JOURNAL SIBEMIR
https://www.journalsibemir.com

ARTÍCULO ESPECIAL

Desbloqueando el Código Fuente de la Creatividad Una Inmersión tecnológica en la IA Generativa

Unlocking the Source Code of Creativity: A Technological Dive into Generative AI

Mauricio J. SOULLIER [a]✉

[a] Head of Product Development – Data Science & AI
Scalene Solutions, Melbourne, Australia

INFORMACIÓN DEL ARTÍCULO

Historia del Artículo:
Recibido: 01 01 2025
Aceptado: 01 04 2025

Palabras claves:
IA Generativa. GPT. Aprendizaje Profundo. Cooperación Humano-Máquina.

Key words:
Generative AI. GPT. Deep Learning. Human-Machine collaboration

RESUMEN

El lanzamiento de ChatGPT en 2022 desencadenó una revolución tecnológica sin precedentes, llevando la inteligencia artificial generativa desde los laboratorios de investigación hasta el centro de atención global. Esta innovadora tecnología, capaz de generar contenido sorprendentemente humano, ya no es solo un concepto futurista, sino una realidad palpable que desafía nuestras nociones preconcebidas sobre la creación y el pensamiento.

A medida que exploramos las profundidades de los modelos de lenguaje de última generación y las técnicas de aprendizaje profundo que los impulsan, nos adentramos en un territorio emocionante y, a la vez, desconcertante. ¿Cómo funcionan estos sistemas que parecen emular la inteligencia humana? ¿Cómo podemos garantizar que esta tecnología sea accesible, confiable y ética?

En este artículo, nos sumergiremos en el corazón de la IA generativa, desentrañando los avances técnicos más recientes y las complejidades subyacentes. Exploraremos el estado del arte, las implicaciones éticas y las perspectivas futuras, ofreciendo una visión profunda y equilibrada de esta revolución tecnológica. Los invito a descubrir cómo la IA generativa está descodificando el *código fuente de la creatividad* y redefiniendo los límites de lo posible en la colaboración entre humanos y máquinas.

ABSTRACT

The launch of ChatGPT in 2022 sparked an unprecedented technological revolution, propelling generative artificial intelligence from research laboratories into the global spotlight. This innovative technology, capable of generating stunningly human-like content, is no longer merely a futuristic concept, but a palpable reality that challenges our preconceived notions about creation and cognition.

As we explore the depths of cutting-edge language models and the deep learning techniques that drive them, we venture into exciting yet perplexing territory. How do these systems that seem to emulate human intelligence operate? How can we ensure that this technology is accessible, reliable, and ethical?

In this article, we delve into the heart of generative AI, unravelling the latest technical advancements and underlying complexities. We will explore the state-of-the-art, ethical implications, and future prospects, offering an in-depth and balanced view of this technological revolution. I invite you to discover how generative AI is decoding the "source code of creativity" and redefining the boundaries of human-machine collaboration.

✉ **Autor para correspondencia**
Correo electrónico: maurice.soullier@scalenesolutions.com

https://doi.org/10.63706/jsibemir.v1i1.17
e-ISSN: 3087-2367/

INTRODUCCIÓN

En el presente artículo, se ofrece una exploración de la Inteligencia Artificial (IA) Generativa, desglosando su naturaleza intrínseca y su evolución histórica. Se destaca la importancia de los modelos generativos en el desarrollo y la configuración de la IA contemporánea. Mediante un análisis conciso pero generalizado, se proporciona un contexto histórico y se examinan los avances significativos que han contribuido a la prominencia de la IA Generativa en los tiempos que corren.

Este trabajo no solo pretende ser introductorio sino también fundacional, permitiendo a los lectores familiarizarse con los principios esenciales de la IA Generativa y servir de guía en caso que se desee profundizar. Se enfatiza la relevancia de comprender los modelos generativos como piedra angular para el avance de la IA en el procesamiento de lenguaje natural y, por consiguiente, el impacto en la interacción Humano-Máquina.

LLEGADA DE CHATGPT Y SU IMPACTO

La llegada de ChatGPT, desarrollado por OpenAI y lanzado en noviembre de 2022, marcó un hito significativo en el campo de la IA. ChatGPT, un modelo de lenguaje de gran escala basado en la arquitectura del Transformador Generativo Pre-entrenado (GPT, por sus siglas en inglés), demostró capacidades sin precedentes en la comprensión y generación del lenguaje natural**[1].** Su habilidad para generar texto coherente y contextualmente relevante (aunque no siempre verídico) ha revolucionado diversos sectores, incluyendo el servicio al cliente, la educación y la creación de contenido **[2].** La adopción generalizada de ChatGPT no solo ha demostrado que mejoró la productividad, sino que también ha suscitado debates sobre las implicaciones éticas y los riesgos potenciales asociados con las tecnologías de IA **[2].**

DEFINICIÓN DE IA GENERATIVA

La IA generativa se refiere a un subconjunto de la inteligencia artificial que se centra en la creación de nuevo contenido, como texto, imágenes, audio y vídeo, basándose en patrones aprendidos a partir de datos existentes **[3].** A diferencia de la IA tradicional, la IA generativa utiliza modelos de aprendizaje profundo para generar resultados novedosos que imitan la creatividad humana **[4].** Estos modelos, incluyendo los modelos de lenguaje de gran escala como GPT, son entrenados con vastos conjuntos de datos y pueden producir contenido indistinguible del creado por humanos **[3]**. La versatilidad y las potenciales aplicaciones de la IA generativa la han convertido en un punto focal de la investigación y desarrollo contemporáneo en IA **[4]**. La Inteligencia Artificial Generativa abarca diversos tipos, cada uno adaptado para tareas específicas o formas de generación de medios. Los siguientes son algunos de los tipos más conocidos: Redes Generativas Adversarias (GANs, por sus siglas en inglés), Modelos basados en Transformadores (TRMs, por sus siglas en inglés), Autoencodificadores Variacionales (VAEs, por sus siglas en inglés) y Modelos de Difusión (DMs, por sus siglas en inglés), por mencionar algunos **[5][6][7][8].** Las secciones siguientes discutirán estos en mayor detalle. Pero antes, hagamos un breve repaso por el contexto histórico.

CONTEXTO HISTÓRICO

Breve Historia del Desarrollo de la IA

El desarrollo de la inteligencia artificial (IA) se remonta a mediados del siglo XX, con el trabajo fundamental de pioneros como Alan Turing y John McCarthy. El artículo seminal de Turing, "Computing Machinery and Intelligence" (1950) introdujo el concepto de inteligencia de las máquinas y propuso el Test de Turing como criterio para determinar la capacidad de una máquina de exhibir un comportamiento inteligente **[9]**. En 1956, la Conferencia de Dartmouth, organizada por McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon, es frecuentemente considerada como el nacimiento de la IA como campo de estudio **[10].** Esta conferencia sentó las bases para la subsiguiente investigación y desarrollo en IA.

A lo largo de las décadas de 1960 y 1970, la investigación en IA se centró en la IA simbólica y los sistemas expertos, que tenían como objetivo codificar el conocimiento humano en programas informáticos. Los logros notables durante este período incluyen el desarrollo del Solucionador General de Problemas (GPS, por sus siglas en inglés) por Newell y Simon y la creación del primer sistema experto, DENDRAL, para el análisis químico **[11][12]**. Sin embargo, las limitaciones de estos sistemas tempranos condujeron a un período de reducción de financiación e interés en la IA, conocido como el "invierno de la IA", durante la década de 1980 **[13].**

Hitos Clave conducentes la IA Generativa

El resurgimiento de la IA a finales del siglo XX y principios del XXI puede atribuirse a varios hitos clave. El advenimiento del aprendizaje automático, particularmente el desarrollo de las redes neuronales, marcó un cambio significativo de los sistemas basados en reglas a los enfoques basados en datos. La introducción de la retropropagación por Rumelhart, Hinton y Williams en 1986 permitió el entrenamiento de redes neuronales multicapa, allanando el camino para el aprendizaje profundo **[14].**

En la década de 2000, el desarrollo de las redes neuronales convolucionales (CNN, por sus siglas en inglés) por LeCun y colegas revolucionó la visión por computador, mientras que la introducción de las redes neuronales recurrentes (RNN, por sus siglas en inglés) y las redes de memoria a corto y largo plazo (LSTM, por sus siglas en inglés) por Hochreiter y Schmidhuber avanzaron el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y la predicción de secuencias **[15].** Estos avances sentaron las bases para la IA generativa, que se centra en la creación de nuevo contenido basado en patrones aprendidos a partir de datos existentes.

El Avance de los Modelos de Lenguaje de Gran Escala

El avance de los modelos de lenguaje de gran escala (LLMs, por sus siglas en inglés) representa un salto significativo en las capacidades de la IA generativa. La introducción de la arquitectura del transformador por Vaswani et al. en 2017 revolucionó el Procesamiento del Lenguaje Natural (NLP) al permitir que los modelos procesen y generen texto con una precisión y coherencia sin precedentes **[16][17].** Los Transformers, caracterizados por sus mecanismos de auto-atención, permitieron escalar los modelos a miles de millones de parámetros, llevando al desarrollo de modelos como BERT (Representaciones de Codificador Bidireccional de Transformadores) y GPT (Transformador Generativo Pre-entrenado) **[18].**

El lanzamiento de GPT-3 por OpenAI en 2020 marcó un momento decisivo en el campo de la IA. Con 175 mil millones de parámetros, GPT-3 demostró una notable competencia en la generación de texto similar al humano, realizando una amplia gama de tareas lingüísticas e incluso exhibiendo una suerte de habilidades de razonamiento rudimentarias, a pesar de no poseer una arquitectura compatible. El éxito de GPT-3 ha estimulado una mayor investigación y desarrollo en LLMs, con esfuerzos continuos para mejorar sus capacidades, eficiencia y consideraciones éticas.

Modelos impulsores clave: GANs, VAEs, Modelos de Difusión y Transformers

Como fue mencionado, el desarrollo de estos impulsores marcó el camino al estado del arte actual. La dinámica aportada por estas tecnologías trajo mejoras significativas en la generación de imágenes, videos y texto, al mismo tiempo que produjo asombro e incertidumbre. Revisemos brevemente cada uno de ellos.

Redes Generativas Adversarias (GAN)

Una red generativa adversaria (GAN, por sus siglas en inglés) es una clase de aprendizaje automático basado en redes neuronales. El aspecto novedoso en esta configuración de red generativa adversaria es que no depende de datos de entrenamiento etiquetados. Además, la arquitectura que ofrece es bastante única en comparación con las Redes Neuronales Profundas convencionales **[19].** De hecho, consta de dos componentes principales denominados **Generador** y **Discriminador**. La operación principal del generador es seguir generando datos falsos utilizando el ruido, mientras que el propósito del discriminador es distinguir si la imagen generada es real o falsa. El discriminador se entrena utilizando las imágenes reales del dominio que el generador está tratando de producir sintéticamente, y el único propósito del discriminador es identificar si la salida producida por el generador es falsa o no. El sistema general se basa en la dinámica de juegos de suma cero, el ganador permanecerá sin cambios y el modelo perdedor cada vez tiene que modificar sus parámetros, continuará haciendo esto hasta que el discriminador sea incapaz de detectar si la salida del generador es falsa o no **[5].** El único propósito de esto es construir un modelo generador potente que genere datos sintéticos que parezcan reales.

En su forma simplificada, el generador tiene como objetivo engañar al discriminador proporcionando la imagen generada sintéticamente con el objetivo de que se pruebe que es real. El discriminador discierne entre imágenes genuinas y falsificadas y genera la señal de salida. Esta señal de salida luego va tanto al generador como al discriminador, permitiendo al generador producir una mejor salida sintética (aprendizaje). Y, en caso de que el discriminador falle en probar que la imagen es falsa, también utiliza la señal para cambiar sus pesos para dar mejores predicciones. En toda esta arquitectura, es importante notar que solo el discriminador tiene acceso a la imagen real, la imagen sintética y su propia señal de salida, mientras que el generador solo aprende de la señal de salida del discriminador **[19]**.

Modelos de Difusión

Los Modelos de Difusión han sido diseñados para mejorar el rendimiento de la Red Generativa Adversaria Simple; la técnica fue introducida por Salimans et al **[20].** En una etapa posterior, Kingma et al introdujeron una variante del modelo de difusión llamada Flujo Autorregresivo Inverso (IAF, por sus siglas en inglés) como un bloque de construcción para modelos generativos **[21].** IAF es un tipo de flujo normalizador. Este es un tipo de modelo generativo que tiene como objetivo aprender distribuciones de probabilidad complejas mediante la transformación de una distribución base simple en la distribución objetivo a través de una serie de transformaciones invertibles. Técnicamente, se inspiran en la termodinámica de no equilibrio **[22].** Definen una cadena de Markov de pasos de difusión para agregar lentamente ruido aleatorio a los datos y luego, aprenden a revertir el proceso de difusión para construir muestras de datos deseadas a partir del ruido. A diferencia de los modelos VAE, los modelos de difusión aprenden con un procedimiento fijo y la variable latente tiene alta dimensionalidad (igual que los datos originales). En resumen, repasemos en algunos conceptos clave:

- **Cadena de Markov**: Es una secuencia de estados donde la probabilidad de cada estado depende solo del estado inmediatamente anterior.
- **Pasos de difusión**: El proceso gradual de agregar ruido a los datos originales. Esto se realiza en múltiples pasos, cada uno añadiendo una pequeña cantidad de ruido.
- **Proceso de reversión**: El modelo aprende a invertir el proceso de difusión, eliminando gradualmente el ruido para reconstruir los datos originales.
- **Alta dimensionalidad**: La variable latente en los modelos de difusión tiene la misma dimensión que los datos originales, lo que contrasta con otros modelos generativos que suelen utilizar espacios latentes de menor dimensión.
- **Procedimiento fijo**: A diferencia de otros modelos que pueden tener arquitecturas variables, los modelos de difusión siguen un proceso de entrenamiento más estandarizado.

Autoencodificadores Variacionales (VAE)

Otro modelo en el campo de la IA generativa son los Autoencodificadores Variacionales (VAEs, por sus siglas en inglés), introducidos por Kingma et al. **[9]**. En términos generales, un VAE consta de dos partes principales: un **codificador** (o modelo de inferencia) y un **decodificador** (o modelo generativo). El codificador toma los datos de entrada y los transforma en una representación dimensional inferior, llamada *espacio latente*, generalmente a través de una red neuronal. Esta representación latente se modela como una distribución de probabilidad, típicamente una distribución gaussiana multivariante. El decodificador, por su parte, toma muestras de la representación latente y las transforma de nuevo en el espacio de datos original, también a través de una red neuronal.

El objetivo principal es lograr una salida con una media y varianza similares a la entrada dada después de la introducción de varianza. Esto proporciona una forma estructurada de aprender representaciones significativas de los datos y luego generar nuevas muestras a partir de dicha distribución **[23].**

Transformers

El modelo de Transformer, introducido por **[17]** es una arquitectura de red neuronal que utiliza mecanismos de atención para mejorar la velocidad y la calidad de la generación de texto. En primer lugar, cada palabra en la entrada se codifica como un vector de alta dimensión utilizando un incrustado (del inglés, *embedding)* de palabras **[24].** Además, se añade un incrustado posicional para capturar la información de la posición de las palabras en la secuencia. El modelo luego aplica una operación llamada *atención de producto escalado*, que permite a cada palabra en la secuencia considerar todas las otras palabras al generar una representación. Esto se realiza a través de tres vectores derivados del *embedding* de la palabra: el vector de consulta, el vector de clave y el vector de valor **[24]**. Las representaciones resultantes se pasan a través de redes de alimentación hacia adelante (del inglés, *feed-forward networks*), que son simplemente capas densas de perceptrones. Finalmente, la salida de la red se decodifica en palabras en el vocabulario objetivo. Esto se hace a través de una capa de *softmax*, que genera una distribución de probabilidad sobre todas las palabras posibles.

Transformers como ChatGPT se entrenan para predecir la siguiente palabra en una secuencia, dada todas las palabras anteriores en la secuencia, llamada *contexto* (esto se conoce como modelado de lenguaje **autorregresivo**). Durante el entrenamiento, se ajustan los pesos de la red neuronal para minimizar la diferencia entre las predicciones del modelo y las palabras reales en el texto de entrenamiento.

Es importante destacar que el Transformer es un modelo de aprendizaje profundo no supervisado, lo que significa que se entrena en grandes cantidades de texto sin etiquetar y aprende a

generar texto que se asemeja a su entrada de entrenamiento. Una característica clave del modelo es que, a diferencia de las redes neuronales recurrentes, puede procesar todas las palabras de la entrada simultáneamente, lo que lo hace altamente paralelizable y, por lo tanto, más eficiente para entrenar con el hardware moderno.

ESTADO ACTUAL DE LA IA GENERATIVA

El campo de la IA generativa ha experimentado avances rápidos, con modelos actuales que demuestran capacidades sin precedentes en diversos dominios Esta sección ofrece una visión general de los modelos destacados, sus capacidades, limitaciones y aplicaciones en el mundo real. Los LLMs dominan el espacio de IA generativa actualmente, seguidos muy de cerca por los generadores de imágenes y, más recientemente, video.

Ejemplos notables son GPT-3 con sus 175 mil millones de parámetros, PaLM, publicado por Google y con 540 mil millones de parámetros y BLOOM, un modelo multilenguaje con 176 mil millones de parámetros. Todos ellos tienen una arquitectura base similar (Transformer) y fueron desarrollados para trabajar a gran escala **[25] [26] [27].** La diferencia principal reside en los datos de entrenamiento y el foco de generación: mientras GPT-3 está orientado a aplicaciones comerciales y de investigación general (abierto al público), PaLM utiliza un conjunto de datos más diverso y se enfoca en capacidades de razonamiento y resolución de problemas multimodo (utilizado solo por Google). BLOOM por su parte, hace foco en capacidades multi idioma y programación.

En el dominio de generación de imágenes, entre los modelos más prominentes encontramos DALL-E capaz de generar imágenes a partir de descripciones textuales, Stable Diffusion y Midjourney, más enfocados en generaciones artísticas **[28][29] [30].**

Capacidades y limitaciones

Los modelos de IA generativa poseen una amplia gama de capacidades que han impactado significativamente en varios campos. Los GANs son particularmente efectivos en la generación de imágenes, videos e incluso música de alta calidad **[5]**. Los VAEs son valiosos para tareas que requieren la generación de nuevas muestras de datos que se asemejen a la entrada original, como la síntesis de imágenes y la detección de anomalías **[21]**. Los LLMs, como GPT-3, sobresalen en la generación de lenguaje natural, traducción, resumen y respuesta a preguntas **[17]**.

A pesar de sus impresionantes capacidades, los modelos de IA generativa también tienen limitaciones notables. Los GANs pueden ser difíciles de entrenar debido a problemas al producir poca variedad en sus salidas **[5]**. Los VAEs a menudo luchan con la generación de imágenes de alta resolución en comparación con los GANs **[21]**. Los LLMs, aunque potentes, requieren grandes cantidades de datos y recursos computacionales para el entrenamiento, y a veces pueden producir salidas sesgadas o sin sentido (problema conocido como alucinación). Además, las implicaciones éticas de la IA generativa, como el potencial de mal uso en la creación de deepfakes o la generación de contenido dañino, siguen siendo preocupaciones significativas**[31]**. Presentaremos en una sección más adelante respecto a estos desafíos.

Aplicaciones en el mundo real

La IA generativa ha encontrado aplicaciones en una amplia gama de industrias, demostrando su potencial transformador y disruptivo. En el sector de la salud, los modelos generativos se utilizan para el descubrimiento de medicamentos, diagnostico por imágenes y la medicina personalizada **[32].** Por ejemplo, los GANs se han empleado para generar imágenes médicas sintéticas para entrenar modelos de diagnóstico, mientras que los VAEs se utilizan para identificar anomalías en los datos médicos **[33]**. En la industria del entretenimiento, la IA generativa se utiliza para crear efectos visuales realistas, generar música y desarrollar contenido para videojuegos **[34]**. Los LLMs como GPT-3 se utilizan ampliamente en el servicio al cliente, la creación de contenido y la educación, proporcionando respuestas automatizadas, generando artículos y asistiendo en el aprendizaje de idiomas **[25].** En el sector financiero, los modelos de IA generativa se aplican a la detección de fraudes, el comercio algorítmico y la gestión de riesgos**[3]**. Por ejemplo, los GANs pueden generar datos financieros sintéticos para mejorar la robustez de los sistemas de detección de fraudes, mientras que los LLMs ayudan en el análisis de informes financieros y la generación de perspectivas de mercado. La industria manufacturera se beneficia de la IA generativa a través de la optimización de los procesos de diseño, el mantenimiento predictivo y la gestión de la cadena de suministro **[3]** .En general, la versatilidad y el potencial de la IA generativa continúan impulsando la innovación y la eficiencia en varios dominios.

CONSIDERACIONES ÉTICAS

El vertiginoso avance de las tecnologías de IA generativa ha puesto en primer plano una serie de consideraciones éticas. Esta sección examina los principales desafíos éticos asociados con la IA generativa, centrándose en el sesgo y la equidad, las preocupaciones de privacidad, los problemas de propiedad intelectual y el potencial para la desinformación.

Sesgo y Equidad en los Sistemas de IA

El sesgo y la equidad en los sistemas de IA se han convertido en problemas críticos a medida que estas tecnologías influyen cada vez más en los procesos de toma de decisiones en varios dominios. Los sistemas de IA pueden perpetuar e incluso exacerbar los sesgos existentes presentes en los datos de entrenamiento, lo que lleva a resultados que pueden ser considerados injustos. Por ejemplo, se ha demostrado que los sistemas de reconocimiento facial exhiben tasas de error más altas para los grupos minoritarios, como destacaron Buolamwini y Gebru en 2018 **[35]**. De manera similar, se ha encontrado que los algoritmos de contratación de personal discriminan contra ciertos grupos demográficos, reforzando las desigualdades existentes **[36]**. Abordar estos sesgos requiere un enfoque multifacético, que incluye mejorar la calidad de los datos, diseñar algoritmos justos e implementar métricas de evaluación robusta **[37]**. Los investigadores han propuesto varias estrategias de mitigación, como el preprocesamiento de datos para eliminar sesgos, ajustes algorítmicos para garantizar la equidad y técnicas de post-procesamiento para corregir resultados sesgados **[38]**.

Preocupaciones de Privacidad y Protección de Datos

La proliferación de tecnologías de IA ha suscitado preocupaciones significativas sobre la privacidad, particularmente en lo que respecta a la recopilación, almacenamiento y uso de datos personales. Los sistemas de IA a menudo requieren grandes cantidades de datos para funcionar de manera efectiva, lo que puede llevar a la erosión de la privacidad individual. El Reglamento General de Protección de Datos (GDPR) en la Unión Europea representa un marco integral destinado a proteger los datos personales y garantizar la privacidad **[39]**. Otros países alrededor del mundo están siguiendo sus pasos e incluso, estos temas están siendo discutidos en el ámbito político. Sin embargo, los riesgos únicos que plantea la IA, como la posibilidad

de re-identificación de datos anonimizados y el uso de la IA para la vigilancia, requieren nuevos enfoques para la protección de datos **[40].** Los investigadores han enfatizado la necesidad de transparencia en los sistemas de IA, abogando por divulgaciones algorítmicas y la implementación de técnicas que preserven la privacidad, como la privacidad diferencial, el aprendizaje federado y la encriptación homomórfica **[41].**

Problemas de Propiedad Intelectual y Derechos de Autor

La propiedad intelectual (PI) y los derechos de autor también están siendo afectados de manera trascendente. Las obras generadas por IA, como textos, imágenes y música, plantean preguntas sobre la autoría y la propiedad. Las leyes de derechos de autor tradicionales, que se basan en la creatividad humana, luchan por acomodar las obras creadas por IA **[42]**. En los Estados Unidos, la Oficina de Derechos de Autor ha dictaminado que las obras creadas únicamente por IA no son elegibles para la protección de derechos de autor, enfatizando la necesidad de autoría humana **[43].** Por el contrario, la Unión Europea ha explorado el concepto de reconocer a la IA como titular de derechos *sui generis*, aunque este enfoque sigue siendo controvertido. El panorama legal para las obras generadas por IA continúa evolucionando, con debates en curso sobre el equilibrio adecuado entre la protección de los derechos de PI y el fomento de la innovación **[40].**

Desinformación y Deepfakes

La aparición de *deepfakes* y otros medios de desinformación generados por IA plantea desafíos éticos y sociales significativos. Los deepfakes, que son videos falsos hiperrealistas creados usando GANs, se han utilizado para difundir información falsa, manipular la opinión pública y dañar reputaciones **[41]**. El potencial de los deepfakes para socavar la confianza en los medios e instituciones ha llevado a llamados a medidas regulatorias y al desarrollo de tecnologías de detección **[40].** Los investigadores han destacado el doble papel de la IA tanto en la creación como en la lucha contra la desinformación, con herramientas impulsadas por IA que se están desarrollando para detectar y mitigar la propagación de deepfakes **[44].** Las consideraciones éticas incluyen el equilibrio entre combatir la desinformación y preservar la libertad de expresión, así como la necesidad de cooperación internacional para abordar la naturaleza global de este tipo de *amenaza* **[38].**

Responsabilidad y transparencia

En el ámbito de la IA generativa, la **responsabilidad** y la **transparencia** son dos pilares críticos que garantizan el uso ético y responsable de estas tecnologías.

La **responsabilidad** en la IA generativa se refiere a la necesidad de que los sistemas de IA sean responsables de sus acciones, decisiones y los resultados que producen. La responsabilidad de estos modelos a menudo está vinculada a sus datos de entrenamiento. Por ejemplo, si un modelo generativo produce contenido inapropiado o dañino, es crucial rastrear hasta los datos de entrenamiento y entender qué condujo a esta salida. Sin embargo, este proceso no es sencillo debido a la complejidad y opacidad de estos modelos. Por lo tanto, el desarrollo de métodos para la interpretabilidad y aplicabilidad del modelo es un área de investigación significativa para garantizar la responsabilidad **[45].**

La **transparencia**, por otro lado, implica hacer que el funcionamiento de los sistemas de IA sea claro y comprensible para los humanos. En el contexto de la IA generativa, la transparencia podría significar proporcionar información clara sobre la arquitectura del modelo, los datos de entrenamiento y los algoritmos utilizados para generar la salida **[46].**

Sin embargo, lograr la transparencia en la IA generativa es un desafío debido a la naturaleza de "caja negra" de estos modelos. El funcionamiento interno de modelos como las GANs y los VAEs implica espacios latentes de alta dimensión y transformaciones no lineales, que son difíciles de interpretar y entender.

A pesar de estos desafíos, se están haciendo esfuerzos para mejorar la transparencia en la IA generativa. Se están explorando técnicas como la **visualización de características**, la **reducción de dimensionalidad** y la **simplificación del modelo** para hacer estos modelos más interpretables **[47].**

Debe quedar claro así, que la responsabilidad y la transparencia en la IA generativa están estrechamente entrelazadas. Mejorar la transparencia puede llevar a una mejor responsabilidad ya que permite una mejor comprensión de cómo funciona el sistema de IA, lo que puede ayudar a identificar cuándo y por qué el sistema de IA podría cometer un error. A la inversa, garantizar la responsabilidad a menudo requiere un nivel de transparencia sobre el funcionamiento interno del sistema de IA **[48].**

En conclusión, a medida que la IA generativa continúa avanzando y se integra más en la sociedad, la importancia de la responsabilidad y la transparencia solo crecerá. Estos principios son esenciales para construir confianza en estos sistemas, asegurar su uso ético y mitigar cualquier daño potencial que puedan causar.

IMPACTO SOCIAL

A esta altura de las circunstancias, es evidente que la IA generativa tiene implicancias profundas en varios aspectos de la sociedad. Esta sección examina el impacto social de la IA generativa, centrándose en la transformación del trabajo y el empleo, las implicancias educativas, los efectos en las industrias creativas y la democratización de la tecnología.

Transformación del Trabajo y Empleo

La integración de la inteligencia artificial (IA) en la fuerza laboral está transformando fundamentalmente la naturaleza del trabajo y el empleo. Las tecnologías de IA, particularmente la automatización y el aprendizaje automático, están remodelando los roles laborales, los requisitos de habilidades y los patrones de empleo. Los estudios indican que, si bien la IA puede desplazar ciertos trabajos, también crea nuevas oportunidades y mejora la productividad **[31].** El nuevo mantra parece ser "la IA no va a reemplazarte, pero alguien que la use lo hará" **[49].** Por ejemplo, la automatización impulsada por la IA puede realizar tareas repetitivas y mundanas, permitiendo a los trabajadores humanos concentrarse en actividades más complejas y creativas **[50].** Sin embargo, este cambio requiere una recapacitación y actualización de habilidades significativas de la fuerza laboral para adaptarse a nuevos roles que requieren habilidades técnicas y cognitivas avanzadas **[51].** El impacto de la IA en el empleo es multifacético, con posibles beneficios como el aumento de la eficiencia y la innovación, así como desafíos como el desplazamiento laboral y la desigualdad.

Implicancias Educativas

La IA está preparada para revolucionar la educación al personalizar las experiencias de aprendizaje, automatizar tareas administrativas y mejorar los resultados educativos. Las herramientas impulsadas por la IA pueden proporcionar contenido educativo personalizado, plataformas de aprendizaje adaptativas y retroalimentación en tiempo real, mejorando así el compromiso y el rendimiento de los estudiantes **[52].** Por ejemplo, los sistemas de tutoría inteligentes pueden adaptarse a

los estilos y ritmos de aprendizaje individuales, ofreciendo soporte y recursos personalizados. Además, la IA puede ayudar a los educadores al automatizar la calificación, gestionar tareas administrativas e identificar a los estudiantes que necesitan ayuda adicional. Sin embargo, la integración de la IA en la educación también plantea preocupaciones éticas, como la privacidad de los datos, el sesgo algorítmico y la brecha digital **[53].** Asegurar un acceso equitativo a las tecnologías educativas impulsadas por la IA es crucial para evitar exacerbar las desigualdades existentes.

Visto desde otro ángulo, si nos enfocamos en los estudiantes, se presentan varios riesgos, como el mal uso académico (plagio o la dependencia excesiva de los estudiantes en estas herramientas), subestimación del papel del profesor (socavando la autoridad y el estatus de los profesores), e integridad académica (reconocimiento del uso potencial de estas herramientas en detrimento del aprendizaje) **[54].**

Es esencial que las instituciones educativas implementen políticas y orientaciones adecuadas para garantizar el uso seguro y responsable de la IA generativa. Además, se necesita más investigación para entender completamente estos riesgos y desarrollar estrategias efectivas para mitigarlos.

Efectos en las Industrias Creativas

Las industrias creativas están experimentando una transformación profunda debido al empleo de IA generativa crear arte, música, literatura y otras formas de contenido creativo. Estas tecnologías pueden aumentar la creatividad humana al generar ideas novedosas, mejorar los flujos de trabajo creativos y automatizar tareas rutinarias. Por ejemplo, el arte y la música generados por IA han ganado reconocimiento y éxito comercial, demostrando su potencial colaborador creativo **[55].** Sin embargo, el auge de la IA en los campos creativos también plantea preguntas sobre la autoría, la originalidad y el valor de la creatividad humana, como mencionamos en la sección anterior. Actualmente, estas herramientas no son completamente autónomas, por lo que necesitan de una correcta guía humana para producir el contenido. El valor de la "co-creación" reside en acentuar la capacidad creativa humana, principalmente mediante lo que denominamos una "exploración guiada hacia la creación de valor".

Si bien la IA puede mejorar la productividad y la innovación, es esencial equilibrar su uso con la preservación de la expresión artística humana y el patrimonio cultural.

Democratización de la Tecnología

La democratización de la IA se refiere a hacer que las tecnologías de IA sean accesibles y beneficiosas para una población más amplia, más allá de los confines de los expertos especializados y las grandes corporaciones. Esto implica desarrollar herramientas de IA fáciles de usar, promover marcos de IA de código abierto y garantizar un acceso equitativo a los recursos de IA **[55].** La democratización de la IA puede empoderar a individuos y comunidades al permitirles aprovechar la IA para diversas aplicaciones, desde la atención sanitaria y la educación hasta el gobierno local y la innovación social **[56].** Sin embargo, lograr una verdadera democratización requiere abordar desafíos como la alfabetización digital, las disparidades de infraestructura y las consideraciones éticas previamente mencionadas. Asegurar que el desarrollo y despliegue de la IA sean inclusivos y participativos es crucial para maximizar los beneficios sociales de la IA y mitigar los posibles daños.

UNA RÁPIDA VISTA AL FUTURO

Los avances no terminan aquí, y el crecimiento exponencial de la IA generativa parece no desacelerarse. Si bien el hincapié en las mejoras debe darse en cuestiones más sutiles y muchas veces difusas, como la ética y el impacto social, el ámbito académico y comercial parecen haber finalmente unido fuerzas para cooperar, al menos en el dominio técnico y de aplicación. Nuevas alianzas surgen a diario, promoviendo áreas de investigación y utilización impensadas en los últimos años.

Áreas de Investigación Emergentes

Un área emergente típica es el desarrollo de modelos específicos de dominio, que están adaptados a industrias o aplicaciones particulares. Estos modelos buscan mejorar el rendimiento y reducir los requisitos computacionales al centrarse en tareas especializadas; quizás, de alguna forma, aprendiendo de las limitaciones del modelo de Newell y Simon y del viejo refrán "quien mucho abarca, poco aprieta" **[11].** Un claro ejemplo es la innovación llevada a cabo por Apple recientemente con Apple Intelligence, donde se intenta llevar el poder de la IA generativa directamente a sus dispositivos y aplicaciones, ofreciendo mejoras en la privacidad y procesamiento local **[57].** Esto implica la optimización de las arquitecturas de modelos para mejorar la eficiencia y la escalabilidad, con técnicas como la poda de modelos, la cuantización y la destilación de conocimientos para reducir la huella computacional (y energética) de los modelos grandes sin comprometer el rendimiento y la privacidad del usuario.

Otro avance es la integración de capacidades multimodales, que permiten a los modelos procesar y generar contenido en diferentes modalidades, como texto, imágenes y audio; aquí, hay promesas de aplicaciones en áreas como: realidad virtual, realidad aumentada y HCI **[58].**

Integración con Otras Tecnologías (Robótica, IoT, ..., y más)

La integración de la IA generativa con otras tecnologías, como la robótica y el Internet de las Cosas (IoT), está preparada para revolucionar varias industrias. En robótica, la IA generativa puede mejorar los sistemas autónomos al permitir capacidades de percepción, planificación y control más sofisticadas **[59].** Por ejemplo, los modelos generativos se pueden utilizar para producir simulaciones realistas para el entrenamiento de robots, mejorando su capacidad para navegar por entornos complejos. En el contexto del IoT, la IA generativa puede aumentar la analítica predictiva al analizar grandes cantidades de datos generados por dispositivos interconectados, lo que lleva a obtener información más precisa y oportuna. Esta sinergia entre la IA generativa y el IoT puede impulsar innovaciones en ciudades inteligentes, atención sanitaria y automatización industrial **[59].**

CONCLUSIÓN

La IAG ha emergido como una fuerza revolucionaria en el dominio de la IA, propulsada por progresos significativos en las redes neuronales, el aprendizaje profundo y los modelos de lenguaje. El desarrollo histórico de la IA resalta hitos fundamentales como la aparición de las Redes Generativas Antagónicas (GANs), los Autoencodificadores Variacionales (VAEs), los Transformers y los Modelos de Difusión, que han impulsado colectivamente las capacidades de la IA generativa. La infraestructura técnica de estos modelos, incluyendo sus arquitecturas y procesos de entrenamiento, enfatiza la complejidad y las demandas computacionales implicadas en su desarrollo.

Los sistemas de IA generativa pueden ser considerados "creativos" en el sentido de que son capaces de producir contenido novedoso y original. Esta creatividad se facilita a menudo por su habilidad para combinar y recombinar patrones

existentes de formas innovadoras. Por ejemplo, las GANs pueden generar imágenes realistas aprendiendo de un conjunto de datos de imágenes existentes y creando nuevas que son indistinguibles de las fotos reales. De manera similar, los modelos de lenguaje de gran escala (LLMs) pueden generar texto creativo al basarse en vastos corpus de material escrito para producir nuevas historias, poemas o artículos, incluso con diversos estilos de escritura. Sin embargo, es importante destacar que la creatividad de la IA generativa no es equivalente a la creatividad humana, que implica intencionalidad, profundidad emocional y experiencia subjetiva.

Este artículo, sin intención filosófica, busca ilustrar las complejidades del paradigma de la IA generativa y brindar una fuente de referencia para quien desee ahondar en este vasto universo. Esta tecnología, al proporcionar nuevas herramientas y perspectivas, potencia la creatividad humana y redefine la colaboración humano-máquina, convirtiéndose en nuestro "socio creativo" para amplificar nuestras formas de expresión.

Mirando hacia el futuro, la colaboración Humano-Máquina está preparada para jugar un papel fundamental. La sinergia entre la inteligencia humana y las capacidades de la IA puede impulsar la innovación, mejorar la toma de decisiones y abordar desafíos complejos. Los sistemas de IA sobresalen en el procesamiento de grandes cantidades de datos e identificación de patrones, mientras que los humanos aportan pensamiento crítico, creatividad y juicio ético. Este enfoque colaborativo puede llevar a aplicaciones de IA más efectivas y éticas, fomentando un futuro en el que la IA aumente el potencial humano en lugar de reemplazarlo. Al adoptar la colaboración Humano-IA, la sociedad puede aprovechar las fortalezas de ambas entidades para lograr resultados que ninguna podría alcanzar por sí sola.

REFERENCIAS BIBLIOGRÁFICAS

1. *OpenAI*. *GPT-4 Technical Report*. 2023. Disponible en: https://www.openai.com/research/gpt-4.
2. Foster L. *Generative Deep Learning*. O'Reilly Media, Inc.; 2023.
3. Gozalo-Brizuela R. A survey of Generative AI Applications. *arXiv*. 2023;2306.02781.
4. Sengar SS. Generative Artificial Intelligence: A Systematic Review and Applications. *arXiv*. 2024;2405.11029.
5. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. In: *Advances in Neural Information Processing Systems*; 2014 Dec 8-13; Montreal, Canada. p. 2672-80.
6. Vasanthi P. Multi-head self-attention based transformer for multi-scale object detection. *Multimedia Tools and Applications*. 2023;1-27.
7. Kingma DP, Welling M. Auto-Encoding Variational Bayes. In: *Proceedings of the 2nd International Conference on Learning Representations*; 2014 May 7-9; Banff, Canada.
8. Chen M, Radford A, Child R. Diffusion Models for Generative AI. *arXiv*. 2024;2404.07771.
9. Turing AM. Computing Machinery and Intelligence. *Mind*. 1950;59(236):433-60.
10. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*. 1955;76(1):12-14.
11. Newell A, Shaw J, Simon HA. The Logic Theory Machine: A Complex Information Processing System. *IRE Transactions on Information Theory*. 1959;IT-2(3):61-79.
12. Buchanan BG, Shortliffe EH. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley; 1984.
13. Crevier D. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books; 1993.
14. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533-6.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44.
16. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*; 2017 Dec 4-9; Long Beach, CA. p. 5998-6008.
17. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*; 2017 Dec 4-9; Long Beach, CA. p. 5998-6008.
18. Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Conference of the North American Chapter of the Association for Computational Linguistics*; 2019 Jun 2-7; Minneapolis, MN.
19. Creswell A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*. 2018;35(1):53-65.
20. Salimans T, Ho J, Xu X, et al. Improved techniques for training GANs. In: *Advances in Neural Information Processing Systems*; 2016 Dec 5-10; Barcelona, Spain. p. 2234-42.
21. Kingma DP, Salimans T, Welling M. Variational Inference with Normalizing Flows. In: *Proceedings of the 5th International Conference on Learning Representations*; 2017 Apr 24-26; Toulon, France.
22. Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. In: *Proceedings of the 32nd International Conference on Machine Learning*; 2015 Jul 17-23; Lille, France. p. 2256-64.
23. Bergmann D, Liao R, Czarnecki WM, et al. Generating realistic synthetic data with variational autoencoders. In: *Proceedings of the 2024 Conference on Computer Vision and Pattern Recognition*; 2024 Jun 16-22; Seattle, WA.
24. Radford A, Narasimhan K, Salimans T, et al. Learning Transferable Visual Models From Natural Language Supervision. In: *Proceedings of the 2021 Conference on Computer Vision and Pattern Recognition*; 2021 Jun 19-25; Nashville, TN. p. 10632-45.
25. Brown T, Mann B, Ryder N, et al. Language Models are Few-Shot Learners. *arXiv*. 2020;2005.14165.
26. Chowdhery A, Narasimhan K, Devlin J, et al. PaLM: Scaling Language Modeling with Pathways. *arXiv*. 2022;2204.02311.
27. Le Scao T, Catasta M, Radford A, et al. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *arXiv*. 2022;2211.05100.
28. Ramesh A, Pavlov M, Goh G, et al. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv*. 2022;2204.06125.
29. Rombach R, Blattmann A, Lorenz D, et al. High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv*. 2022;2112.10752.
30. Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. arXiv:2206.02904 .
31. Bender EM, Gebru T, McMillan-Major A, et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *ACM Conference on Fairness, Accountability, and Transparency*; 2021 Mar 18-21; Virtual Conference. p. 610-623.
32. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2019;542(7639):115-8.
33. Yi X, Walia E, Babyn P. Generative adversarial networks for medical image analysis. *Artificial Intelligence Review*. 2019;52(1):28-39.
34. Elgammal A, Liu B, Elhoseiny M, Mazzoni A. CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. In: *Proceedings of the 8th International Conference on Computational Creativity*; 2017 Jun 26-30; Paris, France.
35. Buolamwini J. Algorithmic Bias Detectable in Facial Recognition Technologies. *Proceedings of the 2018 ACM Conference on Fairness, Accountability, and Transparency*; 2018 Mar 23-24; New York, NY.
36. Dastin J. Hiring Algorithms and Bias. *Reuters*. 2018 Jul 25. Disponible en: https://www.reuters.com/article/us-amazon-com-jobs-idUSKBN1AB1F1
37. Mehrabi N. Fairness and Bias in AI Systems. *Communications of the ACM*. 2021;64(9):54-61.

38. Ferrara E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *arXiv*. 2023;2304.07683.
39. Voigt P. General Data Protection Regulation (GDPR). *European Journal of Law and Technology*. 2017;8(1):1-20.
40. King J. Rethinking Privacy in the AI Era. *Stanford HAI Policy White Paper*. 2024.
41. Lu J. Data Privacy, Human Rights, and Algorithmic Opacity. *California Law Review*. 2023;61:61-123.
42. Zhuk M. Intellectual Property and AI-Generated Works. *IP Journal*. 2023.
43. Lim R. Copyright and AI-Generated Works in the US. *Harvard Journal of Law & Technology*. 2023.
44. Ishii S. AI and Deepfake Detection Technologies. *Journal of Information Security*. 2019;10(4):123-34.
45. Novelli R. Accountability in artificial intelligence: what it is and how it works. *AI & Society*. 2023;38(3):379-92.
46. Ammanath B. Building Transparency into AI Projects. *Harvard Business Review*. 2022 Jun. Disponible en: https://hbr.org/2022/06/building-transparency-into-ai-projects
47. Zhong L, Haonan. Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution. *arXiv*. 2023;2303.09272.
48. Xia B, Qu L. Towards a Responsible AI Metrics Catalogue: A Collection of Metrics for AI Accountability. In: *IEEE/ACM 3rd International Conference on AI Engineering*; 2024 Apr 15-18; San Francisco, CA.
49. Lakhani K. AI Won't Replace Humans — But Humans With AI Will Replace Humans Without AI. *Harvard Business Review*. 2023 Aug. Disponible en: https://hbr.org/2023/08/ai-wont-replace-humans-but-humans-with-ai-will-replace-humans-without-ai
50. Brynjolfsson E. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W.W. Norton & Company; 2014.
51. Bessen J. AI and Jobs: The Role of Demand. *NBER Working Paper No. 24235*. 2019.
52. Luckin R. *Intelligence Unleashed: An Argument for AI in Education*. Pearson; 2016.
53. Selwyn N. *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press; 2019.
54. Victoria Department of Education. *Australian Framework for Generative Artificial Intelligence in Schools*. 2023. Disponible en: https://www2.education.vic.gov.au/pal/generative-artificial-intelligence/policy
55. Anantrasirichai N. Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*. 2022;55(4):589-656.
56. Zhou E. Generative artificial intelligence, human creativity, and art. *PNAS Nexus*. 2024;3(3).
57. Apple. *Introducing Apple Foundational Models*. 2024. Disponible en: https://machinelearning.apple.com/research/introducing-apple-foundation-models
58. *OpenAI*. *ChatGPT can now see, hear, and speak*. 2023. Disponible en: https://openai.com/index/chatgpt-can-now-see-hear-and-speak/
59. Wang Y. IoT in the Era of Generative AI: Vision and Challenges. *arXiv*. 2024;2401.01923.