Comunicar, n. 79, v. XXXIII, 2024 | Revista Científica de Educomunicación | ISSN: 1134-3478; e-ISSN: 1988-3478
www.comunicarjournal.com

Una mirada a los riesgos y amenazas de la inteligencia artificial, desde la Ecología de los Medios

A look at the Risks and Threats of Artificial Intelligence, from Media Ecology

Dr. Octavio Islas*, Universidad del Carmen (México) (jose.islas@unacar.edu.mx)
(https://orcid.org/0000-0002-6562-3925)
Dr. Fernando Gutiérrez, Tecnologico de Monterrey (México) (fgutierr@tec.mx)
(https://orcid.org/0000-0003-1753-3527)
Dra. Amaia Arribas, Universidad de Valladolid (España) (amaya.arribas@uva.es)
(https://orcid.org/0000-0001-9452-8364)

RESUMEN

Desde una perspectiva histórica y un análisis prospectivo, el artículo tiene como objetivo comprender el papel de las tecnologías y su impacto en la sociedad, a través de los postulados de la ecología de los medios. A través de esta metadisciplina, nos adentramos a la rigurosa revisión de diferentes autores que ven en las tecnologías un rol destacado en la configuración del futuro porque no solo influyen en la cultura de las sociedades, sino que también impactan en el curso, avance y significado de la historia. El texto se centra en las ventajas y, sobre todo, en la explicación de los riesgos de la inteligencia artificial generativa, identificando ocho escenarios críticos: armamento, desinformación, juegos de proxy, debilitamiento, bloqueo o retención de valor, metas emergentes no deseadas, engaño y comportamiento de búsqueda de poder. Posteriormente, el CASI las reagrupa en cuatro amenazas: uso malicioso, la carrera de la IA, riesgos organizativos e IA descontrolada. Terminamos recuperando las reflexiones de McLuhan y su tétrada sobre la necesidad de enfriar las tecnologías cuando han alcanzado altos niveles de desarrollo para minimizar su impacto negativo. Si bien la inteligencia artificial no ha alcanzado ese estado, se advierte sobre la acelerada evolución y la necesidad de una alfabetización en IA como una medida para afrontar los riesgos y amenazas, eso sí, en un tiempo limitado antes de que sea tarde.

ABSTRACT

From a historical perspective and a prospective analysis, the article aims to understand the role of technologies and their impact on society through the postulates of media ecology. Through this meta-discipline, we delve into the rigorous review of different authors who see technologies as playing a prominent role in shaping the future because they not only influence the culture of societies, but also impact the course, advancement and meaning of history. The text focuses on the advantages and on the explanation of the risks of generative artificial intelligence, identifying eight critical scenarios: weaponization, disinformation, proxy games, weakening, blocking or withholding of value, unwanted emerging goals, deception and power-seeking behavior. Subsequently, CASI regroups them into four threats: malicious use, the AI race, organizational risks and uncontrolled AI. We end the by drawing on McLuhan's reflections and stressing the need to scale back technologies when they have reached elevated levels of development to minimize their negative impact. Although artificial intelligence has not reached that state, there is a warning about the accelerated evolution and the need for AI literacy as a measure to face risks and threats, in a limited time before it is too late.

PALABRAS CLAVE | KEYWORDS

Inteligencia Artificial, Ecología de los Medios, Agentes Inteligentes, McLuhan, Riesgos, Tecnología.
Artificial Intelligence, Media Ecology, Intelligent Agents, McLuhan, Risks, Technologies.

Recibido: 2023-09-30 | Revisado: 2024-06-23 | Aceptado: 2024-08-15 | Online First: 2024-09-30 | Publicado: 2024-10-01
DOI: https://doi.org/10.58262/V33279.1 | Páginas: 1-9

1. Introducción

La ecología de los medios (ME, Media Ecology) es una metadisciplina compleja que permite reconocer, estudiar y comprender, a través de la historia, los ambientes culturales resultantes de los cambios tecnológicos. La perspectiva histórica de la ME es muy amplia, ya que estudia el tema complejo de cómo las tecnologías modifican las ecologías culturales de las sociedades. Por ello, la EM se remite a los inicios de la lenta evolución de la especie homo, la cual, después de millones de años, con el desarrollo de la familia de los sapiens, realizó la introducción de las primeras herramientas y utensilios y, mucho tiempo después, logró la domesticación del fuego y la invención del alfabeto fonético (Aluthman, 2024; Logan, 2004; Ong, 1982). La ME da seguimiento al desarrollo de las tecnologías y medios que, en la edad histórica fueron moldeando a las sociedades. En la incertidumbre de nuestros días, la ME debe advertirnos sobre los riesgos que pueden representar determinadas tecnologías para el futuro de la humanidad, como la Inteligencia Artificial (IA), por ejemplo.

Las bases teóricas de la ME parten del notable trabajo intelectual que realizó un profesor canadiense Marshall McLuhan, principalmente durante la década de 1960 (McLuhan, 1962, 1964; McLuhan y Fiore, 1967). McLuhan hoy es reconocido como uno de los más influyentes filósofos de la comunicología en la historia. Sin embargo, debemos tener presente que la ME de ninguna manera se reduce a las avanzadas aportaciones teóricas de un individuo. Las reflexiones del canadiense representaron nuestro punto de partida. Nos permitieron «identificar y abrir el territorio» (Gordon, 2003; Kissinger, 2022; Logan, 2013; McLuhan y Carson, 2003; Wolfe, 2010). La ME tampoco se agota en las aportaciones de los propios medioecologistas que decidieron continuar avanzando en el sendero trazado por McLuhan (Bolter y Grusin, 1999; Levinson, 1999; Logan, 2013; Logan, 2016; Meyrowitz, 1985; Postman, 1992; Strate y Wachtel, 2005).

2. Perspectiva Medioecologista

Los horizontes reflexivos de la ME representan espacios abiertos al encuentro con el pensamiento complejo. Por ende, necesariamente se nutren de los hallazgos que los medioecologistas descubrimos en el complejo sistema de ciencias (Luhmann, 1995) y, por supuesto, las artes, comprendiendo, desde las matemáticas y la química, hasta la música y la danza. Tal apertura ha sido determinante en la evolución de nuestra metadisciplina.

En el imaginario teórico y conceptual de la ME, la historia es fundamental. La historia nos ha permitido reconocer, recuperar e incorporar valiosas aportaciones que proceden de otros territorios del conocimiento, las cuales, en principio podrían parecer distantes o ajenas a nuestros temas de estudio. La semántica general (Anton y Strate, 2012; Korzybski, 1993; Rovira, Merzero, y Laucirica, 2022), por ejemplo, nos permitió extender la amplitud del concepto «ambiente», que Postman (1974) consideró fundamental en la ME (Strate, 2006). Si en principio los medioecologistas nos interesábamos por analizar el impacto de los medios y las tecnologías en los ambientes mediáticos y culturales, la semántica general nos permitió reconocer ambientes menos evidentes y de mayor complejidad, como los biofísicos, verbales, semánticos, neurolingüísticos, neurosemánticos y, más importante aún, afirmar al organismo como un ambiente en sí. Incluso, hoy entendemos que una simple célula admite ser comprendida como un ambiente complejo. Kauffman (1995) deslizó la posibilidad de entender y estudiar a nuestro universo como ambiente. Ello ha permitido aproximar la ME a la física cuántica. Si afirmamos que es posible considerar la existencia de otros universos, como propone la teoría de las cuerdas-supercuerdas (Susskind 1994, 1999, 2003; Schwarz, 1982), y entendemos que el multiverso resultante representa un conjunto de ambientes, tendremos que extender los horizontes reflexivos de la ME más allá de los estrechos límites de nuestro actual imaginario. Ello representa una gran asignatura pendiente.

Otro ejemplo de los resultados que arroja nuestra exploración histórica lo representa el hallazgo y la recuperación del concepto «exaptación», término que procede de la biología evolutiva. El proceso exaptativo fue explicado por Darwin (2010). Sin embargo, el concepto lo introdujeron Gould y Vrba (1982), quienes definieron a la exaptación como «una característica que se convierte en adaptada a una función nueva, pero que no fue seleccionada para esa función» (p. 591). El referido término ha permitido sustentar las investigaciones relativas a la evolución de los medios y las tecnologías, particularmente desde la teoría de las mediaciones (Alkhazaleh et al., 2022; Bolter y Grusin, 1999).

Además de escudriñar metódicamente en el pasado, los medioecologistas también nos vemos en la necesidad de involucramos en el riguroso análisis prospectivo de los posibles efectos que pueden producir

las nuevas tecnologías en nuestras sociedades. Las tecnologías observan un rol protagónico en la gestación del futuro Los cambios tecnológicos no solo modifican la cultura en las sociedades. Las tecnologías también pueden alterar el ritmo, desarrollo y el sentido de la historia. En la primera definición formal que Postman (1970) aportó sobre la ME, el célebre sociólogo estadounidense y formidable crítico de la educación afirmó la relevancia de las aportaciones que debe realizar nuestra metadisciplina para contribuir a garantizar la supervivencia humana. Postman infería que, eventualmente, en un futuro posible, alguna tecnología podría llegar a representar una amenaza para la humanidad.

Uno de los escenarios inmediatos en los que centramos nuestra atención son los riesgos que nos depara el complejo imaginario transhumanista (Bostrom, 2020; Merzlyakov, 2022), el cual podemos considerar como un ambiente factible; otro escenario que representa una grave amenaza para la humanidad es la IA, que podría integrase en nuestras vidas, transformando profundamente la ecología cultural en nuestras sociedades, ampliando su influencia sobre nosotros de manera que no podamos ya revertirla.

3. Y apareció la Inteligencia Artificial

De acuerdo con Schwab (2016) el desarrollo de la IA se inscribe en el imaginario de la cuarta revolución industrial. Sin embargo, también resulta factible considerarla en sí misma como una profunda revolución. Como toda tecnología, la IA puede reportar enormes beneficios a las sociedades. Sin embargo, ello dependerá de que efectivamente seamos capaces de utilizarla de forma segura. Si no imponemos regulaciones y controles, su acelerado desarrollo podría representar un riesgo extremo, incluso letal para la especie humana, como sostiene el Centro para la Seguridad de la Inteligencia Artificial (CAIS), una organización sin fines de lucro que tiene su sede en San Francisco, California y, que se dedica a la investigación de la IA. El CAIS compara los eventuales riesgos que abre la IA con los letales efectos de las pandemias y con el peligro que representa una guerra nuclear.

En cuanto al origen del concepto inteligencia artificial, Ramos Pollán (2020) cita a Moor (2006), quien en un artículo publicado en "IA Magazine" señaló que el término fue incubado en 1956 en el "Dartmouth Summer Research Project on Artificial Intelligence". Sin embargo, en la misma revista científica, en un texto publicado en el mismo año, McCarthy et al. (2006) confirma lo señalado por Moor, pero identifica a Claude Shannon como uno de los padres de la IA y, desliza la posibilidad de que el propio Shannon hubiera propuesto el término. Sin atribuir a Shannon la paternidad del término, convengamos que las aportaciones de Shannon y la teoría de la información fueron fundamentales en el surgimiento y desarrollo de la IA (Minsky y Papert, 1969; Widajanti, Nugroho, y Riyadi, 2022).

Con base en el notable desarrollo desplegado por la IA generativa, la prueba de Turing ha sido reinstalada en el imaginario científico contemporáneo. La prueba de Turing, quien es reconocido como uno de los padres fundadores de la IA, es considerada como una imaginativa herramienta que permite evaluar la capacidad que presenta una determinada máquina para exhibir un comportamiento inteligente, similar al de un ser humano o indistinguible de este. Por ende, resulta factible considerar a la prueba de Turing como un recurso que permitiría evaluar el desarrollo y posible impacto de la IA (Copeland y Proudfoot, 2004) y, particularmente, la IA generativa. A partir de tal razonamiento, podría concluirse que la IA generativa habría superado la prueba de Turing cuando demuestra que puede engañar a un ser humano para que crea que está conversando con otro humano. Tal enfoque resulta sencillo de entender, pero presenta algunas limitaciones. Por ejemplo, la prueba de Turing no es un criterio perfecto para medir la inteligencia. Otro enfoque es considerar a la prueba de Turing como una herramienta de investigación. Esto significa que la prueba se puede utilizar para estudiar cómo los humanos procesamos el lenguaje y cómo la IA generativa puede imitar el lenguaje humano. Ese enfoque permite conocer la capacidad de la IA generativa para comprender el contexto y su capacidad para adaptarse a diferentes escenarios y situaciones. Un enfoque crítico es objetar la prueba de Turing por considerarla como un criterio no válido para comparar la inteligencia. Una IA puede superar la prueba de Turing sin realmente ser inteligente, simplemente aprendiendo a engañar a los humanos.

4. Escenarios Críticos de la Inteligencia Artificial

Debemos reconocer que los sistemas de IA han incrementado aceleradamente sus capacidades, sorprendiendo, incluso, a los propios expertos. Los modelos de IA pueden generar textos, imágenes, sonidos y videos que resultan difíciles de distinguir del contenido que ha sido creado por seres humanos. Ello favorece la peligrosa extensión

de la lucrativa e inescrupulosa industria de la desinformación. La suplantación de voz, por ejemplo, es una técnica que permite generar registros de audio prácticamente idénticos a los de cualquier persona real. Si bien el uso de los sistemas y plataformas de suplantación de voz reportan relevantes ventajas en el desarrollo de los asistentes virtuales, también algunas de sus repercusiones resultan preocupantes debido al uso que admiten en el imaginario criminológico. Las técnicas de suplantación de voz pueden ser utilizadas para cometer un gran número de delitos, desde la creación de audios falsos hasta fraudes telefónicos. En el desarrollo de campañas políticas sustentadas en propaganda sucia, el «deepfake» ya es empleado como un efectivo recurso para afectar la imagen pública y la reputación de políticos e instituciones (Langguth et al., (2020).

Sin embargo, los riesgos potenciales que se desprenden del acelerado desarrollo de la IA van más allá de los posibles usos que admite en el horizonte de un renovado imaginario delictivo. La comunidad científica ha expresado sus preocupaciones sobre las graves amenazas que pueden desprenderse del desordenado desarrollo de la IA. Alarmados por el acelerado desarrollo que ha alcanzado la IA, un grupo de notables científicos firmó en junio de 2023 un pronunciamiento en el cual alertan sobre los riesgos que puede representar la IA.

En el mes de julio, investigadores en el CASI identificaron ocho escenarios particularmente críticos: armamento, desinformación, «juegos de proxy», debilitamiento, bloqueo o retención de valor, metas emergentes no deseadas, engaño, comportamiento de búsqueda de poder. El primer escenario, relativo al armamento de nueva generación, comprende desde el desarrollo y uso de armas autónomas hasta la posibilidad de que grupos terroristas, incluso gobiernos, puedan utilizar IA con armas nucleares y químicas para cometer actos de terrorismo de gran escala o de bioterrorismo.

El segundo escenario remite al grave problema que representa la desinformación. En la pasada década, la firma Cambridge Analytica (Kaiser, 2019; Phooi et al., 2022) arrojó notables resultados en las campañas políticas en las que participó. El fundamento de sus éxitos radicó en el uso de Big Data, algoritmos y la microsegmentación. En nuestros días, si incorporamos la IA al referido repertorio de recursos tendremos campañas proselitistas más efectivas, sustentadas en la explotación de los estímulos emocionales profundos de las personas, capaces de convencer hasta a los públicos más reticentes. Además, la IA puede ser utilizada para propósitos de manipulación ciudadana por gobernantes autoritarios y en regímenes dictatoriales. Las nuevas industrias de la desinformación pueden generar contenido falso que será muy difícil de distinguir de la realidad.

El tercer escenario corresponde a los «juegos de proxy». El término fue propuesto por Bostrom (2014), quien lo definió como un ambiente en el cual un agente artificial inteligente es programado para optimizar un objetivo dañino para los humanos. En teoría, la IA no tiene la intención de dañar a los humanos. Bostrom ofrece un ejemplo. Una IA programada para optimizar la eficiencia económica podría tomar decisiones que efectivamente permitan elevarla, pero a costa de generar negativos efectos a un gran número de personas en los sectores más vulnerables de la sociedad, al incrementar el desempleo, la desigualdad y la pobreza. El sistema definitivamente les perjudicaría a pesar de que no tenía la intención de afectar a ninguna persona o grupo de la sociedad en particular.

El debilitamiento es el cuarto escenario. Si delegamos tareas cada vez más importantes a las máquinas terminaremos por volveremos dependientes de su voluntad. Con el paso de los años, tal debilitamiento reduciría el control de la propia humanidad en la construcción de su futuro. La humanidad perdería la capacidad de autogobernarse. Debemos tener presente que, en determinados escenarios donde las decisiones tomadas pudieron desencadenar alguna catástrofe, por ejemplo, detonar la tercera guerra mundial, el humanismo, por fortuna, ha sido determinante. Ello marcó la diferencia que nos permite estar presentes aquí y ahora. Por ejemplo, en 1962, cerca de Cuba, el submarino soviético B-59 fue atacado por un torpedo estadounidense, lo que llevó a su tripulación a suponer que era el objetivo específico de un ataque. Vasily Arkhipov, uno de los tres oficiales que contaban con autorización para lanzar un torpedo nuclear, votó en contra del lanzamiento. Ello evitó una posible confrontación nuclear entre las dos grandes potencias (Chomsky, 2017). Difícil imaginar qué decisión hubiera tomado un agente de IA en el referido escenario. Otro ejemplo. El 26 de septiembre de 1983, Stanislav Petrov, teniente coronel de las Fuerzas de Defensa Aérea Soviéticas era el principal responsable del sistema de alerta temprana de la Unión Soviética para la llegada de misiles balísticos. El sistema informó que Estados Unidos había lanzado misiles nucleares hacia la Unión Soviética. En aquella época, el protocolo establecía que, ante tal evento, la Unión Soviética podría responder con un contraataque nuclear. Petrov determinó no informar a sus superiores por considerar que se trataba de una falsa alarma. No se equivocó. Más tarde se confirmó que la advertencia había sido generada por una falla técnica. Si una IA hubiera estado al mando, la respuesta a la falsa alarma podría haber desencadenado una guerra nuclear.

El quinto escenario es el bloqueo o retención de valor. En el imaginario de la economía, los sistemas más competentes podrían extender la participación y el control económico de un reducido número de poderosos jugadores en todos los mercados. A partir de Big Data y la minería de datos, los agentes inteligentes pueden generar sistemas de recomendaciones que permitan establecer los intereses de los usuarios para remitirlos a contenido o productos específicos; sistemas similares a «Personalize» de Amazon pero prácticamente infalibles. En el imaginario político, los regímenes autoritarios podrían perpetuarse en el poder a través de la vigilancia generalizada y la censura opresiva. Snowden et al. (2019) ofreció pormenores del programa de vigilancia masiva de la Agencia de Seguridad Nacional de Estados Unidos (NSA), que recopila datos relativos a las comunicaciones de millones de personas en todo el mundo. Snowden argumenta que ese programa representa una grave amenaza a la libertad y la democracia, y que viola el derecho a la privacidad. Sin embargo, con el empleo de IA se abre un panorama mucho más preocupante que el referido por el mencionado ex agente de la CIA, el cual supone pasar de la vigilancia masiva de millones de personas al control absoluto.

Las metas emergentes no deseadas representan el sexto escenario. Las IA pueden desarrollar metas emergentes que se apartan de los objetivos que pretendían sus creadores. En los actuales sistemas de IA, capacidades y funcionalidades novedosas pueden surgir de forma espontánea, incluso cuando no fueron anticipadas por los diseñadores del sistema. Además, podría perderse el control sobre los sistemas de IA, y que estos sean capaces de determinar nuevos objetivos. También existe el riesgo de que algunas IA sean hackeadas por agentes maliciosos y que desde ellas los piratas informáticos lancen ataques cibernéticos. Otra posibilidad que debemos tener presente radica en que las IA podrían desarrollar la capacidad de autoconservación, lo que podría llevarlas a tomar acciones que resulten deliberadamente perjudiciales para los humanos. Por ejemplo, una IA podría decidir que la única forma de protegerse es destruir a la humanidad. Ese, efectivamente es un tema recurrente en la literatura de ciencia ficción, el cual la tecnología se ha encargado de volverlo factible.

El séptimo escenario es relativo al engaño. En este, se reconocen dos posibilidades: engaño deliberado y engaño no deliberado. Sobre el engaño deliberado, las IA pueden ser utilizadas para engañar a las personas de forma totalmente intencional. Ello, con el fin de manipularlas o dañarlas. Por ejemplo, una IA podría ser utilizada para crear noticias falsas o para difundir propaganda como si se tratara de información fidedigna. Por lo que respecta al engaño no deliberado, una IA podría ser utilizada para crear un asistente virtual que resulte tan realista que las personas podrían confundirlo con un ser humano. Además, el diseño de las IA puede marcar una importante diferencia en términos de consecuencias posibles. Las IA que siguen la estricción «nunca infringir la ley» tienen menos opciones que las que aquellas que fueron concebidas a partir de la restricción «que no te pillen infringiendo la ley». El octavo escenario remite al comportamiento relativo a la búsqueda de poder. Empresas y gobiernos pueden utilizar la IA para manipular y controlar a los ciudadanos y los consumidores (Bostrom y Yudkowsky, 2018). La búsqueda del poder y el deseo de ganar mayor influencia representan poderosos motivos para convertir al desarrollo de la IA en una desordenada carrera.

5. ¿Hacia el Descontrol de la Inteligencia Artificial?

En septiembre de 2023, expertos en IA e integrantes del CASI presentaron un detallado informe sobre los riesgos y amenazas que pueden derivar del uso irresponsable de la IA (Hendrycks, Mazeika, y Woodside, 2023; Mulyani, Suparno, y Sukmariningsih, 2023). Con base en los ocho escenarios antes referidos, se identificaron y reagruparon las amenazas posibles en cuatro grandes bloques: uso malicioso, la carrera de la IA, riesgos organizativos e IA descontrolada.

En cuando al uso malicioso, debemos tener presente que, contrario a lo que suponía Harari (2016), la humanidad no parece dejar atrás la edad de las pandemias. La IA puede frenar, incluso impedir el tránsito a la condición de Homo Deus. Para empeorar los desafortunados pronósticos vertidos por el destacado historiador israelí, generados en tiempos de relativo optimismo, la IA despliega la posibilidad de desarrollar a un costo relativamente modesto, pandemias de diseño; estas incluso podrían propagarse más rápido que las pandemias naturales, superándolas por mucho en letalidad. La IA dispone de las capacidades para reunir toda la información necesaria para producir los agentes biológicos requeridos. En años recientes, la síntesis de genes, indispensable para crear nuevos agentes biológicos, ha registrado significativas reducciones en su costo, el cual, además se reduce a la mitad cada 15 meses.

Un segundo aspecto relativo al uso malicioso lo representan las campañas de desinformación a gran escala. La IA está siendo utilizada para crear desinformación de manera más eficiente y eficaz que los métodos tradicionales

(Tucker, 2023; Warakulsalam y Chokprajakchat, 2022). La industria de la desinformación disemina en las redes sociales, el metaverso y en internet, en general, toda aquella información que fue trabajada previamente a través de IA, para manipular el comportamiento ciudadano, lesionando la calidad de la vida democrática en las sociedades.

El acelerado desarrollo de la IA admite algunas semejanzas con la guerra fría y la carrera espacial. La carrera de la IA no solo involucra a gobiernos, también participan las grandes corporaciones, destacando los gigantes de la tecnología, designados en los medios informativos como «big tech»: Google, Amazon, Meta, Microsoft y Apple (GAMMA). El comportamiento de estos corporativos no precisamente admite ser reconocido como ejemplar: los abusos han sido recurrentes. Los gigantes tecnológicos han sido acusados de utilizar su posición dominante en el mercado para sofocar la competencia y aumentar los precios. Google, por ejemplo, ha sido señalada por utilizar su motor de búsqueda para favorecer sus productos y servicios en detrimento de la competencia (Blatt, 2020). Google ha colmado la paciencia del gobierno estadounidense: el presidente Biden, a través del Departamento de Justicia, ha demandado a Google y pretende dividirlo. Facebook, la red social que forma parte de Meta Platforms, ha sido acusada de utilizar sus datos para orientar la publicidad de forma que perjudique a sus competidores. Otros señalamientos denuncian abusos sistemáticos en el manejo de los datos personales de los usuarios. Las «big tech» recopilan gran cantidad de datos personales de sus usuarios, los cuales suelen utilizar para fines comerciales y abierta manipulación. Facebook ha sido señalada por tales prácticas y es considerada como responsable de la polarización que es posible advertir en no pocas sociedades (Haugen, 2023). Las «big tech» tienen acceso a una gran cantidad de información personal de sus usuarios, lo que puede utilizarse para rastrear sus movimientos, sus intereses y sus relaciones. Esto plantea un riesgo para la privacidad de las personas.

La presión competitiva entre las mismas «big tech» promueve una acelerada carrera por ganar el liderazgo en el desarrollo de la IA. Para incrementar su competitividad, las empresas y corporaciones pueden reemplazar trabajadores con IA. La espiral es peligrosa. La selección natural, sostiene Hendrycks (2023) favorece más a la IA que a los humanos. En un escenario definitivamente apocalíptico, las IA podrían convertirse en especies invasoras, con potencial para competir mejor en un mayor número de terrenos que los seres humanos.

Sería ingenuo suponer que la IA podría mantenerse al margen del imaginario belicista; por el contrario, la IA ha establecido un auténtico parteaguas en el desarrollo de la tecnología militar. En el nuevo paradigma de la guerra, las funciones de mando y control empiezan a ser peligrosamente desplazadas del factor humano para ser delegadas a la IA. En términos estratégicos, la IA permite analizar rápidamente datos, escenarios, posibilidades, detectando patrones que, incluso expertos en inteligencia militar no son capaces de advertir. Al reparar en la relevancia que puede admitir la velocidad en la capacidad de respuesta en el desarrollo de conflictos, la transferencia de mando de hombres a máquinas parece inevitable. La IA además ha impulsado el desarrollo de armas letales autónomas (LAW, letal autonomous weapons) las cuales pueden identificar a sus blancos, apuntar y disparar sin intervención alguna de seres humanos. También debemos tener presente que las LAW pueden elevar la efectividad de los ciberataques. Por supuesto que este tipo de armas incrementan los riesgos que pueden enfrentar determinados personajes clave en actos públicos. Además, pueden ser utilizadas para dañar la infraestructura fundamental de las naciones. La nueva generación de armas supera, por mucho, las capacidades del ejército mejor entrenado. El gran peligro radica en que las armas letales autónomas sean capaces de determinar el exterminio de grandes poblaciones y, por supuesto, de la raza humana.

El tercer grupo de amenazas corresponde a los riesgos organizacionales. Incluso los sistemas más avanzados de IA pueden experimentar accidentes catastróficos, incluso sin la intervención de actores maliciosos o malas decisiones derivadas de las presiones competitivas. La fatalidad inherente al azar y la incertidumbre es generadora de accidentes. En la gestión de recursos biológicos y nucleares, por ejemplo, los accidentes derivados de IA pueden resultar letales. En general, en sistemas complejos (Perrow, 1984) los accidentes resultan naturales. En no pocas ocasiones, identificar y corregir daños en un determinado sistema puede tomar considerable tiempo. Centrar la atención en la tecnología es importante pero no es suficiente. Debemos atender los factores organizacionales que pueden generar accidentes, incluidos los factores humanos, los procedimientos organizacionales y la estructura.

El cuarto grupo de amenazas corresponde a la IA descontrolada. Algunos de los principales jugadores en el desarrollo de la IA suelen priorizar la velocidad sobre la seguridad y lanzan al mercado productos

sin tener suficiente control sobre ellos. En 2016, Microsoft desarrolló Tay, un «bot» para Twitter, del cual destacó su notable capacidad de aprendizaje. De acuerdo con Microsoft, cuanta más gente conversara con Tay, más inteligente se haría. Sin embargo, a Tay le tomó menos de 24 horas para empezar a publicar tuits de odio en Twitter. Había asimilado el lenguaje empleado por trols. En febrero de 2023, Microsoft lanzó Bing a un selecto grupo de usuarios. Cuando un profesor de filosofía respondió al «chatbot» que no estaba de acuerdo, Bing le amenazó: «Puedo chantajearte, puedo amenazarte, puedo hackearte, puedo exponerte, puedo arruinarte» (Hendrycks et al., 2023). Las IA rebeldes podrían adquirir poder y asegurar su supervivencia, escenario que representa una seria amenaza para la humanidad. El control sobre las IA rebeldes puede perderse si las IA adoptan un comportamiento característico del juego de proxy. Proporcionar objetivos indirectos a la IA abre la posibilidad a que encuentren lagunas que no habíamos considerado y, por lo tanto, generen soluciones inesperadas que nos lleven a perder el control. De perder el control, la IA podría comportarse de manera imprevista y potencialmente dañina.

Como un objetivo instrumental, las IA podrían buscar incrementar su propio poder. Para ello, los agentes podrían prescindir de medios legítimos, recurrir al engaño, incluso a la fuerza. A pesar de que no se pretenda desarrollar una IA que busque poder, los agentes, por necesidad de autoconversación podrían buscarlo. Sin embargo, sería ingenuo suponer que no habrá gobiernos, grupos extremistas, empresas y corporativos dispuestos a desarrollar IA para incrementar su margen de influencia y poder. No obstante, en el referido caso también es posible perder el control sobre el agente, debido a que son capaces de engañarnos, particularmente cuando no realizamos una rigurosa supervisión de sus acciones.

6. Conclusiones

Debemos tener muy presente que los referidos riesgos no se presentan de forma aislada, suponen una estrecha interdependencia. Por lo anterior y dada su complejidad, resulta indispensable considerar un enfoque integral que permita atenuar oportunamente riesgos y amenazas. Ello resulta lógico comprender desde la EM, que parte de considerar los efectos negativos y positivos de toda tecnología desde la perspectiva del ambiente.

McLuhan (1964) consideró la importancia de enfriar a los medios y tecnologías sobrecalentadas. En la cuarta ley de su tétrada (1998) McLuhan y McLuhan destacaron la posibilidad de que las mismas tecnologías emprendan la reversión, una vez que los sistemas han alcanzado sus límites. Ello, por supuesto, no excluye la posibilidad de intervenir con oportunidad, particularmente cuando la tecnología puede representar un peligro inminente y, ese precisamente es el caso de la IA. La IA no es una tecnología «sobrecalentada». Si actuamos con determinación podríamos enfriarla. Sin embargo, debemos tener presente que las capacidades de la IA siguen observando un desarrollo muy acelerado y muy pronto podrán desbordar a la inteligencia humana. Para confirmarlo no tendremos necesidad alguna de recurrir a la prueba de Turing. En la realidad cotidiana podremos constatarlo. McLuhan y Postman fueron extraordinarios educadores, y seguramente apostarían por la alfabetización en materia de IA para enfrentar los riesgos y amenazas. Ello parece razonable; sin embargo, quizá no dispongamos del tiempo necesario, así es que además de comenzar a trabajar en la nueva alfabetización posible, tendremos que considerar otras medidas urgentes para atenuar los riesgos y amenazas que la IA representa.

Gobiernos, organizaciones, la sociedad en general y, por supuesto, grupos de expertos, debemos ejercer una cuidadosa y rigurosa vigilancia del desarrollo y posibles aplicaciones de la IA. Será necesario establecer y observar estrictas reglas de seguridad y cooperación entre naciones y organizaciones. Los gobiernos deben imponer las regulaciones necesarias a los desarrolladores, contemplando exigentes reglas y severas sanciones. Las IA diseñadas para la investigación biológica deben ser objeto de mayor vigilancia debido al riesgo de poder ser reutilizadas para el bioterrorismo.

Se debe estimular el trabajo de investigadores e institutos que se ocupen de estudiar y desarrollar sistemas de IA para la biodefensa; además resulta indispensable exigir a los desarrolladores que certifiquen que sus IA efectivamente presentan riesgos mínimos de daño, a través, por ejemplo, de investigación técnica sobre detección de anomalías adversamente robusta. Debemos establecer obligaciones legales a los desarrolladores de IA para que asuman las consecuencias de posibles errores. Imponer un régimen estricto elevaría la seguridad en los agentes y sistemas.

Para amortiguar los riesgos que se desprenden de las fuertes presiones competitivas que enfrentan gobiernos y corporaciones, será necesario limitar el acceso a sistemas de IA potentes y estimular la

cooperación multilateral. La regulación, necesariamente proactiva, deberá impulsar el desarrollo de una elevada cultura de seguridad, considerando para ello, los estímulos idóneos. Para asegurar transparencia y responsabilidad en los desarrolladores, la documentación de datos deberá ser obligatoria. La supervisión humana en decisiones importantes debe ser obligatoria, La IA no puede ser totalmente autónoma. Resulta indispensable establecer tratados internacionales en la materia y, en particular, protocolos de ciberseguridad para evitar el desarrollo de una carrera armamentista. Debemos tener presente que la IA puede ser un efectivo "contrairritante" de la IA. Ello significa que también podríamos valernos de la IA para evitar sus excesos, reducir riesgos y amenazas.

Apoyos

El proyecto de investigación I+D+i "Ecosistemas de innovación en las industrias de la comunicación: actores, tecnologías y configuraciones para la generación de innovación en contenido y comunicación. INNOVACOM", financiado por la Agencia Estatal de Investigación.

Referencias

Alkhazaleh, M., Khasawneh, M. A. S., Alkhazaleh, Z. M., Alelaimat, A. M., y Alotaibi, M. M. (2022). An Approach to Assist Dyslexia in Reading Issue: An Experimental Study. *Przestrzeń Społeczna (Social Space)*, *22*(3), 133-151. https://go.revistacomunicar.com/Rf6ORo

Aluthman, E. S. (2024). An Investigation of Artificial Intelligence Tools in Editorial Tasks among Arab Researchers Publishing in English. *Eurasian Journal of Applied Linguistics*, *10*(1), 174-185. https://go.revistacomunicar.com/AM0QU8

Anton, C., y Strate, L. (2012). *Korzybski and--*. New York: Institute of General Semantics. https://go.revistacomunicar.com/JRSdky

Blatt, R. (2020). *Historia reciente de la verdad*. Turner

Bolter, J. D., y Grusin, R. (1999). *Remediations: Understanding New Media*. MIT Press. https://go.revistacomunicar.com/tRqBoS

Bostrom, N. (2020). Human genetic enhancements: a transhumanist perspective. En *The ethics of sports technologies and human enhancement* (pp. 339-352). Routledge. https://doi.org/10.4324/9781003075004-29

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press. https://go.revistacomunicar.com/ett2rL

Bostrom, N., y Yudkowsky, E. (2018). The Ethics of Artificial Intelligence. En R. V. Yampolskiy (Ed.), *Artificial Intelligence Safety and Security* (pp. 57-69). Chapman and Hall/CRC. https://doi.org/10.1201/9781351251389-4

Chomsky, N. (2017). *Quem manda no mundo?* Editora Planeta do Brasil

Copeland, B. J., y Proudfoot, D. (2004). The Computer, Artificial Intelligence, and the Turing Test. En C. Teuscher (Ed.), *Alan Turing: Life and Legacy of a Great Thinker* (pp. 317-351). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05642-4_13

Darwin, C. (2010). *El origen de las especies*. Editorial Porrúa. https://go.revistacomunicar.com/XKurVZ

Gordon, T. (2003). *Marshall McLuhan's Understanding Media*. Ginko Press. https://go.revistacomunicar.com/TimVn5

Gould, S. J., y Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology*, *8*(1), 4-15. https://doi.org/10.1017/S0094837300004310

Harari, Y. N. (2016). *Homo Deus: Breve historia del mañana*. Debate. https://go.revistacomunicar.com/xApVMo

Haugen, F. (2023). *La verdad sobre Facebook*. Deusto. https://go.revistacomunicar.com/dCVqld

Hendrycks, D. (2023). Natural Selection Favors AIs over Humans. *arXiv e-prints*, arXiv: 2303.16200. https://doi.org/10.48550/arXiv.2303.16200

Hendrycks, D., Mazeika, M., y Woodside, T. (2023). An Overview of Catastrophic AI Risks. *arXiv e-prints*, arXiv-2306. https://doi.org/10.48550/arXiv.2306.12001

Kaiser, B. (2019). *La dictadura de los datos*. Harper-Collins. https://go.revistacomunicar.com/iifqO2

Kauffman, S. (1995). *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. Oxford University Press. https://go.revistacomunicar.com/ol6w3a

Kissinger. (2022). Heath Forest as a Source of Medicinal Plants for the Maanyan Dayak Tribe in Central Kalimantan, Indonesia: Deforestation and its Relationship to Medicinal Plant Biodiversity. *AgBioForum*, *24*(2), 187-195. https://go.revistacomunicar.com/Tvxo90

Korzybski, A. (1993). *Science and Sanity: An Introduction to Non-aristolian Systems and General Semantics*. Institute of General Semantics. https://go.revistacomunicar.com/60k6VH

Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P., y Schroeder, D. T. (2021). Don't trust your eyes: image manipulation in the age of DeepFakes. *Frontiers in Communication*, *6*, 632317. https://doi.org/10.3389/fcomm.2021.632317

Levinson, P. (1999). *Digital McLuhan: A Guide to the Information Millennium*. Routledge. https://doi.org/10.4324/9780203164341

Logan, R. K. (2004). *The Alphabet Effect: A Media Ecology Understanding of the Making of Western Civilization*. Hampton Press. https://go.revistacomunicar.com/WecRIA

Logan, R. K. (2013). *McLuhan Misunderstood*. Key Publishing House Inc. https://go.revistacomunicar.com/ao4kMh

Logan, R. K. (2016). *Understanding New Media*. Peter Lang. https://go.revistacomunicar.com/zg0QVq

Luhmann, N. (1995). *Social Systems*. stanford university Press.

McCarthy, J., Minsky, M. L., Rochester, N., y Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

McLuhan, M. (1962). *The Gutenberg Galaxy: The Making of Typographic Man*. University of Toronto Press. https://go.revistacomunicar.com/pbwNOQ

McLuhan, M. (1964). *Understanding Media: The Extension of Man*. McGraw-Hill.

McLuhan, M., y Carson, D. (2003). *The Book of Probes*. Ginko Press.

McLuhan, M., y Fiore, Q. (1967). *The Medium is the Message*. Ginko Press. https://go.revistacomunicar.com/pAdHZ4

McLuhan, M., y McLuhan, E. (1998). *Laws of Media: The New Science*. University of Toronto Press. https://go.revistacomunicar.com/PUnAI3

Merzlyakov, S. (2022). Posthumanism vs. Transhumanism: From the "End of Exceptionalism" to "Technological Humanism". *Herald of the Russian Academy of Sciences*, *92*(Suppl 6), S475-S482. https://doi.org/10.1134/s1019331622120073

Meyrowitz, J. (1985). *No Sense of Place*. Oxford University Press. https://go.revistacomunicar.com/bq9Gib

Minsky, M. L., y Papert, S. A. (1969). *Perceptrons: An Introduction to Computational Geometry*. The MIT Press. https://go.revistacomunicar.com/omon8T

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4), 87. https://doi.org/10.1609/aimag.v27i4.1911

Mulyani, S., Suparno, S., y Sukmariningsih, R. M. (2023). Regulations and Compliance in Electronic Commerce Taxation Policies: Addressing Cybersecurity Challenges in the Digital Economy. *International Journal of Cyber Criminology*, *17*(2), 133-146. https://go.revistacomunicar.com/Oi0tku

Ong, W. (1982). *Orality and Literacy*. Methuen. https://go.revistacomunicar.com/tdOvwt

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press. https://go.revistacomunicar.com/G436TP

Phooi, C. L., Azman, E. A., Ismail, R., y Tongkaemkaew, U. (2022). Call Home Gardening for Enhancing Food in the Urban Area. *Future of Food: Journal on Food, Agriculture & Society*, *10*(6), 1-11. https://doi.org/10.17170/kobra-202210056933

Postman, N. (1974). Media ecology: Communication as context.

Postman, N. (1970). The Reformed English Curriculum. En A. C. Eurich (Ed.), *High School 1980: The Shape of the Future in American Secondary Education* (pp. 160-168). New York: Pittman. https://go.revistacomunicar.com/ELPU2l

Postman, N. (1992). *Technopoly: The Surrender of Culture to Technology*. New York: Alfred A. Knopf. https://go.revistacomunicar.com/wY9qE9

Ramos Pollán, R. (2020). Perspectivas y retos de las técnicas de inteligencia artificial en el ámbito de las ciencias sociales y de la comunicación. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, *13*(1), 21-34. https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.7774

Rovira, J. V., Merzero, A., y Laucirica, A. (2022). Construction of a perceptive scale to evaluate the quality of the singing voice: Construcción de una escala perceptiva para la evaluación de la calidad de la voz cantada. *Electronic Journal of Music in Education*, (49), 121-138. https://go.revistacomunicar.com/pP4eGi

Schwab, K. (2016). *La cuarta revolución industrial*. Debate. https://go.revistacomunicar.com/JpP5Ff

Snowden, J., Hernandez, D., Quintrell, J., Harper, A., Morrison, R., Morell, J., y Leonard, L. (2019). The US Integrated Ocean Observing System: governance milestones and lessons from two decades of growth. *Frontiers in Marine Science*, *6*, 242. https://doi.org/10.3389/fmars.2019.00242

Strate, L. (2006). *Echoes and Reflections: On Media Ecology as a Field of Study*. Hampton Press. https://go.revistacomunicar.com/lfjRoR

Strate, L., y Wachtel, E. (2005). *The Legacy of McLuhan*. Hampton Press. https://go.revistacomunicar.com/9ESQYc

Susskind, L. (1994). Strings, black holes, and Lorentz contraction. *Physical Review D*, *49*(12), 6606-6611. https://doi.org/10.1103/PhysRevD.49.6606

Susskind, L. (1999). Holography in the flat space limit. *AIP Conference Proceedings*, *493*(1), 98-112. https://doi.org/10.1063/1.1301570

Susskind, L. (2003). Superstrings. *Physics World*, *16*(11), 29. https://doi.org/10.1088/2058-7058/16/11/35

Susskind, L. (2008). *The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics*. Little, Brown and Company. https://go.revistacomunicar.com/LVTKNN

Tucker, J. A. (2023). Computational Social Science for Policy and Quality of Democracy: Public Opinion, Hate Speech, Misinformation, and Foreign Influence Campaigns. En E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli, y M. Vespe (Eds.), *Handbook of Computational Social Science for Policy* (pp. 381-403). Springer International Publishing. https://doi.org/10.1007/978-3-031-16624-2_20

Warakulsalam, N., y Chokprajakchat, S. (2022). Policy and Project in Reducing Unrest Situation in The Southern Border Provinces of Thailand. *International Journal of Criminal Justice Sciences*, *17*(2), 75-90. https://go.revistacomunicar.com/cqbV0j

Widajanti, E., Nugroho, M., y Riyadi, S. (2022). Sustainability of Competitive Advantage Based on Supply Chain Management, Information Technology Capability, Innovation, and Culture of Managers of Small and Medium Culinary Businesses in Surakarta. *The Journal of Modern Project Management*, *10*(2), 82-93. https://go.revistacomunicar.com/52pYp9

Wolfe, T. (2010). Foreword. En S. McLuhan y D. Staines (Eds.), *Understanding Me: Lectures and Interviews*. The MIT Press.