

Capítulo 2

Inteligencia Artificial: ni muy automatizada, ni muy ecológica

*Antonio de Jesús García Chávez
José Antonio García Macías*

*Todos ven lo que tú aparentas; pocos advierten lo que eres.
—Nicolás Maquiavelo*

<https://doi.org/10.61728/AE24001021>



Introducción

Tras un largo periodo con relativamente poca actividad y resultados, en años recientes se ha generado un incremento en la cantidad de publicaciones científicas relacionadas con la inteligencia artificial (IA), con un crecimiento del 100 % desde 2010 (Stanford University, 2023). Hoy en día nos encontramos en una revolución en relación con la IA, y sus efectos repercuten mucho más allá de los confines de los laboratorios de investigación y las empresas tecnológicas, abarcando también los sectores industriales y de gobierno. Esto se vuelve claro al analizar el aumento en el número de incidentes y controversias relacionados al uso indebido de la IA; los últimos reportes indican que los casos de mal uso ético de la IA se han multiplicado por 26 desde el año 2012 (Stanford University, 2023).

Si bien es cierto que la investigación en IA ha aumentado en todas sus áreas, hay que denotar que desde el 2017 particularmente las áreas de reconocimiento de patrones, aprendizaje automático y visión por computadora son aquellas con la mayor cantidad de publicaciones científicas (Stanford University, 2023). Sin embargo, en los últimos cinco años la IA se ha colocado no solo como un tema de investigación y desarrollo comercial; ya que en gran medida aplicaciones y sistemas con base en IA, como lo son filtros para fotografías y video, así como chatbots, se han encargado de embarcar a más de uno en un viaje por el reino de las IA. Lo anterior ha sido posible en gran medida gracias al encanto de la IA generativa y, en particular, de los enigmáticos Grandes Modelos de Lenguaje (LLM o Large Language Models), los cuales han captado la imaginación y expectativas del mundo de manera inusitada. Estos LLM, tales como GPT o PaLM, representan un cambio de paradigma que trasciende la mera computación y se adentra en las intrincadas facetas de la cognición, la creatividad y la comunicación humanas. Por lo que no ha de parecernos extraño que, en un futuro no muy lejano, el área del Procesamiento del Lenguaje Natural abandone su posición como sexta área de investigación con más publicaciones científicas y se posicione dentro de las tres áreas con mayor número

de publicaciones, siendo estas Reconocimiento de patrones, Aprendizaje de máquina y Visión por computadora.

El crecimiento y atención mediática de estos Modelos de Lenguaje, no solo ha sido un esfuerzo proveniente de la academia, lo cual sería fácil de suponer ya que hasta 2014 la mayoría de los modelos de aprendizaje automático más significativos procedían del mundo académico. Sin embargo, en los años siguientes, la industria ha asumido un papel dominante. En 2022 el sector industrial había producido treinta y dos modelos de aprendizaje automático dignos de mención, mientras que el mundo académico solo había contribuido con tres (Stanford University, 2023). El desarrollo de sistemas de IA, de vanguardia ahora, exige datos sustanciales, recursos computacionales e inversiones financieras, todos los cuales están más fácilmente disponibles para los actores de la industria que para las organizaciones sin fines de lucro y las instituciones académicas.

Inteligencia artificial generativa

Aunque la IA abarca una amplia gama de aplicaciones, un área en concreto que ha acaparado gran atención en los últimos años es la inteligencia artificial generativa (IAG), la cual en el último año ha dominado el panorama de la IA de forma espectacular, constituyendo un auténtico acontecimiento inesperado y transformador. Esto ha ocurrido ya que la IAG ha demostrado un amplio abanico de capacidades que superan con creces las de cualquier sistema de IA anterior, y que abarcan numerosas funciones con beneficios económicos y humanos tangibles (Eloundou, Manning, Mishkin y Rock, 2023).

Imaginemos por un momento un mundo futurista en el que los ordenadores no solo puedan entender nuestro lenguaje, sino también componer historias, crear obras de arte e incluso componer música que resuene con nuestras emociones. La IA generativa hace posible este mundo, que, hasta hace no mucho, resultaba como un sueño aparentemente lejano. Es interesante notar que algunas de las capacidades actuales de estos modelos de IA generativos solo se consideraban alcanzables en el marco de sistemas de IA que poseyeran una competencia general de nivel humano, lo cual ha desafiado las expectativas de muchos expertos en el área (Goertzel y Singularitynet, 2023).

En el panorama de la IAG, centraremos nuestra atención en el ámbito del lenguaje, particularmente en los grandes modelos del lenguaje o bien “Large Language Models (LLM) y más concretamente en los modelos conocidos como GPT (Generative Pre-trained Transformers). Pero ¿qué es exactamente lo que hace tan cautivadores a estos LLM? Estos modelos al ser entrenados con grandes cantidades de datos aprenden a imitar en gran medida la creatividad humana, obteniendo de esta forma, el potencial para escribir artículos convincentes, generar imágenes realistas e incluso entablar conversaciones naturales. Estos LLM actualmente se emplean en diversas aplicaciones industriales, desde la búsqueda en la Web y los chatbots conversacionales incluso hasta en el análisis de documentos médicos y financieros (Pandu Nayak, 2019; Joshua Broyde, 2021; Thewsey, 2021; Lee, 2023). Debido a esta amplia aplicabilidad, estos LLM se han denominado foundation models (que pueden traducirse como “modelos de cimentación”) en el área del Procesamiento del Lenguaje Natural (Bommasani et al., 2021).

Hasta este momento hemos contextualizado lo impresionante y atractivos que resultan estos LLM, así como lo versátiles que son al impactar en distintas áreas. Por lo que a continuación focalizaremos nuestros esfuerzos en arrojar luz en algunos de los conceptos e ideas que rodean las distintas arquitecturas y procesos de entrenamiento que estos LLM necesitan para lograr estos comportamientos y capacidades tan intrigantes. Consideremos por cuestiones pedagógicas la siguiente analogía, imaginemos que los LLM son bibliotecas que contienen una vasta y amplia cantidad de libros, cada uno de estos libros, representará un aspecto diferente del lenguaje natural y la expresión humana, algunos se enfocan en esclarecer cuestiones sintácticas, semánticas, gramaticales y pragmáticas, por mencionar algunos. De manera similar en esta biblioteca, podrías encontrar libros donde la expresión escrita del lenguaje natural se emplea en debates, noticias, artículos científicos, libros de ficción, poemas y enciclopedias; consideremos entonces todos los formatos y usos del lenguaje en internet, páginas web, blogs, redes sociales, etc. Emplearemos entonces esta analogía para explicar a grandes rasgos los componentes principales de estos LLM.

Consideremos que los libros contienen en cierta medida, el conocimiento colectivo del mundo con respecto al uso del lenguaje natural. Los

LLM devoran este festín textual, con un objetivo similar al de un erudito, el cual es comprender los matices del lenguaje y convertirse en expertos en diversas materias. Hay que denotar que las capacidades de estos LLM se ven potenciadas con respecto al tamaño de su biblioteca, por ejemplo BERT (Devlin, Chang, Lee y Toutanova, 2019) uno de los primeros y más populares LLM fue entrenado con un tamaño de biblioteca de alrededor de 24 GB. Por otro lado, GPT 3 (Brown et al., 2020) fue entrenado con alrededor de 285 000 000 hojas de texto plano o bien aproximadamente 114 000 veces la obra completa de Shakespeare, lo que equivale a 570 GB de información; por lo regular la cantidad de datos que suele emplear estos LLM suele oscilar entre 15 GB y 5 TB de texto plano (Rae et al., 2021; Hoffmann et al., 2022).

¿Entonces qué papel cumple el bibliotecario en esta analogía?, el bibliotecario es aquel que ayuda en la comprensión de las conexiones y los patrones de información que contiene su biblioteca, de tal forma que cuando se le pregunta por algún título, información o sugerencia, este trabaja de prisa en recordar y reunir los libros adecuados, para de esta forma otorgar una respuesta pertinente y útil. Este bibliotecario en el área de la inteligencia artificial, se le ha conocido con distintos nombres, cibernética en los años 1940-1960, conexionismo en los años 1980-1990, o bien quizá lo reconocamos con su nombre más popular redes neuronales (Goodfellow, Bengio y Courville, 2016). Por lo que, en esta analogía, el bibliotecario representa la arquitectura de red neuronal sobre la cual los LLM fueron construidos, esta arquitectura lleva por nombre “Transformer” [15], la cual se ha mantenido prácticamente sin cambios desde 2018. Estos bibliotecarios son los encargados de leer y analizar los libros, aunque debemos mencionar que su forma de dar lectura es diferente a la nuestra; estos son entrenados de una de las siguientes maneras, o bien se entrenan para predecir una palabra enmascarada en un enunciado (es decir, palabras que hay que rellenar), o bien para predecir la palabra siguiente en un enunciado (similar a como los teclados de nuestros teléfonos predicen la siguiente palabra que escribiremos a manera de recomendación). Por lo que el bibliotecario colabora con los parámetros del modelo para dar sentido a la información contenida en los textos, resaltando de esta manera los detalles importantes, clasifican la información y la hacen accesible para su consulta.

Si bien estos LLM constan de más que solo libros, un sistema de catalogación y un bibliotecario (datos de entrenamiento, parámetros, Transformers), podemos decir que al haber explorado estos fundamentales bloques de nuestra biblioteca, hemos sido capaces de apreciar en mayor medida las capacidades de estos LLM, y si de algo estamos seguros, es que veremos cada vez más la adopción de estas tecnologías en nuestras vidas.

Humanos como ayudantes de las máquinas

Suele ser común que las historias que perduran y vencen la prueba del tiempo son aquellas que cuentan los ganadores, o bien, aquellas narrativas que los involucrados buscan imponer y mantener para futuras generaciones; esto ocurre por igual en las historias y narrativas alrededor del desarrollo de la IA. Si bien es cierto que actualmente no es tan difícil encontrar opiniones e investigaciones que argumentan un caso negativo respecto de las IA (Chomsky, 2023; Jem Bartholomew, 2023). Estas suelen encontrarse abrumadoramente opacadas por el sensacionalismo, el cual suele emplear afirmaciones dramáticas sobre temas como las IA con conciencia (Roose, 2023) y el desplazamiento laboral (Gareth Corfield, 2022).

Por otro lado, medios de comunicación, así como futurólogos emplean afirmaciones y títulos exagerados, por ejemplo “¿Podría ChatGPT escribir mi libro y alimentar a mis hijos?” (Heritage, 2022). Por lo anterior no es extraño que una versión de la historia que suele olvidarse o bien dejar de lado, es la de los trabajadores fantasmas. Estas personas son los campeones ocultos de la IA, ya que proporcionan el trabajo tras bambalinas que sustenta gran parte de los avances logrados en las aplicaciones de aprendizaje automático en todo el mundo hoy en día. Cuando los algoritmos de Google funcionan como se supone que deben, mostrando la información que buscabas y no tonterías o porno raro o propaganda o consejos médicos peligrosos, es gracias a estos trabajadores fantasmas. Sin embargo, hay una cierta ironía cruel, ya que a medida que la tecnología con base en IA, por ejemplo, LLMs, realiza su aparición en la escena mediática y económica, los más adecuados para mantenerla funcionando son estos trabajadores fantasmas, que resultan ser también los más precarios en las empresas que los necesitan. Numerosos estudios han demostrado que los

trabajadores están mal pagados y reciben un trato injusto (Horton, 2011; International Labour Organization, 2016; Nast, 2017), un análisis reveló que el salario medio por hora de los trabajadores era de solo 2 USD (Hara et al., 2018).

Esta cifra es muy inferior al salario mínimo de EE. UU. (7.25 USD). Lastimosamente, historias como esta no resultan atípicas. Una investigación realizada por la revista *Time* reveló que a contratistas con sede en Kenia se les pagaba menos de 2 dólares la hora, por revisar contenidos para ChatGPT y muchos de estos contenidos eran tan tóxicos que les dejaban traumatizados (Perrigo, 2023). Aunque son varias las empresas que ofrecen este servicio del tipo *crowdsourcing*, siendo las principales *Mechanical Turk* y *Appen*, resulta una constante los problemas a los que estos trabajadores se enfrentan; por ejemplo, un *rater* (evaluador) en una entrevista para la revista *Los Angeles Times* comentó lo siguiente “suele ser común que de entre las tareas que evaluó pueda encontrar insultos racistas, intolerancia y violencia”, sin embargo esto no resulta lo peor, “A veces he visto parte del contenido gráfico reproducido en mis sueños. Por eso nunca trabajó hasta tarde por la noche. En diez años he visto dos veces pornografía infantil, pero gracias a Dios es muy raro. Lo dejaría” (Merchant, 2023).

Sería sencillo e inocente argumentar que los evaluadores son aquellos con la última decisión, al ser estos los que seleccionan las tareas o proyectos en los cuales trabajar, sin embargo, la realidad es otra. La mayoría de las plataformas de *crowdsourcing* permiten a los solicitantes (*requesters*) crear de manera flexible sus tareas o proyectos, así como establecer los precios correspondientes. Además, los solicitantes tienen la capacidad de evaluar el desempeño de los evaluadores para garantizar la calidad de sus respuestas, verificando si cumplen con los requisitos mediante la revisión de calificaciones (Hara et al., 2018), o bien rechazando las tareas enviadas que no cumplan sus criterios (Bederson y Quinn, 2011). Por otro lado, a los evaluadores se les suele proporcionar una cantidad limitada de información (Irani y Silberman, 2013), por ejemplo, suele ser común que las plataformas de *crowdsourcing* solo ofrezcan detalles esenciales como el precio de la tarea, el nombre del solicitante, un título y una breve descripción. Esta escasez de información supone un reto considerable para los *raters* que buscan tareas de buena calidad.

No obstante, no todo está perdido, hasta hace poco, las acciones colectivas eran poco frecuentes entre los trabajadores de alta tecnología porque estaban bien pagados y satisfechos con sus condiciones de trabajo (Nedzhvet-skaya, 2019). Sumado a esto, cuando una empresa tecnológica toma decisiones poco éticas en relación con el desarrollo y la implantación de la IA, sus empleados pueden encontrar una incoherencia entre su identidad moral individual y su percepción de la imagen de la organización (Alahmad y Robert, 2020).

Por lo que el activismo es una de las principales formas que tiene el público y trabajadores de presionar a una organización para que cambie sus políticas y prácticas (Kalodimos y Leavitt, 2020). No obstante, es crucial considerar que la capacidad de influencia de los activistas puede variar según los costos que conlleva para la organización el no abordar las inquietudes y solicitudes de estos (Belfield, 2020).

Si bien el escenario de los rateros, como acabamos de observar, es bastante desolador, la llegada de estos LLM, ha generado una nueva y más especializada versión de estos rateros, los cuales se concentran en tareas más sofisticadas, ajenas a tareas monótonas de etiquetado de datos como imágenes y video (Metz, 2023). Esta nueva clase de evaluadores nace en el momento que los investigadores notan que el tamaño de estos LLM (parámetros del modelo), no significa que sean mejores a la hora de seguir e interpretar la intención de los usuarios (Ouyang et al., 2022). Por lo que compañías como OpenAI emplearon técnicas de Aprendizaje por refuerzo a partir de retroalimentación humana (Stiennon et al., 2020), con el objetivo de subsanar este problema. Por lo que estos evaluadores actúan como tutores, proporcionando a estos LLM retroalimentación más profunda y específica, en un esfuerzo por mejorar sus respuestas. Esto resulta de vital importancia, ya que a medida que los sistemas de IA se vuelvan más potentes y se les asignan tareas cada vez más importantes, es probable que los errores que cometan se vuelvan más sutiles y críticos en relación con la seguridad (Stiennon et al., 2020).

Estos evaluadores suelen encontrarse en edades de entre los 19 y 62 años, y sus cualificaciones van desde títulos técnicos hasta doctorados. Por otro lado, estos evaluadores se diferencian de los rateros clásicos, también en el salario, con un pago por hora de entre 15 y 30 dólares, esto tan solo

en Estados Unidos, ya que trabajadores en otros países ganan notoriamente menos (Metz, 2023). En gran medida la retroalimentación que realizan estos nuevos evaluadores es la que ha permitido que modelos como GPT, puedan aproximar sus Chatbots a una experiencia conversacional más fluida y por turnos, en lugar de limitarse a ofrecer una única respuesta. De manera similar, esta tutoría que realiza ayuda a empresas como OpenAI a reducir la desinformación, la parcialidad e información tóxica que pueden llegar a producir estos LLM.

Una industria hambrienta de recursos

“Todo necio confunde valor y precio” es una frase atribuida inicialmente a Francisco de Quevedo y después a Antonio Machado, misma que viene muy bien a colación con la IAG. A decir verdad, hay muchos sistemas online que utilizamos en la vida cotidiana y para los cuales no pagamos un costo monetario, sean estos servicios de correo electrónico, mensajería instantánea, canales de video como YouTube, redes sociales, almacenamiento de archivo, por mencionar algunos. A dichos sistemas le atribuimos un valor significativo pues en ocasiones nos facilitan tareas cotidianas y en otras simplemente nos brindan entretenimiento o evasión. A cambio, el precio que pagamos por la utilización de dichos servicios es el de nuestra privacidad (Veliz, 2021), pues nuestra información va a dar a los servidores de quienes proveen los servicios y de ahí, mediante autorización por contratos que nunca leemos o entendemos, utilizan la información para analizarla, procesarla, empaquetarla y venderla al mejor postor. También podrá tener efectos más nocivos como manipular comportamientos, alterar procesos democráticos e impulsar el consumismo con sus efectos medioambientales. Así lo detalla la Dra. Zuboff en su libro *La era del capitalismo de vigilancia* (Zuboff, 2020). Hasta ahora, ChatGPT ha seguido el conocido modelo freemium, donde los usuarios utilizan de forma gratuita una versión básica del sistema, pero deben pagar para poder utilizar funciones avanzadas. Algo similar ocurre con otros sistemas populares de IAG tales como Midjourney y DALL·E. No obstante, vale la pena detenerse a reflexionar sobre los costos ocultos que conllevan la utilización de tales servicios.

A primera vista se puede tener la percepción de que los usuarios de sistemas de IAG producen las imágenes, textos, o cualquier otro producto solamente con la capacidad que sus recursos caseros o de oficina le proveen. Sin embargo, este costo es despreciable si se compara con lo que está sucediendo en el lado de la infraestructura de nube de los proveedores de servicios y –sobre todo– de lo que tuvo que suceder antes de que se pudiera proveer ese servicio. En efecto, para que el servicio fuera puesto a disposición de los usuarios, los modelos de cimentación (*foundation models*) tuvieron que ser entrenados con cantidades de datos colosales; esto se realiza en grandes centros de datos, los cuales consumen también cantidades exorbitantes de energía eléctrica, poder de cómputo, almacenamiento de memoria y fluidos para enfriamiento.

Un estudio realizado para evaluar el ciclo de vida de entrenamiento de varios modelos de IA grandes y comunes (Strubell, 2019) determinó que el proceso puede emitir más de 626.000 libras de dióxido de carbono, lo cual sería equivalente a casi cinco veces las emisiones que tendría un automóvil estadounidense promedio durante su vida útil (incluyendo la fabricación del automóvil en sí). Por otra parte, Microsoft recibió fuertes críticas tras haberse dado a conocer que, justo en medio de una sequía de más de tres años, estaba utilizando enormes cantidades de agua para el sistema de refrigeración de uno de sus centros de datos utilizado mientras entrenaban el modelo GPT-4 de OpenAI.

Por supuesto, a la par de Microsoft están otros impulsores de la IAG tales como Google, consumiendo billones de litros de agua en sus centros de datos, con una tendencia al alza (Adarlo, 2023). En una época donde los sistemas hídricos se encuentran en constante estrés, entre otros efectos atribuidos del cambio climático, no es ocioso preguntarse por qué destinar dichos recursos escasos a actividades que no están relacionadas con la supervivencia humana.

El procesamiento de lenguaje natural es un área de estudio que, tras algunas décadas sin producir resultados notables, en tiempos recientes ha tenido avances fenomenales, sobre todo en lo correspondiente a los modelos de lenguaje o LLM. Estos avances recientes se deben en gran medida a la disponibilidad de grandes volúmenes de datos, aunados a la posibilidad de procesamiento adecuado para manipularlos. Sin embargo,

la tendencia para seguir con estos avances parece ser la producción de modelos cada vez más grandes, que implican costos y riesgos asociados tales como: costos ambientales (que normalmente soportan aquellos que no se benefician de la tecnología resultante); costos financieros, que a su vez levantan barreras de entrada, limitando quién puede contribuir a esta área de investigación y qué idiomas pueden beneficiarse de las técnicas más avanzadas; el costo de oportunidad, a medida que los investigadores desvían sus esfuerzos de direcciones que requieren menos recursos; y el riesgo de daños humanos sustanciales, incluidos estereotipos, denigración, aumento de la ideología extremista y arrestos injustos (Bender, 2021).

Conforme crece la demanda de recursos energéticos, empresas como Microsoft han visto que esta tendencia no va a parar, por lo que han decidido hacer fuertes inversiones en producción de energía nuclear (Bennet, 2023) e incluso han publicado una oferta de trabajo para un gerente de programa de tecnología nuclear, encargado de diseñar una estrategia de reactor “para alimentar los centros de datos en los que residen la nube de Microsoft y la IA”. Si se excluyen las actividades relacionadas con criptomonedas, el consumo de energía eléctrica por parte de los centros de datos ha estado relativamente estable en alrededor del 1 % del consumo global en los últimos años (de Vries, 2023). Sin embargo, dada la carrera desenfrenada por posicionarse como líderes de la IAG que han emprendido las grandes empresas, no es descabellado suponer que dicho consumo puede aumentar de manera significativa. Conforme la IAG se incorpore cada vez más en sistemas de cómputo que se utilizan por usuarios de todos los sectores, la creciente demanda traerá también por consecuencia una mayor cantidad de recursos.

Hay estudios que matizan el impacto ambiental de las actividades en torno a la IA, por ejemplo, haciendo comparativas entre actividades llevadas a cabo por sistemas de IA y contrastando con humanos realizando las mismas actividades. En uno de tales estudios (Tomlison et al., 2023) se presenta un análisis comparativo de las emisiones de carbono asociadas con diversos sistemas de IA (ChatGPT, BLOOM, DALL-E2, Midjourney) y los individuos humanos que realizan tareas equivalentes de escritura e ilustración; en sus resultados encuentran que los sistemas de IA emiten entre 130 y 1500 veces menos CO₂e por página de texto generada en

comparación con los escritores humanos, mientras que los sistemas de ilustración con IA emiten entre 310 y 2900 veces menos CO₂e por imagen que sus homólogos humanos. Resultados como estos podrían utilizarse para justificar el reemplazo de trabajos asignados a personas en favor de sistemas automatizados, aunque las decisiones de este tipo no estarían considerando los altos costos derivados de alterar la estabilidad económica y social de regiones geográficas enteras.

Comentarios finales

El avance desenfrenado de la investigación y el desarrollo en torno a diversos aspectos de la IA ha provocado movimientos importantes en la industria de la alta tecnología. En unos pocos meses, OpenAI pasó de ser una minúscula empresa desconocida a una de las líderes de su segmento, con una valuación de mercado de más 80 000 millones de dólares. La gran demanda por GPUs y chips especializados para la IA, ha dado un fuerte impulso a empresas como Nvidia, que otrora ocuparan un nicho especializado de mercado. Ante la gran demanda de procesamiento, Microsoft le ha apostado muy fuerte al hardware lanzando su propia línea de chips para IA (Nellis, 2023) con el fin de posicionarse en ese mercado y no depender de proveedores externos. La industria creada en torno a la IA se ha concentrado de tal manera que un puñado de empresas determina cómo los desarrolladores de tecnología fabrican y acceden a la potencia computacional; esto influye en el comportamiento incluso de las empresas de IA más grandes cuando se enfrentan a los efectos de la escasez de poder de cómputo.

Un informe reciente de la firma de capital de riesgo Andreessen Horowitz describe la computación como “un factor predominante que impulsa la industria hoy”, señalando que las empresas han gastado “más del 80 % de su capital total en recursos informáticos” (Appenzeller, 2023). Esta concentración en computación también incentiva a los proveedores de infraestructura de nube a actuar de manera que protejan su posición dominante en el mercado, compitiendo para lanzar productos antes de que estén listos para su uso generalizado. Todas estas tendencias no apuntan más que hacia una industria pujante y ávida de recursos energéticos, lo cual trae consigo impactos al ya de por sí afectado medioambiente global.

Entre los muchos recursos que demanda el desarrollo actual de la IA se encuentran los llamados recursos humanos, término que ya de por sí resulta deshumanizante. Se antoja, en cierta manera paradójica, que una industria que tiende a automatizar todo, necesite de trabajo humano para afinar mecanismos, limpiar datos y en general para hacer el “trabajo sucio” pero indispensable para que los sistemas produzcan resultados por lo menos útiles. A menos que se cambie radicalmente de paradigma, los sistemas de IA seguirán desarrollándose en el futuro mediato bajo un modelo de *human in the loop* (humano en el ciclo), necesitando de intervención humana para poder funcionar adecuadamente (Mosqueira-Rey et al., 2023).

A fin de cuentas, la situación de fondo es que las tecnologías de IAG, al igual que muchas otras en la industria informática, siguen un modelo de diseño centrado en las tecnologías y no en las personas, tal como lo aclara el experto en usabilidad y diseño Donald Norman: “Muchos de los sistemas, procedimientos y dispositivos actuales están centrados en la tecnología, diseñados en torno a las capacidades de la tecnología y se les pide a las personas que completen las partes que la tecnología no puede hacer”. Aclara también que “Centrarse en las personas significa cambiar esto, empezando por las necesidades y capacidades de las personas. Significa considerar a todas las personas involucradas, teniendo en cuenta la historia, la cultura, las creencias y el entorno de la comunidad. La mejor manera de hacerlo es dejar que quienes viven en la comunidad proporcionen las respuestas” (Norman, 2019).

Actualmente se sigue debatiendo cuál va a ser la aplicación que justifique el desarrollo de las IAG, pues el hacer la tarea de los estudiantes, o bien generar contenido (textos, imágenes, videos, etc.), sin atribuir las fuentes usadas, no parecen ser las *killer applications* de este sector. Una de las aplicaciones que más presumen quienes quieren vender sistemas de IAG son los ayudantes virtuales que, según prometen, ayudarán a los expertos a llevar a cabo sus tareas de maneras más seguras y confiables. Esto se pone en duda tras resultados de estudios serios recientes, como el uso de chatbots como auxiliares en aplicaciones médicas. En dicho estudio (Lee et al., 2023), se encontró que los auxiliares cibernéticos mostraban respuestas con: a) alta variabilidad y poca similaridad, b) baja confiabilidad de alrededor de 41 % y c) potencial de incurrir en daños. Resulta claro que,

hasta ahora, es difícil confiar en un asistente virtual que ante una situación crítica puede dar respuestas no deterministas, poco confiables y con potencial de causar daños. No obstante, los avances en el área son constantes y a un ritmo fulgurante, por lo que se puede sospechar que habrá mejoras significativas que justifiquen el uso de sistemas IAG en un número creciente de tareas.

Parece haber un consenso en predecir que la IAG seguirá mostrando avances e impactos trascendentales en el futuro inmediato. Como ejemplo, la influyente publicación MIT Technology Review en sus cinco predicciones sobre IA, para el año 2024, indica que cuatro de ellas tendrán que ver con IAG, siendo la quinta relativa a desarrollos de robots generalistas. El gran reto que se presenta es que los avances en torno a la IAG se presentan con una tasa de cambio mucho mayor de lo que son capaces de absorber y asimilar individuos, organizaciones y marcos regulatorios. Si no somos capaces de comprender cabalmente las implicaciones de la IAG en los diversos ámbitos ¿cómo vamos a ser capaces de usarla, controlarla y regularla?

Bibliografía

- Adarlo, S. (2023, septiembre 26). *Critics Furious Microsoft Is Training AI by Sucking Up Water During Drought*. *Futurism*. <https://futurism.com/critics-microsoft-water-train-ai-drought>
- Alahmad, R. y Robert, L. P. (2020). *Artificial intelligence (ai) and its identity: Antecedents identifying with ai applications*. ArXiv, abs/2005.12196. <https://api.semanticscholar.org/CorpusID:218870057>
- Appenzeller, G., Bornstein, M. y Casado, M. (2023, abril 27). *Navigating the High Cost of AI Compute*. Andreessen Horowitz. <https://a16z.com/navigating-the-high-cost-of-ai-compute>
- Bederson, B. B. y Quinn, A. J. (2011). Web workers unite! Addressing challenges of online laborers. In *Conference on human factors in computing systems - proceedings* (pp. 97–105). Association for Computing Machinery. <https://dl.acm.org/doi/10.1145/1979742.1979606> doi: 10.1145/1979742.1979606
- Belfield, H. (2020). Activism by the ai community: Analysing recent achievements and future prospects. In *Proceedings of the aaai/acm conference on ai, ethics, and society* (p. 15–21). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3375627.3375814> doi:10.1145/3375627.3375814
- Bender, E. M., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
- Bennet, D. (2023, septiembre 29). Microsoft Sees Artificial Intelligence and Nuclear Energy as Dynamic Duo. Bloomberg. <https://www.bloomberg.com/news/newsletters/2023-09-29/microsoft-msft-sees-artificial-intelligence-and-nuclear-energy-as-dynamic-duo>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Liang, P. (2021). *On the opportunities and risks of foundation models*. ArXiv. <https://crfm.stanford.edu/assets/report.pdf>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Amodei, D. (2020, may). *Language models are few-shot learners*. In *Advances in neural information processing systems* (Vol. 2020-Decem). Neu-

- ral information processing systems foundation. <https://arxiv.org/abs/2005.14165v4>
- Chomsky, N. (2023). *Noam Chomsky: The False Promise of ChatGPT*. *NY times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
- De Vries, A. (2023). *The growing energy footprint of artificial intelligence*. *Joule*, 7(10), 2191-2194.
- Devlin, J., Chang, M. W., Lee, K. y Toutanova, K. (2019, oct). *BERT: Pre-training of deep bidirectional transformers for language understanding*. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies- Proceedings of the Conference, 1, 4171–4186. en <http://arxiv.org/abs/1810.04805>
- Eloundou, T., Manning, S., Mishkin, P. y Rock, D. (2023, marzo). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. <https://arxiv.org/abs/2303.10130v5>
- Gareth Corfield, M. F. (2022). *Meet ChatGPT, the scarily intelligent robot who can do your job better than you*. [telegraph.co.uk](https://www.telegraph.co.uk/business/2022/12/05/meet-scarily-intelligent-robot-who-can-do-your-job-better-than-you). <https://www.telegraph.co.uk/business/2022/12/05/meet-scarily-intelligent-robot-who-can-do-your-job-better-than-you>.
- Goertzel, B. y Singularitynet, . (2023, sep). *Generative AI vs. AGI: The Cognitive Strengths and Weaknesses of Modern LLMs*. <https://arxiv.org/abs/2309.10371v1>
- Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C. y Bigham, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Conference on human factors in computing systems - proceedings* (Vol. 2018-April). Disponible en <https://doi.org/10.1145/3173574.3174023> doi:10.1145/3173574.3174023
- Heritage, S. (2022). Could ChatGPT write my book and feed my kids? [thetimes.co.uk](https://www.thetimes.co.uk/article/could-chatgpt-write-my-book-and-feed-my-kids-7972vx0xp). <https://www.thetimes.co.uk/article/could-chatgpt-write-my-book-and-feed-my-kids-7972vx0xp>.
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Sifre, L. (2022). *Training Compute-Optimal Large Language Models*. *Advances in neural information processing systems* (Vol. 35). Neu-

- ral information processing systems foundation. <https://arxiv.org/abs/2203.15556v1>
- Horton, J. J. (2011). *The condition of the Turking class: Are online employers fair and honest?* *Economics Letters*, 111(1), 10–12. <https://arxiv.org/abs/1001.1172v1> doi: 10.1016/j.econlet.2010.12.007
- Hui, X., Reshef, O. y Zhou, L. (2023). *The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market*. Available at SSRN: <https://ssrn.com/abstract=4527336> or <http://dx.doi.org/10.2139/ssrn.4527336>.
- International Labour Organization. (2016). *Non-Standard Employment Around the World: Understanding challenges, shaping prospects* (Vol. 44, No. 29). <http://www.ilo.org/global/publications/books/WCMS534326/lang?en/index.htm>
- Irani, L. C. y Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 611-620). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470742> doi: 10.1145/2470654.2470742
- Jem Bartholomew, D. M. (2023). *How the media is covering Chat-GPT*. CJR. <https://www.cjr.org/towcenter/media?coverage?chatgpt.php>.
- Joshua Broyde, C. P. (2021). *Build a medical sentence matching application using BERT and Amazon Sage-Make*. Amazon Web Services. <https://aws.amazon.com/es/blogs/machine-learning/build-a-medical-sentence-matching-application->
- Kalodimos, J. y Leavitt, K. (2020). Experimental shareholder activism: A novel approach for studying top management decision making and employee career issues. *Journal of Vocational Behavior*, 120, 103429. Disponible en <https://www.sciencedirect.com/science/article/pii/S0001879120300543> doi: <https://doi.org/10.1016/j.jvb.2020.103429>
- Lee, A. (2023). *What Are Large Language Models and Why Are They Important?* NVIDIA. blogs.nvidia.com.
- Lee, P., Bubeck, S. y Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, 388(13), 1233-1239.
- Merchant, B. (2023). Column: Minimum wage 'ghosts' keep Google and Microsoft's AI arms race from becoming a nightmare. *Los Angeles Times*.

- [https://www.latimes.com/business/technology/story/2023-02-16/column-google-microsoft-](https://www.latimes.com/business/technology/story/2023-02-16/column-google-microsoft)
- Metz, C. (2023). The Secret Ingredient of ChatGPT Is Human Advice. *New York times*. <https://www.nytimes.com/2023/09/25/technology/chatgpt-rlhf-human-tutors.html>.
- Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. y Fernández-Leal, Á. (2023). *Human-in-the-loop machine learning: A state of the art*. *Artificial Intelligence Review*, 56(4), 3005-3054.
- Nast, C. (2017). *Amazon's Turker Crowd Has Had Enough*. *Wired*. <https://www.wired.com/story/amazons-turker-crowd-has-had-enough/>.
- Nayak, P.(2019). Understanding searches better than ever before. *Google Blog*, 1–6. <https://blog.google/products/search/search-language-understanding-bert/>
- Nedzhvetskaya, N. y Tan J. T. (2019). What we learned from over a decade of tech activism. *The Guardian*. [https://www.theguardian.com/commentisfree/2019/dec/22/tech-worker-activism-2019-what-](https://www.theguardian.com/commentisfree/2019/dec/22/tech-worker-activism-2019-what)
- Nellis, S. (2023, nov 15) Microsoft introduces its own chips for AI, with eye on cost. *Reuters*. <https://www.reuters.com/technology/microsoft-introduces-its-own-chips-ai-with-eye-cost-2023-11-15/>
- Norman, D. (Julio 23, 2019). *The Four Fundamental Principles of Human-Centered Design and Application*. <https://jnd.org/the-four-fundamental-principles-ofhuman-centered-design-and-application/>
- Perrigo, B. (2023). OpenAI Used Kenyan Workers on Less Than 2Per-Hour. *Time*. [https://time.com/6247678/openai-chatgpt-kenya-](https://time.com/6247678/openai-chatgpt-kenya)
- Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, H. F. e Irving, G. (2021). *Scaling language models: Methods, analysis & insights from training gopher*. CoRR, abs/2112.11446 . <https://arxiv.org/abs/2112.11446>
- Roose, K. (2023). Bing's A.I. Chat: 'I Want to Be Alive. *New York Times*. <https://www.nytimes.com/2023/02/16/technology/bing-chat-bot-transcript.html>.
- Stanford University. (2023). Artificial Intelligence Index. *AI Index Report 2023*. Retrieved 2023-10-22, from <https://aiindex.stanford.edu/report/>
- Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., . . . Chris- tiano, P. (2020). Learning to summarize from human feedback.

- In Proceedings of the 34th international conference on neural information processing systems. *Red Hook*, NY, USA: Curran Associates Inc.
- Strubell, E., Ganesh, A. y McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. arXiv preprint arXiv:1906.02243.
- Thewsey, A. (2021). Bring structure to diverse documents with Amazon Textract and transformer-based models on Amazon SageMaker. *Amazon Web Services*. <https://aws.amazon.com/es/blogs/machine-learning/bring-structure-to-diverse-documents>
- Tomlinson, B., Black, R., Patterson, D. and Torrance, A.W. (2023, marzo 23). *The Carbon Emissions of Writing and Illustrating Are Lower for AI than for Humans*. SSRN. <https://ssrn.com/abstract=4399923>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017, jun). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem, 5999–6009. <http://arxiv.org/abs/1706.03762>
- Véliz, C. (2021). Privacidad es poder: Datos, vigilancia y libertad en la era digital. *Debate*.
- Zuboff, S. (2020). *La era del capitalismo de vigilancia*. Paidós.

