



Esta y las demás imágenes Autoría Borja Handel. usando DALLE-3 model

Borja García Haendler

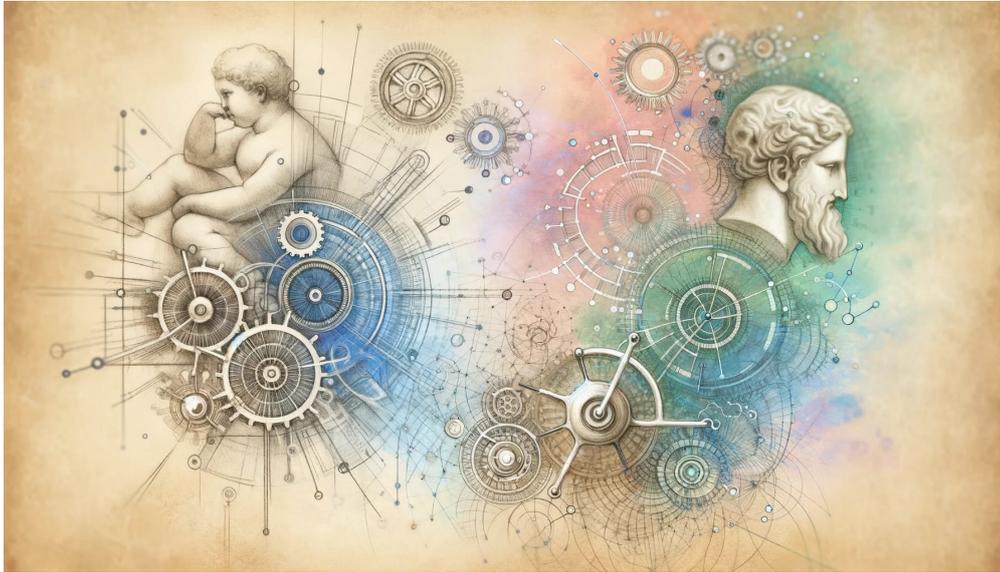
Jefe de MRPC Asia y científico de datos CRO Asia Innovation AI Lab.

En la vanguardia de la innovación tecnológica de los últimos años, la Inteligencia Artificial Generativa (IAG) representa un campo en rápida evolución y sobre todo de gran atención. Empresas como OpenAi, antes desconocida para el gran público, se ha convertido de la noche a la mañana en una potencial amenaza para las grandes tecnológicas, y es que sus productos, basados en IAG, pueden cambiar las reglas del juego. Similarmente,, <https://leonardo.ai/>, son sólo unos ejemplos de esta revolución.

Como todo avance significativo en una herramienta tecnológica, su desarrollo y aplicación plantean cuestiones éticas y morales significativas, que en esencia, son las mismas cuestiones morales que se plantean con cualquier uso y posible abuso de una tecnología.

En este artículo voy a examinar cuáles podrían ser unos estándares éticos fundamentales para la IAG, abogando por un enfoque multidisciplinario y centrado en el ser humano que respeta tanto los valores universales como las normativas específicas de cualquier entorno

social:



Seguridad y Fiabilidad

La tecnología debe ser desarrollada y probada rigurosamente para garantizar que no cause daño y funcione de manera predecible. Esto es especialmente importante en sectores de alto riesgo como la salud, las finanzas y los vehículos autónomos. Por ejemplo, los sistemas de IAG en la detección de retinopatía diabética deben ser precisos y fiables, dado que un diagnóstico erróneo podría tener graves consecuencias para la salud. Por ello, es recomendable que la IAG sea supervisada por un humano experto en el sector

Transparencia y Explicabilidad

La transparencia y la explicabilidad en la IAG no solo debieran ser imperativos éticos sino también prácticos. Los sistemas de IAG deben ser diseñados para que sus procesos y decisiones sean comprensibles para los usuarios. Esto es crucial en aplicaciones como el diagnóstico médico y el comercio algorítmico, donde los usuarios deben poder comprender y confiar en las decisiones tomadas por la IA. El problema fundamental aquí es que, si bien conocemos como crear una IAG, todavía desconocemos la manera de explicar los resultados de esta tecnología con suficiente detalle. Por ejemplo, si una IAG produce un contenido que ataca información privada, no somos capaces de encontrar, en qué lugar de la red neuronal se encuentra dicha información para eliminarla. Según la Comisión Europea, la

explicabilidad es uno de los cuatro principios éticos fundamentales para garantizar una IA digna de confianza. La explicabilidad en IA (XAI) se refiere a la necesidad de que los procesos de IA sean transparentes y que las decisiones sean bien comunicadas. Este principio es crucial para mantener la confianza en la IA y asegurar que las decisiones tomadas por sistemas inteligentes sean comprensibles y justificables

Privacidad y Protección de Datos

Y hablando de información privada, la IAG requiere el manejo de grandes volúmenes de datos, algunos pueden ser personales y sensibles. La protección de la privacidad y los datos es fundamental, lo que implica el manejo seguro de la información, la obtención del consentimiento para su uso y el cumplimiento de las regulaciones de protección de datos como el GDPR y el CCPA. El problema, es que es tal cantidad de datos de los que estos modelos se nutren, que garantizar que no haya ninguna información sensible, es tarea extremadamente complicada.

No todas las etapas del ciclo de vida de una IA implican el tratamiento de datos personales. Cuando una IA procesa datos personales, como en la elaboración de perfiles de personas físicas, debe cumplir con normativas como el RGPD y la LOPDGDD. Los principios de limitación y precisión son fundamentales para asegurar que los datos personales se utilicen adecuadamente y se mantengan por el tiempo necesario. La 45ª Asamblea Global de Privacidad enfatiza que la IA generativa debe basarse en principios de protección de datos, privacidad, control humano, transparencia y valores democráticos. Los desarrolladores deben incorporar estas consideraciones desde la concepción y operación de sistemas de IA generativa.

Equidad y No Discriminación

Los sistemas de informáticos, y por ende la IAG deben ser diseñados y entrenados para evitar sesgos y discriminación. Esto incluye garantizar conjuntos de datos diversos y representativos y realizar evaluaciones regulares para detectar resultados sesgados. En el reclutamiento o en los modelos de puntuación crediticia, por ejemplo, es vital garantizar que los sistemas de IA no perpetúen discriminaciones históricas o sociales. Otro ejemplo interesante, es la interpretación de la historia que se puede observar en la IAG toma partido respecto, la razón, sencilla, la mayor cantidad de texto que se está utilizando está en inglés, por consiguiente tiene más peso las ideas escritas en ese idioma que en otro. Esto

tiene consecuencias de fijación de la verdad.

La IAG ética se enfoca en el impacto social de los sistemas de IA y su equidad percibida. Es importante entender los impactos sociales de la integración de sistemas de IA en la sociedad y la necesidad de generar confianza en estos sistemas



El problema del Deep Fake

El problema ético de la inteligencia artificial generativa, especialmente en el contexto de los «deepfakes», es multifacético y profundamente preocupante. Un «deepfake» es un video o audio creado con inteligencia artificial que imita a una persona real, haciendo parecer que dijo o hizo algo que nunca ocurrió en realidad. Esta capacidad de generar “deepfake” es más preocupante por lo accesible y fácil uso para cualquier persona en cualquier lugar del planeta.

- **Fraude y Engaño:** Los deepfakes pueden utilizarse para crear engaños convincentes. Esto incluye desde estafas personales hasta grandes fraudes financieros. Por ejemplo, un video falso de un CEO haciendo un anuncio financiero falso podría manipular los mercados.
- **Desinformación y Impacto en la Democracia:** En el ámbito político, los deepfakes pueden ser herramientas poderosas para la desinformación. Videos falsificados de políticos diciendo o haciendo cosas comprometedoras podrían usarse para influir en la opinión pública y alterar el resultado de elecciones. Esto representa una amenaza directa

a la integridad de los sistemas democráticos. (veamos que pasa este año 2024, con las elecciones que vienen, en este aspecto). Un ejemplo notable fue un video deepfake del presidente ucraniano Volodymyr Zelenskyy, que fue desacreditado rápidamente pero mostró cómo estas tecnologías pueden ser usadas en contextos políticos para desinformar y polarizar aún más a las sociedades

<https://news.northeastern.edu/2022/04/01/deepfakes-fake-news-threat-democracy/>

- **Violación de la Privacidad y Consentimiento:** La creación de imágenes o videos de personas sin su consentimiento, especialmente en contextos comprometedores o difamatorios, es una grave violación de la privacidad. Esto no solo afecta a figuras públicas, sino también a individuos privados.
- **Erosión de la Confianza:** A medida que los deepfakes se vuelven más sofisticados, la capacidad de diferenciar entre contenido real y falso disminuye. Esto podría llevar a una erosión generalizada de la confianza en los medios digitales, complicando aún más la búsqueda de la verdad y la verificación de hechos.
- **Abuso y Acoso:** Los deepfakes pueden ser utilizados para crear contenido abusivo o de acoso, incluyendo la venganza pornográfica y el acoso en línea. Esto plantea serios problemas de seguridad y bienestar para las víctimas.
- **Implicaciones Legales y de Política Pública:** La regulación de los deepfakes presenta desafíos legales significativos. Determinar cuándo y cómo limitar su creación y distribución implica equilibrar la libertad de expresión con la protección contra el daño.

En respuesta a estos desafíos, se están desarrollando tanto tecnologías para detectar *deepfakes* como marcos legales y políticas para regular su uso. Sin embargo, la carrera entre la creación de deepfakes y su detección es constante, y las implicaciones éticas continúan evolucionando a medida que avanza la tecnología.

Derechos de Propiedad Intelectual

Un tema importante dentro de la IAG, son los Derechos de Autor de Contenido Generado por IA: Determinar quién posee los derechos de autor del contenido generado por la IA es complejo. El debate principal es si la propia IA puede ser titular de derechos de autor, o si los derechos deberían pertenecer al desarrollador de la IA, al usuario que activó la IA, o incluso al propietario de los datos en los que fue entrenada la IA.

Cuando una IA genera contenido que es derivado de obras creadas por humanos (por ejemplo, crear una nueva obra de arte basada en una pintura existente), se plantean

preguntas sobre la infracción de los derechos del creador original. Los sistemas legales en muchas jurisdicciones todavía están lidiando con cómo aplicar las leyes de derechos de autor existentes a tales obras derivadas. Desde mi opinión, creo que todo artista humano, en su labor creativa y de exploración, expresa con su obra, en parte su experiencia personal e individual, pero también incluye contenido derivado de otras obras. Picasso en su Guernica, tiene un componente original, una parte de experiencia personal, y finalmente un conocimiento de arte y de pintura de otros artistas. Es evidente que no hubiésemos tenido un Picasso sin la influencia que hicieron en él todos los artistas anteriores y contemporáneos.

Uso Justo y Uso Transformador: En algunos casos, el uso de la IA para crear obras derivadas podría defenderse bajo las doctrinas de uso justo o uso transformador. Sin embargo, estos suelen ser juzgados caso por caso, y hay una falta de precedente claro para obras generadas por IA.

Derechos Morales: Estos derechos, como el derecho de atribución e integridad, también son un punto de preocupación. Hay preguntas sobre cómo se aplican cuando una IA transforma significativamente una obra creada por humanos.

Muchos países están comenzando a examinar sus leyes de derechos de autor para ver cómo pueden adaptarse o reformarse para abordar los desafíos planteados por la IA. Esto incluye discutir nuevos tipos de derechos de autor o cambios en el marco existente para acomodar el contenido generado por IA.



Responsabilidad

La responsabilidad es crucial en la IAG. Debe haber una clara atribución de responsabilidad por las decisiones y acciones de los sistemas de IA, con roles bien definidos para desarrolladores, usuarios y operadores.

Desarrolladores y Proveedores de IA: Las empresas y equipos que desarrollan y proporcionan sistemas de IA son responsables de diseñar y mantener estos sistemas de manera ética y segura. Esto incluye la implementación de medidas para prevenir usos indebidos y garantizar que la IA se comporte de manera predecible y conforme a las leyes aplicables.

Usuarios: Los individuos o entidades que utilizan sistemas de IA son responsables de cómo emplean estas herramientas. Esto incluye seguir las pautas establecidas por los desarrolladores, respetar las leyes de derechos de autor y privacidad, y no utilizar la IA para actividades ilegales o dañinas.

IAG con valores Centrados en el Ser Humano

La IAG debe alinearse con los valores humanos, éticos y culturales. Esto significa que la IA debe potenciar y no reemplazar las capacidades y decisiones humanas, respetando la dignidad, los derechos y la diversidad cultural. No olvidemos que esta herramienta, está para servir y ayudar al ser humano

Sostenibilidad Ambiental

La IAG debiera desarrollarse y operarse de manera ambientalmente sostenible, minimizando su huella ecológica. Esto incluye el uso de algoritmos eficientes en términos de energía, la utilización de energías renovables y la optimización del hardware. Algo de momento lejano, además debido a su actual popularidad, muchas compañías y universidades entrenan IGAs con cada vez más datos y con más modelos para buscar mejoras.

Cooperación Global

Dada la naturaleza global de la IA, es esencial la cooperación internacional para desarrollar y hacer cumplir estándares éticos globales y mejores prácticas. Esto es particularmente

relevante en el contexto de la IAG, donde las innovaciones y los desafíos no conocen fronteras. De nada sirve que el 99% de la población se ajuste a estos estándares mientras un 1% utilice el potencial de la IA con fines dañinos (algo que lamentablemente, y conociendo al ser humano, será un problema nuevo que tendremos que lidiar a partir de ahora)

Monitoreo y Evaluación Continuos

Los estándares éticos para la IAG deben ser continuamente monitoreados, evaluados y actualizados en respuesta a nuevos desafíos y avances tecnológicos. Esto garantiza que la IA se mantenga alineada con las normativas éticas y sociales en evolución. Similarmente a como hace la sociedad con respecto a los individuos que se salen de los estándares de una sociedad.

Conclusión

La Inteligencia Artificial Generativa (IAG) se presenta como una revolución tecnológica con un potencial inmenso para transformar múltiples sectores. No obstante, su rápido desarrollo trae consigo una serie de desafíos éticos y morales que deben ser abordados con un enfoque multidisciplinario y centrado en el ser humano. Estos desafíos incluyen asegurar la seguridad y fiabilidad, mantener la transparencia y explicabilidad, proteger la privacidad y los datos personales, garantizar la equidad y no discriminación, respetar los derechos de propiedad intelectual, asignar adecuadamente la responsabilidad y alinear la tecnología con los valores humanos, culturales y ambientales.

La cooperación global y el monitoreo continuo son esenciales para establecer y mantener estándares éticos que se ajusten a los avances y desafíos emergentes de la IAG. Al abordar estos aspectos con seriedad y compromiso, podemos maximizar los beneficios de la IAG, asegurando que su impacto sea positivo y sostenible, y que enriquezca a la sociedad respetando al mismo tiempo los derechos y valores humanos fundamentales. La IAG no es solo un avance tecnológico; es una oportunidad para reflexionar y reafirmar nuestros principios éticos en la era digital.

Para citar esta entrada

García Haendler, Borja. Jefe de MRPC Asia y científico de datos CRO Asia Innovation AI Lab. En *Niaia*, consultado el 28/02/2024 en [Estándares Éticos en la Inteligencia Artificial](#)

Generativa

Creemos en el libre flujo de información. Republique nuestros artículos libremente, en impreso o digital, bajo licencia Creative Commons, citando la fuente



La Web de NIAIÁ y sus publicaciones (salvo aquellas en las que se especifique de otra manera) están bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)