

Inteligencia artificial: empezamos

Artificial intelligence: we start

Hernán Mavrommatis

Universidad Nacional de La Matanza

mavromou@gmail.com

Resumen

¿La inteligencia artificial es como nos muestran las películas y series de ciencia ficción? ¿La inteligencia artificial es comparable con la inteligencia humana? ¿Es tarde o temprano para preocuparnos por nuestro futuro? ¿Existen riesgos actuales a los que debemos prestar atención? ¿De qué manera las organizaciones y las personas están adoptando la inteligencia artificial? En este trabajo se prevé dialogar con estas preguntas brindando ejemplos sencillos con un lenguaje claro y ameno para aquellas personas que no poseen conocimientos previos en materia de tecnología. Se intentará que las personas que por su experiencia y/o profesión no tienen formación en tecnología, y que observan desde “afuera” los últimos avances en inteligencia artificial, puedan comprender de una manera simple las implicancias de dichos avances, sin dejar de identificar las principales discusiones que actualmente se encuentran dando los referentes del tema.

Palabras clave: Inteligencia artificial; tecnología; futuro

Abstract

Is artificial intelligence what science fiction movies and series show us? Is artificial intelligence comparable to human intelligence? Is it too late or too late to worry about our future? Are there current risks we should pay attention to? How are organizations and people adopting artificial intelligence? This work plans to discuss these questions by providing simple examples with clear and enjoyable language for those people who do not have prior knowledge of technology. An attempt will be made to ensure that people who, due to their

experience and/or profession, do not have training in technology, and who observe the latest advances in artificial intelligence from the "outside", can understand in a simple way the implications of said advances, without failing to identify the main ones discussions that the referents of the topic are currently giving.

Keywords: Artificial intelligence; technology; future

Sección: Ensayos

Recibido: 29/09/2023

Aceptado: 15/03/2024

Introducción

¿La inteligencia artificial con la que hoy interactuamos es como nos muestran las películas y series de ciencia ficción? ¿La inteligencia artificial hace cosas increíbles, pero aun así su inteligencia es comparable con la inteligencia humana? ¿Es tarde o temprano para preocuparnos por nuestro futuro? ¿Existen riesgos actuales a los que debemos prestar atención ya? ¿De qué manera las organizaciones y las personas están adoptando la inteligencia artificial? Ocuparnos de alcanzar algunas respuestas, así como ser conscientes de realizarnos éstas y otras preguntas, forman parte de lo que cada vez más personas empezamos a hacer y debemos seguir haciendo en materia de inteligencia artificial.

Las discusiones que ya empezamos a tener

Para aquellas personas que viven desde “afuera” este desenfadado desarrollo de un conjunto de tecnologías que se agrupan bajo el concepto de inteligencia artificial (IA). Es decir, para aquellas personas que por su experiencia o profesión no pertenecen a la rama de la ciencia y tecnología, y se encuentran permanentemente con aplicaciones y sus creaciones, probablemente las perciben como si fuera magia... pero claramente no lo son. Para ustedes están escritas estas líneas.

Si la persona que lee pertenece a ese grupo en el que probablemente su primer acercamiento con la inteligencia artificial fueron las series y películas. Dependiendo de la edad, puede haber experimentado desde HAL 9000 en *2001: Odisea del Espacio* (1968), pasando por *Terminator* (1984), *I.A. Inteligencia Artificial* (2001), hasta la serie *Mejores que nosotros* (2018), y la lista sigue...

Pero existe una particularidad. Cuando empezamos a conocer la IA desde el mundo de la ciencia ficción, hay un enfoque muy marcado, una asociación con los robots, mejor dicho, con los androides, porque en la mayoría de estos ejemplos se basan en robots que poseen cuerpos humanos. Igualar androides al concepto de IA está muy bien para las películas y series, pero en la vida real existe un concepto que se llama *embodiment* (Pfeifer e Iida, 2004), su traducción sería corporización o encarnación. El *embodiment* es una perspectiva que requiere trabajar con sistemas físicos del mundo real, es decir, robots (p. 4). En otras palabras, el *embodiment* consiste en vincular un modelo de IA entrenado en un mundo simulado en un

cuerpo robótico, generalmente de forma humana. Existen en la actualidad ejemplos de la empresa Google DeepMind (Santana y González, 2023). En realidad, la esencia tecnológica de las IA's no requieren necesariamente robots o androides como surge de la ciencia ficción, las inteligencias artificiales se basan en líneas de código escritas en algún lenguaje de programación. Uno de los más utilizados es Python.

Entonces lo que “ya empezamos”, para parafrasear el título de este documento (si pronunciamos la “y” griega como una vocal), es justamente empezar a comprender que las IA's son líneas de código que se ejecutan en una computadora. Pero no como cualquier programa de computación, sino con la particularidad que este código permite ser entrenado creando modelos de redes neuronales artificiales tan complejos que difícilmente sus mismos desarrolladores pueden explicar cómo dichos modelos llegan a sus resultados. Esto último es el caso del *Deep Learning* (Aprendizaje Profundo), una de las tecnologías de IA que más desarrollos tuvo y está teniendo en la actualidad.

Desde hace algunos años, y particularmente en estos últimos dos años, hemos visto y oído muchas cosas que, aun cuando muchos de nosotros estamos acostumbrados a ver la velocidad exponencial con la que evoluciona la tecnología (Kurzweil, 2015), nos sigue pareciendo increíble. Pero no solo la velocidad que tienen los desarrollos de IA, sino la cantidad de cosas increíbles que estas tecnologías pueden hacer.

Hace veinticinco años que soy profesor, comencé enseñando conceptos duros como hardware de PC, electrónica, teoría de circuitos y ese tipo de cosas, pero en los últimos quince años me volqué a conceptos blandos como la innovación, la motivación, la creatividad, el humor, etc. Nunca me alejé de la tecnología, parte de mi actividad profesional se mantuvo y mi vuelco académico se replicó en otros ámbitos y pase de ser de aquellas personas que trabajan con equipos y dispositivos de hardware y lenguajes de programación de bajo nivel, a centrarme en la gestión de la tecnología. Así fue como me convertí en master de gestión de servicios tecnológicos. Cuento esto porque en mis veinticinco años de escribir apuntes para clases, posteo de blogs y recientemente artículos en revistas, papers, capítulos de libros y otros documentos de trabajo, jamás me pasó lo que me pasó colaborando en la escritura del documento *Recomendaciones para una Inteligencia Artificial Fiable*¹ (Argentina JGM, 2023),

¹ Guía de acceso libre:

https://www.argentina.gob.ar/sites/default/files/2023/06/recomendaciones_para_una_inteligencia_artificial_fiable.pdf

que toma como referencia los principios de la UNESCO sobre la ética de la IA (UNESCO, 2022), que fueron acordados por sus estados miembros (más de 190 países).

Cuando hablamos de inteligencia artificial no estamos hablando de una inteligencia que se compare con la inteligencia humana en todo su potencial y en el amplio espectro cognitivo humano. Las IA's que hoy conocemos y con la que muchos de nosotros interactuamos son estrechas, es decir, hacen muy bien una tarea muy puntual. Por ejemplo, una IA que juega al ajedrez y puede ganarle hasta al más experimentado humano, no puede jugar al Go (juego de tablero de origen asiático). De la misma manera, las IA's que pueden chatear no pueden conducir autos, identificar patrones o anomalías en imágenes satelitales... y así la lista sigue. Este concepto de estrechez generalmente está asociado a una tarea específica y ésta involucra de la misma manera un tipo de dato específico. Por ejemplo, las IA's que reconocen imágenes se las entrena con imágenes, las que reconocen el lenguaje natural se las entrena con lenguaje natural, o sea texto escrito, las que reconocen la voz humana se las entrena con audios de voces de diferentes personas, y así sucesivamente. Frecuentemente asocio el concepto de inteligencia estrecha con cada uno de los distintos tipos de *inteligencias múltiples* (Gardner, 2005). Para que se entienda el concepto, Gardner propuso reformular el concepto de inteligencia humana pasando de una medida escalar, conocida como coeficiente de inteligencia, que suele medir aspectos lógico-matemáticos, correspondientes con una de las múltiples inteligencias que definió dicho autor, la lógico-matemática. Pero la inteligencia humana no empieza y termina en esta parte del espectro cognitivo humano, es mucho más amplia y se pueden definir otros tipos de inteligencias, la lingüística, la musical, espacial, etc. Las personas poseemos un mix de estas inteligencias desarrolladas en diferentes grados. En cambio las inteligencias artificiales ocupan uno o dos tipos de inteligencias pero no todas. Asimismo, algunas inteligencias como la intrapersonal y la interpersonal son muy difíciles de sintetizar dado que las IA's no poseen conciencia, agencia, conocimiento del mundo y sensibilidad (Vallor, 2022).

Pero mientras estaba escribiendo este documento, decía líneas arriba, Open AI (empresa de Elon Musk) había presentado a Chat GPT-4² a través de una demostración en vivo y la publicación de un paper de más de noventa hojas. En esta nueva versión, el gran modelo de procesamiento natural de lenguaje (que es la categoría donde se incluyen este tipo de IA's),

² Acceso a través de <https://chat.openai.com/>

ahora era *multimodal*. La multimodalidad es una propiedad que, no hace mucho, comenzaron a tener las IA's (Li et al., 2023) y se basa en la capacidad de realizar tareas con dos tipos de datos diferentes. Por ejemplo, en el caso de Chat GPT-4 lenguaje natural escrito e imágenes (OpenAI, 2023), incluso hasta puede reconocer el texto dentro de las imágenes, ¡una locura! No hubo otra opción que reescribir lo que estaba explicando, ajustándome a la reciente evolución de la tecnología. Es decir, incluso las personas que, como yo, hace muchos años estamos cerca de la tecnología, ya comenzamos a ver y oír cosas que nos parecen increíbles a una velocidad aún más rápida a lo que estábamos acostumbrados.

Pero el hecho de que las tecnologías evolucionen de manera exponencial no es algo nuevo. Si observamos en un amplio periodo histórico de miles de años atrás, por ejemplo, desde el surgimiento de la agricultura, y graficamos los inventos tecnológicos a partir de entonces, vamos a ver que cada vez son más frecuentes y complejos mostrando una curva exponencial. Esto último fue advertido por expertos de distintas disciplinas, historiadores, filósofos, tecnólogos, líderes de empresas tecnológicas, académicos y demás actores (Asimov, 2007; Harari, 2014; Kurzweil, 2015). Rápidamente cada uno de ellos, desde su enfoque particular, empezó a ver los potenciales peligros de la IA (Cortés, 2017). Sus preocupaciones comenzaron a divulgarse y no tardaron en aparecer libros, charlas y congresos que abordaron el tema. Uno de los más conocidos es la Conferencia de Asilomar, donde muchos de los referentes antes mencionados acordaron distintos principios (Future of Life Institute, 2017) que buscan mitigar estos eventuales peligros. Por supuesto, tampoco tardaron en aparecer dos bandos bien definidos: aquellos que tienen una mirada más apocalíptica de la IA y otros que son más optimistas. Entre los distintos peligros potenciales que se especulan está el de la creación de lo que se llama la *IA general*, también conocida como *IA fuerte*, que a diferencia de la IA estrecha (o débil), sí es comparable con la inteligencia humana. Algunos expertos especulan que entre 2040 y 2070 podría aparecer este tipo de IA, lo que podría a su vez acelerar la existencia de algo temido por los expertos: la *Super Inteligencia Artificial* (Bostrom, 2016).

Pero estas son, por ahora, solo especulaciones (siempre hay que agregar un paréntesis que diga “por ahora” cuando se habla de evolución de IA). Por otro lado, si nos basamos en la historia de las predicciones de muchas personas pioneras en el ámbito de las IA que se pronosticaron en los años 70, no deberíamos preocuparnos (Lerner y Mavrommatis, 2021). No obstante, no por ello son menos probables o peligrosas. Existen muchos experimentos

mentales que sirven para reflexionar sobre lo que podría llegar a ocurrir en caso de que exista esta Super IA, que incluso hasta podríamos no crearla nosotros sino ser una creación, o evolución de la IA general o fuerte. Pueden buscar en internet estos experimentos mentales, algunos son divertidos e interesantes y otros aterradores (Tierraseca Blesa, 2023). Pero más allá de todo, muchos expertos acuerdan que desde el momento en que se desarrollan estas tecnologías se debe buscar la manera de incluir los valores humanos, de manera tal de que cualquiera sean los objetivos que estas IA's evolucionadas tengan, no vayan en contra de los valores humanos (Future of Life Institute, 2017). Estas cuestiones están al alcance de cualquier persona curiosa que quiera saber más sobre este tema, y cada vez hay una mayor divulgación de estos peligros, por lo cual muchos de nosotros ya empezamos a preocuparnos por nuestro futuro.

Pero además existen riesgos en el diseño y uso de las IA, que representan actuales “peligros”, no tan apocalípticos como los de antes pero sí que pueden perjudicar en distinto grado a un conjunto de personas, o bien, favorecer a unas personas sobre otras. Estos riesgos están comprendidos básicamente por dos problemas, los *sesgos* y la *explicabilidad* (Argentina JGM, 2023). Los sesgos que pueden tener las IA's, son parecidos a nuestros sesgos cognitivos, y tiene lógica: después de todo, los datos que usamos para entrenar a las IA's y crear modelos de redes neuronales artificiales no son otra cosa que el registro digital de las acciones y creaciones humanas. Con esto último podemos ver claramente que la calidad de los datos que se utilicen para entrenar a estas IA's es clave. Hay una frase que se suele usar en la jerga y que resulta una analogía muy esclarecedora: “entra basura, sale basura”. Se refiere a que si las IA's se entrenan con mala calidad de datos, los resultados serán igual de malos.

Existen muchos ejemplos de sesgos. Los más divulgados son aquellos que reflejan problemas de diversidad de razas y/o género en los datos de entrenamiento. Es decir, no responden bien con ciertas razas (BBC, 2015) y/o tipos de género (Logically, 2019) y brindan respuestas erráticas, que muchas veces son injustas, además de inútiles. Anteriormente hicimos referencia a la explicabilidad cuando decíamos que es muy difícil para los desarrolladores de modelos de IA explicar cómo esos modelos llegan a esos resultados. Dada una determinada audiencia, la explicabilidad se refiere a los detalles y razones que da un modelo para que su funcionamiento sea claro o fácil de entender (Arrieta et al., 2020). En otras palabras, los modelos de redes neuronales artificiales pueden llegar a ser tan complejos que resulta muy difícil, por no decir imposible (hasta ahora), encontrar una explicación Sin

embargo, existen maneras de mejorar la explicabilidad, como por ejemplo podría ser el caso de uso de algoritmos más sencillos de lo que se conoce como *Machine Learning* (Aprendizaje de máquina o aprendizaje automático). Ahora las personas lectoras se preguntarán: ¿y por qué no utilizan algoritmos más explicables? Porque resulta que los algoritmos más explicables pueden hacer menos cosas y/o poseen menos potencia. A medida que el uso de las IA's se va haciendo masivo, cada vez son más las personas que están empezando a padecer estos riesgos. En consecuencia, corresponde que además de las divulgaciones de las bondades y potencialidades que poseen las IA's empecemos a concientizar respecto de estos riesgos.

Está claro que cada vez somos más personas que ya empezamos a adoptar IA para nuestro entretenimiento, para trabajar en nuestra profesión, así como también en distintos ámbitos de nuestra vida, nuestra educación, nuestra salud y nuestra seguridad. Es por ello que resulta interesante identificar algunas maneras en que tanto las organizaciones como las personas adoptamos la IA (Lerner y Mavrommatis, 2021). El modelo de adopción más conocido, y temido –hay que decirlo–, es la automatización. Este modelo busca reemplazar tareas humanas que suelen estar tipificadas, ser repetitivas, tienen muy baja variabilidad, se realizan siempre con los mismos parámetros de entrada y con los mismos propósitos. El problema es: ¿qué pasa con las personas cuyo trabajo principal se basa en este tipo de tareas? Esta pregunta se puede responder de varias maneras. Las mejores prácticas apuntan a un concepto que se llama *reskilling* (recapitación o reentrenamiento). La idea consiste en capacitar a esas personas desplazadas por la automatización, para poder reasignarlas, brindándoles conocimientos y desarrollándose en habilidades tales como el pensamiento crítico, el pensamiento creativo, la resolución de problemas nuevos, etc.

Otro modelo de adopción se lo conoce como *augmented intelligence* (inteligencia aumentada). Este modelo busca el trabajo colaborativo humano-máquina tratando de que la IA amplifique la inteligencia de las personas. Esto implica un diseño específico de IA's con las que las personas puedan interactuar para mejorar sus tareas. Un ejemplo, entre miles, puede ser un profesional de la medicina que debe dar un diagnóstico. La IA podría analizar distintas imágenes de tomografías, radiografías y/o resonancias magnéticas para encontrar anomalías indistinguibles para el ojo humano o descubrir patrones escondidos que respondan a una patología específica, y sugerir estos resultados una persona con conocimientos médicos. Esta persona, entonces, puede acceder a información muy valiosa que de no ser por una IA no tendría (o al menos no tan rápido y/o de forma precisa). Este profesional, por ejemplo un

médico, podría además comprender el contexto de su paciente, empatizar con él, revisar su historia clínica, charlar y observar su ánimo, comportamiento, movimientos, quejas y demás cuestiones propias de los humanos que se relevan con las prácticas médicas tradicionales y estar en una mejor posición para brindar un diagnóstico y tratamientos ajustados. Cada integrante de la dupla humano-máquina hace lo que mejor sabe hacer colaborando entre sí para llegar a un mejor resultado si se lo compara con el trabajo por separado (solo humano o solo máquina).

Conclusiones

Desde hace decenas de miles de años la explosión cognitiva permitió a los humanos comenzar a utilizar el lenguaje, trabajar de manera colaborativa y desarrollar tecnología. Desde entonces las personas no hemos dejado de crear. Esas creaciones se fueron complejizando a lo largo del tiempo, se convirtieron en distintas tecnologías que convergen unas con otras dando lugar a verdaderas revoluciones tecnológicas que cambiaron la historia humana para siempre. A partir de allí hemos tenido una interacción con la tecnología que fue convirtiéndose en una interdependencia, los humanos creamos tecnología y ella moldea la manera en que nosotros vivimos.

La tecnología de IA no es como cualquier otra tecnología (Lerner y Mavrommatis, 2021). Sus particularidades nos presentan un desafío pragmático como humanos y filosófico como especie, algunos referentes del tema sostienen que podría peligrar la propia supervivencia de la humanidad, o mínimamente nuestro dominio sobre el planeta. Es muy difícil hacer pronósticos al respecto, solo sabemos que existen bondades y riesgos actuales, así como potenciales ventajas y peligros. Solo podemos actuar de la mejor manera posible con la información que hoy tenemos para evitar futuros distópicos, y aunque no seamos conscientes de cómo construir nuestro futuro con las IA's, ya empezamos.

Referencias bibliográficas

Argentina. Jefatura de Gabinete de Ministros. (2023). *Recomendaciones para una inteligencia artificial fiable*. Secretaría de Innovación Pública.

https://www.argentina.gob.ar/sites/default/files/2023/06/recomendaciones_para_una_inteligencia_artificial_fiable.pdf

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Benetot, A., Tabik, S., Barbado, A.... y Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>

Asimov, I. (2007). *Historia y cronología de la ciencia y los descubrimientos*. Grupo Planeta (GBS).

BBC. (julio 2015). Google apologizes for Photos app's racist blunder. BBC News. <https://www.bbc.com/news/technology-33347866>

Bostrom, N. (2016). *Superinteligencia: Caminos, peligros, estrategias*. Tell.

Cortés, J. (diciembre 2017). Nick Bostrom: “No tendremos una segunda oportunidad con la inteligencia artificial”. *El País*. https://elpais.com/retina/2017/12/02/tendencias/1512231406_905237.html

Future of Life Institute. (agosto, 2017). *AI Principles, future of life*. <https://futureoflife.org/open-letter/ai-principles/>

Gardner, H. (2005). *Inteligencias múltiples*. Paidós.

Harari, Y. N. (2014). *Sapiens. De animales a dioses: Una breve historia de la humanidad*. Debate.

Kurzweil, R. (2015). *La singularidad está cerca: cuando los humanos transcendamos la biología*. Lola Books.

Lerner, A. M. y Mavrommatis, H. (2021). Inteligencia Artificial (IA): perspectivas para integrar el valor de la innovación y la creatividad con la gestión en las organizaciones. En *Inteligencia Artificial una mirada interdisciplinaria* (141-170). Academia Nacional de Ciencias Morales y Políticas. Buenos Aires, Argentina.

Li, J., Li, D., Savarese, S. y Hoi, S. (2023). *Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models*. arXiv. <https://arxiv.org/pdf/2301.12597.pdf>

Logically. (julio 2019). *5 Examples of Biased Artificial Intelligence*. Logically. <https://www.logically.ai/articles/5-examples-of-biased-ai>

OpenAI. (2023). *GPT-4 Technical Report*. ArXiv. <https://doi.org/10.48550/arXiv.2303.08774>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2022). *Recomendación sobre la ética de la inteligencia artificial*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Pfeifer, R. y Iida, F. (2004). Embodied artificial intelligence: Trends and challenges. *Lecture notes in computer science*, 1-26. https://link.springer.com/chapter/10.1007/978-3-540-27833-7_1

Rehana, H., Çam, N. B., Basmaci, M., Zheng, J., Jemiyo, C., He, Y., Özgür, A. y Hur, J. (2023). Evaluation of GPT and BERT-based models on identifying protein-protein interactions in biomedical text. arXiv. <https://arxiv.org/ftp/arxiv/papers/2303/2303.17728.pdf>

Santana, C. y González, D. (mayo 2023). *Los robots de DeepMind ya juegan al fútbol [Video]*. Youtube. <https://youtu.be/9NFiE3HdP5g?feature=shared>

Tierraseca Blesa, L. (2023). *VOYAGE. Concept art y artbook para un videojuego de ciencia ficción* [Trabajo Final de Grado, Universitat Politècnica de València]. <https://riunet.upv.es/handle/10251/199729>

Vallor, S. (2022). *Innovation with Intelligence: Challenges for Living Well with AI*. Centre for technomoral future, University of Edinburgh.