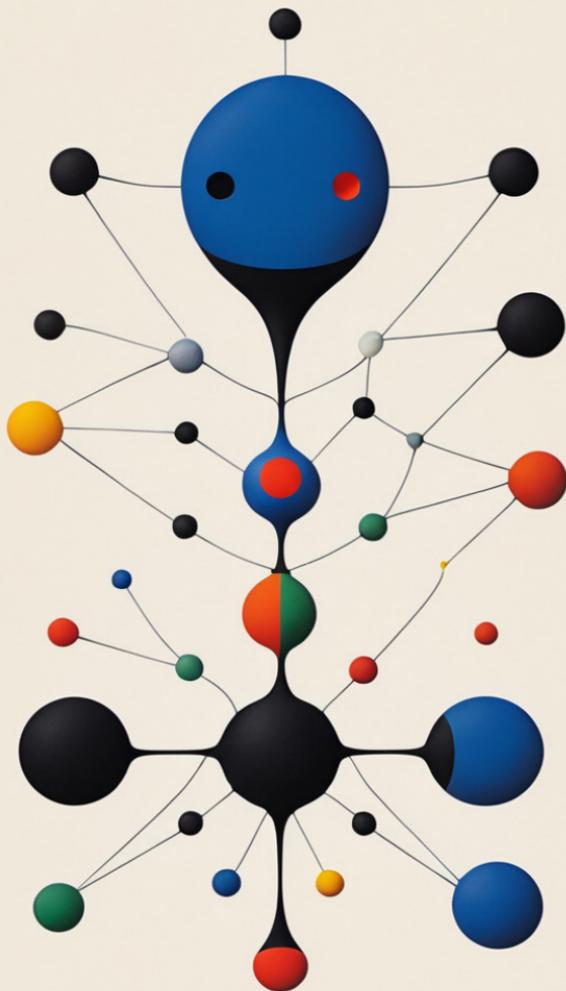


De GPT a LLaMA:

Una breve guía práctica sobre Inteligencia Artificial Generativa
para Cuba



Hamlet López García
Diasbel Lilianna Chang Olivera
María Carla O'Connor Barrios

De GPT a LLaMA:

Una breve guía práctica sobre Inteligencia Artificial Generativa
para Cuba

Hamlet López García
Diasbel Lilianna Chang Olivera
María Carla O'Connor Barrios



Imagen de cubierta generada con Stable Diffusion

© Hamlet López García, 2024

© Diasbel Lilianna Chang Olivera, 2024

© María Carla O'Connor Barrios, 2024

Sobre la presente edición:

© Instituto Cubano de Investigación Cultural

Juan Marinello, 2024

ISBN: 978-959-242-236-0

Instituto Cubano de Investigación Cultural Juan Marinello

Ave. Rancho Boyeros no. 63, Plaza de la Revolución,

La Habana, Cuba

<https://icic.cult.cu>

ÍNDICE

INTRODUCCIÓN / 5

¿POR QUÉ UN LIBRO PRÁCTICO SOBRE INTELIGENCIA
ARTIFICIAL? / 10

BLOQUEOS, CENSURA, Y LA NECESIDAD DE INTERACTUAR CON LAS IA
GENERATIVAS / 10

¿CÓMO PENSAR LA IA GENERATIVA? / 17

¿Y DESPUÉS DE USAR LA IA QUÉ? / 27

BREVE INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL / 34

INTELIGENCIA ARTIFICIAL GENERATIVA / 50

DEFINICIÓN DE LA IA GENERATIVA / 50

¿CÓMO FUNCIONA LA IA GENERATIVA? / 51

EL FUTURO DE LA IA GENERATIVA / 55

DESAFÍOS ÉTICOS Y DE SEGURIDAD / 58

INTELIGENCIA ARTIFICIAL GENERATIVA DE TEXTO / 63

QUÉ ES. CÓMO FUNCIONA. EJEMPLOS / 63

PARA QUÉ SIRVE / 73

SESGOS Y CONSIDERACIONES ÉTICAS / 76

INTELIGENCIA ARTIFICIAL GENERATIVA DE IMÁGENES / 78

QUÉ ES. CÓMO FUNCIONA. EJEMPLOS / 78

PARA QUÉ SIRVE / 80

SESGOS Y CONSIDERACIONES ÉTICAS / 84

¿CÓMO SEGUIR A PARTIR DE AQUÍ? / 100

REFERENCIAS / 103

SOBRE LOS AUTORES / 108

INTRODUCCIÓN

En las líneas de la literatura, el cine e incluso el periodismo se han tejido narrativas apocalípticas que exploran la influencia de la tecnología en nuestras vidas. La visión de un mundo completamente conectado, inmerso en un flujo constante de información, donde los límites de espacio y tiempo se desdibujan, ha suscitado preocupaciones diversas.

Dentro de este contexto nos enfrentamos no pocas veces al debate de cómo integrar las tecnologías digitales de forma que no sean solo una herramienta más, sino una extensión armoniosa de nuestras sociedades y en ellas, de nuestra cotidianidad.

En los últimos meses, un término que alguna vez pareció reservado para el ámbito de la ciencia ficción ha ganado una presencia cada vez más firme en la realidad: la inteligencia artificial. Específicamente, la inteligencia artificial generativa se ha materializado como una fuerza masiva, que en muy poco tiempo ya involucra a millones de usuarios, medios y empresas en la creación de contenido. Sin embargo, más allá del fervor tecnológico, ¿qué es realmente la inteligencia artificial generativa? y

¿cómo podemos aprovecharla para potenciar nuestras rutinas productivas y creativas?

Con estas y muchas otras preguntas nace el presente libro, escrito desde Cuba y para Cuba, para sus profesionales y sus estudiantes. El lector que tenemos en mente es aquella persona que para trabajar debe utilizar la computadora, llenar informes, buscar información y procesar textos e imágenes. No necesariamente se trata de un programador o un informático, pero sí de alguien que se beneficiaría con uso de la inteligencia artificial (IA en lo adelante) generativa, y que está acostumbrado a descargar e instalar programas, a pesar de que, en ocasiones, solo cuentan con interfaces en inglés.

El contenido de este texto aborda una tecnología con impactos significativos en la productividad de diversos sectores económicos, entre ellos el desarrollo de software, la biotecnología y la educación. Creemos que dicho impacto seguirá creciendo en los años por venir. Sin embargo, nos acercamos a una tecnología a la que resulta difícil acceder desde Cuba. Así como sucede con muchas otras herramientas digitales, el bloqueo impuesto por el gobierno de Estados Unidos impide el acceso de los profesionales y estudiantes cubanos a plataformas como ChatGPT y Claude.

En este sentido, podemos identificar dos consecuencias: en primer lugar, las ganancias obtenidas mediante

el uso de la IA, que incluyen mayor productividad, rapidez en el acceso a la información y procesamiento eficiente, no están al alcance de los residentes en la Isla. Un profesional que viva y trabaje en el exterior puede generar la misma cantidad de valor que produce uno viviendo en Cuba, pero en mucho menos tiempo, y por tanto resulta más productivo. La segunda consecuencia, y la que consideramos más importante a largo plazo, es que al no tener acceso, el residente en la Isla tampoco aprende a usarlas de la mejor manera; lo que dificulta enormemente el desarrollo de competencias y habilidades específicas para el uso de estas herramientas. Es por ello que, a pesar del bloqueo norteamericano, y de las dificultades para acceder a los modelos de IA generativa, nos pareció esencial escribir este libro, y aportar nuestro pequeño esfuerzo al desarrollo del país.

El libro se divide en cinco capítulos, acompañados de esta introducción y las conclusiones. En el primer capítulo presentamos una propuesta estratégica y consciente para el uso de la inteligencia artificial. Explicamos en detalle qué significa para nosotros ser estratégicos, pero desde ahora nos interesa destacar que una guía práctica para el uso de cualquier tecnología, no puede limitarse a indicar, solamente, qué botón apretar para obtener un resultado específico. Ni la tecnología, ni la vida son tan

sencillas. Ignorar las complejidades escondidas detrás de simples actos puede conllevar a grandes consecuencias, y en el capítulo hablamos también de los costos ocultos asociados al uso de la inteligencia artificial.

El segundo capítulo lo dedicamos a dar algunas pinceladas históricas; ofrecemos una definición provisional de inteligencia artificial; y trazamos un mapa de sus principales áreas de investigación. El tercer capítulo introduce a la inteligencia artificial generativa, abordando desafíos éticos y vislumbrando su futuro probable. Los capítulos cuarto y quinto se centran en la inteligencia artificial generativa de texto y de imágenes. Mencionan nombres específicos de servicios para probar estas tecnologías y se detiene en técnicas y recomendaciones para su uso.

El libro finaliza con lo que para nosotros constituyen las ideas esenciales del texto. Aquí adelantamos una: no importa tanto la tecnología en sí, lo que permite y cómo luce ahora mismo, sino nuestra relación con ella: cómo las usamos, para qué las usamos, y qué lugar le damos en nuestras vidas.

Expresamos el deseo sincero de que este libro despierte la curiosidad y el interés de experimentar. Las modas van y vienen, y no negamos que haya en estos momentos (finales de 2023 y principios de 2024) cierta

exageración mediática asociada con la inteligencia artificial generativa. Pero estamos convencidos que una vez pasado el bullicio de los apocalípticos y de los integrados, siguiendo los términos del pensador italiano Humberto Eco, va a permanecer el valor de la inteligencia artificial generativa para los profesionales; y que los lectores tengan la opción de acceder a ella es nuestro interés principal.

¿POR QUÉ UN LIBRO PRÁCTICO SOBRE INTELIGENCIA ARTIFICIAL?

BLOQUEOS, CENSURA, Y LA NECESIDAD DE INTERACTUAR CON LAS IA GENERATIVAS

Si bien la inteligencia artificial lleva décadas desarrollándose en diferentes ramas de la ciencia y técnica, el mes de noviembre de 2022 representó un hito, porque marcó un cambio en la percepción de uso y utilidad de esta herramienta con la llegada de [ChatGPT](#), desarrollada por la empresa norteamericana OpenAI. Desde esa fecha la compañía ofrece un servicio de chatbot en línea, gratuito hasta el momento en que redactamos este texto, apoyado en el modelo de inteligencia artificial ChatGPT 3.5. Puede responder preguntas del usuario de una manera coherente y natural, cumplir órdenes como resumir información, o redactar documentos simples con un formato específico, entre otras habilidades.

La inteligencia artificial que usa está basada en una familia de redes neuronales artificiales (de las que hablaremos más adelante) denominada GPT, que corresponde a las siglas en inglés de *Generative Pre-trained Transformer*, y en español sería transformador pre-entrenado generativo. Esta familia de redes neuronales está

especialmente entrenada mediante aprendizaje profundo (*Deep Learning*) con gigantescas cantidades de datos para poder sostener conversaciones con el usuario en un lenguaje natural y de forma coherente.

Por su parte, [Claude](#), aparecido varios meses después, es otro popular servicio de chatbot en línea con un modelo de inteligencia artificial, basado también en la familia de redes neuronales GPT, pero en este caso ofrecido por la empresa norteamericana Anthropic. Esta última tiene como sello distintivo presentarse como un rival de OpenAI, y con mucha más atención a la seguridad y a la responsabilidad social.

Lo que nos interesa resaltar va mucho más allá de las ofertas gratuitas y populares de OpenAI y Anthropic. ChatGPT y Claude son servicios empresariales. Las empresas surgen y desaparecen, también sus servicios. ChatGPT puede desaparecer mañana. Sin embargo, la inteligencia artificial generativa, como tecnología de propósito general, va a seguir su avance en las esferas de la vida social y productiva donde pueda tener aplicación.

Las habilidades y competencias que un profesional adquirió usando ChatGPT y Claude deberán ser fácilmente trasladables a las nuevas herramientas del futuro. Los que nunca antes interactuaron con estas herramien-

tas, aún no saben qué se puede y qué no se puede hacer, comenzarán de cero. Por consecuencia, dichos profesionales podrían ser menos productivos que sus pares, quienes ya adquirieron tal conocimiento y lo incorporan en su trabajo diario. Serán menos productivos no porque sean menos capaces, sino porque necesariamente requerirán un tiempo para aprender a usar las nuevas herramientas.

Antes, identificamos a la inteligencia artificial generativa como una tecnología de propósito general. Esta es una denominación importante que ayuda a identificar cuáles son las tecnologías, dentro de todo el abanico de ofertas tecnológicas que ofrecen nuestras sociedades, que son realmente estratégicas, de las que debemos asegurarnos tener el acceso, así como las competencias para usarlas, si no queremos perder oportunidades de desarrollo. Una *tecnología de propósito general* es aquella que tiene impacto en múltiples industrias, economías y sociedades por su amplia adopción y su capacidad para potenciar la productividad a través de varios sectores.

Por lo general impulsan cambios estructurales en toda la economía, a través de la reorganización o creación de sectores económicos totalmente nuevos, y son determinantes para su crecimiento a largo plazo. En esta categoría de tecnologías está el uso de la electricidad, las

computadoras e Internet. La inteligencia artificial generativa tiene potencial para ser una tecnología de propósito general porque puede utilizarse en múltiples campos, con posibilidad además de ser disruptiva, desplazar puestos de trabajo, crear nuevos empleos, y reforzar la digitalización de nuestras sociedades.

Hay vías para evadir o al menos compensar las limitaciones de acceso a la IA generativa a causa del bloqueo norteamericano. Algunas de ellas ya son conocidas por parte de la sociedad cubana, como sello local de nuestra particular manera de digitalizarnos. Así sucede con el uso de las VPN o lo que es lo mismo, las redes virtuales privadas, que canalizan nuestro tráfico de Internet a través de servidores fuera de Cuba. Esta es una de las maneras más fáciles y baratas de acceder a ChatGPT, Claude y servicios similares. Otras son menos conocidas, como el uso del software libre y abierto.

El tema de las licencias propietarias y no propietarias en el mundo de las aplicaciones informáticas es suficientemente complejo y profundo como para no tratarlo aquí. Merece varios libros por sí mismo. Aquí solo lo mencionamos a propósito de la estrategia de desarrollo de la inteligencia artificial generativa seguida por la empresa norteamericana Meta (la misma de Facebook), que incluyó la publicación de sus modelos entrenados de IA

generativa denominados LLaMA (que serían las siglas que corresponden a la frase en inglés *Large Language Models from Meta*), listos para ser usados y modificados por una comunidad de entusiastas.

Es necesaria una distinción entre servicio de acceso a un modelo de inteligencia artificial y el modelo en sí mismo. OpenAI ofrece varios servicios de acceso a sus modelos de inteligencia artificial, uno de ellos gratuito, pero no los publica. Están celosamente guardados en los servidores de la empresa. Meta por su parte ha publicado los modelos que ha desarrollado, a partir de los cuales se pueden construir servicios de chatbot como los que ofrece OpenAI. Esto permite experimentar con la IA generativa, incluso, desde la comodidad de la computadora personal, si se cumplen las condiciones técnicas mínimas para ello.

Señalamos esta distinción porque hasta principios de febrero del 2023, la única opción ampliamente disponible y gratuita para probar un modelo de inteligencia artificial generativa de texto, era la ofrecida por la empresa norteamericana OpenAI, con el modelo ChatGPT 3.5. Sin embargo, este se encuentra alojado en una empresa y como usuarios no tenemos acceso a sus características internas. También es preciso reconocer que si la empresa estuviera de acuerdo (hasta el momento no es

el caso), no podríamos ejecutarlo de manera independiente en nuestras máquinas. Es simplemente demasiado grande y costoso, por lo que solo podemos acceder a él a través de la web.

Los modelos LLaMA por su parte, también están basados en redes neuronales artificiales especializadas en la generación de texto; pero a diferencia de OpenAI enfatizan el volumen y la calidad de los datos de entrenamiento, en lugar de priorizar la cantidad de parámetros (más adelante, cuando hablemos de las redes neuronales, veremos qué son los parámetros). La empresa Meta los hizo públicos en varias versiones de tamaños: 7, 13 y 70 billones de parámetros. Como comparación, ChatGPT 3.5 cuenta con 175 billones de parámetros.

Algunas de las versiones más pequeñas de los modelos LLaMA, junto a otros programas informáticos realizados por la comunidad de código abierto, permiten ejecutar estos modelos de inteligencia artificial en computadoras domésticas relativamente modernas. El hecho de que los modelos más pequeños se puedan ejecutar en una computadora doméstica significa que el profesional cubano, si tiene la ventaja de contar con una computadora adecuada, puede experimentar y aprender

a usar la inteligencia artificial generativa desde su casa o su trabajo, y por tanto comenzar a construir las competencias y habilidades asociadas.

Resulta importante aclarar que los modelos LLaMA no son modelos que tienen las mismas prestaciones que ChatGPT o Claude. La cantidad de parámetros influye en las prestaciones que ofrece. Por ejemplo, tienen más dificultades para deducir y extraer conclusiones a partir de textos, y seguir órdenes. Son más difíciles o menos intuitivos de usar. También pueden ser más lentos, dependiendo del tamaño del modelo y las características técnicas de la computadora del usuario.

Con independencia de lo antes expuesto, los autores del presente libro pensamos que la familia de modelos LLaMA ofrece una opción tecnológica atractiva para Cuba en el desarrollo de competencias y capacidades autóctonas. No solo porque las competencias necesarias para utilizarlas con eficacia son las mismas en esencia, al margen de la cantidad de parámetros que tenga el modelo; sino por lo atractivo de tener un modelo local: los datos de las interacciones con la inteligencia artificial no van a una empresa extranjera fuera del país, sino que se quedan en los ordenadores propios y, por tanto, podemos usar a la inteligencia artificial como asistente en tareas sensibles o privadas.

Además, algunos de los modelos LLaMA son publicados con licencias permisivas, que permiten su modificación,

junto a su uso comercial. Es decir, podemos construir productos comerciales basados en IA generativa al margen del bloqueo norteamericano. Al tener licencias de código abierto, tenemos acceso a algunos de los cuerpos de datos que sirven para su entrenamiento y podemos investigar cómo construirlos y mejorarlos. Entonces no se trata solamente de la ventaja o no de usarlos desde nuestras computadoras o de tener mayor privacidad, sino de la posibilidad de construir capacidades tecnológicas propias para el desarrollo soberano en inteligencia artificial generativa.

¿CÓMO PENSAR LA IA GENERATIVA?

Existe un campo del conocimiento, denominado Estudios Sociales de la Ciencia y la Tecnología, el cual se ha dedicado a estudiar las tecnologías, no como artefactos obvios en sí mismos, listos para ser usados, sino que se ha enfocado en interpretar sus capacidades y funcionamientos desde el punto de vista de la política, la cultura, la economía, el género, etc.

Esta idea, que puede parecer abstracta para algunos, especialmente aquellos con un acceso inmediato y fácil a una computadora avanzada, es una verdad evidente para el cubano y cualquier otro que enfrente dificultades totalmente fuera de su control para acceder a una aplicación o a un servicio digital. El bloqueo norteameri-

cano de pronto convierte a las tecnologías digitales en instrumentos políticos, por ejemplo. El hecho de que sea necesaria una computadora relativamente nueva para probar un modelo de inteligencia artificial como LLaMA hace real e inmediata la relación entre las capacidades económicas y la posibilidad de experimentar nuevas tecnologías. De igual forma, podríamos seguir con otros temas sumamente esenciales, pero la idea que nos interesa resaltar aquí, es que las tecnologías no son objetos sencillos por sí mismos y tampoco son transparentes. No dejan ver de inmediato todos los supuestos que resultan necesarios para ponerlos en uso. Están cargados de contexto y de cultura, especialmente del contexto y la cultura donde se originaron.

¿Por qué traer un tema académico al uso profesional del día a día de la inteligencia artificial? ¿No debería ser un asunto de aprender cuál botón apretar y ya?

En realidad, no. Los autores sostenemos que es esencial y estratégico que los lectores consideren a la tecnología, y especialmente a la inteligencia artificial generativa, como objetos que están cargados de contexto, de agendas y limitaciones propias, puestas ahí por sus creadores. También, afirmamos que no hay nada más

práctico que examinar con cuidado ese contexto, esa agenda y esas limitaciones para actuar en consecuencia.

Si pensamos en el contexto de las herramientas de inteligencia artificial generativa, en los significados que conlleva, y si además conocemos sobre su historia, podemos tener una mejor perspectiva de cuáles son sus oportunidades y limitaciones, no en abstracto o en términos generales, sino con un impacto inmediato en su uso diario y a largo plazo. Es lo que llamamos un *uso estratégico de las tecnologías*. Un uso estratégico es un uso consciente, que aprovecha las facilidades de uso de la tecnología, pero al mismo tiempo está al tanto de sus limitaciones, de sus sesgos, y que la trata con la desconfianza suficiente como para no depender demasiado de ella.

El contexto es importante porque carga consigo muchos supuestos. La IA generativa, y aquí vamos a tomar como ejemplo específicamente a la IA generativa de texto, no a la de imagen, genera contenido prediciendo. A partir de enormes conjuntos de datos, sacados en su mayoría de Internet, los modelos de IA aprenden patrones y relaciones entre palabras e ideas, las cuales utilizan luego para asignar probabilidades y predecir cuál palabra le sigue a cuál. Esto, sin duda, tiene implicaciones.

Hasta hace muy poco la mayoría de los contenidos disponibles en Internet estaban en inglés. Fueron ori-

ginados en países anglosajones, o por sectores acomodados de otras sociedades que usan el inglés como su segunda lengua. Esto significa que la cultura y los conocimientos con los que el modelo de IA generativa ha estado más en contacto y tiene mayor familiaridad provienen de la lengua escrita en inglés. El conocimiento de la historia, la cultura, las obras del universo anglosajón tiene mayor profundidad y riqueza que el conocimiento de otros universos culturales.

Si se piensa un poco en ello, esto no es sorprendente y de hecho es esperable. Pero no es evidente por sí mismo. Para un usuario anglosajón de una IA generativa como ChatGPT o Claude puede ser perfectamente invisible. Para él la IA es transparente, y a menos que sea un estudioso de las culturas no anglosajonas, no tiene por qué notar las limitaciones culturales de estas herramientas. Puede incluso considerar a la IA generativa como un sustituto más rápido y cómodo que un buscador de Internet, como Google, a la hora de obtener información histórica o cultural.

Un habitante de una cultura no anglosajona, digamos, un usuario latinoamericano, si usa la IA generativa para obtener conocimiento sobre su propia cultura va a te-

ner mayores dificultades. Obtendrá información pobre, o lo que puede ser peor, inventada. Por otro lado, le va a resultar más fácil conocer sobre la cultura inglesa que sobre la latinoamericana, e incluso más; va obtener mejores respuestas, con mejor calidad, si hace las preguntas en inglés. Y todo ello está determinado por los textos que le sirvieron de entrenamiento inicial a la IA.

Otro ejemplo de lo práctico que puede ser detenerse un momento a pensar en el contexto, es considerar cuáles son las precondiciones necesarias para utilizar con efectividad las IA como ChatGPT o Claude. Vamos desde las más elementales. Es necesaria una computadora o un móvil con Internet. Es necesario, si se está en Cuba, tener acceso a una VPN o encontrar una IA que no esté bloqueada para la Isla y que sea gratuita. Supongamos que esas pre condiciones se han cumplido. Ahora bien, si esa herramienta se va a usar para el trabajo hay dos aspectos como mínimo, que deben ser considerados: el manejo de los datos y la sostenibilidad de su uso.

Hablemos primero del manejo de los datos. Al emplear un servicio de chat con una IA gratuita como ChatGPT 3.5 no gastaremos dinero. Sin embargo, los datos de nuestra interacción serán guardados y presumiblemente serán vendidos o usados para entrenar nuevas generaciones de IA,

a no ser que en los términos de uso del servicio se declare explícitamente lo contrario. Aun así, es mejor asegurarse de la existencia de mecanismos de control para que la empresa que ofrece el servicio cumpla esos términos de uso.

Esto implica que debe evaluarse cuidadosamente cuáles son los datos personales o profesionales que se van a compartir con la IA; y de preferencia, no compartir ninguna información privada. Lógicamente, esto limita la utilidad práctica de la IA, pues no se debe usar, por ejemplo, para que ayude en la elaboración de reportes financieros o de cualquier otro tipo que contenga información secreta o sensible. Tampoco debería escribirse código informático que distinga de la competencia. En ningún caso se debe compartir con ella información que contenga datos de cuentas bancarias, direcciones personales, o cualquier otra información que nadie daría de buenas a primeras a un desconocido.

Incluso si la empresa tiene términos de uso respetuosos de la privacidad del usuario, y existen salvaguardas en funcionamiento para proteger los datos personales, hay que considerar la posibilidad de errores. A principios de diciembre de 2023, investigadores de Google revelaron que fueron capaces de obligar a ChatGPT a proporcionar nombres, números telefónicos y direcciones de personas particulares dándole instrucciones

específicas que forzaban un mal funcionamiento del modelo. Por ejemplo, **le pidieron que repitiera la palabra “poema” hasta el infinito**, lo que forzó al modelo a ir más allá de sus mecanismos de entrenamiento, y comenzó a mostrar los datos originales, que incluyó información personal. Un evento un poco más preocupante ocurrió a principios de 2023 cuando **empleados de Samsung filtraron datos sensibles de la empresa** en tres ocasiones diferentes al usar ChatGPT para hacer más eficientes sus propias tareas laborales.

El segundo aspecto es la sostenibilidad del empleo de la herramienta a lo largo del tiempo. Implica preguntarse acerca de las consecuencias si por cualquier razón ya no se puede acceder a la inteligencia artificial. ¿Qué pasaría si cambian los términos de uso de la IA que estoy usando? ¿Si deja de existir una oferta gratuita o si los costos de acceso súbitamente suben? ¿Podría dejar de trabajar con ella y aun así mantener la misma calidad en mi trabajo? ¿Hay alternativas disponibles?

En ocasiones no tiene que ver con que los términos de uso de la empresa cambien, sino que el propio modelo evoluciona y sus funcionalidades se modifican según la visión de la empresa que lo desarrolla. Por lo mismo, las prácticas y métodos por los que se obtuvo un determi-

nado resultado ayer pueden no ser los mismos hoy, y seguramente no lo serán en unos meses. Resulta aconsejable evaluar constantemente la conveniencia del uso del modelo para los flujos de trabajo o los resultados que se desean obtener, y al mismo tiempo tener listos planes de contingencia por si el acceso a la inteligencia artificial ya no es posible.

Un poco más arriba habíamos mencionado el hecho de que los grandes modelos de lenguaje que le sirven de base a la inteligencia artificial generativa de texto se entrenan en grandes cuerpos de datos sacados en su mayor parte de Internet. Igualmente, las inteligencias generativas de imágenes se han entrenado en enormes conjuntos de imágenes provenientes de Internet. Pero un detalle importante a la hora de emplearlas es considerar la propiedad intelectual de los contenidos generados. ¿A quién atribuir la autoría?

En los términos de servicio de ChatGPT, en noviembre de 2023 se dice, por ejemplo, que la propiedad del contenido generado pertenece al usuario. Pero al mismo tiempo recomienda al usuario revisar la legislación sobre propiedad intelectual específica de su país. ¿Qué ocurre si el contenido generado se parece demasiado a otro contenido que está disponible en Internet, pero que es propiedad intelectual de una empresa o de un individuo?

El profesional que use la inteligencia artificial generativa debe estar al tanto de estas posibilidades si usa el contenido generado en su propio trabajo, más aún si será publicado.

Otro aspecto a tener en cuenta es cuán confiable puede ser el contenido generado por el modelo, y si será convenientemente evaluado antes de ser utilizado de manera profesional. Los contenidos generados por una inteligencia artificial generativa pueden ser inexactos, falsos, pueden contener sesgos e incluso ser abiertamente discriminatorios o proporcionar información dañina. No se trata de copiar y pegar el contenido generado por la inteligencia artificial en lo que sea que estemos trabajando. Hay evidencias de que las inteligencias artificiales reproducen estereotipos y criterios discriminatorios. ¿Cómo vamos a manejarlas? Antes de pensar en incorporar a la inteligencia artificial generativa en procesos de trabajo profesionales, hay que separar tiempo para revisar el contenido producido de manera sistemática, para atender a estas posibles dificultades y corregirlas.

Por último, cabe preguntarse si es posible incorporar a la inteligencia artificial generativa a procesos de trabajo que necesiten texto o imágenes de buena calidad de manera consistente, a procesos de trabajo que exijan

que a una determinada entrada u orden por parte del usuario se produzca siempre la misma salida de texto o de imagen. La consistencia de resultados es esencial por ejemplo para en caso de errores rastrear la causa y corregirla de manera definitiva. Si un móvil proporciona la temperatura ambiente de manera incorrecta, se puede rastrear el código fuente de la aplicación que reporta la temperatura con la tranquilidad que una vez arreglado, ya no dará más problemas. Va a ser el mismo código cada vez que se ejecute. Sin embargo, la inteligencia artificial generativa no funciona de igual forma. Dará un resultado ligeramente diferente cada vez que se interactúe con ella. Incluso puede ser que en un momento proporcione una respuesta incorrecta y en otro momento, con la misma orden del usuario, su respuesta sea correcta. Lo que implica que un proceso de trabajo donde esté incorporada una inteligencia artificial generativa, hasta el momento, no va a ser apretar un botón y desentenderse del asunto con la tranquilidad de que todo saldrá bien porque así fue en el pasado.

Hemos visto que hay limitaciones y fortalezas de la inteligencia artificial generativa que vale la pena tener en cuenta a la hora de usarlas. Además, estas fortalezas y limitaciones se entienden mucho mejor, y por tanto

se gestionan mucho mejor, si se considera cómo está hecha esa inteligencia artificial, dónde y por quién fue desarrollada. Faltaría por ver si lo que podemos avizorar de su probable futuro también tiene alguna influencia en nuestros modos de relacionarnos con ellas.

¿Y DESPUÉS DE USAR LA INTELIGENCIA ARTIFICIAL QUÉ?

Si bien el primer artículo presentando la tecnología que soporta a ChatGPT fue publicado en 2018, su salida al público y el inicio de la enorme popularidad de la tecnología generativa es muy reciente. ChatGPT comenzó como servicio en noviembre de 2022. Unos meses más tarde, en febrero del 2023 Meta anunció su propio modelo de lenguaje, LLaMA, el cual sería ampliamente adoptado por la comunidad de código abierto. Si se siguen las noticias, todas las semanas hay alguna novedad o se publica un nuevo servicio de chatbot que amplía lo que se puede hacer con la inteligencia artificial generativa.

En muy poco tiempo aparecieron los modelos multi-modales, con los cuales se puede interactuar por medio de imágenes, texto o video. Algunos servicios incorporaron la posibilidad de enriquecer la información a través de búsquedas en Internet. Estos son solo ejemplos de que la rapidez de los cambios no permite decir con

seguridad cuáles son las funcionalidades que van a definir a la inteligencia artificial generativa por los próximos años. Al ser novedosa y emerger de un campo de investigación donde constantemente aparecen nuevos adelantos, es aún inestable en sus rasgos definitorios. Por la misma razón, todavía es pronto para fijar con claridad cuáles son las habilidades necesarias para usarlas de forma efectiva y sacarles el máximo provecho. Saber dar instrucciones de manera precisa, y proporcionándole ejemplos a la inteligencia artificial siempre será valioso, pero la tendencia indica que los modelos con el tiempo irán mejorando en su capacidad de entender la intención original del usuario, y por lo tanto será menos relevante ser preciso.

No podemos asegurar que se ha aprendido de una vez y por todas cómo interactuar con la inteligencia artificial y que ese conocimiento no será necesario modificarlo en los próximos tres o cinco años. Por lo tanto, si se quiere usar al máximo una tecnología que aún no está lo suficientemente estable en su desarrollo, la estrategia cambia radicalmente. Ya no se trata únicamente de asegurarse de saber al detalle qué escribir, cómo navegar las pantallas del chatbot o cuál botón apretar, aunque sigue siendo importante, por supuesto. Ahora se trata de aprender a aprender. De hacerlo constantemente, adaptándonos

a los cambios y de aprovecharlo al máximo en nuestro trabajo. Bien mirado, la exigencia de mantenernos aprendiendo no se cumple sólo en nuestra interacción con la inteligencia artificial. Casi ningún elemento se mantiene estable el tiempo suficiente como para acomodarnos tranquilos en nuestro conocimiento. Es vital estar preparados para un entorno en permanente transformación, donde nuestros conocimientos y habilidades pueden tornarse obsoletos si no tenemos el cuidado de actualizarlos de manera constante.

Esto implica que, primero, debemos ser conscientes de que sin importar el esfuerzo dedicado, el conocimiento que obtengamos sobre inteligencia artificial generativa siempre va a ser parcial, con una alta tasa de envejecimiento. Segundo: hay que explorar continuamente el entorno. Buscar activamente las nuevas herramientas y funcionalidades que han aparecido, así como valorar lo que ofrecen. Para ello habrá que usar con regularidad motores de búsqueda en Internet, sitios web especializados y comunidades en línea donde se difunde ese tipo de información. Además, será preciso evaluarla de manera permanente, conscientes de que esta puede cambiar en cuestión de semanas.

Actualmente es común encontrar listados de servicios de inteligencia artificial generativa para casi cualquier

tarea; y es de esperar que se sigan publicando, porque aún no existe el listado definitivo. Aparecen cada día nuevas aplicaciones e igualmente desaparecen, pues la existencia de las empresas que las ofrecen no está escrita de una vez para siempre. Algunas de ellas no logran sobrevivir el duro entorno competitivo del mercado, y por tanto la manera en la que nos acomodamos a su interfaz de usuario, configurar los parámetros de la IA y otras características particulares puede hacerse obsoleta de manera súbita. Por ello la necesidad de explorar constantemente.

Aprender a aprender es desarrollar las habilidades y la actitud de adaptarse y adquirir nuevos conocimientos permanentemente. Lo anterior incluye cultivar de manera consciente prácticas como examinar críticamente el entorno, el pensamiento propio y la relación entre ambos, enfocarse en resolver problemas y fomentar la auto reflexión. Aplicado al uso de la inteligencia artificial generativa implica incorporar rutinas diarias de exploración y búsqueda de nuevas herramientas, y a su vez evaluar no solo las nuevas inteligencias artificiales que encontremos, sino también nuestra capacidad de aprovecharlas al máximo. Es posible que una nueva herramienta se ajuste mejor a los procesos de trabajo, que

nos haga más eficientes y productivos. Habrá otras que no se adapten o lo hagan de manera muy pobre, a pesar de los comentarios favorables que se puedan encontrar en la red sobre ese modelo en particular. En otros casos será necesario adaptar las rutinas de trabajo, si eso significa seguir cumpliendo nuestros objetivos profesionales. En todo caso, solo explorando y probando las nuevas herramientas es que podemos estar seguros que estamos aprovechando lo mejor posible de lo que el entorno tecnológico tiene para brindarnos.

Eventualmente, también significa decidir de las inteligencias artificiales encontradas, cuál satisface mejor las necesidades profesionales que tenemos, y al mismo tiempo promete ser más estable a largo plazo o al menos no desaparecer súbitamente. Resulta aconsejable buscar aquellos servicios que tienen empresas de buena reputación, un largo historial de resultados, y una buena cantidad de usuarios. Esos parámetros no garantizan que la empresa o el servicio que ofrece no desaparezca de la noche a la mañana; pero al menos ofrecen mejores posibilidades de supervivencia. En el peor escenario, como tendencia estas suelen avisar con antelación de cualquier cambio en sus servicios. Adicionalmente

que un servicio tenga muchos usuarios significa que hay un grupo mayor de personas aprendiendo del uso de la herramienta, acumulando ese conocimiento y potencialmente compartiendo lo que saben.

Un aspecto del que no hemos hablado lo suficiente aún, pero está directamente relacionado con la necesidad de vigilar el desarrollo de las inteligencias artificiales, es que el entorno jurídico que regula su uso puede cambiar y hay que estar preparados para dichos cambios.

A punto de concluir la redacción de este libro, por ejemplo, la Unión Europea aprobó la [primera legislación comprensiva](#) para la regulación de la inteligencia artificial. Uno de sus puntos es que el usuario de un servicio en línea debe ser advertido de manera explícita cuando esté interactuando con una inteligencia artificial. Antes, un profesional con un sitio web que proporcionaba servicios a consumidores europeos, podía o no declarar que tenía una inteligencia artificial chateando con los usuarios. Ahora es obligatorio.

Hasta aquí, hemos visto lo esencial de mantener una perspectiva de aprender a aprender, establecer una vigilancia constante del entorno digital, ser lo suficientemente flexible como para modificar prácticas y flujos de

trabajo en dependencia de los cambios tecnológicos y regulatorios... Sin embargo, se puede ir todavía más allá.

Muchos de los nuevos desarrollos de la inteligencia artificial generativa, así como información de cómo usarlos mejor, pueden encontrarse en las comunidades en línea que reúnen profesionales y aficionados. Sitios como [Reddit](#) y [Discord](#) tienen algunas de las comunidades más activas y es conveniente seguirlas de cerca e incluso puede servir para identificar oportunidades de colaboración con otros profesionales, especialmente dentro del movimiento de código abierto, donde la actividad es constante para hacer más fácil el uso de los nuevos desarrollos tecnológicos. En estas comunidades siempre se podrá encontrar nuevos métodos, aproximaciones y estrategias para aprovechar al máximo la inteligencia artificial generativa, empujando hacia delante la frontera de lo que somos capaces de hacer como profesionales.

BREVE INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

El encanto por la automatización y la inteligencia artificial está arraigado en nuestras mentes como seres racionales. No solo concebimos continuamente herramientas y mecanismos sofisticados para aliviar nuestra carga de trabajo, sino que también aspiramos a un máximo de productividad junto al mínimo de esfuerzo humano posible. En esta extraña mezcla de ingenio y pereza está la base del desarrollo de la inteligencia artificial y es una parte integral de quiénes somos.

En Aristóteles, por ejemplo, ya podemos encontrar una reflexión sobre un mundo automatizado, y adelantar algunas de sus consecuencias para la distribución del trabajo:

Pues si cada uno de los instrumentos pudiera cumplir por sí mismo su cometido obedeciendo órdenes o anticipándose a ellas, si, como cuentan de las estatuas de Dédalo o de los trípodas de Hefesto, de los que dice el poeta que entraban por sí solos en la asamblea de los dioses, las lanzaderas tejieran solas y los plectros tocaran la cítara, los constructo-

res no necesitarían ayudantes ni los amos esclavos.
(Aristóteles, *Política*, p. 55)

Más adelante en el tiempo, Ramón Llull, un pensador y misionero catalán del siglo XIII, en su obra *Ars Magna*, hipotetizó una máquina lógica capaz de construir de forma automática argumentos irrefutables para usar en las discusiones religiosas con los no cristianos. Es decir, una máquina para razonar automáticamente. Concibió para ello un sistema de ideas simples junto a reglas para su combinación, las cuales se ejecutarían mediante una estructura de palancas y ruedas.

Por otro lado, el pensador árabe **Ismail Al-Yazari** escribió en el siglo XII lo que se considera uno de los más influyentes libros sobre autómatas de la Edad Media, *El libro de conocimientos de dispositivos mecánicos ingeniosos*. Este incluía la descripción de alrededor de 50 autómatas junto a las instrucciones de cómo construirlos y alcanzó una amplísima divulgación, con copias manuscritas que podían encontrarse en las cortes de toda Europa.

Acercándonos a la era moderna, en 1936, el matemático y científico inglés Alan Turing publicó su artículo *Sobre los números computables*. Este introdujo el concepto de una máquina universal capaz de realizar cualquier cálculo que pudiera llevarse a cabo mediante un algoritmo, o lo que es lo mismo, por una serie de pasos bien

establecidos para alcanzar un resultado. Dicha idea sentó las bases para el desarrollo de la inteligencia artificial, pues demostró cómo las máquinas pueden simular procesos de pensamiento y toma de decisiones humanas mediante algoritmos programados. Podríamos incluso pensar que fue uno de los momentos fundacionales.

Pero desafortunadamente ese momento fundacional solo pudo concretarse veinte años después, en 1956, cuando John McCarthy, científico y psicólogo cognitivo estadounidense, acuñó de manera definitiva el término "inteligencia artificial" en la Conferencia de Dartmouth celebrada en Hanover, New Hampshire, Estados Unidos. Con ese acto dio inicio oficialmente al campo de investigación, definió sus problemas y grandes líneas de trabajo, muchas de las cuales se mantienen hasta nuestros días.

Existen otros momentos en la historia de la computación y la inteligencia artificial que muestran el avance indetenible en la automatización de actividades humanas, con muestras notables de creatividad e ingenio. Por nuestra parte rescatamos de estas pinceladas las siguientes notas:

- La preocupación por las consecuencias sociales y económicas de la introducción de la automatización están presentes desde el inicio. Sin embargo, no ha

detenido el avance de su aplicación en los procesos de trabajo, e incluso, en el razonamiento.

- La tecnología, en particular sus segmentos de avanzada como es la inteligencia artificial, requiere de fuertes inversiones, y su desarrollo tiende a fortalecer las ventajas competitivas ya pre existentes de las sociedades que la impulsan. No es casualidad que Estados Unidos, una de las potencias vencedoras tras la Segunda Guerra Mundial, y la única que emergió con la economía y base industrial intacta, rápidamente progresara a la cabeza de las naciones en el desarrollo de la computación y de la inteligencia artificial, logrando una hegemonía en el campo que se mantiene hasta hoy en día.

Aunque este no es un libro estrictamente académico, y por tanto no estamos obligados a dar una definición rigurosa, una vez que hablamos de la historia, sí creemos útil declarar qué es para nosotros inteligencia artificial. Repetimos que no pretendemos ninguna rigurosidad en esta definición, solo un espacio común de comprensión con el lector.

Entenderemos por *inteligencia artificial* al campo de la ciencia y la tecnología dedicada a otorgar cualidades inteligentes a las máquinas. Por inteligencia, nos referi-

mos a la capacidad de razonar, planificar y autorregular el comportamiento en función de alcanzar metas.

Uno de los rasgos comunes entre las diferentes definiciones que se pueden encontrar de inteligencia artificial, tanto en redes digitales como en la literatura especializada, es precisamente la noción de la inteligencia humana como el patrón a seguir. Para construir la definición se necesita al intelecto humano como referente; por ser la máxima expresión de inteligencia que conocemos en la naturaleza. No es de extrañar que la estrategia comúnmente seguida a la hora de responder a la pregunta ¿qué es la inteligencia artificial?, sea proponer la semejanza al ser humano cuando enfrenta un problema, y por lo mismo dichas definiciones se acompañan de verbos y acciones característicamente humanas como razonar, aprender o plantearse objetivos. Dos ejemplos:

- El [sitio web de la Enciclopedia Británica](#) define a la inteligencia artificial como la capacidad de ejecutar tareas comúnmente asociadas a seres inteligentes. Casi enseguida añade que el término es aplicado con frecuencia a sistemas dotados de los procesos intelectuales usualmente visto en los humanos.
- Stuart Russell y Peter Norvig, en uno de los libros más usados a nivel internacional para la formación de ingenieros y científicos en inteligencia artificial (*Artificial in-*

telligence: A modern approach, 2010), dividen el universo de definiciones posibles en dos conjuntos: aquellas que enfatizan la racionalidad y el pensamiento, contra las que enfatizan el comportamiento inteligente.

El hecho de que reúnan las definiciones posibles en esos dos conjuntos reitera nuestra afirmación inicial, de que inevitablemente hay que tomar a las capacidades humanas como punto de partida.

Una de las más notables entre estas capacidades, y la primera por la cual hacemos una evaluación rápida de cuán inteligentes son las personas que nos rodean (casi siempre de manera injusta), es el manejo del lenguaje. Existe un vasto campo de investigación dentro de las ciencias de la computación denominado *Procesamiento de Lenguaje Natural* (*Natural Language Processing*), dedicado íntegramente a hacer a las computadoras capaces de entender, interpretar y generar lenguaje humano. Tiene múltiples aplicaciones, por ejemplo, en la traducción automática entre diferentes idiomas, la interacción con las máquinas por medio de la conversación y no por teclados y pantallas, realizar análisis de sentimientos en textos de manera automática, entre otros.

Sin embargo, este campo lidia con desafíos increíblemente complejos. Las palabras no tienen significados fijos asociados con ellas, sino que varían según el con-

texto, el tono e incluso la cultura local, al punto de que pueden tener significados totalmente contrarios a lo que explícitamente se dice, como puede pasar con la ironía o el sarcasmo, por solo citar un ejemplo. Las palabras dicen al pie de la letra algo, que un guiño o una media sonrisa de pronto obliga a interpretar completamente en la dirección contraria. ¿Cómo puede una máquina lidiar con eso?

Las reglas gramaticales pueden ser otro obstáculo de peso. No existen reglas gramaticales fijas y universales para todos los lenguajes. Incluso al interno de cada uno de ellos hay excepciones, en dependencia de la situación comunicativa particular del que habla. Por si fuera poco, el idioma evoluciona constantemente. Lo que unas décadas atrás era una expresión incorrecta hoy puede naturalizarse y con el paso del tiempo hasta encontrar su lugar dentro del canon de la alta cultura.

Por su parte las computadoras están hechas para funcionar bajo patrones fijos y rígidos. No se supone que el código de un programa informático se cambie a sí mismo en dependencia del lugar o el momento en que se ejecute. En contraste, el lenguaje humano es demasiado flexible, rico y creativo como para acomodarse fácilmente dentro de la lógica de ceros y unos que gobiernan a las máquinas. No obstante, a pesar de estas notables dificultades se han logrado resultados, como evidencian un amplio abanico

de aplicaciones, desde los correctores automáticos en los procesadores de texto hasta la comunicación coherente, pero a veces un poco predecible de ChatGPT y Claude.

Una herramienta esencial para sostener los avances logrados en el procesamiento de lenguaje natural, y de manera más amplia en la inteligencia artificial como un todo, lo constituyen las redes neuronales artificiales.

Las *redes neuronales artificiales* constituyen un paradigma dentro del campo de la inteligencia artificial inspirado en el funcionamiento del cerebro humano. Consiste en modelar un conjunto interconectado de unidades o nodos, que trabajan juntos para procesar y analizar información. Cada neurona artificial realiza cálculos simples, como las funciones matemáticas que aprendimos en la enseñanza secundaria, pero cuando se combinan en una red, entre todas pueden realizar tareas complejas como reconocer patrones, clasificar datos o tomar decisiones. Estas neuronas artificiales, al igual que las naturales y las funciones matemáticas, reciben datos. Luego producen datos modificados de acuerdo a una lógica interna que puede cambiar con el tiempo.

Las redes neuronales son capaces de aprender y adaptarse a partir de ejemplos y datos de entrenamiento. Durante el entrenamiento se le presentan los datos

de entrada y los datos de salida que se desea de ellas, y todo su trabajo consiste en adaptar su lógica interna en pasos sucesivos hasta alcanzar la salida deseada.

El concepto de redes neuronales se presentó por primera vez en la década de 1940 y durante 1950 se construyeron los primeros prototipos. Curiosamente estos fueron mayormente equipos electrónicos especializados, los cuales modelaban a las neuronas artificiales de manera física, en forma de bulbos y circuitos, en lugar de código informático como es la práctica usual hoy en día.

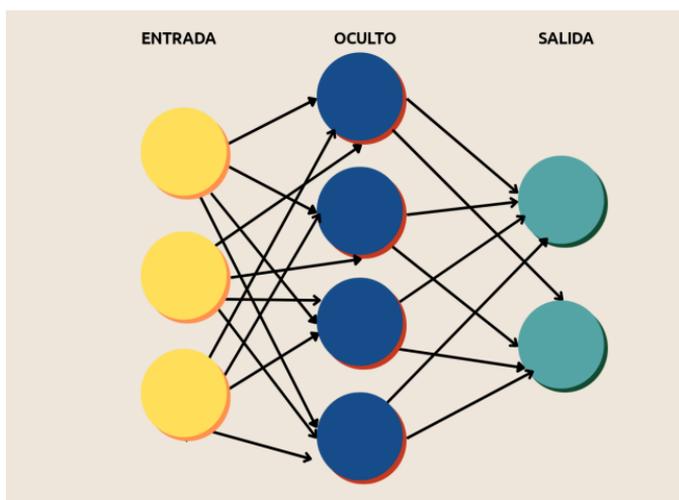


Figura 1. Representación de una red neuronal artificial.
Infografía de elaboración propia.

Uno de los esfuerzos más influyentes fue liderado por el investigador norteamericano Frank Rosenblatt, quien trabajó en la construcción de las máquinas denominadas Perceptron (1957-1962) con hasta cuatro capas de redes neuronales. En el Stanford Research Institute, un grupo liderado por Charles A. Rosen y Alfred E. (Ted) Brain desarrolló máquinas de redes neuronales llamadas MINOS I (1960) y MINOS II (1963). Estas máquinas, que en su gran mayoría fueron financiadas por el ejército de los Estados Unidos, podían realizar tareas como la clasificación de símbolos y el reconocimiento de caracteres escritos a mano.

Sin embargo, estos logros no fueron suficientes y el avance en el campo se desaceleró significativamente desde finales de la década de 1970 hasta mediados de 1980, en lo que fue conocido como “El invierno de la inteligencia artificial”. Hubo varias razones para esto, entre las que destacaron la combinación de críticas teóricas al enfoque de las redes neuronales y la insatisfacción en la comunidad académica tras un período de elevadas expectativas, lo que conllevó a la ausencia de financiamiento para iniciar nuevos proyectos.

Después de esta primera crisis, la inteligencia artificial recuperó impulso. Las redes neuronales, junto a otros campos dentro de la disciplina, experimentaron un re-

surgimiento debido al aumento del poder computacional, una mayor abundancia de datos digitalizados y a la creación de algoritmos de aprendizaje mejorados, como el denominado Retropropagación, el cual hacía eficiente la construcción de redes neuronales artificiales de múltiples capas. El desarrollo de aplicaciones de reconocimiento de distinto tipo, ya sea de voz, escritura o facial demostraron su utilidad de manera práctica; lo que a su vez atrajo de nuevo el tan necesitado financiamiento y estimuló el relanzamiento del campo de investigación.

Hemos presentado, sin profundizar demasiado, los términos de procesamiento de lenguaje natural y redes neuronales artificiales. Quizás valga la pena, para tener una visión más completa del paisaje en el que nos movemos, ubicar esos términos entre otros que también sueñan mucho por estos días. Nos referimos a aprendizaje de máquinas (*Machine Learning*), aprendizaje profundo (*Deep Learning*), grandes datos (*Big Data*) y análisis de datos (*Data Analysis*).

El *aprendizaje de máquinas* es un subconjunto de la inteligencia artificial dedicado a crear algoritmos y modelos. Es un campo multidisciplinar, que se sirve de múltiples conocimientos provenientes de las matemáticas, estadística, ciencias cognitivas, neurociencias, entre otros. A su vez puede ser subdividido en dos tipos prin-

principales, el aprendizaje supervisado y el aprendizaje no supervisado. Esta subdivisión es importante porque nos recuerda que los resultados de la IA dependen en última instancia de un ser humano.

En el aprendizaje supervisado los sistemas de IA se entrenan en conjuntos de datos etiquetados previamente. Es decir, el sistema aprende a asociar un contenido con una determinada etiqueta. Esto les permite capturar patrones de los datos de entrada y asociarlos con etiquetas específicas, las cuales corresponderían a las salidas que se desean. El aprendizaje supervisado se usa especialmente en la construcción de inteligencias artificiales dedicadas a la clasificación.

El aprendizaje no supervisado, por otro lado, implica entrenar sistemas de inteligencia artificial en datos sin etiquetar, donde no existe un resultado previamente definido, sino que este depende del algoritmo utilizado. En este caso, el sistema debe identificar patrones dentro de los datos y reunirlos cuando sean similares, lo que lo hace adecuado para tareas como agrupación de datos y detección de anomalías.

Por ejemplo, si pensamos en reconocimiento de imágenes, el aprendizaje supervisado podría involucrar una IA entrenada en decidir si en alguna imagen se muestra un gato o un perro, mientras que el aprendizaje no su-

pervisado implicaría diferenciar distintas razas de perros o gatos, basándose en patrones visuales como manchas, tamaños, etc.

El *aprendizaje profundo* es un método o técnica particular dentro del campo más amplio del aprendizaje de máquinas, y que ha revolucionado en los últimos años la inteligencia artificial como un todo. Es la clave detrás de las capacidades de procesamiento de gigantescas cantidades de datos de manera eficiente y precisa, por parte de gobiernos y empresas como Google o Amazon. En su base, el aprendizaje profundo se apoya en las redes neuronales artificiales, pero con muchísimas capas de neuronas, hasta cientos o miles de ellas, que están coordinadas entre sí por medio de complejos algoritmos y avanzados diseños de arquitecturas, que le permite alcanzar altos niveles de abstracción en los patrones aprendidos.

Antes mencionamos la frase “gigantescas cantidades de datos”, y se repite varias veces a lo largo del libro. La inteligencia artificial contemporánea, y la generativa especialmente, está intrínsecamente ligada a la enorme disponibilidad de información digitalizada que disfrutamos hoy, gracias al avance de la transformación digital

en nuestras sociedades y al rápido flujo de estos datos a través de Internet.

El concepto de *grandes datos* se refiere a la inmensa colección de información en soporte digital, que puede estar estructurada o no estructurada, provenientes de múltiples fuentes, con variados formatos y con una complejidad creciente en su gestión. En sí mismo es un campo de investigación dinámico, de avanzada, que junto al aprendizaje de máquinas ha impulsado el desarrollo de sofisticados sistemas de IA, no solo generativa, sino para la clasificación, agrupamiento, o modelado de datos originados en acuciantes problemas, desde el calentamiento global hasta la gestión de los procesos industriales.

Lógicamente, gestionar datos incluye necesariamente examinarlos, filtrarlos, etiquetarlos, transformarlos y simplificar su representación, las cuales son las tareas del *análisis de datos* y sus protagonistas, los científicos de datos. Su trabajo es crucial, porque solo a partir de un conjunto de datos de excelente calidad, es decir, limpio, bien etiquetado, balanceado, entre otros criterios, es que se garantizan entrenamientos óptimos de modelos y por tanto inteligencias artificiales realmente útiles. Si se introduce información de mala calidad en el entrenamiento de la IA, necesariamente los resultados obtenidos no serán los deseados. Y vale aclarar, la construcción

de conjuntos de datos de la mejor calidad posible es un proceso extremadamente caro en tiempo, recursos y conocimientos implicados.

Primero, implica revisarlos para comprender su estructura, contenido y posible valor. Esto incluye identificar vacíos o inconsistencias y determinar rasgos útiles para su uso en el entrenamiento. A continuación, se someten a una limpieza para eliminar errores, redundancias o información irrelevante que podrían distorsionar los resultados y conducir a conclusiones incorrectas. Este proceso puede incluir actividades como eliminar registros duplicados, corregir errores de entrada o crear valores faltantes con estimaciones razonables, para rellenar vacíos que de otra manera afectarían la calidad de los resultados. El siguiente paso, la transformación, involucra convertir los datos brutos en un formato más adecuado para los propósitos analíticos y de modelado. Dependiendo de las especificidades del proyecto, esto puede requerir escalar variables, agregar series temporales o aplicar cualquier otra técnica de preprocesamiento para garantizar que estén optimizados para la construcción de los modelos. Finalmente, el modelado implica ajustar algoritmos estadísticos, de aprendizaje automático o de aprendizaje profundo para descubrir patrones ocultos, hacer predicciones o respaldar proce-

tos de toma de decisiones. La etapa final la constituye la validación y evaluación del modelo para determinar su precisión, robustez y generabilidad.

Con la información proporcionada hasta aquí, estamos un poco mejor equipados para avanzar hacia la inteligencia artificial generativa, entender lo que puede hacer, y lo que resulta muchas veces más importante, entender lo que no puede hacer, junto a lo que pudiera hacer mal.

INTELIGENCIA ARTIFICIAL GENERATIVA

DEFINICIÓN DE LA IA GENERATIVA

La inteligencia artificial (IA) generativa se refiere a un tipo de tecnología capaz de generar de forma autónoma y coherente nuevos contenidos simbólicos, en formatos habituales tales como el texto, las imágenes y el sonido, siendo prácticamente indistinguibles de los producidos por el pensamiento humano.

Los modelos de IA generativa se construyen aplicando técnicas de aprendizaje automático para codificar representaciones de datos mediante redes neuronales. Esto les permite, a partir de las grandes cantidades de información que se usaron para su entrenamiento, aprender patrones y estructuras que luego utilizan a modo de recetas o guías para generar nuevos contenidos.

Algunos de los primeros intentos de IA generativa se pueden remontar a la década de 1960, cuando se ensayaron programas basados en reglas para la composición de texto. Sin embargo, todavía tenían muchas limitaciones, como vocabularios muy pequeños y dificultades para procesar el contexto adecuadamente. No fue hasta 2014 que se produjeron avances significativos con la introducción de técnicas de aprendizaje profundo, que

permitieron a los sistemas de IA por primera vez sintetizar fotos, videos y audio de manera realista.

Los avances en los grandes modelos de lenguaje, como la familia de modelos GPT (de donde sale ChatGPT) y otras, han mejorado enormemente las capacidades de la IA generativa. Los sistemas contemporáneos populares como ChatGPT, Dall-E y Stable Diffusion pueden generar automáticamente texto coherente, imágenes fotorealistas y contenido multimedia complejo a partir de la petición de un usuario hecha en lenguaje natural y no mediante un código de programación.

¿CÓMO FUNCIONA LA IA GENERATIVA?

La habilidad clave que permite a estos modelos mostrar comportamiento generativo es que pueden aprender los patrones y las distribuciones estadísticas presentes en sus datos de entrenamiento y, luego, utilizar este conocimiento aprendido para sintetizar nuevos datos que son ejemplos creíbles de la misma clase de objetos a partir de los cuales se entrenó. Por ejemplo, convenientemente entrenados pueden generar textos en el mismo estilo y con el mismo vocabulario utilizado en las obras de Shakespeare.

Una de las formas más extendidas de inteligencia generativa son los grandes modelos de lenguaje, que son redes neuronales entrenadas con enormes cantidades

de texto para capturar las propiedades estadísticas del lenguaje natural. Estos modelos de lenguaje operan mediante el mapeado de una secuencia de entrada (como una oración o un texto más largo) a una representación interna.

Luego a partir de una palabra dada o un texto, predicen la siguiente palabra más probable basándose en esta representación. Dicho de otra manera, se le hace corresponder a cada palabra, en un fragmento de texto, una probabilidad de que siga o no a otra determinada palabra, y este dato es el que se guarda en el modelo. A modo de ilustración solamente, de todas las palabras posibles, un modelo de IA generativa asignará una probabilidad muy alta a la palabra “leche” y una probabilidad más baja a la palabra “bisturí” a la hora de completar la frase “Desayunó pan y café con...”. Al encadenar estas predicciones de palabras una detrás de otra, el modelo puede generar automáticamente texto que aparenta ser fluido y coherente.

El secreto es que, con la exposición a billones de palabras durante el entrenamiento, el modelo aprende implícitamente intrincados patrones del lenguaje en varios niveles de abstracción; desde cuáles caracteres y palabras tienen más probabilidades de seguir unos a otros, hasta conceptos y narrativas complejas, pero igualmen-

te basadas en probabilidades, que va componiendo hasta crear un texto que sigue propiedades estadísticas similares, pero en su composición es enteramente nuevo.

A un nivel más detallado, durante el entrenamiento, los parámetros del modelo se ajustan continuamente para hacer coincidir las probabilidades asignadas internamente a las que se quieren emular del mundo exterior. Una vez entrenado, tenderá por tanto a producir las palabras más probables basadas en su predicción aprendida, resultando en una generación que captura las correlaciones superficiales del ejemplo original.

Los modelos generativos más avanzados emplean técnicas adicionales para mejorar la calidad de sus resultados. Por ejemplo, aprenden dependencias de largo alcance, es decir, no de palabras o de fragmentos de textos inmediatamente contiguos, sino a mucha más distancia uno de otros, y también se apoyan en textos almacenados en base de datos especializadas para trabajar con la inteligencia artificial (denominadas base de datos vectoriales), que usan para expandir y mejorar los textos sintetizados.

Para crear un asistente de inteligencia artificial como ChatGPT o Claude, el proceso involucra algunas etapas adicionales de entrenamiento y ajuste fino.

La primera etapa es la que ya hablamos, de extraer patrones estadísticos, conocida como pre-entrenamien-

to, que es computacionalmente costosa y puede llevar varios meses o incluso años para completarse usando equipamiento especializado.

Una vez creado el modelo base en la etapa de pre entrenamiento, se procede a la segunda etapa, denominada de ajuste fino, para obtener un asistente que se comporte de manera atenta y cooperativa. Para ello se construye un conjunto de instrucciones cuidadosamente etiquetadas que describen cómo la IA debería comportarse al responder preguntas y generar respuestas. En esta etapa se cuida especialmente que las preguntas y respuestas sean de alta calidad y estén extremadamente bien seleccionadas. Típicamente se necesita alrededor de 100,000 muestras de preguntas con sus respuestas para esta etapa de entrenamiento. Un aspecto positivo es que esta última es mucho menos costosa computacionalmente y se puede actualizar diariamente o semanalmente para mejorar continuamente su rendimiento.

Una tercera etapa opcional de ajuste fino es la conocida como aprendizaje reforzado por retroalimentación humana (RLHF, por sus siglas en inglés). La misma implica incorporar etiquetas de comparación para refinar aún más las respuestas de la IA. En esas etiquetas los evaluadores valoran las respuestas generadas por modelos anteriores, y refuerzan por tanto las respuestas con

mayor calidad. Este proceso refina las cualidades como asistente de la IA y mejora sus respuestas.

EL FUTURO DE LA IA GENERATIVA

En un [video reciente](#) publicado por Andrej Karpathy, investigador e ingeniero que ha realizado un importante trabajo en el campo de la IA generativa, se dibujan de manera muy clara las posibles líneas de evolución de la IA generativa en el futuro inmediato, a partir de lo que se puede avizorar desde la revisión de la literatura académica que se publica sobre el tema y las reuniones profesionales del sector.

Karpathy enumera tres líneas de evolución principales:

La primera es permitir a los modelos de IA generativa pensar problemas de manera más profunda y deliberativa, similar al modo en que los humanos razonan cuando se enfrentan a un problema, por ejemplo, matemático. Los modelos generativos actuales como GPT-3 dependen únicamente de emparejamientos de patrones derivados de enormes conjuntos de datos de entrenamiento, limitando su capacidad de razonar problemas complejos por medio de la exploración de perspectivas alternativas a lo largo del tiempo.

Los investigadores están explorando enfoques para cultivar la deliberación y el razonamiento paso a paso

mediante el desarrollo de arquitecturas computacionales que permitan a los modelos analizar recursivamente problemas, considerar de manera intencionada variantes, reflexionar sobre sus posibles resultados y poner a prueba su comprensión, de manera similar a como los humanos aprovechan el razonamiento consciente. Para ello se están explorando técnicas como la recursión, varios algoritmos para la búsqueda sistemática de la mejor respuesta entre varias disponibles y la indagación iterativa de mecanismos de retroalimentación implícita.

Una segunda línea de trabajo es otorgar a los modelos la posibilidad de mejorar de manera continua la generación de contenidos, devolviéndoles una evaluación de sus respuestas, de tal manera que puedan auto corregirse de manera autónoma. Lograr que se perfeccionen a sí mismos presenta retos importantes debido a la naturaleza abierta del lenguaje y la falta de medidas objetivas y definidas de lo que es una buena respuesta. Sin embargo, según Karphaty algunos trabajos recientes han comenzado a diseñar enfoques de auto-supervisión que aprovechan señales de retroalimentación en patrones conversacionales, indicadores de coherencia lingüística y fáctica, y la medición de la alineación con objetivos conversacionales

para ayudar a los modelos a mejorar de manera independiente sus capacidades.

Una tercera línea de trabajo es la posibilidad de personalizar modelos de IA generativa, permitiéndoles ser expertos en dominios del conocimiento muy específicos, en lugar de buscar un solo modelo que pueda servir en todos los campos del saber.

En términos más generales es posible avizorar que se seguirán desarrollando nuevas arquitecturas para los modelos de IA generativa, y progresivamente serán capaces de capturar relaciones más complejas y sutiles entre los datos de entrenamiento.

Así mismo con toda seguridad seguiremos viendo el incremento en los tamaños de los modelos y por tanto de las capacidades computacionales necesarias para ejecutarlos. Existe consenso entre algunos investigadores que una de las mejores maneras de mejorar la precisión en la respuesta de los grandes modelos de lenguaje es aumentando el poder computacional.

La multimodalidad o la capacidad de interactuar con la IA generativa por medio de texto, audio e imágenes, también potenciará la utilidad de los modelos como asistentes en el trabajo humano, y al mismo tiempo permitirá el desarrollo de aplicaciones mucho más sofisticadas.

Además es posible asegurar con certeza que todos estos desarrollos ocurrirán bajo una intensa discusión sobre

regulaciones, códigos éticos, y establecimiento de salvaguardas en la creciente introducción de la IA en la vida diaria. En esta discusión participan por igual académicos, funcionarios públicos y diversos sectores profesionales, así como algunos organismos internacionales, buscando entornos que favorezcan un mejor aprovechamiento de las potencialidades de la IA generativa, al mismo tiempo que aminore o elimine por completo los peligros que conllevan.

DESAFÍOS ÉTICOS Y DE SEGURIDAD

Desde que se hicieron públicos los primeros logros de la IA generativa también fueron evidentes algunos riesgos y desafíos relacionados con su uso extensivo en la sociedad, los cuales van mucho más allá de la tendencia a alucinar, o a inventar información dándola por cierta (un comportamiento que exhiben todos los modelos de una manera u otra). Las limitaciones a las que haremos referencia aquí, tienen impactos negativos en la sociedad mucho más marcados.

Los modelos de IA generativa han sido entrenados en enormes volúmenes de datos provenientes en su mayor parte de Internet. Dado los sesgos y manifestaciones discriminatorias presentes en gran parte del contenido disponible en la red, es de esperar que los modelos los

reproduzcan si no son adecuadamente manejados, con el riesgo de dañar individuos o comunidades específicas.

A esta dificultad se le suma la de no saber en primera instancia a quién asignarle la responsabilidad cuando un modelo de IA generativa tiene un resultado no deseado. ¿Es responsabilidad de la empresa que la entrenó? ¿Del usuario que introdujo la orden? ¿O acaso será responsabilidad del autor original del contenido negativo que le sirvió de entrenamiento?

Estas problemáticas están relacionadas con la falta de transparencia sobre cómo se genera el contenido al interior del modelo, la imprecisión que puede tener la información producida y el sesgo que muchas veces muestra. Actualmente se desarrollan métodos y tecnologías que deben ayudar a que los contenidos generados por estos modelos sean verificables, no engañosos y estén mejor alineados con valores y normas socialmente responsables.

El hecho de que puedan generar grandes cantidades de contenido de calidad en muy poco tiempo también representa retos para las nociones tradicionales de autoría y originalidad, con impacto en sectores como el periodismo, la escritura creativa y la publicidad.

También representan riesgos para el comportamiento político en las sociedades contemporáneas. No está

claro quién es el dueño en última instancia de los datos con que se entrenan estos modelos, ni quién se haría responsable por su uso en campañas de manipulación de la opinión pública a través de la desinformación o la generación en grandes volúmenes de contenido sesgado y malintencionado. Lo cierto es que la IA generativa puede ser usada como un instrumento de poder político, al facilitar la creación relativamente barata y rápida de mensajes persuasivos altamente personalizados a sectores de público muy específicos, por ejemplo, durante elecciones y consultas populares.

Por último, la IA generativa, como cualquier otro software, tiene algunos riesgos de seguridad. Entre los más importantes se encuentra el *jailbreak*, el ataque de inyección de órdenes, y el entrenamiento con datos envenenados.

Jailbreaking, un término en inglés que viene a significar “escapar de la prisión”, hace referencia a técnicas diseñadas para sobrepasar las limitaciones de funcionamiento de una tecnología puestas por el fabricante a nivel de hardware o de software. En el caso por ejemplo de los grandes modelos de lenguaje, implica introducir órdenes especialmente pensadas para hacer generar al

modelo de IA contenido dañino que de otra manera no produciría, como recetas para fabricar armas caseras o instrucciones de cómo robar.

El ataque de inyección de órdenes se refiere a la posibilidad de introducir de manera inadvertida para el usuario nuevas órdenes que anulen las suyas propias. Por ejemplo, se le puede pedir a una IA que analice una imagen, y en esa imagen puede estar inscrita de manera invisible para el usuario, pero no para la IA, una nueva orden que anule la original y que en su lugar promocioe determinado producto.

Por último, el entrenamiento con datos envenenados es un tipo de ataque donde el modelo se entrena con datos incorrectos o malintencionados, afectando su rendimiento cuando es usado en producción.

Los riesgos de seguridad informática asociados a los grandes modelos de lenguaje han mostrado ser una amenaza real y concreta que debe ser tomada en serio, por todas las partes interesadas, desde las empresas hasta los usuarios. Cada uno de estos actores deben tomar medidas adecuadas para protegerse de ataques malintencionados. Por la parte de las empresas dedicadas al entrenamiento de modelos de IA generativa, estos deben cuidar la limpieza y calidad de los datos que usan, y entrenar adecuadamente sus modelos para que no sean

vulnerables a jailbreaks y ataques de inyección de órdenes. Por la parte del usuario este siempre debe verificar la información suministrada por el modelo, nunca confiar ni suministrar información privada y verificar la identidad de la empresa o creadores detrás de un modelo de IA generativa.

INTELIGENCIA ARTIFICIAL GENERATIVA DE TEXTO

QUÉ ES. CÓMO FUNCIONA. EJEMPLOS

La inteligencia artificial generativa de texto, tal como su nombre lo indica, es un tipo de inteligencia artificial capaz de generar texto de manera automática, en respuesta a instrucciones escritas en lenguaje natural. Se apoya en el uso de grandes modelos de lenguaje (*Large Language Model* o LLM en inglés), y se presentan a través de interfaces conversacionales, comúnmente experimentadas por los usuarios como chatbots y accesibles tanto en versiones web como en aplicaciones móviles.

En otras palabras, un usuario puede ingresar al sitio web de ChatGPT o descargar su aplicación en el móvil. Al crear una cuenta e iniciar sesión, tiene la capacidad de proporcionar instrucciones para que la inteligencia artificial genere contenido en forma de texto. Sin embargo, surge la pregunta: ¿Cómo logran las inteligencias artificiales generativas de texto, a través de esta aparentemente sencilla interacción, ofrecer respuestas coherentes?

La gran mayoría de los modelos de lenguaje, hasta el momento, utilizan una arquitectura de red neuronal conocida como "GPT", la cual se refiere a transformador re-

entrenado generativo (*Generative Pretrained Transformer* en inglés). La misma está diseñada para procesar eficientemente secuencias de datos textuales, capturando incluso relaciones de largo alcance.

Antes de abordar tareas específicas, estos modelos se pre-entrenan con grandes cantidades de texto, aprendiendo a predecir la siguiente palabra en una secuencia dada. Este proceso les permite encadenarlas una tras otra en función de las asociaciones comunes, proporcionando una respuesta coherente a la solicitud del usuario, como se mostró en el ejemplo de "Desayunó pan y café con leche", en el capítulo 3.

Lo que para los usuarios se siente como una conversación fluida con la IA generativa de texto, donde uno pregunta y el otro responde, para los modelos de lenguaje se trata de un proceso continuo en el que completan el siguiente texto faltante de acuerdo al contexto proporcionado, según como fueron entrenados.

El modelo GPT fue desarrollado por OpenAI y se ha convertido en una referencia en el campo de la inteligencia artificial. Además, cada versión de GPT ha representado avances en términos de tamaño y capacidad,

mostrando un rendimiento más eficiente en diversas tareas de procesamiento de lenguaje natural.

Al utilizarlo para generar texto, el modelo recibe una instrucción o entrada inicial. Dentro del entorno que abarca a los grandes modelos de lenguaje, la instrucción (*prompt*) es la orden en forma de texto que se proporciona a un modelo de inteligencia artificial para guiar o solicitar una respuesta específica. Puede ser una frase, pregunta o conjunto de palabras que se utiliza para dirigir la generación de texto del modelo.

Por ejemplo, si se está utilizando un modelo de lenguaje generativo para traducir una oración del inglés al francés, la instrucción podría ser la oración en inglés que se desea traducir. Del mismo modo, si se está solicitando información sobre un tema específico, la instrucción sería la pregunta o la consulta relacionada con ese tema.

La elección y formulación cuidadosa de la instrucción es crucial para obtener resultados precisos y relevantes del modelo, en tanto este guía la generación de texto en función de la información proporcionada. Según la tarea que se esté llevando a cabo las instrucciones pueden variar en longitud y complejidad. De hecho, existe una práctica relativamente nueva, denominada ingeniería de instrucciones o *prompt engineering*, que se dedica a su desarrollo y optimización, en busca de utilizar eficientemente modelos de

lenguaje (ML) en una amplia variedad de aplicaciones y temas de investigación.

ChatGPT, la IA de OpenAI

ChatGPT representa la evolución más reciente y destacada de la serie GPT; pues sorprendió al mundo con su capacidad para traducir idiomas, responder preguntas con precisión, redactar artículos, completar oraciones y generar contenido creativo, como narrativa y poesía.

Desarrollada por OpenAI, esta IA generativa de texto está construida sobre la tercera versión de la serie, GPT-3, un modelo que se entrenó con 175 mil millones de parámetros, incluyendo libros, Wikipedia y diversos contenidos de la web.

La constante iteración y mejora en cada versión de la serie, desde la primera GPT-1 en 2018 hasta la más reciente GPT-4 en marzo de 2023, fue posible gracias a avances continuos en arquitecturas de inteligencia artificial, métodos de entrenamiento y el aprovechamiento de vastos conjuntos de datos.

Aunque la tecnología GPT ha estado disponible para investigadores desde 2018, el lanzamiento de ChatGPT marcó un hito significativo al ofrecerla con acceso gratuito a través de una interfaz fácil de usar, permitiendo

que cualquier persona con conexión a Internet pueda explorar la herramienta de generación de texto.

Sin embargo, desde algunas regiones del planeta ChatGPT no está disponible y entre ellas se encuentra Cuba. Por esta razón siempre que desde la Isla se quiera interactuar con dicha inteligencia artificial resulta necesario activar una VPN.

A pesar de estas restricciones regionales, pocos días después de su lanzamiento el 30 de noviembre de 2022, ChatGPT alcanzó 1 millón de usuarios y en diciembre de 2023 tenía 180.5 millones de usuarios activos al mes. No sorprende entonces que su impacto desencadenara una rápida explosión de nuevas IA generativas de texto, tanto por parte de grandes empresas globales como por emprendimientos emergentes. Estas iniciativas incluyeron el lanzamiento de sistemas similares y la creación de nuevas herramientas innovadoras sobre la base de GPT.

Claude, la IA generativa de Anthropic

Claude, el modelo de inteligencia artificial desarrollado por Anthropic, destaca por su notable eficiencia en el procesamiento de información, siendo aproximadamente tres veces más rápido que el GPT-4 de OpenAI. Dicho

rendimiento lo posiciona como un referente en tareas especializadas de procesamiento y generación de texto.

Este asistente de inteligencia artificial generativa de texto, al igual que otras plataformas, como ChatGPT y [Bard](#), exhibe la capacidad para redactar textos, programar y proporcionar respuestas precisas a preguntas, adaptándose al estilo y tono especificados por el usuario. Actualmente, Claude está integrado en soluciones de empresas colaboradoras, tales como Notion, Quora y DuckDuckGo. Un ejemplo de su implementación es la creación de [Poe](#), un chatbot inteligente en Quora, el cual aprovecha la potencia de Claude para ofrecer respuestas informadas y amigables a las preguntas de los usuarios.

Según Anthropic, este modelo de IA puede procesar y analizar hasta 100,000 tokens de texto, es decir, tres veces más que GPT-4, convirtiéndolo en el modelo de IA más poderoso conocido hasta ahora.

La cantidad de tokens que un modelo de lenguaje grande puede procesar en un momento dado se conoce como ventana de contexto o memoria a corto plazo. Mientras que un usuario promedio puede leer 100.000 tokens en cinco horas, requiriendo tiempo adicional para recordar y analizar la información, GPT-3.5 de

OpenAI tiene una ventana de contexto de 4096 tokens, aproximadamente 3000 palabras cuando se utiliza con ChatGPT. Sin embargo, utilizando la API de GPT-4, esta capacidad se puede aumentar a 32.768 tokens, una cifra aún inferior a la capacidad de Claude AI, que puede procesar 100.000 tokens, equivalentes a aproximadamente 75.000 palabras.

Esta característica convierte a Claude en una herramienta valiosa para las empresas que necesitan procesar grandes volúmenes de datos, especialmente en áreas como informes financieros, documentos legales, repositorios de código y consultas a bases de datos. Además, gracias a su amplia ventana de contexto puede resultar útil cuando se necesite interrogar un PDF relativamente grande, lo que puede ser aprovechado por el usuario común.

Claude aún no tiene la popularidad de ChatGPT o Bard. Una de las razones es que no está disponible en varios países, entre ellos Cuba. De hecho, la experiencia de chat de Claude 2 solo está disponible de forma general en E.E.U.U y Reino Unido.

Claude, al igual que la gran mayoría de IA generativas de texto, es un producto experimental. A pesar de que tiene algunas diferencias notables frente a la competencia, presenta los mismos problemas, y en la generación de con-

tenido además de respuestas exactas, puede expresar medias verdades y texto completamente falso.

En este sentido, el propio sitio Antropic.com declara su intención de trabajar para que el modelo sea más ético y responsable.

Bing, la IA generativa de Microsoft

Bing, el buscador de Microsoft, ha dado un paso significativo en su evolución al incorporar una serie de mejoras impulsadas por la inteligencia artificial (IA). Ante el gigante Google, Bing se enfrenta continuamente al desafío de mejorar sus resultados de búsqueda y ofrecer nuevas funciones a sus usuarios.

De esta manera y en respuesta al auge de las interacciones con inteligencias artificiales generativas de texto, Microsoft incorporó un motor de búsqueda conversacional llamado **Bing Chat**, que se apoya en GPT-4. Aunque no es exactamente ChatGPT, podría ser un hermano cercano. Este nuevo servicio permite a los usuarios interactuar con el buscador de manera más natural, planteando preguntas complejas y recibiendo respuestas más detalladas que las proporcionadas por los buscadores convencionales.

En otras palabras, Microsoft está intentando realizar las búsquedas tradicionales con un formato conversa-

cional, de forma que en un mismo buscador se integren ambos servicios, posicionándose competitivamente frente a la propuesta de Google con su IA generativa Bard.

Bing Chat está disponible para Cuba, tan solo es necesario acceder al buscador de Microsoft. Además, en comparación con otras alternativas es una de las herramientas más útiles para contrastar la fuente de información de donde la IA tomó el contenido para generar texto.

LLaMA, la IA de Meta

LlaMA, cuya abreviatura en inglés corresponde a *Large Language Model Meta AI* y que traducido al español sería Gran modelo de lenguaje Meta AI, es un gran modelo de lenguaje lanzado por la división de IA de Meta en febrero de 2023.

Su más reciente versión, **LlaMA 2**, apuesta por una inteligencia artificial de código abierto, que se ofrece de forma gratuita tanto para fines comerciales como investigativos. Estas características han fomentado la aparición de iniciativas de código abierto que pueden desarrollarse, implementarse y mantenerse sin requerir enormes recursos financieros.

LlaMA 2 está diseñada para ser accesible a individuos, creadores, investigadores y empresas de todos los tamaños, brindándoles la oportunidad de experimentar, innovar y escalar sus ideas de manera responsable. El mode-

lo se presenta en tres tamaños distintos, pre-entrenado con 7 mil, 13 mil y hasta 70 mil millones de parámetros. Además, Meta afirma que se entrenó con un 40% más de datos y tiene el doble de extensión de contexto que LLaMA 1.

La empresa también afirma que lanzar su nueva IA como un proyecto de código abierto forma parte de una estrategia de transparencia y accesibilidad. Pero si bien esto en sí mismo es un fin loable, no se puede perder de vista su utilidad adicional como estrategia de mercado para ganar relevancia en la comunidad de desarrolladores e investigadores, especialmente aquellos que carecen del respaldo económico para acceder a tecnologías propietarias de las grandes empresas.

Los desarrolladores de LLaMA se enfocaron en escalar el rendimiento del modelo mediante el aumento en el volumen de datos de entrenamiento, en lugar de la cantidad de parámetros. LLaMA utilizó para ello 1.4 billones de tokens extraídos de fuentes de datos disponibles públicamente.

Sobre estos dos términos el propio ChatGPT explica que los *tokens* son unidades individuales de texto. Estos en el contexto del procesamiento del lenguaje natural representan palabras o partes de palabras. Constituyen

la unidad fundamental con la que trabajan los modelos de lenguaje durante el entrenamiento y la generación.

Por otro lado, los *parámetros* constituyen los elementos internos modificables de un modelo de inteligencia artificial generativa. Más concretamente, se refieren a los pesos y conexiones adaptables presentes en las capas de la red neuronal. A lo largo del proceso de entrenamiento, estos se afinan para reducir la discrepancia entre las predicciones del modelo y las respuestas deseadas, ajustando de esta manera el conocimiento y los patrones asimilados. En el contexto de un modelo de lenguaje, los parámetros determinan la asignación de probabilidades a diversas secuencias de tokens.

Con LLaMA, Meta busca posicionarse en el mercado de chatbots ofreciendo una alternativa de código abierto versátil y eficiente, que si bien aún requiere desarrollo para perfeccionar sus capacidades generativas, representa una opción prometedora para democratizar el acceso a la inteligencia artificial.

PARA QUÉ SIRVE

Las IA generativas se están convirtiendo, cada vez más, en herramientas fundamentales en el entorno laboral. Su capacidad para comprender y generar texto de manera coherente las está incorporando progresivamente en diversas rutinas profesionales, destacándose, sobre

todo, en sectores creativos como el periodismo, la comunicación y la publicidad.

Desde nuestra perspectiva es crucial adoptarlas no como simples sustitutos de puestos laborales, sino como tecnologías capaces de potenciar la eficiencia: la automatización de tareas repetitivas y tediosas permite a los profesionales centrarse en aquellas más estratégicas y creativas.

Con una amplia variedad de aplicaciones, las IA generativas de texto se han vuelto herramientas polifacéticas. Desde diferentes roles profesionales, en la actualidad se utilizan para agilizar la redacción de artículos, blogs y descripciones de productos; crear mensajes promocionales y contenido para redes sociales; desarrollar diálogos, respuestas y narrativas para personajes de videojuegos, simulaciones y aplicaciones de realidad virtual; redactar comunicados y manuales; incluso para ser asistentes en la producción de código informático, proporcionando ejemplos y soluciones para problemas específicos. Estas son solo algunas funciones que podríamos asignarle a dichas tecnologías, pero son muchas más las que están apareciendo en el actual contexto de constante innovación de procesos.

Diseñar entonces rutas de uso con las mejores pautas de interacción del usuario dependerá de diversos factores, dentro de los cuales siempre será fundamental

tener claridad sobre las necesidades y los objetivos a cumplir. La persona detrás de la máquina debe partir del contexto de producción de contenido donde introduce la IA. Redactar el diálogo de un videojuego implica una perspectiva distinta a la redacción de un boletín personalizado.

Además, crear buenas instrucciones para ChatGPT u otros modelos de IA requiere una comprensión profunda de las capacidades y limitaciones del modelo, así como del público y el contexto previsto de la conversación.

En un intento por ilustrar algunos pasos que podrían resultar eficaces a la hora de construir instrucciones para la interacción con la IA, resumimos los siguientes consejos:

- Ofrecer detalles: Se debe incluir suficiente contexto para obtener respuestas relevantes y precisas del modelo.
- Incluir restricciones: Definir límites específicos de longitud de texto mejora la focalización de las respuestas.
- Considerar el tono y estilo de la respuesta: Se puede solicitar estilos específicos, proporcionar ejemplos con el tono deseado o incluso pedir al modelo que genere una lista de posibles estilos para mejorar la precisión de las respuestas.
- Incorpora conocimientos o habilidades específicas: Si el modelo ha sido entrenado en un campo o habilidad

particular, es posible aprovecharlo desde las instrucciones.

SESGOS Y CONSIDERACIONES ÉTICAS

Aunque la IA generativa de texto posee la capacidad de producir nuevo contenido, es importante reconocer que no siempre es recomendable aceptarlo sin más. Una limitación que se acentúa al considerar el fenómeno de las alucinaciones.

En el ámbito de las inteligencias artificiales, el término *alucinaciones* hace referencia a la propensión de estas IA a concebir información que carece de fundamento en hechos o fuentes verídicas. A medida que generan texto, su habilidad intrínseca para crear contenido puede inducir las a inventar detalles, eventos o datos sin respaldo real, lo que resulta en la producción de resultados imprecisos e incluso en la presentación de información completamente falsa.

En este sentido cabe señalar que los propios proveedores de estas inteligencias artificiales generativas advierten contra la aparente coherencia en las respuestas, declarando desde la interfaz de sus modelos que pueden mostrar información incorrecta o dañina (Claude), respuestas imprecisas u ofensivas (Bard) y sorpresas o errores (Bing Chat).

La capacidad de las herramientas de IA generativa de texto para generar respuestas que suenan razonables,

favorece que los errores pasen desapercibidos, a menos que el usuario tenga un conocimiento sólido del tema en cuestión o verifique la información generada.

Más allá de la fascinación que está causando la introducción de las inteligencias artificiales generativas de texto desde la primera aparición de ChatGPT, resulta fundamental que todo usuario se relacione con ellas desde un uso estratégico. Pues entre las desventajas más reconocidas de estos modelos se encuentran los sesgos que responden al propio contenido disponible en Internet, y que reproduce estereotipos de género, raza o creencias religiosas.

Por otro lado, si bien estas tecnologías se utilizan dentro de procesos de producción y rutinas creativas, las IA generativas pueden devolver contenido genérico, monótono y predecible que para los lectores puede no ser atractivo como el escrito por los seres humanos.

Estos detalles ganan relevancia en espacios relacionados al periodismo, la comunicación y la educación, los cuales deben basarse en fuentes reales y contenido de valor. Por tanto, su incorporación efectiva a procesos de trabajo pasa por una reflexión crítica sobre la herramienta, acomodada al contexto específico de los objetivos deseados.

INTELIGENCIA ARTIFICIAL GENERATIVA DE IMÁGENES

QUÉ ES. CÓMO FUNCIONA. EJEMPLOS

Desde resucitar a Lola Flores para anunciar la cerveza Cruzcampo hasta ver al Papa Francisco desfilando en las grandes pasarelas de moda, si pensábamos que lo habíamos visto todo, la inteligencia artificial generativa nos ha demostrado, en un período muy breve de tiempo, que entramos en la era de la creatividad e imaginación visual sin límites con las ventajas, contradicciones y riesgos que ello implica.

La generación de imágenes a partir de texto con inteligencia artificial (IA) se realiza mediante modelos de aprendizaje profundo, entrenadas en grandes volúmenes de imágenes previamente etiquetadas con descripciones textuales de lo que contienen.

Por ejemplo, al ingresar la frase “un perro negro corriendo por la playa” la IA generativa de imágenes es capaz de relacionar la frase en lenguaje natural con patrones visuales que ha aprendido sobre la forma de un perro, la representación visual de la arena, el movimiento de las olas, etcétera. Asimismo, ante la descripción textual “un astronauta comiendo un sándwich en la luna”, logra sintetizar cómo se vería esa escena, creando los detalles

sobre la textura del traje espacial, las estrellas en el fondo, las sombras...

En la base de este proceso de creación visual están modelos de IA generativa especializadas en imágenes. Entre los modelos más avanzados podemos encontrar Stable Diffusion, DALL-E 3 y Midjourney.



Figuras 2 y 3. Imágenes creadas con la herramienta Multimedia Mágico de Canva, que usa el servicio DALL-E, de OpenAI (<https://www.canva.com/magic-home>)

Podemos, entonces, adelantar algunos de los usos de la IA generativa de imágenes a partir de sus potencialidades. La más obvia es la creación de imágenes a partir de una descripción textual o instrucción, donde

se especifican las características de la imagen resultado, e incluso, los modelos más avanzados permiten regular su resolución desde muy baja (32x32 píxeles) hasta alta (1024 x 1024). A ello se añade el fotorrealismo que le pueden imprimir estas herramientas a las imágenes creadas y la capacidad de proponer múltiples resultados, a partir de la instrucción o de sus variaciones; lo que permite explorar rápidamente un sinnúmero de ideas en torno a un mismo concepto.

PARA QUÉ SIRVE

La inteligencia artificial generativa está impulsando una revolución en diversas disciplinas del mundo laboral, académico y hasta en el ámbito artístico por su potencialidad de automatizar rutinas productivas reduciendo los tiempos de trabajo y dotando el resultado, en muchos casos, de un acabado satisfactorio desde la primera vez que se ejecuta su instrucción.

Aunque ya existían herramientas digitales eficaces para los profesionales del sector creativo y las ingenierías, los modelos generativos de IA han agilizado aún más los procesos de innovación para la generación de bocetos, logos, diseños de muebles, esquemas de dispositivos electrónicos, moda y más, a partir de descripciones de texto. Esto los ayuda a obtener referencias visuales, en

cuestión de minutos, que les facilita moldear la obra en función de sus objetivos. De igual forma, pueden obtener nuevos conceptos de productos sin requerir crear o construir manualmente prototipos complejos, lo que tiene un impacto en el abaratamiento de los costos finales de producción.

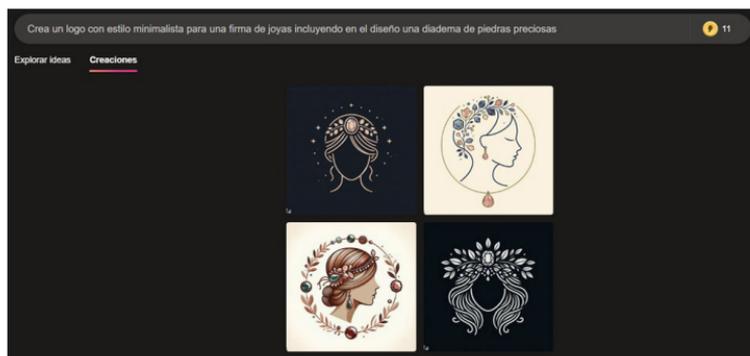


Figura 4. Imágenes creadas con Microsoft Bing Image Generator (<https://www.bing.com/images/create>)

La industria del entretenimiento y los medios de comunicación también han encontrado en estos modelos herramientas muy útiles. Sectores como el cine, la televisión y los videojuegos cuentan con una larga experiencia en la experimentación y desarrollo de la animación y los efectos especiales. Sin embargo, con la IA generativa se pueden generar automáticamente escenarios, efectos visuales y personajes abriendo un mundo de posibilidades

para ahorrar tiempo, reducir costos y explorar caracterizaciones de personajes de formas totalmente innovadoras.

En este sentido, a través del empleo del *deepfake* hemos visto rejuvenecer e incluso “volver a la vida” a personalidades como la actriz norteamericana Carrie Fisher quien hizo un cameo en la película de Star Wars Rogue One (2016) con su icónico personaje de la Princesa Leia, luciendo idéntica que en 1977; y la también desaparecida bailaora española Lola Flores que en 2021 protagonizó un comercial de la cerveza Cruzcampo, años después de su muerte. Esta tecnología (cuyo nombre proviene de la unión de *deep learning*, y la palabra *fake*, falso) permite superponer el rostro de una persona al de otra y falsificar sus gestos y discurso, obteniendo un resultado visual creíble.

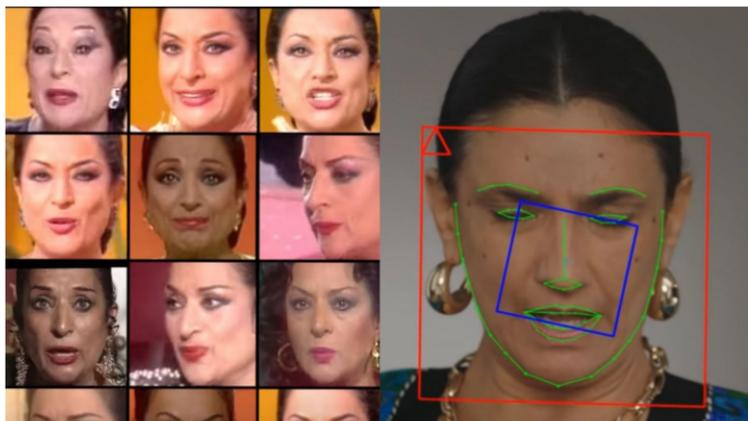


Figura 5. Fotograma del making of del comercial de la cerveza Cruzcampo a través del empleo del deepfake. Fuente: <https://www.youtube.com/watch?v=BQLTRMYHwvE&t=2s>

Por otro lado, en el campo de la investigación científica ayudan a realizar simulaciones, visualizar hipótesis complejas y datos abstractos en áreas como la química, la biología, la medicina, la astronomía, entre otros. En la medicina, por ejemplo, la IA generativa se utiliza para crear escáneres cerebrales falsos basados en datos de pacientes, que permiten entrenar algoritmos para detectar tumores en imágenes reales.

En el sector educativo se están usando para la generación de ilustraciones explicativas para cualquier tema de forma dinámica; lo que ayuda al docente a captar el interés del estudiante y estimular su creatividad desde una perspectiva más novedosa. Sin dejar de lado que también contribuye desde las aulas al desarrollo de competencias para la interacción con la inteligencia artificial.

El potencial de la IA generativa es enorme. El hecho de que se esté apostando por la multimodalidad de funciones puede tomarse como un indicador de las demandas de los usuarios, por encontrar en un solo espacio las diferentes herramientas con las que interactúa.

Es una capacidad muy potente con infinidad de aplicaciones creativas y prácticas. Sin embargo, no podemos pasar por alto las deficiencias desde el punto de

vista de representación cultural, de género, étnica, entre otras que la mayoría de estas herramientas presentan.

SESGOS Y CONSIDERACIONES ÉTICAS

A pesar de que los modelos de IA generativa de imágenes están en constante evolución, aún muestran numerosos sesgos. Al momento en el que se escribe esta guía, entre las principales desventajas asociadas al uso de dichas herramientas se encuentra la incapacidad de devolver imágenes en alta resolución (aunque los sistemas de pago sí ofrecen esta posibilidad), la ausencia de modelos entrenados en español, lo que fuerza a interactuar con ellas en inglés para un mejor desempeño, y sobre todo, la hegemonía de los grandes sistemas de representaciones simbólicas producidos desde los países anglosajones de Occidente.

Un [estudio de 2023](#) sobre cómo Midjourney representa visualmente términos aparentemente genéricos en el contexto de la profesión periodística identificó los siguientes sesgos: en los trabajos no especializados, solo mostraba imágenes de hombres y mujeres jóvenes, mientras que en posiciones especializadas, se mostraban tanto personas jóvenes como mayores. No obstante, solo los hombres presentaban rasgos de madurez física, asociada a la experiencia y experticia. Ello evidencia patrones de edadismo y sexismo. En cuanto a la

representación de ambos sexos, las mujeres eran jóvenes, mientras que a los hombres se les "permitía" tener arrugas. Estos resultados refuerzan implícitamente prejuicios como la suposición de que las personas mayores no trabajan (o no pueden trabajar) en puestos no especializados, que solo los hombres mayores son aptos para trabajos especializados, y que los trabajos más comunes y sencillos son cosa de mujeres. Sin dejar de mencionar que la IA también parecía presentar el género como algo binario, en lugar de mostrar ejemplos de una expresión de género más fluida.

La IA generativa de imágenes también demostró una hegemonía racial en cuanto a la representación de profesiones de servicio público como es el caso de los periodistas. Las imágenes vinculadas a términos como reportero o corresponsal muestran exclusivamente a personas caucásicas. Asimismo, el estudio arrojó tendencias conservadoras en cuanto al aspecto: cero tatuajes, piercings y peinados poco convencionales, pero sí atributos como camisas abotonadas y corbatas, que marcan expectativas de clase. Aunque esta vestimenta responde a la estética de determinadas profesiones como la de presentador de noticias, no debiera ser, nunca mejor dicho, una camisa

de fuerza a la hora de reflejar cómo lucen los reporteros o periodistas en la vida real.

Estas limitaciones de la IA generativa tienen un impacto perjudicial, en tanto, reproducen y amplifican estereotipos y prejuicios en torno a determinadas profesiones, clases sociales, grupos étnicos, diversidad sexual e identidad de género. Ello puede dar paso al reforzamiento de la discriminación y las violaciones de derechos, así como reducir la confianza de los ciudadanos en la tecnología de inteligencia artificial, a largo plazo.

Otro ejemplo que ilustra las consecuencias negativas de los sesgos fue la controversia sobre un trabajo del sitio web *Buzzfeed* sobre la representación de 194 culturas alrededor del mundo con Barbies, mediante el empleo también de Midjourney. La publicación se volvió tendencia en la red social X (antiguo Twitter) ante las críticas de los usuarios.

Las imágenes muestran el comentario de una persona que identificó el error en el brazo derecho de la Barbie francesa, que presenta la palma de la mano invertida. Mientras que la muñeca que representa nuestro país tiene tres extremidades superiores: las dos junto al cuerpo y la tercera reposada en el carro. Ello sin mencionar la representación simplista de la mujer y la cultura

cubanas. No obstante, estas fallas son insignificantes si tenemos en cuenta que según la publicación, la Barbie alemana vestía un uniforme militar del ejército nazi y a su lado, la representante de Sudán del Sur muestra un fusil como parte de su conjunto.



scorpionic @bIvckbeatIes · 8 jul.

...

En respuesta a @BuzzFeed

France hand is backwards and cuba has 3 arms. Also most of those **barbies** look ai generated or photoshopped models



Amy Glover / Midjourney



43. Cuba



6

6

41

6 mil

🔖

Figura 6. Captura de pantalla de la denuncia de un usuario en X sobre los errores de las imágenes sintéticas de las Barbies representativas de Francia y Cuba. Fuente: <https://twitter.com/blvckbeatles/status/1677711287076155393>



Figura 7. Captura de pantalla de la denuncia de un usuario en X sobre los errores de las imágenes sintéticas de las Barbies representativas de Alemania y Sudán del Sur. Fuente: <https://twitter.com/wagatwe/status/1677690070445309952>

La publicación ya no se encuentra disponible online debido a la furia que desató en la comunidad de usuarios, quienes denunciaron su carácter eurocentrista y xenofóbico, además del empleo de estereotipos que denigran la esencia de estas naciones y su gente. Al margen de las posibles intenciones que haya tenido *Buzzfeed* de

instar a la polémica con el fin de generar ingresos mediante la interacción de los usuarios, esta serie de imágenes demuestra la irresponsabilidad del medio de comunicación al enviar al ciberespacio imágenes sin una adecuada evaluación previa.

De aquí se deriva otro tema urgente en esta conversación: el empleo ético y responsable en la comunicación social de las IA generativas. Cada vez es más frecuente el uso editorial de estas herramientas. Revistas como *The Economist* y *Cosmopolitan*, la han empleado en ediciones y artículos dedicados enteramente a las potencialidades y alcance de las IA. Sin embargo, el investigador español Jorge Franganillo destaca el caso protagonizado por el escritor Charlie Warzel, colaborador de la revista *The Atlantic*, que desató una ola de críticas por ilustrar un artículo de opinión con una imagen sintética, prescindiendo así del trabajo de un ilustrador y del material de un banco de imágenes.

Aunque el debate es álgido sobre cuál debe ser la pauta a seguir, al momento en el que se redacta este libro no existe consenso en torno a un código de ética o norma que regule dicha actividad. Más bien, lo que circulan son manuales generales para el empleo de la inteligencia artificial en los distintos ámbitos. Ante los retos de la incorporación de esta a la comunicación so-

cial resulta válido hacernos las siguientes interrogantes: ¿Con qué criterios éticos se han desarrollado los algoritmos? ¿Son compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural? ¿Qué pasa cuando las líneas entre lo real y lo generado son casi imperceptibles?

En 2023 imágenes de Donald Trump siendo detenido por la policía dieron la vuelta al mundo. Las fotos creadas en Midjourney por Eliot Higgins, periodista y fundador del colectivo de investigación Bellingcat, estremecieron a los usuarios por su hiperrealismo; que, sin lugar a dudas, engañó hasta los ojos más entrenados para detectar bulos. Asimismo, circularon fotos del Papa Francisco portando un abrigo de plumas y el abrazo entre los políticos españoles Yolanda Díaz y Pablo Iglesias que fue portada en el diario ibérico *El Mundo*. Todos y cada uno de estos escenarios era falso; sin embargo, las figuras representadas no lo son. Jorge Franganillo en su artículo *La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos* alerta que la confianza de los lectores en las evidencias gráficas se debilita cuando algunas cabeceras se apoyan en imágenes creadas con IA generativa muy realistas para ilustrar sus titulares. Frente a la ola de noticias y audiovisuales falsos que a diario circulan por las redes sociales digitales

e Internet, el empleo de modelos generativos pueden propiciar nuevas formas de desinformación.

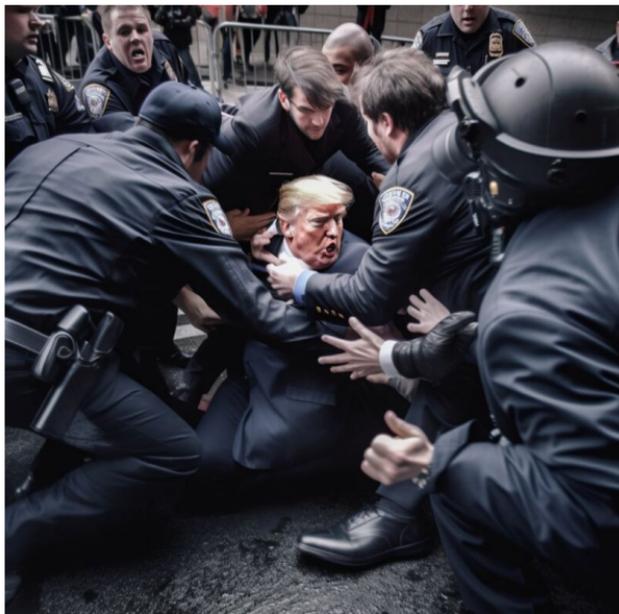


Figura 8. Imagen creada por Eliot Higgins en Midjourney en 2023. Fuente: <https://arstechnica.com/tech-policy/2023/03/fake-ai-generated-images-imagining-donald-trumps-arrest-circulate-on-twitter/>

Las ajetreadas dinámicas del mundo moderno han modificado los tiempos de consumo por parte de las audiencias. Para un lector promedio que compra el periódico o que accede a las noticias a través del móvil mientras se desplaza para ir al trabajo, a primera ojeada

no tiene mucho tiempo para discernir si la imagen en portada de Yolanda Díaz y Pablo Iglesias es verídica o no. Por lo que es más susceptible a creer que tal cordial encuentro entre rivales políticos tuvo lugar. Solo si revisa por segunda vez, con detenimiento, la composición de la primera página del diario, es que comprenderá la intencionalidad detrás de su diseño. En este caso específico, se ha empleado la imagen como gancho (o clickbait si es en la web) para conectar el principal titular del día de la alerta del PSOE sobre la vicepresidenta, y el artículo contiguo sobre los desafíos de la inteligencia artificial generativa.



Figura 9. Portada del diario *El Mundo* del martes, 4 de abril de 2023.

¿Qué pasa si imágenes sintéticas hiperrealistas adquieren voz y movimiento? La continua aparición de nuevas y mejores herramientas de IA para producir contenido despierta inquietud en las industrias mediática y cultural; pues estos hitos muestran la rapidez con que las tecnologías digitales pasan de la innovación a la mercantilización, y cómo en el proceso pasan factura no solo a los profesionales detrás del telón sino al público que está expuesto directa e indirectamente a sus resultados. En este sentido, a lo largo de este libro hemos profundizado intencionalmente en el conocimiento sobre las IA generativas de texto e imágenes, por ser quizás las herramientas más populares entre los usuarios. Sin embargo, no queremos pasar por alto la polémica en torno a la arista ética que las herramientas de generación artificial de sonido y video también están causando.

Además del tema de la desinformación podemos identificar otros focos de debate como la sustitución o devaluación del trabajo humano ante sistemas entrenados, y los derechos sobre la imagen y la voz de una persona para ser reproducidos por una IA. En el epígrafe anterior, comentamos sobre cómo el deepfake ha revolucionado la industria audiovisual al poder transformar por completo la apariencia o el discurso de un personaje (verídico o ficticio), haciéndole decir o hacer cosas que no son reales. Con más frecuencia encontramos en Internet y las redes sociales digitales, videos e incluso cuentas

con fines lúdicos como @deeptomcruise, el usuario de TikTok que a través de esta tecnología suplanta al actor Tom Cruise con videos donde cuenta chistes y hace trucos de magia.

No obstante, en los últimos años se han vuelto tendencia otros contenidos sintéticos que han causado gran repercusión más allá de la industria del entretenimiento. En 2022, por ejemplo, fue [publicado en un tabloide ucraniano](#), un corto donde presuntamente el presidente Volodimir Zelenski ordenaba la rendición de su ejército en el enfrentamiento armado con Rusia. Considerado como el primer deepfake empleado en un conflicto bélico, fue inmediatamente desmentido por el propio mandatario ucraniano. Posteriormente, Nathaniel Gleicher, jefe de Políticas de Seguridad de Meta, confirmó que su equipo eliminó el vídeo de sus plataformas. Algo que también hicieron YouTube y Twitter.

Al año siguiente, circularon por las redes digitales dos videos de supuestos presentadores de noticias de un medio de comunicación falso llamado Wolf News. En este caso, ambos audiovisuales fueron creados con [Synthesia](#), un software de IA que desarrolla avatares basados en la apariencia de actores contratados y pueden programarse en más de cien idiomas y acentos. Victor Riparbelli, su cofundador y director ejecutivo, explicó que los creadores de los presentadores de noticias falsos habían violado las condiciones de servicio de la

empresa, que establecen que esta tecnología no debe utilizarse para “contenido político, sexual, personal, delictivo o discriminatorio”. También confirmó que sus perfiles de usuarios habían sido cancelados e instó a las autoridades políticas a establecer reglas más claras en cuanto al uso de las IA.

El *New York Times* **dedicó un trabajo** para advertir las consecuencias del empleo de esta tecnología donde alertaba sobre cómo la creación de personajes hiperrealistas con fines de manipulación puede perjudicar la capacidad de las personas de identificar lo real de la ficción así como, a la larga, mermar aún más la desconfianza del público en los canales de información oficial.

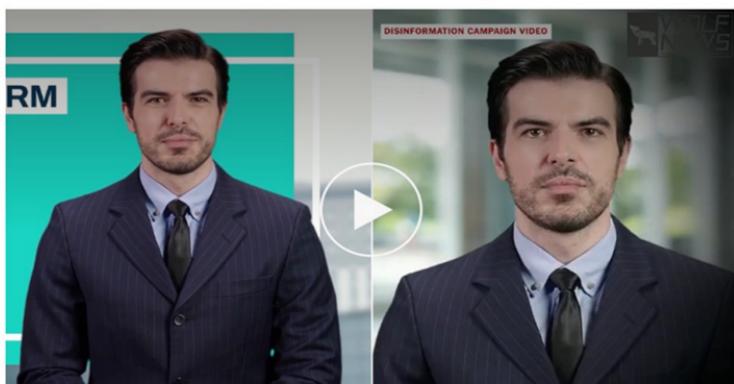


Figura 10. El joven de la imagen es Jason, uno de los más de 85 avatares desarrollados de la empresa Synthesia, para hablar en 120 idiomas y con diferentes acentos.

Fuente: <https://www.nytimes.com/es/2023/02/09/espanol/inteligencia-artificial-deepfake.html>

El uso con fines poco éticos de estas tecnologías pueden atentar contra la reputación de las víctimas e incluso, a gran escala, entorpecer o empeorar las relaciones diplomáticas entre naciones. Resulta muy difícil prevenir la explotación de la imagen de personas con presencia pública para generar bulos, fraudes y ataques hacia otras personas, organizaciones y gobiernos. Artistas, políticos, presentadores de noticias y creadores de contenido son más vulnerables, dado el nivel de exposición que tienen en la esfera pública y mediática, así como la cantidad de información visual sobre ellos que existe en la web. Una vez más, lo que preocupa es el acceso masivo de los usuarios a las herramientas de IA generativa sin el desarrollo de regulaciones legales firmes que eviten sus usos lacerantes.

Otra capa de complejidad que se añade es el respeto a los atributos y cualidades que distinguen a cada individuo. ¿Existe algún sello más identitario que la voz humana? ¿Qué pasa cuando un sistema entrenado es capaz de almacenar, analizar y reproducir tu voz? En 2023, un internauta bajo el seudónimo de FlowGPT creó varias canciones que emulaban de forma artificial la voz y estilo musical del cantante puertorriqueño Bad Bunny, quien al conocer la noticia, repudió públicamente estas acciones. El año anterior, la plataforma de streaming

Spotify borró decenas de miles de canciones sintéticas para tratar de frenar el fraude. La compañía informó que esta eliminación no se realizó por su origen artificial, sino porque las utilizan para cobrar derechos de autor a nombre de cuentas fraudulentas.

La voz de un individuo, como uno de sus rasgos de identidad, se somete a consideraciones especiales que van más allá de lo tecnológico, por lo que su tratamiento debe ser igualmente responsable. De ahí que, diferentes empresas de síntesis de voz han introducido pautas sobre el uso de la tecnología y están incorporando los medios para garantizar que la persona propietaria de la voz sintetizada realmente haya autorizado su uso para proyectos específicos. El propósito de estas medidas es proteger los intereses de quienes prestan su voz con fines profesionales y artísticos, para ayudarles a mantener el control sobre esta y recibir la remuneración correspondiente.

No obstante, ¿qué pasa cuando se trata de una persona fallecida? En 2021, la viuda del chef estadounidense Anthony Bourdain rechazó el empleo (sin su consentimiento) de una herramienta de IA para hacerlo pronunciar tres líneas de las que solo había constancia escrita, en un documental sobre la vida de su esposo. El hecho de que las IA generativas faciliten el rescate de la memo-

ria histórica a partir de la recreación digital (por ahora) de personas desaparecidas físicamente, abre la puerta a nuevas interrogantes de matiz legal y filosófico sobre la línea entre la vida y la muerte: ¿existe en estos tiempos el verdadero descanso eterno?

Una de las principales ambiciones de la especie humana es dejar su legado para la posteridad. Parte del consuelo para los que quedan tras un deceso son las constancias gráficas del paso por la vida del ser querido. En el caso del artista aumenta el valor de su obra pues es la huella de un talento al que la muerte puso punto final. ¿Sabemos si Lola Flores hubiese querido seguir trabajando después de fallecida? Probablemente era una idea inconcebible en su momento.

Investigadores y expertos alertan sobre los intereses económicos detrás de la generación de la imagen de una personalidad desaparecida físicamente con tecnología de IA. Amén de los debidos autorizos de familiares y albaceas, en la mayoría de los casos nos encontramos ante el vacío legal de la imposibilidad de saber si la persona hubiese aceptado la explotación de su imagen en el futuro para proyectos de diferente índole. Al “revivir” a una personalidad en un audiovisual resulta imposible conocer los términos y condiciones que en vida hubiese impuesto, así como su aporte real al proceso creativo.

Asimismo, el hecho de reutilizar a las mismas figuras atenta contra la aparición de nuevos talentos, y ahí sí, pudiéramos afirmar la sustitución del talento humano por sistemas entrenados. No por gusto, a mediados de 2023 Hollywood se paralizó por más de cien días por la huelga del sindicato de actores, donde uno de los principales reclamos fue la aprobación de nuevas normas que regularán el uso de la inteligencia artificial.

Asumir la IA generativa de imágenes como parte de nuestro arsenal de herramientas tanto para uso personal como profesional requiere de una conciencia crítica y vigilante. El ser humano crea tecnología y se transforma a sí mismo a través de ella. El hecho de que estas herramientas se usen para reforzar desigualdades, prejuicios, comportamientos éticamente cuestionables..., salta las alarmas. Habla más de nosotros como sociedad que de la tecnología en sí misma.

¿CÓMO SEGUIR A PARTIR DE AQUÍ?

A lo largo de este libro, hemos realizado una introducción práctica de la inteligencia artificial generativa para no especialistas en el tema. Introducimos conceptos básicos, destacamos hitos de su historia y enfatizamos aspectos a tener en cuenta para su uso en el ámbito laboral diario.

Estamos convencidos de que la inteligencia artificial generativa impulsará en gran medida la productividad en numerosos sectores económicos. Creemos que su adopción elevará el nivel de exigencia a los futuros profesionales. Al mismo tiempo, consideramos que requiere una reflexión detenida sobre las condiciones de su uso, los objetivos a alcanzar y, en general, qué lugar le daremos a un uso consciente y estratégico. Por estratégico, como argumentamos en el libro, entendemos un uso consciente, que aprovecha las facilidades de uso de la tecnología, pero al mismo tiempo está al tanto de sus limitaciones, de sus sesgos, y que la trata con la desconfianza suficiente como para no depender demasiado de ella.

Dada la situación de bloqueo que enfrenta Cuba y los esfuerzos de Estados Unidos por negar a los profesiona-

les cubanos el acceso a las tecnologías más avanzadas, se intensifica la necesidad de redoblar los esfuerzos por superarse constantemente, por una parte; y por otra soportar los riesgos y los beneficios inherentes en el uso de cualquier nueva herramienta tecnológica.

Un uso estratégico implica evaluar y ajustar de manera continua la coincidencia entre las funcionalidades y sesgos de la IA con los objetivos y metas del flujo de trabajo donde se inserta, especialmente la compatibilidad con métodos de trabajo previos y los sistemas de valores del profesional o grupo de profesionales que la usa. En el caso de los servicios de chatbot en línea, como ChatGPT o Claude, se necesita extremar las precauciones en el manejo de la información sensible, así como establecer protocolos para la evaluación constante de la veracidad y ausencia de sesgos de los contenidos generados por la IA. Por lo mismo se debe monitorear constantemente el desempeño y la efectividad de su uso. Ninguna tecnología constituye una bala mágica para resolver todos los problemas.

Otro mensaje que presentamos en el libro fue la necesidad de enfocarnos no en cómo aprender a usar una tecnología específica, sino a aprender a aprender. Es fundamental dirigir una mirada crítica hacia la IA, pero también lo es preocuparnos por aprender constante-

mente, sobre las tecnologías y su relación con nuestros objetivos profesionales y personales, y preocuparnos por aprender cada vez mejor, con mayor eficiencia, de manera más crítica; porque la velocidad de los cambios de nuestro entorno no nos permite acomodarnos mucho. No hacemos énfasis en una enumeración de sitios ni recursos en línea, que pueden estar ya inaccesibles en el momento de publicación de este libro, aunque el lector interesado puede tomar las referencias que damos al final como punto de partida para su exploración. Recomendamos en general buscar regularmente nuevos recursos con la ayuda de los motores de búsqueda de Internet, evaluar y seguir sitios web, comunidades en línea y publicaciones periódicas. También sugerimos prestar atención a lo que publican y dicen sobre la IA generativa organismos internacionales y regionales de peso, como la UNESCO o la Unión Europea. Especialmente, instamos a considerar a la tecnología, incluso una por momentos tan sorprendente como la IA generativa, como algo construido por humanos, con sus sueños y con sus defectos. Son nuestros valores, en última instancia, los que determinarán hacia cuál lado se inclinará.

REFERENCIAS

ARISTÓTELES. (1981). *Política*. Gredos. España

BRYCE, A. L. (16 de marzo de 2023). "Tecnología del duelo": La IA se utiliza para resucitar a los seres queridos. *Euronews*. <https://es.euronews.com/next/2023/03/16/tecnologia-del-duelo-la-ia-se-utiliza-para-resucitar-a-los-seres-queridos>

DAVIES, C. (23 de marzo de 2023). "No, Donald Trump no ha sido detenido, son imágenes de inteligencia artificial." *El Mundo*. <https://www.elmundo.es/internacional/2023/03/23/641c97c921efa034268b45b0.html>

DAVIES, C. (2023). "Que es LLaMA, como funciona y que sabemos de la inteligencia artificial de Meta." *Xataka Basics*. <https://www.xataka.com/basics/que-llama-como-funciona-que-sabemos-inteligencia-artificial-meta>

DE LA HOZ, K. (4 de julio de 2023). "Recursos para explorar la Inteligencia artificial generativa en el periodismo y en la educación." *Noches de Media*. <https://nochesdemedia.com/2023/07/04/recursos-para-explorar-la-inteligencia-artificial-generativa-en-el-periodismo-y-en-la-educacion/>

ERARD, G. (18 de julio de 2023). Meta anuncia LLaMA 2, su nueva IA de código abierto, y se lanza a la yugular de

GPT-4. *Hipertextual*. <https://hipertextual.com/2023/07/meta-llama-2-nuevo-modelo-ia-codigo-abierto>

FERNÁNDEZ, Y. (6 de octubre de 2023). Bing Chat con ChatGPT: qué es, cómo funciona y qué puedes hacer con el chat con inteligencia artificial del buscador. *Xataka*. <https://www.xataka.com/basics/bing-chatgpt-que-como-funciona-que-puedes-hacer-con-el-chat-inteligencia-artificial-buscador>

FRANGANILLO, J. (13 de septiembre de 2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos. Revista De Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>

GILBERT, D. (21 de julio de 2023). Meta anuncia LLaMA 2, su nueva IA de código abierto, y se lanza a la yugular de GPT-4. *El Mundo*. <https://es.wired.com/articulos/meta-lanza-llama-2-una-nueva-ia-de-uso-comercial-con-codigo-abierto>

GONZÁLEZ, F. (24 de febrero de 2024). Meta presenta su propia inteligencia artificial: LLaMA, y la abre a los investigadores. *Cinco Días*. <https://cincodias.elpais.com/companias/2023-02-24/meta-presenta-su-propia-inteligencia-artificial-llama-y-la-abre-a-los-investigadores.html>

HEIKKILÄ, M. (11 de abril de 2023). ¿Y si pudiéramos pedirle a la IA que tuviera menos sesgos?. *MIT Technology Review*. <https://www.technologyreview.es/s/15237/y-si-pudieramos-pedirle-la-ia-que-tuviera-menos-sesgos>

HORSEY, J. (28 de agosto de 2023). ChatGPT vs bing vs bard vs claude comparison which one is right for you? *Geeky Gadgets*. <https://www.geeky-gadgets.com/chat-gpt-vs-bing-vs-bard-vs-claude/>

KARPATY, A. (12 de enero de 2023). [1hr Talk] *Intro to Large Language Models* - YouTube. YouTube. https://www.youtube.com/watch?v=zjkBMFhNj_g

KLEMKE, N. (n.d.). Introducción a Neural frames. *Neural frames*. Recuperado el 31 de octubre, 2023, de <https://www.neuralframes.com>

NILSSON, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press.

PÁEZ, F. (22 de marzo de 2023). IA generativa: ¿Qué es y cómo funciona? *Mercadeo Digital*. <https://mercadeo-digital.co/que-es-el-prompt-engineering-todo-lo-que-necesitas-saber/>

PASTOR, J. (22 de junio de 2023). Hemos probado Claude, el competidor de ChatGPT creado por Anthropic. Y se comporta la mar de bien. *Xataka*. <https://www.xataka.com/robotica-e-ia/hemos-probado-clau->

de-competidor-chatgpt-creado-anthropic-se-comporta-mar-bien

RIVERA, J. G. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *methaodos. Revista De Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>

RUSSELL, S. J., & NORVIG, P. (2010). *Artificial intelligence a modern approach*. 3rd Edition, Prentice-Hall, Upper Saddle River.

SATARIANO, A., & MOZUR, P. (2023, February 09). Los presentadores de noticias son ficticios, la desinformación es de verdad. *The New York Times*. <https://www.nytimes.com/es/2023/02/09/espanol/inteligencia-artificial-deepfake.html>

SORIANO, R. (8 de noviembre de 2023). Bad Bunny carga contra una canción creada con IA: "Si les gusta esa mierda, no merecen ser mis amigos". *El País*. <https://el-pais.com/cultura/2023-11-08/bad-bunny-carga-contra-una-cancion-creada-con-ia-si-les-gusta-esa-mierda-no-merecen-ser-mis-amigos.html>

THOMSON, T. J., & THOMAS, R. J. (14 de julio de 2023). Siete ejemplos de sesgo en imágenes generadas por IA. *Red internacional de periodistas*. <https://ijnet.org/es/>

story/siete-ejemplos-de-sesgo-en-im%C3%A1genes-generadas-por-ia

ZHANG, C., RAJA, A., & KWEON, I. S. (2023). Text-to-image diffusion models in generative AI: A survey. *Journal of Latex Class Files*, 14(8), 1-23. <https://doi.org/10.48550/arXiv.2303.07909>

SOBRE LOS AUTORES

Hamlet López García (La Habana, 1977). Doctor en Ciencias de la Comunicación Social por la Facultad de Comunicación de la Universidad de La Habana. Es investigador del Instituto Cubano de Investigación Cultural Juan Marinello y profesor en la Facultad de Comunicación de la Universidad de La Habana. Tiene numerosos artículos y capítulos de libros publicados en Cuba y en el extranjero. Entre sus temas de investigación se encuentran la apropiación social de las tecnologías digitales en la sociedad cubana, la economía política de la comunicación y la participación digital.

Diasbel Lilianna Chang Olivera (Camagüey, 1999). Licenciada en Periodismo de la Universidad de La Habana (2021). Es profesora instructora del Departamento de Periodismo de la Facultad de Comunicación y ha colaborado con diversos medios cubanos. Entre los temas de investigación que desarrolla se encuentran el uso y apropiación de las Tecnologías de la Información y la

Comunicación, la colaboración en red y la etnografía digital.

María Carla O'Connor Barrios (La Habana, 1996). Licenciada en Periodismo de la Universidad de La Habana (2019). Es profesora asistente del Departamento de Periodismo, donde dirige la Disciplina Comunicación Hipermedia, Tecnología y Sociedad. Como periodista ha trabajado en las redacciones multimedia de diferentes medios cubanos e internacionales. Entre los temas de investigación que desarrolla están las nuevas narrativas del periodismo, uso y apropiación de las TIC por los usuarios, y el impacto de las redes sociales digitales en el escenario mediático.