecanismos de control y represión de los autoritarismos políticos y contaminando incluso la naturaleza liberal de las democracias más consolidadas.

Aunque las dos perspectivas aportaban elementos útiles para una correcta comprensión del impacto de esta tecnología sobre el cambio político, sin embargo, ambas se equivocaron al atribuir una naturaleza determinista a Internet. Lo que ellos

[23] TORRES, Manuel R. «Internet como motor del cambio político: ciberoptimistas y ciberpesimistas», *Revista del Instituto Español de Estudios Estratégicos*, N.º 1 (2013), pp. 127-148, disponible en: http://revista.ieee.es/index.php/ieee/article/view/40/36 Fecha de la consulta 07.5.2018.

entendían como rasgos estructurales, no dejaban de ser las consecuencias de las acciones humanas. Los usuarios son los únicos responsables de la naturaleza del cambio político. Tanto demócratas como autócratas compiten por dominar esta tecnología y ponerla al servicio de sus objetivos. La herramienta tienen un valor neutral, y no hay nada intrínsecamente prodemocrático o liberticida en ella. Por tanto, no es el propio instrumento quien determina el resultado de esta pugna, sino el contexto de organización política y social y, sobre todo, es la habilidad de los diferentes actores lo que determina quién resultará vencedor en cada momento.

En este sentido, la injerencia política apoyada por AI dibuja un nuevo contexto de juego, y un conjunto de reglas que pueden ser aprovechadas en mayor medida por unos u otros actores. Sin embargo, el balance resultante no está determinado de antemano. El resultado tampoco es invariable, y es lógico que se vea sometido a continuas revisiones a medida que los actores políticos encuentran nuevas ventanas de oportunidad o desarrollan estrategias novedosas.

El verdadero riesgo reside en la falta de reflejos o una respuesta tardía a la hora de asimilar los efectos disruptivos que tendrá esta tecnología en todos los órdenes de la sociedad. Uno de los ámbitos que tiene más riesgo de ser desbordado por la rapidez con la que se va a producir este cambio de paradigma es precisamente el marco regulatorio que establece las condiciones y garantías bajo las cuales se produce el proceso electoral en los países democráticos. Con independencia de que algunos actores estratégicos como servicios de inteligencia, administraciones públicas y medios de comunicación sean capaces de adaptarse y responder a las acciones ofensivas de aquellos que pretenden desvirtuar el proceso democrático, es imprescindible adaptar las reglas de juego para poder anticipar los efectos de unas inevitables operaciones de influencia potenciadas por inteligencia artificial.

En un horizonte temporal cercano se vislumbran tres posibles escenarios: 1) el despliegue de una regulación restrictiva que trate de frenar el uso de la AI en cualquier ámbito que tenga impacto en el comportamiento electoral. 2) Forzar la transparencia sobre los algoritmos que alimentan estos sistemas. 3) Auditar el resultado de la aplicación de los algoritmos.

El primer escenario, aunque completamente inútil, no es inverosímil**.** En ocasiones existe la percepción errónea en los legisladores de que la producción de leyes es capaz por sí misma de extirpar de la realidad una influencia no deseada. El peligro no reside

solamente en la escasa capacidad disuasoria de este intento de bloqueo, sino que posiblemente esta acción regulatoria terminase excluyendo o restando eficacia a los actores que pueden confrontar las operaciones hostiles.

El segundo escenario podría resultar útil para alimentar el escrutinio público sobre las plataformas que pueden ser instrumentalizadas para manipular a la opinión pública. El problema reside en que esto solo es efectivo sobre una ciudadanía capaz de entender las implicaciones de estos algoritmos en la forma en la que los ciudadanos eligen y fiscalizan a sus gobernantes.

El tercer escenario atribuye al ámbito público la responsabilidad de auditar y corregir los efectos de los sistemas basados en inteligencia artificial en el proceso político. Si bien esta opción puede ser una solución frente al carácter inescrutable que pueden alcanzar las redes neuronales en las que se basa la AI, abre el camino hacia un creciente intervencionismo público que puede colisionar no solo con innovación tecnológica y la libertad empresarial, pero también la libertad de información.

En definitiva, esta innovación tecnológica no tiene por qué suponer una vulnerabilidad adicional en el sistema democrático. La interacción entre los distintos actores puede terminar nivelando los efectos negativos con las nuevas oportunidades que se abren para crear sociedades dotadas de nuevas herramientas para defenderse de los intentos de manipulación política. De hecho, tampoco existe ningún factor que determine que el resultado final deba ser un nuevo equilibrio. *A priori*, no existe ninguna barrera que impida que la AI termine siendo una gran palanca de cambio que erradique del proceso político la mentira y permita a los ciudadanos disfrutar de las nuevas ventanas de conocimiento y actuación generadas por la capacidad de abarcar cantidades masivas de datos.

*Manuel R. Torres Soriano**
Profesor de Ciencia Política, Universidad Pablo de Olavide, Sevilla.
Miembro del Grupo de Estudios en Seguridad Internacional (GESI)