Los riesgos de la Inteligencia Artificial

¿En qué ámbitos concretos de nuestra sociedad asoman amenazas a raíz de ella? De momento, es importante abordar los problemas éticos que se derivan de su desarrollo y uso: se requieren además regulaciones adecuadas y una rendición de cuentas clara.

SEBASTIÁN RÍOS
Académico de Ingeniería Industrial,
Universidad de Chile

La Inteligencia Artificial (IA) ha experimentado un rápido avance en los últimos años, revolucionando la forma en que interactuamos con la tecnología y transformando diversas industrias. Sin embargo, a medida que la IA se vuelve más sofisticada y se integra en nuestras vidas de maneras cada vez más intensas, también se plantean preocupaciones y riesgos significativos para nuestra civilización.

El potencial de la IA es innegable. Con algoritmos de aprendizaje profundo y modelos generativos de diverso tipo, han aparecido sistemas que muestran comportamientos similares a los de un ser humano. Incluso, en el último tiempo, la IA ha demostrado su capacidad para realizar tareas complejas, como el reconocimiento de imágenes, el procesamiento del lenguaje natural, la toma de decisiones y la conducción autónoma. Estas capacidades tienen el potencial de mejorar muchos ciertos aspectos de nuestra vida cotidiana, desde la atención médica y la movilidad hasta la eficiencia energética y la toma de decisiones empresariales. Sin embargo, a medida que la IA se vuelve más avanzada, también surgen riesgos que debemos tener en cuenta. Mencionamos acá algunos de los principales peligros asociados con esta.

1. RIESGOS DE SEGURIDAD Y MANIPULACIÓN: a medida que la IA se vuelve más avanzada y autónoma, existe el riesgo de que se utilice de manera malintencionada o para fines dañinos. Esto podría incluir el uso de IA en ataques cibernéticos, la creación de armas autónomas o la manipulación de información para engañar, influenciar y manipular a las personas.

2. DESPLAZAMIENTO LABORAL: a medida que la IA se vuelve más capaz de ejecutar tareas que antes eran realizadas por humanos, existe el riesgo de que haya una disrupción en el mercado laboral, lo que podría resultar en la pérdida de empleos en ciertos sectores. Esto podría tener implicaciones económicas y sociales, como la desigualdad de ingresos y la necesidad de reentrenar y reubicar a los trabajadores afectados.

3. SESGOS Y DISCRIMINACIÓN: la IA puede perpetuar y amplificar sesgos y discriminación si no se desarrolla y se implementa de manera ética y responsable. Esto podría incluir sesgos en los datos utilizados para entrenar los modelos de IA, lo que podría llevar a decisiones discriminatorias en áreas como la contratación, la atención médica o la justicia penal, con consecuencias injustas para ciertos grupos de población.

4. FALTA DE TRANSPARENCIA Y RESPONSABILIDAD: la opacidad en el funcionamiento interno de los modelos de IA y la falta de rendición de cuentas de las decisiones tomadas por sistemas autónomos de IA pueden plantear riesgos en términos de ética, responsabilidad y confianza en la tecnología. Es importante asegurar que la IA sea transparente, explicable y responsable en su diseño, desarrollo y despliegue.

5. IMPACTOS EN LA PRIVACIDAD Y LA SEGURIDAD DE LOS DATOS: la IA a menudo requiere grandes cantidades de datos para su entrenamiento y funcionamiento, lo que plantea preocupaciones sobre la privacidad y la seguridad de los datos. La mala gestión de los datos utilizados en la IA podría resultar en violaciones de privacidad, filtraciones de datos y pérdida de confidencialidad.

6. ÉTICA Y GOBERNANZA: la toma de decisiones éticas en el desarrollo y uso de la IA plantea desafíos complejos. Entre ellos, la asignación de responsabilidades en caso de errores o daños causados por sistemas de IA, la equidad en el acceso y uso de la tecnología, y la necesidad de regulaciones y marcos éticos adecuados para guiar su desarrollo y aplicación.

Ética y regulaciones

Es importante abordar estos riesgos y desafíos de manera proactiva, mediante la implementación de marcos éticos y regulaciones adecuadas, la promoción de la transparencia y la responsabilidad en la IA. Asimismo, es relevante considerar los impactos sociales y laborales que provoca, así como el compromiso de diversas partes interesadas. Es importante incluir a la sociedad, en general, en el diálogo y la toma de decisiones relacionadas con la IA.

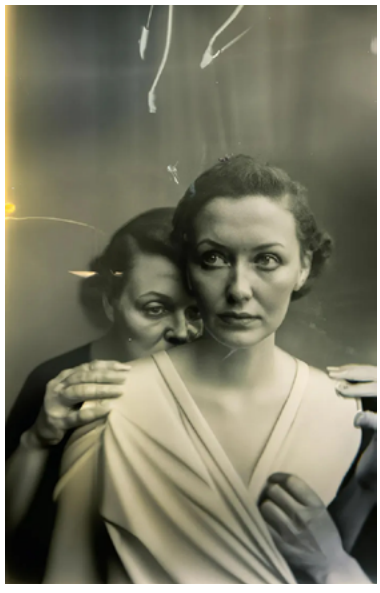
Riesgos para la seguridad

Uno de los riesgos más significativos es la automatización de los ataques cibernéticos. La inteligencia artificial (IA) puede ser entrenada para identificar vulnerabilidades en sis-

temas de seguridad y redes, así como para desarrollar y ejecutar ataques de manera autónoma. Por ejemplo, los atacantes pueden utilizar algoritmos de aprendizaje automático para analizar grandes cantidades de datos en busca de vulnerabilidades en sistemas informáticos, identificar debilidades en contraseñas o desarrollar *malware* personalizado que sea difícil de detectar por parte de las soluciones de seguridad convencionales.

Además, la IA puede ser utilizada para llevar a cabo ataques de ingeniería social, que implican manipular a las personas para obtener información confidencial o acceso a sistemas protegidos. Por ejemplo, los atacantes pueden usar algoritmos de aprendizaje automático para analizar perfiles de redes sociales y crear perfiles falsos que se utilizan para engañar a las personas y así obtener información confidencial, como contraseñas o datos de acceso.

Otro riesgo de seguridad relacionado con la IA es la manipulación de contenido. La IA puede ser utilizada para generar contenido falso o manipulado, como imágenes, videos o texto, con el fin de difundir información errónea o engañosa. Por ejemplo, los llamados «*deepfakes*» son videos o imágenes generados por IA que parecen ser auténticos, pero en realidad son falsos y pueden ser utilizados para difamar a personas, influenciar elecciones o causar daño reputacional.



La reciente imagen ganadora del Sony World Photography Awards , hecha con IA por Boris Eldagsen, que despertó polémica mundial.

Riesgos para el empleo (desplazamiento laboral)

Uno de los principales riesgos de la IA es su impacto en el empleo. A medida que la IA se vuelve más capaz de realizar tareas que anteriormente requerían la intervención humana, existe la preocupación de que los empleos tradicionales sean reemplazados por la automatización, lo que podría resultar en una disrupción masiva en el mercado laboral. Según un informe del Foro Económico Mundial «Future of work. Insights for 2021 and Beyond», se estima que para el año 2025 la automatización podría eliminar 85 millones de empleos en todo el mundo en sectores como la manufactura, el comercio minorista y la administración pública. De igual forma, el mismo estudio estima que surjan 97 millones de nuevos empleos que se adaptarán mejor al nuevo entorno entre humanos, robots y algoritmos.

Pese a esto, surge la inquietud natural asociada a que la tasa de reemplazo de trabajos humanos por IA puede ser mucho mayor a la tasa de reconversión de estos trabajadores. Esto generaría un problema social grave, si no se aborda antes de que esto ocurra. Este cambio podría tener un impacto significativo en la sociedad, con la posibilidad de una mayor desigualdad económica, la pérdida de empleos en comunidades enteras y la necesidad de una reestructuración masiva de la fuerza laboral.

De aquí que es esencial abordar estos desafíos mediante la implementación de políticas y programas que ayuden a la transición de los trabajadores a nuevas oportunidades laborales, como la reeducación y la capacitación en habilidades tecnológicas y de gestión del cambio.

Lamentablemente, en general, los gobiernos son mucho más lentos en reaccionar ante disrupciones tecnológicas y es por ello que urge apresurar el paso en esta dirección. Especialmente en gobiernos de América Latina, donde apenas estamos discutiendo el tema de la digitalización, y en Chile, en particular, donde acaba de publicarse una ley de transformación digital que ya podría venir obsoleta.

Sesgos y discriminación

En primer lugar, la IA puede estar sesgada debido a la calidad y a la representatividad de los datos con los que se entrena. Si los datos utilizados para entrenar a la IA son sesgados o incompletos, la IA puede aprender y perpetuar esos sesgos en sus decisiones. Por ejemplo, si se entrena una IA para contratar empleados utilizando datos de contratación anteriores que tienen sesgos de género o raza, la IA puede perpetuar esos sesgos en su proceso de selección de candidatos, lo que lleva a la discriminación de género u origen racial en la contratación.

En segundo lugar, la IA puede producir sesgos y discriminación si los algoritmos utilizados para desarrollarla no son transparentes o explicables. Muchos algoritmos de IA son cajas negras, lo que significa que sus procesos y decisiones no son claros para los usuarios o desarrolladores. Esto puede dificultar la identificación y mitigación de sesgos y la consiguiente discriminación producida por esta tecnología. Además, si los desarrolladores de la IA no son conscientes de la presencia de sesgos y discriminación en sus algoritmos, es posible que no tomen medidas adecuadas para corregirlos.

En tercer lugar, la IA puede producir sesgos y discriminación, si no se implementan mecanismos de supervisión y regulación adecuados. La falta de regulaciones y estándares claros en el desarrollo y uso de la IA puede permitir la

Si la IA toma decisiones autónomas, ¿quién es responsable cuando esas decisiones son erróneas o perjudiciales? ¿Cómo se asigna la responsabilidad legal y ética en caso de consecuencias negativas?

proliferación de sistemas sesgados y discriminatorios. Por ejemplo, si no se establecen políticas claras para la utilización de la IA en procesos de contratación, préstamos, seguros u otras áreas sensibles, es posible que la IA perpetúe sesgos y discriminación sistemáticamente.

Falta de transparencia y responsabilidad

La inteligencia artificial (IA) puede generar falta de transparencia y responsabilidad en el proceso de toma de decisiones, debido a su complejidad y opacidad. Muchos algoritmos de IA son considerados como «cajas negras», producto de que sus procesos de toma de decisiones no son fácilmente explicables por los desarrolladores o usuarios. Esto significa que puede ser difícil entender cómo se toman las decisiones y por qué se toman de esa manera.

Esta falta de transparencia puede llevar a una falta de responsabilidad en el uso de la IA. Si los usuarios no pueden entender cómo se toman las decisiones, es posible que no puedan identificar errores o sesgos en el proceso de toma de decisiones de la IA. Esto puede llevar a decisiones equivocadas o discriminatorias y puede ser difícil para los desarrolladores o los usuarios asumir la responsabilidad de estos resultados negativos.

Impactos en la privacidad y la seguridad de los datos

La inteligencia artificial (IA) puede producir problemas de impacto en la privacidad y seguridad de los datos, debido a varios factores. La IA a menudo requiere grandes cantidades de datos para entrenarse y funcionar de manera efectiva. Estos datos pueden incluir información sensible y privada de los usuarios, como datos personales, historiales médicos e información financiera, entre otros. El uso de datos sensibles en la IA plantea preocupaciones sobre la privacidad, ya que existe el riesgo de que estos datos sean mal utilizados, accedidos de forma no autorizada o compartidos sin consentimiento, lo que puede resultar en violaciones de la privacidad de los usuarios.

Además, la IA también puede tener implicaciones en la propiedad y control de los datos. Los modelos de IA desarrollados por empresas o instituciones pueden generar datos y conocimientos derivados que pueden tener valor económico. Esto plantea preguntas sobre quién tiene el control y la propiedad de estos datos generados por la IA, así como cómo se utilizan y son compartidos con terceros. La falta de regulaciones y estándares claros en este sentido puede generar preocupaciones sobre la privacidad y seguridad de los datos, así como el uso ético de la IA en la gestión de datos.

Ética y gobernanza

La ética en las decisiones tomadas por la inteligencia artificial (IA) es un tema de preocupación creciente, debido a que la IA tiene la capacidad de tomar decisiones autónomas que pueden tener un impacto significativo en la vida de las personas y en la sociedad en general.

Como se mencionó en el punto 3 de este artículo, los algoritmos pueden generar discriminación en áreas como contratación, préstamos, atención médica y justicia, lo que plantea preocupaciones éticas sobre la equidad y la imparcialidad en las decisiones tomadas por la IA.

Asimismo, el uso de estas herramientas plantea interrogantes sobre la responsabilidad y la rendición de cuentas (punto 4 de este artículo). Si la IA toma decisiones autónomas, ¿quién es responsable cuando esas decisiones son erróneas o perjudiciales? ¿Cómo se asigna la responsabilidad legal y ética en caso de consecuencias negativas? Este tema es especialmente relevante cuando la IA se utiliza en aplicaciones críticas, como vehículos autónomos o sistemas de atención médica, donde las decisiones equivocadas pueden tener graves consecuencias para la seguridad y el bienestar de las personas.

Las «cajas negras»

Finalmente, la falta de transparencia en el proceso de toma de decisiones de la IA plantea preocupaciones éticas. Muchos algoritmos de IA son considerados como «cajas negras», lo que significa que su funcionamiento interno no es fácilmente comprensible para los desarrolladores, usuarios y personas afectadas por las decisiones de la IA. Esto puede dificultar la identificación y corrección de sesgos, errores o discriminación en las decisiones de la IA, lo que, a su vez, puede afectar la confianza pública en la tecnología y plantear interrogantes éticas sobre la transparencia y la responsabilidad en el uso de la IA.

Es importante abordar estos problemas éticos en el desarrollo y uso de la IA, mediante la implementación de principios éticos, regulaciones adecuadas y una rendición de cuentas clara. Solo así podremos garantizar que la IA tome decisiones justas, equitativas y responsables en beneficio de la sociedad en su conjunto. /**M**