



Enfrentando los retos de la inteligencia artificial: ética, transparencia y futuro

Facing the challenges of artificial intelligence: ethics, transparency and the future

Marlene L. Antezana Soria Galvarro¹

¹Docente Investigador, Facultad de Medicina, Universidad Mayor de San Simón. Cochabamba, Bolivia.

<https://orcid.org/0000-0002-3170-5375>

Correspondencia a: Marlene L. Antezana Soria Galvarro

Correo electrónico: arisa9799@gmail.com

Los últimos avances de la tecnología, específicamente la Inteligencia Artificial (IA), no terminan de sorprendernos y ya hay novedades de nuevos adelantos, lo cual es imposible detener y la sociedad, de una u otra forma, es afectado en lo laboral o incluso en su cotidianidad.

El denominativo de Inteligencia Artificial fue aplicado por primera vez por Jhon McCarthy en 1956 en una convención donde se trató sobre el lenguaje natural de algunas maquinas¹.

A lo largo del tiempo e históricamente, hubieron eventos sorprendentes como el Deep Blue que en 1997 ganó un partido de ajedrez al campeón Kasparov², esto es tan solo un ejemplo de los muchos sucesos donde la IA intervino.

Es producto de la IA el Transformador Generativo Preentrenado (GPT) que se lanzó en junio de 2018 con 117 millones de parámetros³, a partir del cual surgieron las diferentes generaciones del GPT, hasta la versión actual. No conviene que se escriba el número de generación, porque que de la hora actual que se escribe este editorial, hasta que se publique, es posible que cambie el GPT a una generación superior.

¿Pero qué es exactamente la IA?, solicitamos definirlo al mismo chatGPT: la inteligencia artificial (IA) se refiere a la capacidad de las máquinas o sistemas informáticos para realizar tareas que normalmente requieren de la inteligencia humana. Estas tareas incluyen el aprendizaje, el razonamiento, la resolución de problemas, la percepción visual, el reconocimiento de voz, el lenguaje natural y la toma de decisiones.

No aplicar la IA en las diferentes áreas del conocimiento implica perderse de las bondades de los adelantos, si, es cierto, pero también es necesario considerar los posibles problemas que se puede tener al utilizar la IA.

Los profesionales y estudiantes de las diversas áreas de la ciencia deben considerar la posibilidad de información errónea, donde la IA puede haberse basado en información inexacta por resultados sesgados⁴.

Plantear una pregunta para que el chatGPT nos responda, debe evitar hasta minúsculas imprecisiones, lo contrario puede llevar a respuestas diferentes de lo correcto.

Las múltiples diferencias entre personas, géneros, países, continentes, tiempos, circunstancias, labores, intereses, etc., pueden no ser vencidas por la IA, y las respuestas obtenidas pueden carecer de coherencia con la realidad y contexto del que solicita la información.

Los diversos modelos de lenguaje pueden aprender y replicar sesgos presentes en los datos de entrenamiento, esto puede generar conflictos en grandes decisiones como institucionales hasta de estado, pudiendo generar reflejando prejuicios hacia diferentes grupos de población o beneficio a quienes no corresponde, ejemplo de estos hay varios, como la selección de personas factibles de alcanzar un crédito bancario, individuos con mas probabilidad de recaer en hechos delictivos, la atribución de posibilidad de x enfermedad basado en grupo sanguíneo, y así hay varios errores que se presentaron en la historia de la ciencia⁵.

Es muy posible que lo negativo planteado en líneas superiores sea salvado, es lo deseado, sobre todo exigiendo y/o forzando la interpretabilidad. Confiar en las decisiones que toman los algoritmos de la IA, es fundamental, pero la confianza será a partir de la posibilidad de interpretar o explicarnos los algoritmos que llevaron a una determinada decisión⁶.

Lo planteado repercutirá en los procesos de enseñanza y aprendizaje. Los estudiantes portan ordenadores como recurso primordial de su persona, en ese contexto, no estamos lejos de que no haya derecho a desconocer temas encomendados, por tener todas las posibilidades de explorar temas específicos de una asignatura. El desafío estará en el trabajo de verificación de datos, para ello es necesario promover cursos sobre

la interpretación de resultados de inteligencia artificial y la verificación de datos⁷.

Es necesario replantear los nuevos recursos de enseñanza, basado en competencias, donde se considere proporcionar a los estudiantes conjuntos de datos de práctica que puedan explorar y analizar. Esto les permitirá aplicar las habilidades adquiridas para evaluar la calidad y autenticidad de los datos⁸.

Les será de utilidad promover en los estudiantes el estudio y análisis de caso éticos que destaquen situaciones donde la falta de transparencia o la manipulación de datos haya llevado a resultados no confiables. Esto ayudará a los estudiantes a comprender los riesgos asociados⁹.

Fomentar la participación en foros de discusión en línea o en el aula para que los estudiantes puedan intercambiar ideas y experiencias sobre la interpretación de resultados de inteligencia artificial y la verificación de datos.

El diseñar proyectos prácticos que desafíen a los estudiantes a aplicar sus habilidades en la verificación de datos en escenarios del mundo real permitirá practicar y aplicar lo aprendido de manera significativa⁹.

En el área de la investigación, así como existen grandes adelantos para obtener información, así también se progresa a la exigencia de la declaración del uso de IA y su interpretabilidad.

Concluyo reconociendo los grandes beneficios de la IA pero también la necesidad de exigirnos el cuestionamiento de los algoritmos de respuestas.

Referencias bibliográficas

1. Oliver N. Inteligencia artificial, naturalmente: un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos. Inteligencia artificial, naturalmente. 2021. Recuperado de: <https://onti.es/es/publicaciones/Inteligencia-artificial%2C-naturalmente>
2. Adaime I. Kasparov vs. Deep Blue. La conflictiva relación hombre-máquina. 2011. <https://comunicacion.sociales.uba.ar/wp-content/uploads/sites/16/2013/02/Adaime.pdf>
3. Beltrán NCB, Mojica ECR. Procesamiento del lenguaje natural (PLN)-GPT-3: Aplicación en la Ingeniería de Software. Tecnología Investigación y Academia. 2020. Recuperado de: <https://revistas.udistrital.edu.co/index.php/tia/article/view/17323>
4. Franganillo J. Contenido generado por inteligencia artificial: oportunidades y amenazas. Anuario ThinkEPI. 2022. Recuperado de: <https://thinkepi.scimagoepi.com/index.php/ThinkEPI/article/view/91555/66313>
5. Lumbreras S. Inteligencia Artificial y medicina: la necesidad de modelos interpretables. TECHNO REVIEW. International Technology, Science and Society Review/Revista Internacional de Tecnología, Ciencia y Sociedad. 2020. Recuperado de: <https://journals.eagora.org/revTECHNO/article/view/2814/1640>
6. Subercaseaux B. Interpretabilidad de modelos de inteligencia artificial con garantías formales. Revista Bits de Ciencia. 2022. Recuperado de: <https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2019/09/3T19-Quarterly-iDANAE.pdf>
7. Ferrante E. Inteligencia artificial y sesgos algorítmicos: ¿Por qué deberían importarnos? Nueva sociedad. 2021. Recuperado de <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>
8. Amo Filvà D, Fonseca Escudero D, Vernet Bellet D, Torres Gómez ED, Muñoz Pastor P, Caballero Codina V, et al. Usos y desusos del modelo GPT-3 entre estudiantes de grados de ingeniería. En: Actas de las XXIX Jornadas sobre la Enseñanza Universitaria de la Informática: Granada, del 5 al 7 de julio de 2023. Universidad de Granada; 2023. <https://blogs.ua.es/faraonllorens/2023/07/06/usos-y-desusos-del-modelo-gpt-3-entre-estudiantes-de-grad-os-de-ingenieria/>
9. López Guillermon JC. Desenmascarando datos: Igualdad e inteligencia artificial. Revista IUS. 2021. Recuperado de: <https://www.scielo.org.mx/pdf/rius/v15n48/1870-2147-rius-15-48-137.pdf>