Conversando con una computadora: ¿Cómo entienden las inteligencias artificiales lo que les pedimos?

por Allan Mejía Berzunza

La evolución humana ha supuesto cambios no solo en lo que podemos hacer, sino en *cómo podemos hacer*. Más precisamente, la necesidad de eficientizar el trabajo ha dictado una aceleración exponencial en la tecnología; no es ninguna sorpresa ver la enorme cantidad de inventos y descubrimientos que han surgido a partir del siglo XIX, especialmente cuando los comparamos con los siglos anteriores (Gregersen, 2020). El sueño humano de la máquina que se mueve por sí sola es sumamente antiguo, con Herón de Alejandría no sólo ideando autómatas ya en la primera mitad del siglo I, sino habiendo documentado la existencia de otros ingenios similares que le precedieron (Ceccarelli, 2007): la prevalencia del término *autómata* siglos antes de las primeras computadoras mecánicas comprueba el latente anhelo de la humanidad por una máquina sirviente (Hockstein et al., 2007).

La forma en que las inteligencias artificiales actuales están cambiando la forma en que los humanos interactúan con la información, con el conocimiento acumulado de la especie, pero, como se explorará a lo largo de este artículo, enseñarle a una máquina a "hablar", a "entender", podría ser un proceso mucho más introspectivo de lo que parece pues al menos por el momento, nuestras creaciones no pueden superarnos en algo que nosotros mismos no entendemos del todo como lo es el lenguaje: desde la psicología cognitiva recurriendo al modelo computacional para el proceso de la información, hasta los modelos lingüísticos de inteligencia artificial, el camino por lograr que una máquina hable y piense está inesperadamente ligado al esfuerzo por entender en cómo el humano mismo habla y piensa.

Hoy más que nunca el término "inteligencia artificial" está llevando a la humanidad a cuestionarse cuánto realmente sabemos sobre nosotros mismos: ¿qué significa "significado"? ¿cuál es el significado de "crear"? A final de cuentas, por más impresionantes que parezcan las inteligencias artificiales, por ahora no son sino un paralelo de lo que sabemos sobre nosotros mismos: mientras no puede negarse la velocidad a la que una computadora puede realizar cálculos o comparar la memoria humana con la enorme cantidad de información a la que los modelos lingüísticos como GPT-3 tienen acceso, los procesos que realizan estas inteligencias artificiales para entender el lenguaje humano son en realidad imitaciones bastante burdas aun de la forma en la que opera el cerebro humano. Pero tampoco podemos negar el potencial que tienen las inteligencias artificiales de ayudarnos a comprender más sobre la forma en que nuestra propia mente funciona.

Breve historia de la comunicación humano-máquina

La evolución del cerebro humano le permitió realizar acciones cada vez más y más complejas, sin embargo, no puede negarse que fue la evolución del lenguaje lo que verdaderamente permitió a nuestros ancestros pasar del uso de herramientas para realizar trabajos y actividades que les permitían mantenerse con vida a la verdadera labor humana de comunicar, registrar y planear (Bickerton, 2009; Boeckx

Imagen recuperada de: https://www.freepik.es/foto-gratis/persona-que-usa-computadora-portatil-usando-inteligencia-artificial-generar-imagenes_74211749.htm#from_view=detail_serie

& Benítez-Burraco, 2014). A partir de ese momento, la sapiencia humana se dedicaría a buscar formas de hacer más con menos y en menos tiempo. Sin embargo, para 1770, año en que el escocés James Watt inventaba la máquina de vapor (Pennock, 2007), ese *menos* implicaba también *menos* humanos.

Pero tan eficientes como podían ser las máquinas, aun no podía removerse del todo el factor humano: era necesario realizar ajustes al proceso, aun se necesitaba reparar o modificar las máquinas para lo cual era necesario a alguien que entendiera cómo funcionaba la máquina y pudiera adaptarla a las necesidades de la situación. Dicho de otra forma: la comunicación con las máquinas se había vuelto el nuevo problema a superar. En 1725, Basile Bouchon inventaría lo que se considera la primera máquina semiautomatizada al utilizar tarjetas perforadas para controlar una máquina tejedora. El principio de las tarjetas perforadas seguiría en uso por más de 100 años gracias a Herman Hollerith, fundador de IBM, quien las ligara a aplicaciones de computación y almacenamiento de información (Kaur et al., 2014). Justamente, la necesidad de procesar y almacenar información (más precisamente de *computarla*) dirigiría los esfuerzos por lograr y mejorar la comunicación humano-máquina.

El siguiente gran paso en los esfuerzos por homologar el lenguaje de las máquinas con el lenguaje humano llegó durante la década de los cuarenta (del siglo XX), cuando la Segunda Guerra Mundial exigió una evolución en la forma en la que podía *ordenársele* a las nuevas máquinas de guerra: atrás habían quedado los tiempos en los que las piezas de artillería podían ser apuntadas por un pequeño equipo de hombres, pues las nuevas armas navales podían disparar mucho más allá del horizonte observable y golpear con precisión a un buque en movimiento. Lo único que se necesitaba era una forma de decirle a las armas *qué* hacer y *cuándo* hacerlo. Para esto se construyeron computadoras electromecánicas, máquinas con cientos de engranes y poleas que le permitían a los usuarios humanos comunicarle al sistema de armas la información del mundo real que la máquina entonces computaría y traduciría en las angulaciones necesarias de las baterías (Bureau of Ordnance, 1949).

Figura 1 Computadora de control de fuego de la USS New Jersey. Obsérvense las manivelas para introducir información sobre el buque y sobre su objetivo. Tomada de "Fire Control" en YouTube (Battleship New Jersey, 2020)

No pasaría mucho tiempo, sin embargo, para que los avances tecnológicos produjeran las primeras computadoras modernas y, a partir de ese momento, las computadoras han impulsado a la humanidad de formas cada vez más complejas y en campos cada vez más variados (Janssen et al., 2019): desde las tareas de comunicación, pasando por la automatización de las líneas de producción, hasta diagnóstico médico con ayuda de inteligencia artificial (Bi et al., 2019). No hay duda de lo que las computadoras han ayudado a la humanidad a lograr pero, aun hay un aspecto en la corta historia de la computación moderna que ha visto poco avance: la comunicación humano-máquina.

Los lenguajes de programación

Los lenguajes de programación surgieron

para establecer la comunicación humano-máquina con las entonces nuevas computadoras digitales, pues éstas ya no tenían perillas o palancas mediante las cuales introducirles información y, si bien el lenguaje binario comparte genealogía con las tarjetas perforadas[1], las computadoras digitales interpretaban y procesaban información mediante el lenguaje binario, es decir, en unos y ceros. El claro ejemplo de esta problemática fue la computadora BINAC, creada en 1949 por la compañía Echert-Mauchly, cuyos programadores eran aun matemáticos y físicos puros a los cuales se recurría por su dominio sobre expresiones lógicas. Al enfrentarse al problema del almacenamiento de la información y la necesidad de comunicarse con un dispositivo completamente digital, estos primeros programadores decidieron utilizar octal[2] para "abreviar" los comandos en binario que la computadora podía entender (Murray Hopper, 1981).

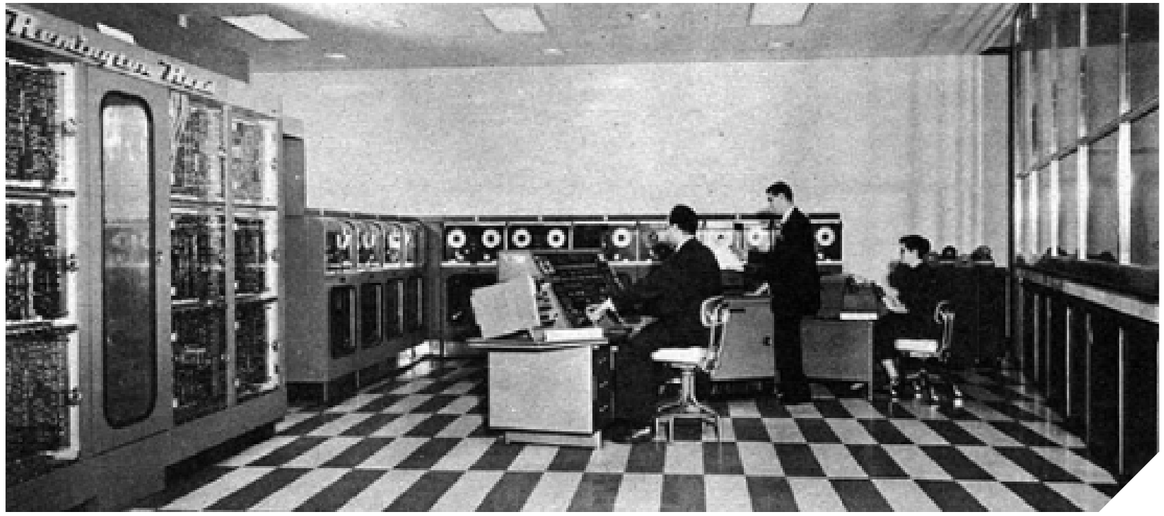


Figura 2 La UNIVAC I sólo podía almacenar el equivalente mil palabras y pesaba un poco más de 7 toneladas (Murray Hopper, 1981). Imagen tomada de "A third survey of domestic electronic digital computing systems" (Weik, 1961)

[1] La presencia o ausencia de una perforación en la tarjeta es similar a la presencia de un 1 o un 0 en binario (Kaur et al., 2014).
[2] Sistema de numeración base 8, utiliza los dígitos del 0 al 7.

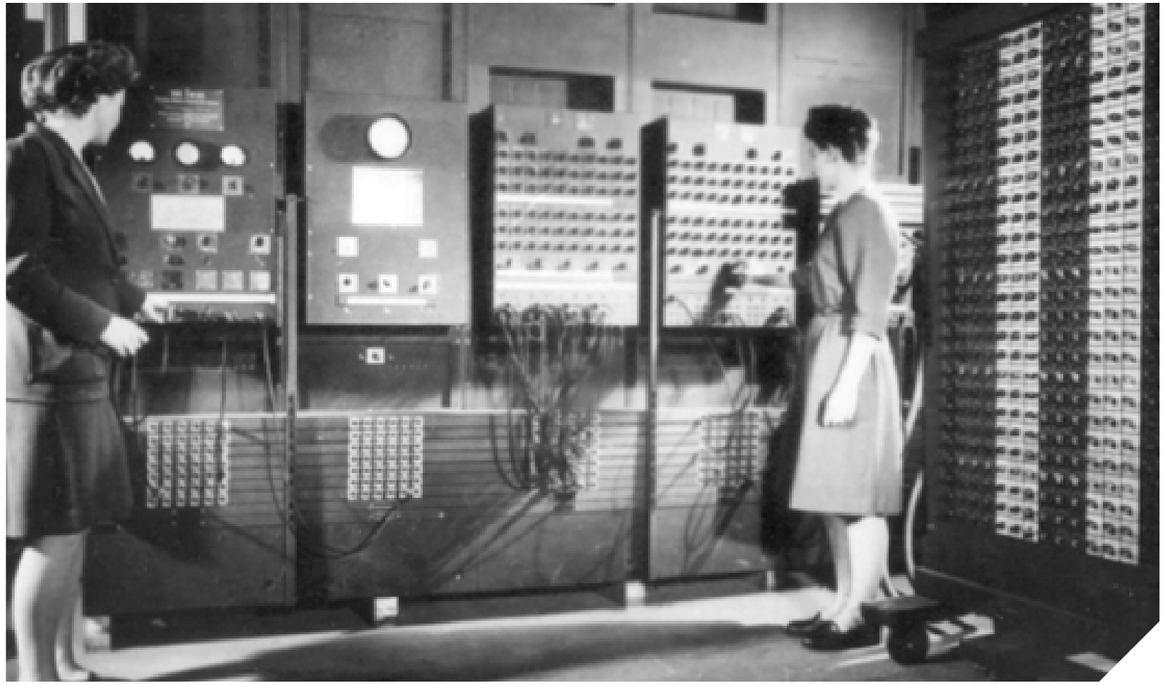


Figura 3 Si bien la ENIAC de 1945 aquí mostrada es considerada la primera computadora programable, ésta carecía de un verdadero lenguaje: su programación se basaba en recombinaciones de cables y bulbos para hacer que la computadora realizara tal o cual operación (Sammet & Holberton, 1981). Imagen tomada de los archivos de la ARL Technical Library.

La necesidad de establecer estándares vanguardistas, acelerada por la nueva organización del mundo tras la Segunda Guerra Mundial, daría lugar a los primeros programas computacionales; fue entonces que, mientras trabajaban en la BINAC, Grace Murray Hopper y su equipo comenzarían a establecer una serie de códigos que sobrepasaban el sistema algebraico que la BINAC utilizaba: ya no se introducían operaciones algebraicas, sino una serie de comandos que la computadora podía emparejar con las operaciones lógicas que

ésta “conocía” (Murray Hopper, 1981).

En la década de 1950 surgen las primeras menciones de “programación automática”, término que en ese entonces se utilizaba para referirse al uso de mnemónicos, una forma en la que computadora y humano podían comunicarse mediante un lenguaje intermedio: no se usaban literalmente ceros y unos, ni tampoco era necesario compilar las operaciones básicas cada vez que se creaba un programa, sino que podían recurrirse a palabras como “SUM” que la computadora entonces interpretaba como la operación matemática de adición. Los mnemónicos eran un lenguaje intermedio, pues no se estaba utilizando un lenguaje propiamente humano, aún tenía que prepararse la memoria de la computadora con aquellos procesos y operaciones lógicas que cada mnemónico habría de invocar además de que se tenía que entrenar a los operadores para sobre el inventario de mnemónicos que la computadora podía entender (Sammet & Holberton, 1981).

Estos lenguajes de programación primitivos funcionaban bien para tareas como la computación de datos y cálculos matemáticos, pero conforme la tecnología avanzaba y las computadoras se volvían más y más potentes, surgía la necesidad de estandarizar los lenguajes de programación: se perdía mucho tiempo y dinero cada que una nueva computadora se ponía en operación, pues la mejora en la tecnología implicaba necesariamente el cambio en su arquitectura y el establecimiento de un nuevo lenguaje ensamblador[3]. Entonces, aquellos programadores tempranos se propusieron crear una especie de *lingua franca* para la comunicación humano-máquina, un lenguaje que pudiera ser utilizado en la mayor cantidad de computadoras posibles sin importar su fabricante y que además fuera fácil de enseñar y aprender para los humanos que usarían dichas máquinas (Sammet & Holberton, 1981); comenzaba la historia de los lenguajes de programación modernos.

Desde la programación orientada a objetos hasta las aplicaciones del internet de las cosas[4], los lenguajes de programación han ido cambiando junto con las computadoras, por mencionar algunos ejemplos[5]: FORTRAN en 1954, ALGOL en 1958, COBOL en 1959, BASIC en 1964, Pascal en 1970, C en 1972, Java en 1995, Python en 1991 y C++ en 1998. Pero el avance de los lenguajes de programación se ha encontrado con una disyuntiva: un lenguaje de programación con el que es fácil escribir un programa, suele ser difícil de leer. Por ejemplo, crear un programa que pueda realizar cálculos con números complejos puede ser muy complicado si se utiliza un lenguaje que no esté específicamente destinado para realizar cálculos de ese tipo; tal es el caso de C++, que requeriría que el usuario “extendiera” el lenguaje para, lo que a su vez podría dificultar que otra persona pudiese entender el código a primera vista (Leendert, 1991).

En otras palabras, lejos de lo que aquellos primeros programadores imaginaban, los lenguajes de programación actuales no están mucho más cerca de permitir a los humanos comunicarse con las computadoras utilizando un lenguaje natural. Además, la verdadera dificultad de hacer que una computadora entienda lo que un humano dice está en el hecho de que aún no entendemos completamente la forma en que el lenguaje humano funciona, pues el mismo Alan Turing decía en 1951: “las computadoras no son sino una imitación del cerebro humano” (Copeland, 2004). Curiosamente, para proponer un esquema sobre la forma en que el cerebro humano procesa, almacena y compara la información que recibe del mundo, la psicología cognitiva se apoyaría en los conceptos que los primeros programadores dedujeron sobre el proceso de la información aun cuando la psicología cognitiva había surgido antes de la primera computadora digital (Leahey, 2003).

No es extraño entonces que los programas computacionales sean imitaciones de la forma en que el cerebro humano procesa la información para solucionar un problema, pues los humanos solemos establecer (muchas veces de manera inconsciente) una serie de pasos para realizar prácticamente cualquier tarea por cotidiana o extraordinaria que parezca: construimos algoritmos que dan solución a nuestros problemas. En ese sentido, un algoritmo es un conjunto de pasos o procedimientos que nos permiten alcanzar un resultado o resolver un problema (Cairó Battistutti, 2005), y esa es justamente la forma en que está construido o escrito un programa de computadora.

[3] **Los lenguajes ensambladores son lenguajes de programación ligados directamente a la arquitectura de las computadoras, por lo que hay un lenguaje único para cada computadora dependiendo su fabricante.**

[4] Internet of Things (IoT) es una tendencia en las ciencias de los sistemas de la comunicación que, durante la última década, ha logrado la interconexión de máquinas con otras máquinas, permitiendo una comunicación continua y discreta de información. Las aplicaciones del Internet of Things son cada vez más variadas y se estima que para el año 2025 habrán más de 50 mil millones de dispositivos conectados de esta manera (Farahmand, 2021)

[5] La lista aquí mostrada es sumamente general y contempla sólo los lenguajes debatiblemente más famosos

Inteligencia artificial y redes neurales

La búsqueda de la inteligencia artificial propiamente dicha no es tan moderna como se puede creer: en el siglo XVII, Thomas Hobbes proponía que "los pensamientos no son expresables en lenguaje escrito o hablado, sino en una dimensión interna" y que, por lo tanto, las operaciones lógicas que implica el *raciocinio* no están limitadas a las matemáticas, sino que son aplicadas a todo aquello que el individuo conoce (Haugeland, 1989). Alan Turing se basaría en ese concepto para responder la pregunta "¿puede una computadora pensar?", pregunta que lo llevaría a establecer el *Test de Turing* entre 1951 y 1952[6] para poder discernir si una computadora es capaz de pensar por sí misma. Pero Turing no pretendió en ningún momento definir el pensamiento, mucho menos la mente, simplemente establecer *criterios* que pudieran diferenciar el verdadero *raciocinio* de la "simple" computación de información (Copeland, 2004). Es aquí donde surge el primer punto a esclarecer: la inteligencia artificial actual no está ni cerca de la verdadera inteligencia, pues el pensamiento lógico-verbal, la solución de problemas como se plantea en la definición del algoritmo, no representa el pensamiento humano completamente[7]. Pero no por esto debe restársele valor a los esfuerzos conseguidos hasta ahora: en su estado actual, el concepto de *inteligencia artificial* se refiere a un programa computacional que recurre a *redes neurales* para aprender y procesar información.

Las redes neurales están basadas en la forma en la que las neuronas del cerebro humano se comunican entre sí para dar lugar a múltiples procesos cognitivos (Copeland, 2004). Cada "neurona" artificial está compuesta por una capa de input, una serie de capas ocultas, y una capa de output, esto quiere decir que cada capa recibe información que procesa según un sub-algoritmo que le dice qué hacer con esa información para después pasarla a la siguiente capa, eventualmente, la "neurona" completa el ciclo a lo largo de sus capas y pasa la información que ha procesado a la siguiente neurona.

Imagen recuperada de: https://www.pexels.com/es-es/foto/hombre-traje-jugando-mesa-8438918/

[6] Una buena parte de la teoría que complementa al Test de Turing no se publicó hasta 1999 (Copeland, 2004).

[7] Para Luria (1989), existe un problema básico para definir el pensamiento humano: las estructuras psicológicas que pretenden describir el pensamiento (como la propuesta de la psicología cognitiva) conciben el mismo como un todo, cuando en realidad, las funciones cognitivas complejas son difusas.

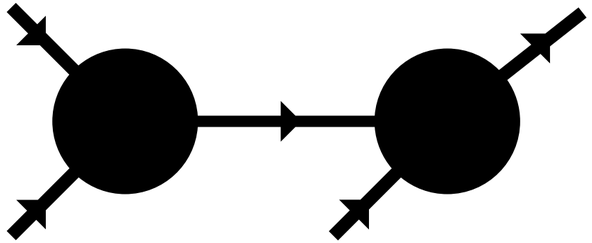


Figura 4 Ejemplo de dos "neuronas" en una red neural procesando información. Adaptada de "What are neural networks?" (IBM, 2021)

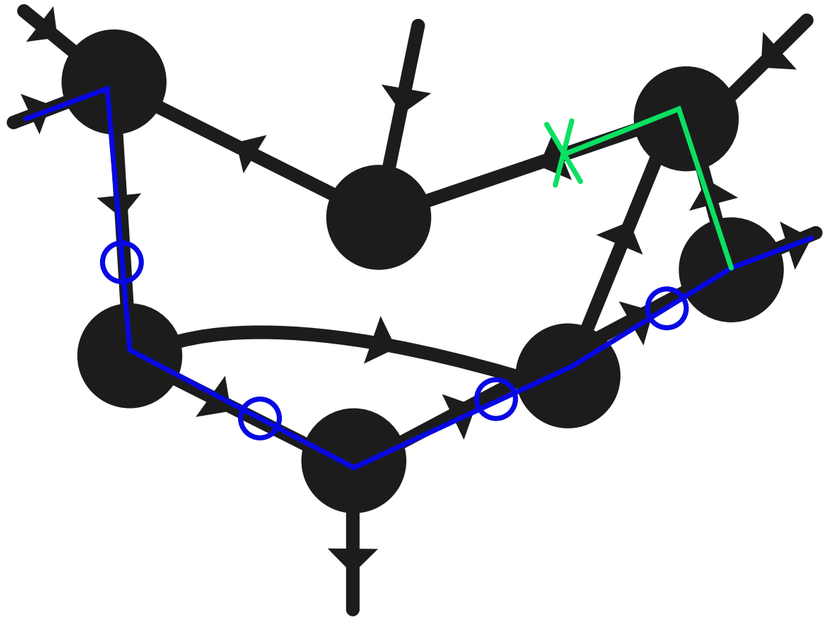


Figura 5 Ejemplo del flujo de la información a través de una red neural. Obsérvese cómo el flujo de datos en color azul continúa su camino por la red mientras que el flujo verde se interrumpe por no cumplir con los parámetros o "weights" establecidos en su programación. Adaptada de "What are neural networks?" (IBM, 2021)

El *cómo* y *cuándo* la información pasa de neurona a neurona depende, desde luego, de la programación de la red neural: cada conexión entre neuronas tiene un *weight* o parámetro, un valor numérico que se establece durante el entrenamiento de la red neural que controla qué tanto el *output* o producción de una neurona afectará el resultado final (IBM, 2021). Las redes neurales son el elemento clave de conceptos como *machine learning, deep learning* y desde luego *inteligencia artificial*. Pero a diferencia de un cerebro humano, el cual continuamente está procesando la información que percibe del exterior, comparándola con la información que tiene en los distintos niveles de memoria y realizando una enorme variedad de juicios y consideraciones para aprender (Leontiev et al., 2004), las redes neurales dependen de los programadores para decirles *qué* hacer con la información que reciben: los antes mencionados *weights*, si bien pueden ser aleatorios durante las primeras iteraciones de la red neural, eventualmente requieren que los programadores los ajusten para obtener resultados consistentes de la red neural (T. Brown et al., 2020; IBM, 2021; Kingma et al., 2021).

Una vez que el modelo comienza a generar *outputs* satisfactorios, los programadores pueden modificar ciertas partes de la red neural para que éste realice distintas acciones, esto es lo que hace la diferencia entre un algoritmo de red neural y el verdadero *deep learning*. Este término suele utilizarse, junto con *machine learning*, de manera intercambiable con *inteligencia artificial*, pero la acepción general es que una red neural con tres o más capas puede considerarse un algoritmo de *deep learning*, pues la palabra *"deep"* se refiere justamente a que el algoritmo tiene un nivel oculto en su red neural, la parte intermedia entre la parte de la red que únicamente prepara la información, la capa de *input*, y la parte que prepara la información para devolverla al usuario, la capa de *output* (IBM, 2021). A su vez, el término *machine learning* suele utilizarse para describir algoritmos de red neural que requieren una intervención más directa por parte de los humanos: mientras que un algoritmo *deep learning* tiene la capacidad de aprender, el *machine learning* no "aprende" a menos que sus desarrolladores explícitamente vuelvan a iniciar el proceso de entrenamiento, es decir, el deep learning es un

paso más cerca hacia la "verdadera" *inteligencia artificial* (IBM, 2020).

Inteligencia artificial y lenguaje humano

Una vez que se tiene la idea general del funcionamiento de las redes neurales es más sencillo apreciar cuán lejos estamos de crear una verdadera inteligencia artificial[8]: los asistentes virtuales (Siri o Alexa) e incluso los chatbots que tanto revuelo han causado últimamente (ChatGPT o Bing Chat), están aún muy lejos de una verdadera inteligencia (IBM, 2020). Hemos hablado de Thomas Hobbes y Alan Turing proponiendo lo que una máquina inteligente tendría que ser capaz de hacer, y a partir de esas propuestas, actualmente se consideran tres tipos de inteligencia artificial (Goertzel, 2014; IBM, 2020):

- **Artificial Narrow Intelligence, ANI:** sistemas que realizan comportamientos "inteligentes" en contextos muy específicos, es decir, pueden llegar a aparentar que se está tratando con un humano real en tanto no se sobrepase su configuración. El estado actual de las inteligencias artificiales se encuentra aún en este nivel.

- **Artificial General Intelligence, AGI:** en este nivel se encontrarían sistemas capaces de realizar una gran variedad de tareas para lograr objetivos en distintos contextos y ambientes, en otras palabras, una AGI podría anticiparse y reaccionar a situaciones y problemas que sus creadores no habrían considerado al momento de su programación inicial. Una AGI tendría ya un cierto nivel de conciencia, pues podría recurrir a sus experiencias para razonar un plan de acción que dé solución a una situación a la que esta inteligencia se encontrara. A manera de ejemplo, ya que no hay ningún sistema actualmente que esté siquiera cerca de este nivel, pueden nombrarse algunas inteligencias artificiales de la ficción: los robots en *"Yo robot"* de Asimov, GladOS de la franquicia de videojuegos *"Portal"* y HAL9000 de la película *"2001: odisea del espacio"*.

- **Artificial Super Intelligence, ASI:** si bien las AGI son teóricamente posibles, las ASI tendrían que superar la sapiencia humana. Desde luego los únicos ejemplos provienen de obras de ficción: La IA de la franquicia de películas *"The Matrix"* y Skynet de la franquicia de películas *"Terminator"*.

Tanto las AGI como las ASI se consideran "inteligencias artificiales fuertes" (IBM, 2020) y, si bien algunos autores consideran que actualmente estamos en el puente entre las ANI y las AGI, la principal limitante es bastante burda: No entendemos cómo funciona la mente humana, no podemos ni siquiera señalar a una parte del cerebro humano y declarar "aquí está la conciencia", por lo tanto, es irreal pretender que pudiésemos imitarla (Goertzel, 2014). Los avances actuales en IA no se han logrado por accidente, como se ha venido planteando a lo largo de este artículo, todo ha sido un largo camino y sería obtuso pretender que el siguiente paso en el desarrollo de la IA sucedería por accidente.

Pero entonces, si las IA no aprenden como los humanos y aun necesitan que sus programadores guíen ese aprendizaje, ¿cómo entienden las IA como ChatGPT, Dall-E o Stable Diffusion lo que les pedimos? Y realmente, los modelos de IA disponibles para el público al momento de escribir este artículo logran interpretar el tan elusivo "lenguaje humano" bastante bien. ChatGPT[9], por ejemplo, es una "IA conversacional que utiliza NPL[10] para generar respuestas parecidas a las que daría un humano (real)" (T. B. Brown et al., 2020). Esto quiere decir que ChatGPT utiliza una serie de redes neurales para entender y generar expresiones escritas: no es que ChatGPT entienda lo que le escribimos, al menos no en el sentido en el que un humano entiende. Cuando un humano observa una palabra escrita en su lengua materna, se dispara un complejo mecanismo cognitivo en su cerebro (pensamiento) que de forma simultánea recupera la pronunciación de dicha palabra (símbolo) y, particularmente, todo aquello que la persona relaciona con esa palabra (referente), producto de su aprendizaje, experiencia e ideología.

Este proceso cognitivo tan complejo ha sido

[8] En adelante "IA".

Imagen recuperada de: https://www.pexels.com/es-es/foto/mujer-tecnologia-flor-sujetando-8438979/
[9] GPT significa "Generative Pre-trained Transformer", literalmente "Transformador Generativo Pre-entrenado".
[10] Natural Language Processing.

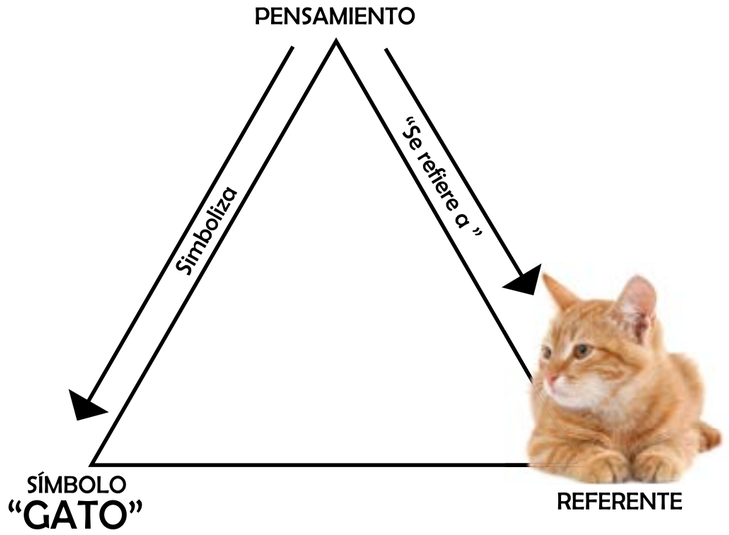


Figura 6 El triángulo de Ogden es uno de los diagramas más básicos para describir el proceso semántico. Adaptada de "Semantics" (Palmer & Frank Robert, 1981).

replicado, desde luego en una escala mucho menor, en la red neural que ordena a la IA en cuestión: primero se tiene que diseñar un algoritmo con la serie de pasos que la IA tendría que seguir al encontrarse con una producción escrita, ese algoritmo entonces involucra su respectiva red neural en la cual se han de asignar los parámetros o *weights*. Evidentemente esto es más complicado de lo que parece, pues aun siendo una recreación bastante simple del proceso cognitivo que toma lugar en el cerebro humano, el entrenamiento del GPT-3, el modelo en el cual se basa ChatGPT[11], tomó varios días ininterrumpidos de GPUs[12] procesando información de entrenamiento. Ahora bien, ¿qué significa "entrenamiento" en el contexto de las inteligencias artificiales? El entrenamiento es, literalmente, el proceso mediante el que los programadores de un modelo de IA le enseñan a las redes neurales a procesar la información, pero sobre todo a establecer una diferencia entre los resultados (outputs) deseables de aquellos considerados errores (Radford et al., 2021). Lo que realmente hace único al modelo GPT-3 es la enorme cantidad de parámetros que posee en su red neural: cerca de 175 mil millones de parámetros (T. B. Brown et al., 2020). Como se ha mencionado anteriormente, al comienzo del entrenamiento la red neural comienza con parámetros aleatorios, derivados de los algoritmos particulares de cada modelo y, según las necesidades u objetivos de los desarrolladores, nuevos parámetros se van añadiendo o se van modificando los existentes.

Modelos de IA como ChatGPT, DALL-E o

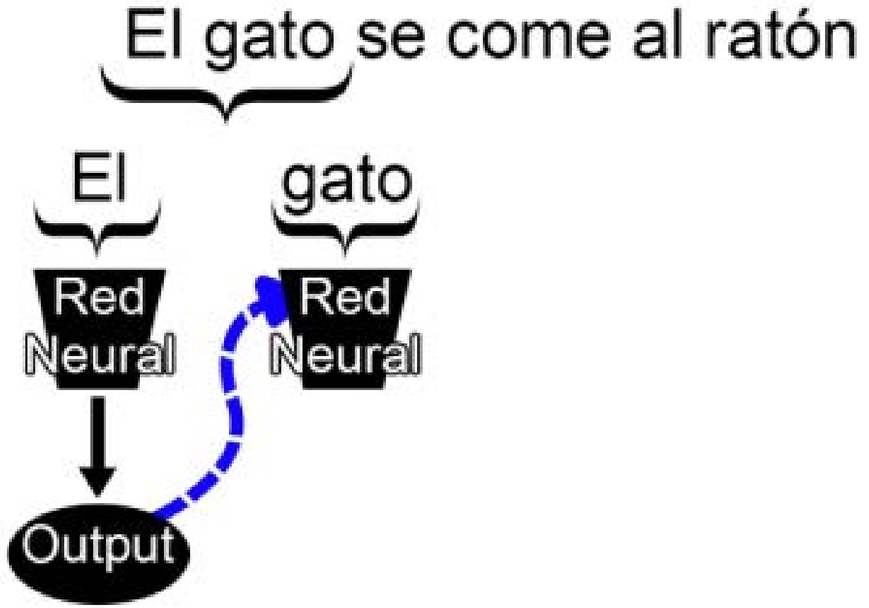


Figura 7 En este diagrama se explica, de forma bastante simplificada, la forma en que un modelo lingüístico como GPT-3 analiza producciones escritas. El proceso es muy similar al análisis sintáctico. (T. B. Brown et al., 2020)

[11] El 14 de marzo de 2023 OpenAI, la empresa responsable de ChatGPT, liberó oficialmente GPT-4 para sus usuarios de paga, no hay información clara, pero se asume que la versión gratuita sigue utilizando GPT-3.

[12] Graphic Processor Unit, componentes principales de las tarjetas de video.

Stable Diffusion[13] están causando mucho revuelo al momento de escribir este artículo porque no sólo entienden lo que sus usuarios les piden, sino que pueden predecir y aprender sin necesidad de que sus programadores intervengan. Los modelos lingüísticos como ChatGPT, por ejemplo, fueron entrenados oponiendo conocimiento "real" con muestras de lenguaje que los programadores utilizaron para "enseñarle" a analizar construcciones escritas: una vez que la red neural logró entender esas construcciones muestra, se fueron añadiendo más capas o modificando las existentes[14] (T. B. Brown et al., 2020; Vaswani et al., 2017), de tal forma que los chatbots que están actualmente a la vanguardia pueden hacer mucho más que los bots tradicionales: las redes sociales y los proveedores de correo electrónico llevan varios años utilizando bots[15] para detectar spam, traducir mensajes e incluso detectar mensajes radicales u ofensivos, mientras que chatbots como ChatGPT o Bing Chat pueden identificar el contexto, las ideas claves e incluso detectar jerga o palabras características de un corpus particular (Kublik & Saboo, 2022). Lo que hace la diferencia entre los bots de detección de spam y los chatbots antes mencionados son los *transformers*, un tipo de red neural propuesto por investigadores de la universidad de Toronto (Kublik & Saboo, 2022) que, habiendo aprendido a identificar las funciones gramaticales de las palabras en distintas muestras de lenguaje durante su entrenamiento, se le ha programado para aplicar lo aprendido en nuevas muestras, que en otras palabras significa que estas redes neurales pueden aplicar lo que sus desarrolladores les han enseñado, mediante horas y horas de ensayo y error, para analizar producciones escritas completamente nuevas; es así como los modelos lingüísticos no solo "entienden" lo que se les pide sino que también pueden producir respuestas a partir de esa expresión inicial con base en los datos duros a los que tiene acceso además de lo que aprendió durante la fase de entrenamiento (T. B. Brown et al., 2020; Douglas Heaven, 2020). Para aclarar un poco más este proceso obsérvese la Figura 7, donde la flecha azul representa la forma en que el *transformer* le permite a la IA comparar lo que sabe sobre la palabra "El" y compararla con el resto de las palabras en una secuencia. En la figura antes mencionada se ha obviado que cada palabra de la oración pasa por el mismo proceso que las primeras dos: los transformers, las redes neurales que como se dijo anteriormente son características de los modelos lingüísticos, procesan las palabras de una producción escrita de manera simultánea, lo que les evita la necesidad de utilizar memoria física para guardar el texto completo durante todos los pasos el proceso y además, evita la interferencia de este durante los procesos intermedios. Esto es lo que caracteriza a los modelos lingüísticos que forman el núcleo de chatbots como ChatGPT o Bing Chat, pues al poder analizar los componentes de un texto dado de manera simultánea pueden entonces aplicar una fórmula matemática para calcular las palabras con mayor probabilidad[16] de aparecer en el texto, a partir de lo cual, pueden incluso predecir la continuación de este (Vaswani et al., 2017). Esto le permite a ChatGPT, por ejemplo, producir una respuesta aceptable pese a no tener conocimiento sobre lo que el usuario le ha preguntado o solicitado; esto, combinado con la masiva red neural que compone el GPT-3, hace que ChatGPT siempre tenga una respuesta "válida".

Para explorar la capacidad de atención y predicción de ChatGPT se realizaron algunos experimentos simples (ver figura 8): para provocar una respuesta incoherente o extraña, se le presentó a ChatGPT una frase en español, lengua que ChatGPT puede "entender" mediante *machine translation*, es decir, la IA traduce la frase que le introducimos (“a mí me gusta el tangananá”) mediante el proceso de análisis antes descrito, y logra predecir el significado de la frase e incluso generar una respuesta escrita en español; sin embargo, como puede apreciarse en la figura, la IA no tiene forma de saber qué significa "tangananá". Continuando con el experimento se introdujo la frase "Pese a la negativa de prensa covfefe" (ver figura 9), una traducción al español de un *tweet* publicado por Donald Trump que de inmediato se volvió viral por su carencia de sentido. En este caso, si bien "covfefe" no tiene significado alguno se esperaba que, por la viralidad del evento, ChatGPT tuviera un registro en su memoria sobre el mismo, de tal forma que la IA pudiera entender el contexto de la frase introducida. Como puede apreciarse

[13] DALL-E y Stable DIffusion son modelos de generación de imágenes a partir de texto, dichos modelos serán explorados más adelante en este artículo.

[14] Los procesos de entrenamiento de los principales modelos de AI no se han revelado de manera detallada por razones de propiedad intelectual.

[15] En este contexto la palabra "bot" se refiere a un algoritmo de machine learning.

[16] Se utilizan fórmulas de cálculo probabilístico que el algoritmo puede aplicar múltiples veces en múltiples pasos del proceso del análisis, desde luego la fórmula es única para cada modelo.

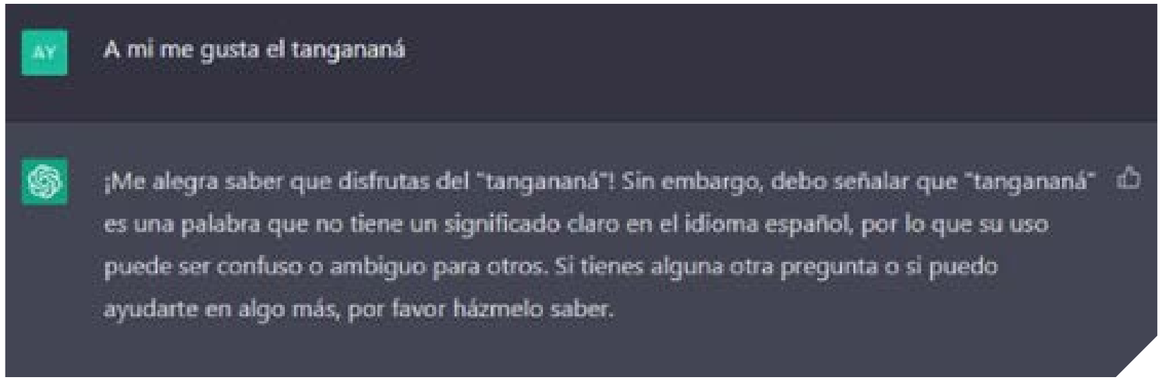


Figura 8 En esta "conversación" con ChatGPT se ha introducido un fragmento de la canción "Tangananica, Tangananá" del programa chileno "31 minutos"

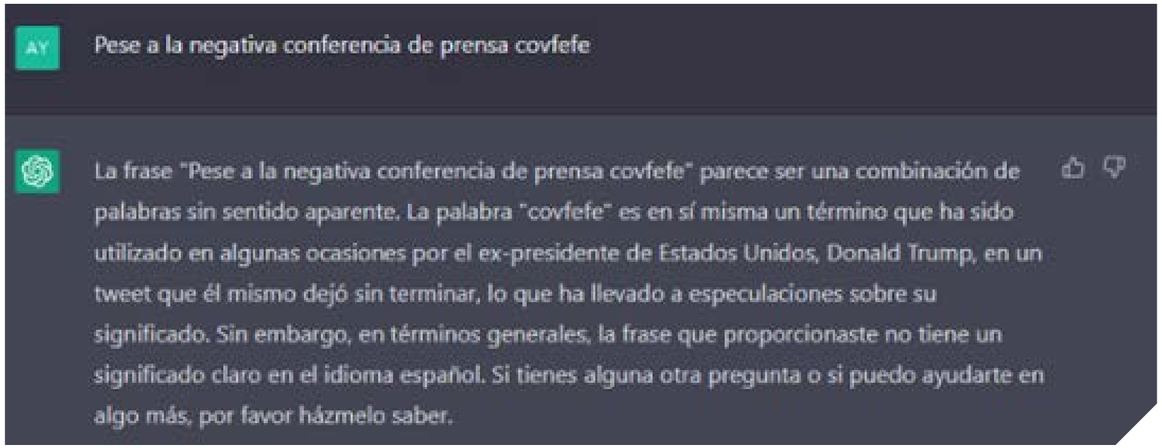


Figura 9 El tweet original leía "Despite the constant negative press covfefe", se asume que fue un error de dedo, pero de inmediato se volvió objeto de burlas y parodias.

en la respuesta generada por el chatbot, éste no entiende la frase, pero logra identificar la palabra "covfefe" y ligarla a información dura en su memoria para dar una respuesta coherente.

Estas dos pruebas, si bien simples, sirven para mostrar lo que los desarrolladores han denominado un "motor de atención": una red neural con parámetros y capas que le permiten a la IA no solo identificar el contexto de una producción y compararlo con la información "dura" que tienen en su memoria, sino a identificar las palabras clave de una construcción, lo que le permite predecir palabras afines mediante un análisis sintáctico único, pues a diferencia de los humanos, la IA no entiende realmente lo que es un verbo o un adjetivo, sino que los compara con valores (scores) que le ha asignado a los componentes de las oraciones usadas durante el periodo de prueba, que en otras palabras significa que la IA identifica la función gramatical de una palabra dependiendo su posición en el texto y comparándola con lo que aprendió durante la fase de entrenamiento (Kublik & Saboo, 2022; Singh, 2023; Vaswani et al., 2017).

El modelo sucesor de GPT-3, GPT-4, se puso al alcance del público el 14 de marzo de 2023. En el reporte oficial no hay una mención clara de lo que hace distintos a GPT-3 de GPT-4, pero se alude a que se han agregado más parámetros (weights), añadiendo varias capas a la red neural. Además, de acuerdo con OpenAI[17] (OpenAI, 2023b), GPT-4 ha sido entrenado directamente con textos referentes a exámenes y cursos estandarizados de una gran variedad de ramas, algunas de las cuales incluyen:

- **SAT Math**
- **Medical Knowledge Self-Assessment Program**
- **AP Art History**
- **AP Biology**
- **AP Chemistry**
- **AP Language and Composition**
- **AP Macro and Microeconomics**
- **Advanced Sommelier**
- **AP Psychology**
- **AP Statistics**

GPT-4 ha logrado obtener puntajes por encima del promedio humano en casi todas las pruebas y exámenes relativos a estos campos. A diferencia de GPT-3, GPT-4 posee

[17] Empresa detrás de GPT y DALL-E.

una cantidad masiva de información que ha estudiado extensivamente durante horas y horas de entrenamiento (OpenAI, 2023b). Pero más impresionante aun es la capacidad de GPT-4 de "entender" imágenes, describirlas, predecir contenido relativo a dichas e incluso producir contenido específico a una rama del conocimiento (véase figura 10). El potencial de este nuevo modelo lingüístico todavía está por verse[18] y no cabe duda de que la competencia con Google y Microsoft impulsará la inteligencia artificial a niveles sorprendentes.

Inteligencias artificiales de generación de imágenes a partir de texto

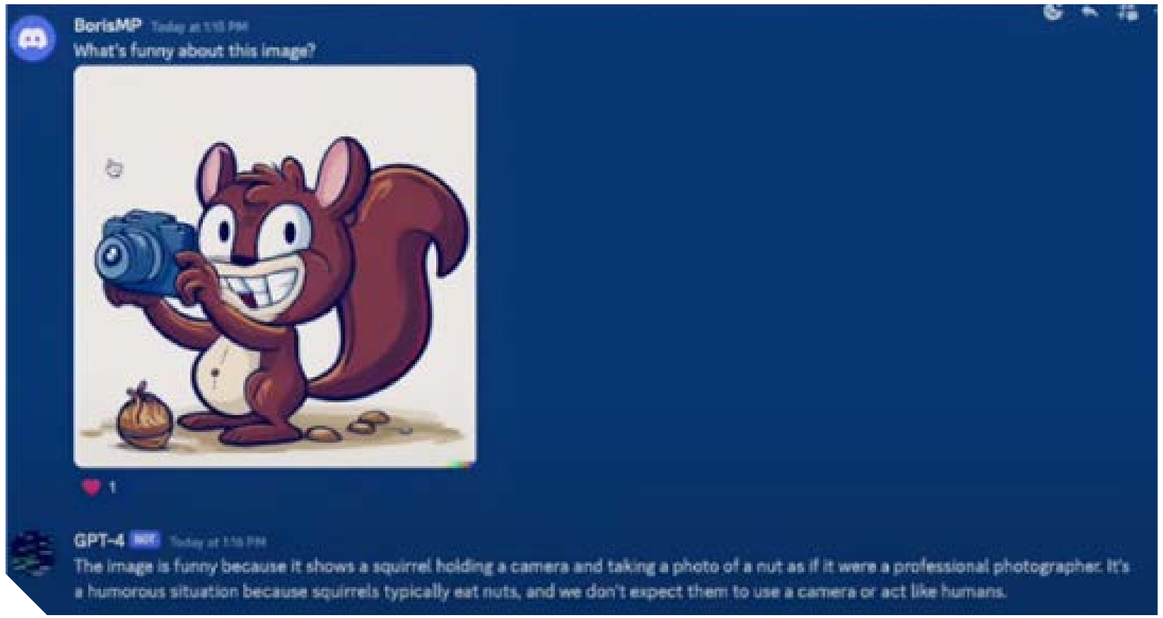


Figura 10 Se le presenta una imagen a GPT-4 y se le pregunta "¿Qué hace chistosa a esta imagen?". La IA entonces analiza la composición de dicha imagen y es capaz de formular una explicación relativamente satisfactoria. Tomada de "GPT-4 Developer Livestream" (OpenAI, 2023a)

El proceso semántico que realizan las inteligencias artificiales como GPT-3 está revolucionando la forma en que los humanos acceden y hacen uso de la información, pero hay también inteligencias artificiales que no sólo pueden interpretar texto, sino que además pueden generar imágenes a partir de ese texto. El principio es el mismo que con los modelos lingüísticos como GPT-3: el texto introducido por el usuario pasa por una red neural con distintos parámetros y genera un *output* de vuelta al usuario, sin embargo, en este caso el output ha de ser una imagen, por lo que hay una serie de pasos extras bastante sustanciales.

Hazañas como la de GPT-4 siendo capaz de extraer información de una fotografía de un dibujo en una servilleta y generar un output sumamente específico a partir de lo que logra predecir de dicha información son sólo una muestra de lo que inteligencias artificiales de generación de imágenes a partir de texto pueden hacer. Actualmente hay dos modelos generativos que han cautivado al mundo: DALL-E y Stable Diffusion, ambos con claras diferencias en su diseño. Comenzaremos por describir la forma en que DALL-E funciona por tratarse de un modelo lingüístico parecido a GPT-3 que ya ha sido explorado anteriormente en este artículo.

El primer problema al momento de conectar lenguaje con imágenes es, como sugiere la figura 6, el siguiente paso en la arquitectura de los *transformers*, la arquitectura de red neural que caracteriza a los modelos lingüísticos. En otras palabras, GPT-3 es muy bueno para identificar, generar y predecir lenguaje escrito, pero no puede ni generar imágenes ni obtener información de imágenes. Ni siquiera GPT-4 que puede "leer" imágenes logra completar esta tarea tan bien como lo haría un modelo expresamente diseñado para esta tarea (Goodfellow et al., 2014; Jay Wang, 2021). Así como los modelos lingüísticos deben ser entrenados para interpretar texto, DALL-E tuvo que ser entrenado para interpretar una imagen, a este proceso se le llama *codificación*. El problema es que, a diferencia de los humanos, una inteligencia artificial no asocia lo visual con lo lingüístico, sino que el lenguaje escrito y las imágenes suponen dos dimensiones completamente distintas: se necesita crear una nueva red neural que le permita a la IA "traducir"

[18] En el livestream "GPT-4 Developer Livestream" (OpenAI, 2023a) disponible en YouTube, Greg Brockman, president y co-fundador de OpenAI le pide a GPT-4 que genere el código java necesario para crear un sitio web que se dibujó en una servilleta.

una imagen en una expresión lingüística a partir de la cual se pueda generar otra imagen, todo esto simplemente para la etapa de entrenamiento.

La *codificación* de imágenes ya lleva varios años siendo utilizada por bots moderadores en redes sociales: se entrenó a las redes neurales poco profundas a detectar patrones de pixeles[19] característicos de contenido sensible que la plataforma en cuestión decide rechazar por defecto. El problema, de manera similar a los bots de detección de spam y traducción, es que estas redes neurales no aprenden, no se adaptan y mucho menos entienden las implicaciones lingüísticas de aquello que analizan. Una de las principales propuestas que marcaría la evolución de las inteligencias artificiales de generación de texto serían los *Visual Auto Encoders* (VAE), redes neurales que convierten una imagen dada en información que la IA pueda analizar mediante *transformers* (Goodfellow et al., 2014; Van Den Oord et al., 2017). Los VAE generan lo que los desarrolladores denominan un "espacio latente", un punto intermedio entre la codificación y la decodificación; en este espacio latente el VAE recurre a un *codebook*, una especie de vocabulario que los desarrolladores programan en el VAE, de tal forma que éste tiene que asignar partes de la imagen inicial a categorías en ese *codebook* (véase figura 11). Es un proceso muy largo que consume una enorme cantidad de recursos y necesita una gran cantidad de GPUs trabajando en conjunto, pero se descubrieron dos ventajas principales (Van Den Oord et al., 2017):

1. Los VAE no necesitan decodificar la imagen input pixel por pixel sino en de 32 por 32 pixeles (tiles o azulejos) y, dependiendo de la capacidad de proceso disponible, se pueden usar múltiplos de esas dimensiones (64x64, 128x128, 256x256, etc.).

2. Se le pueden agregar capas extra al VAE que le permitan mejorar la calidad de la imagen antes de devolverla en el output final (véase figura 12).

Es importante mencionar que DALL-E no utiliza un VAE de forma explícita, OpenAI no explica a detalle cómo funciona este modelo, pero se cree que recurre a un proceso similar durante su entrenamiento, pues OpenAI afirma que este recurre al GPT-3, lo que les permitió entrenar el modelo más rápido y establecer el proceso de decodificación como un proceso discreto (Razavi et al., 2019; Van Den Oord et al., 2017). En otras palabras, lo único que DALL-E tiene que hacer es asignarles categorías a los tiles resultantes de la decodificación de la imagen input[20].
DALL-E no está disponible para el público, al menos no del todo, pues se trata más de una prueba

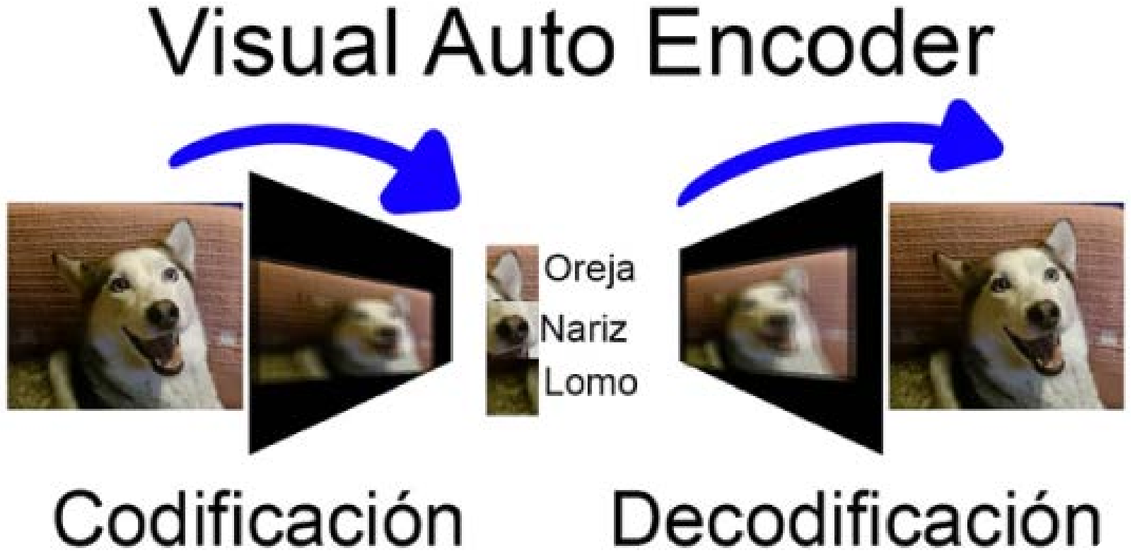


Figura 11 Este diagrama pretende ejemplificar, de una forma muy simple, lo que los VAEs hacen: durante la fase de entrenamiento se ajustan los parámetros de la red neural de tal forma que el VAE pueda reconstruir la imagen original. Adaptada de "Neural discrete representation learning", (Van Den Oord et al., 2017)

Imagen recuperada de: https://www.pexels.com/es-es/foto/mano-oscuro-modelo-textura-7562089/
[19] Una imagen digital está conformada por pixeles, pequeñas extensiones cuadradas que poseen un cierto brillo y color.

[20] Desde luego, por ser parte de lo que hace a DALL-E único, OpenAI no explica este proceso a detalle.

Figura 12 Una de las debilidades de los VAE es que, como si de una fotocopia se tratara, la imagen output suele ser borrosa o tener ruido. Izquierda: Imágenes input. Derecha: Imágenes output. Tomada de "Neural discrete representation learning" (Van Den Oord et al., 2017)

de concepto que de un modelo con una verdadera aplicación como, digamos, GPT-3 o como se verá más adelante, Stable Diffusion. OpenAI ha reconocido que los resultados de DALL-E presentados en sus artículos y reportes de estado están específicamente seleccionados de entre cientos (potencialmente miles) de imágenes generadas carentes de sentido (Jay Wang, 2021), pues, aunque este modelo es capaz de generar imágenes con buena calidad, además de lograr interpretar indicaciones escritas que podrían considerarse imposibles o carentes de sentido (véase figura 13).

Pero hay otro problema: si bien es relativamente sencillo obtener información con la cual

Imagen recuperada de: https://www.freepik.es/foto-gratis/persona-que-usa-herramienta-ai-trabajo_60362997.htm#fromView=search&page=2&position=27&uuid=38fdfeb7-d3bc-4a62-b740-7ef9cc84f970

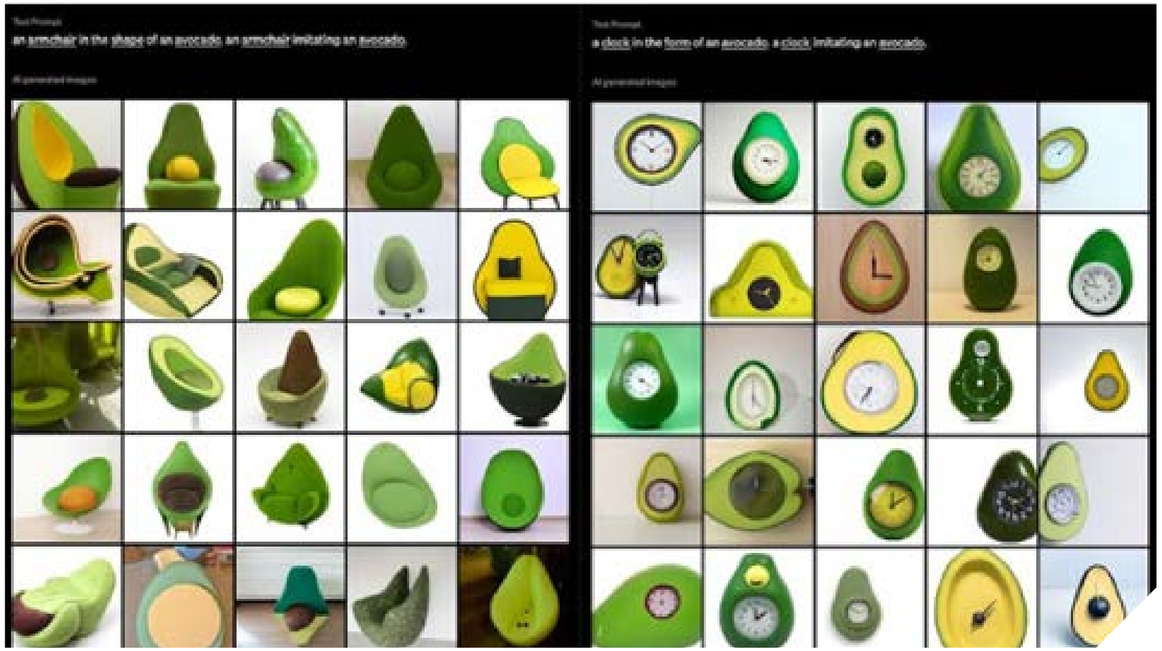


Figura 13 DALL-E es capaz de generar imágenes a partir de indicaciones (prompts) complejos. Sin embargo, debe recordarse que OpenAI ha seleccionado estos resultados de entre miles de imágenes generadas con el mismo prompt. Tomada de "DALL·E: Creating images from text" (Jay Wang, 2021).

entrenar y establecer la base de conocimiento en modelos lingüísticos como GPT-3, los modelos generativos de imágenes como DALL-E necesitan, desde luego, muchas imágenes con las cuales aprender. Más importante aún, cada imagen necesita ir acompañada de una descripción más o menos detallada para que la IA pueda aplicar su entrenamiento y obtener resultados consistentes. A los conjuntos de imágenes acompañadas de texto se les denomina *datasets*, y a lo largo de los años compañías como Microsoft, Amazon y Google han construido sus propios datasets[21], los cuales han sido instrumentales en el entrenamiento de modelos como DALL-E. Como se dijo anteriormente, este entrenamiento

[21] Algunos ejemplos de datasets popularmente usados para entrenar inteligencias artificiales son ImageNet con 14 millones de imágenes (Stanford Vision Lab, 2020); Microsoft COCO con cerca de 330 mil imágenes (Lin et al., 2014) y CIFAR-10 que contiene imágenes de 32x32 pixeles con 10 distintas categorías de objetos (Krizhevsky, 2017).

ha sido mucho más lento y problemático que el de los modelos lingüísticos que no generan imágenes, pues aún cuando se logran ajustar los parámetros de un modelo de manera satisfactoria, no hay ninguna garantía de que el modelo mantenga su consistencia al exponerlo a un dataset nuevo ya que, a diferencia de los humanos, las inteligencias artificiales carecen de un motor cognitivo que enlace experiencias, conocimientos e ideologías con los estímulos del exterior (Luria, 1989), por lo que la IA puede llegar a realizar descripciones extrañas (véase figura 14). Greimas (1987) explicaba (en humanos, desde luego) con su propuesta de los ejes semánticos: el mecanismo semiótico no se da por adición y substracción, sino por una concatenación de lo que el sujeto asume o conoce. Curiosamente, la forma en que las inteligencias artificiales de generación de imágenes como DALL-E parecen realizar un proceso similar al propuesto por Hjelmslev que es más mátematico (Bigot, 2010), aunque no hay menciones directas a Hjelmslev en la literatura referente a DALL-E, Stable Diffusion, VAEs o clasificadores, se asume que la aproximación hjelmsleviana se adapta mejor a las fórmulas probabilísticas que componen el motor lógico de las redes neurales de estas inteligencias artificiales.

A finales de 2022, Stability AI publicó el modelo 1.5 de Stable Diffusion en la página github.com[22]. Aunque Stable Diffusion es comúnmente conocido como un generador de imágenes a partir de texto, este modelo no utiliza un modelo lingüístico como lo hace DALL-E, en su lugar recurre a una combinación de distintos modelos como VAEs y CLIP[23] para ejecutar la síntesis de imágenes y, sobre todo, Stable Diffusion recurre a un proceso que añade ruido a la imagen durante varios pasos

Imagen recuperada de: https://www.freepik.es/foto-gratis/persona-que-usa-computadora-portatil-usando-inteligencia-artificial-generar-imagenes_74211758.htm#fromView=search&page=1&position=32&uuid=5d33c179-76c8-4ffd-974e-a58aee08e432

[22] En diciembre de 2022 la misma compañía liberó el modelo 2.0.

[23] Contrastive Language-Image Pre-Training" (Radford et al., 2021).

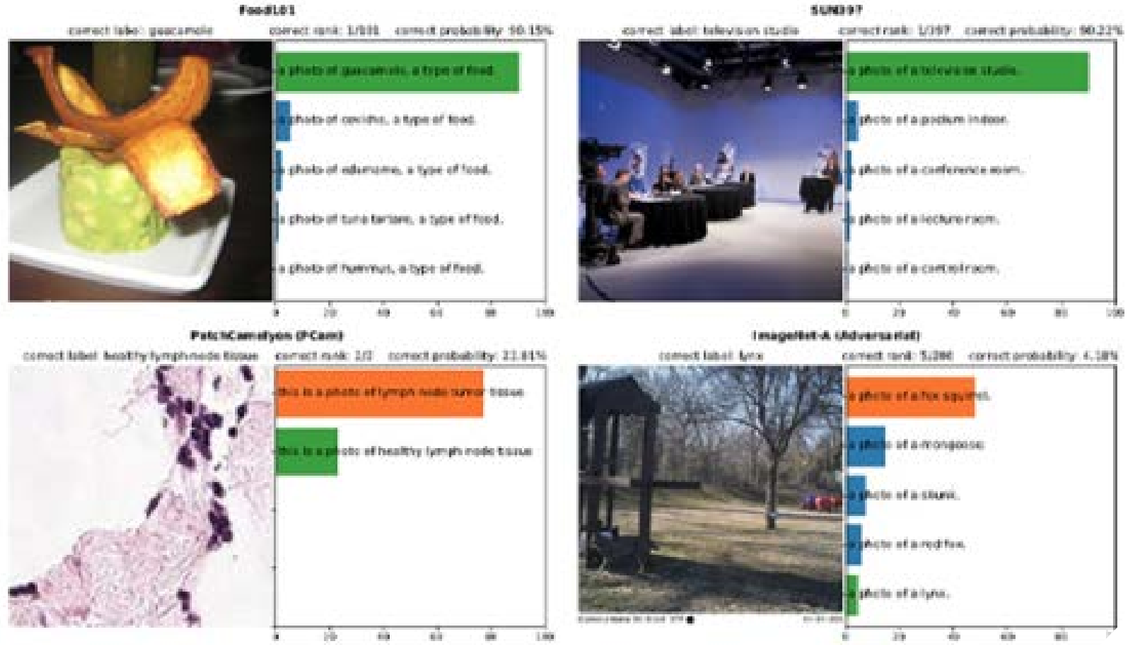


Figura 14 Radford et al. (2021) llevaron a cabo pruebas con su clasificador CLIP, un tipo de red neural que busca poder generar descripciones textuales para imágenes dadas sin necesidad de supervisión humana, con distintos datasets. El nombre del dataset usado se lee en negritas. Las barras verdes señalizan una descripción aceptable y las naranjas una incorrecta o sin sentido. Tomada de "Learning Transferable Visual Models From Natural Language Supervision" (Radford et al., 2021).

(veáse figura 15). Por esta razón este modelo no entiende realmente lo que sus usuarios le piden mediante la caja de indicaciones o prompts (veáse figura 16) como lo haría DALL-E, pues incluso cuando Stable Diffusion recurre a VAEs[24], el proceso de difusión inversa mediante el que este modelo genera imágenes no involucra procesos lingüísticos (Ho et al., 2020; Kingma et al., 2021; Nichol & Dhariwal, 2021; Song et al., 2020).
Para compensar esto, Stability AI ha hecho públicos tanto los parámetros generales de la red neural como

[24] Stable Diffusion utiliza checkpoints, variaciones de su entrenamiento base que pueden omitir la necesidad de VAEs.

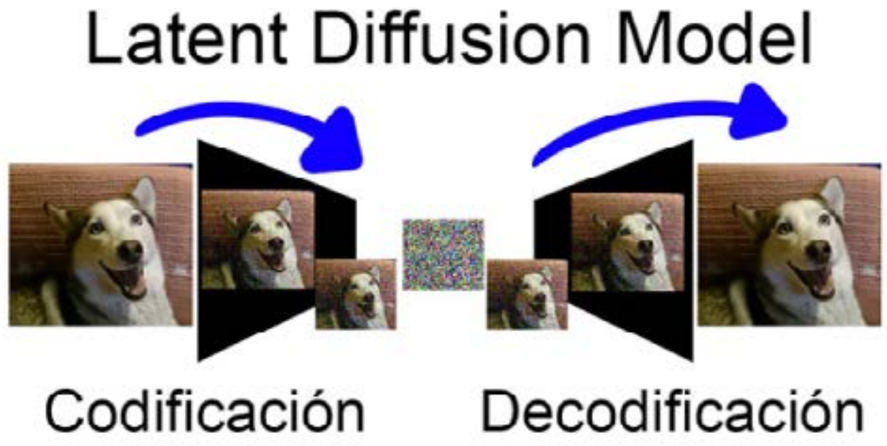


Figura 15 Los modelos de difusión latente (LPM) como Stable Diffusion agregan "ruido" a la imagen input poco a poco hasta tener sólo ruido, a partir del cual reconstruyen la imagen mediante un algoritmo probabilístico.

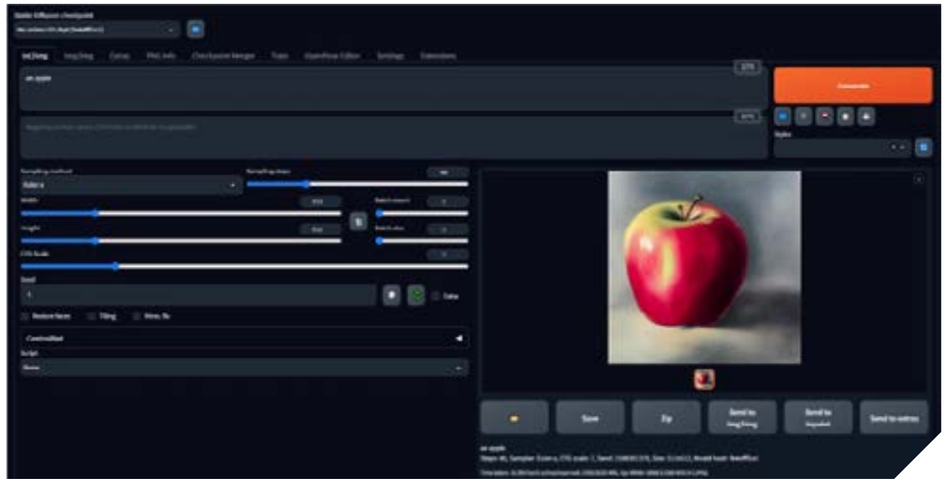


Figura 16 Una de las interfaces gráficas disponibles para Stable Diffusion. Obsérvense las dos cajas de texto en la parte superior.

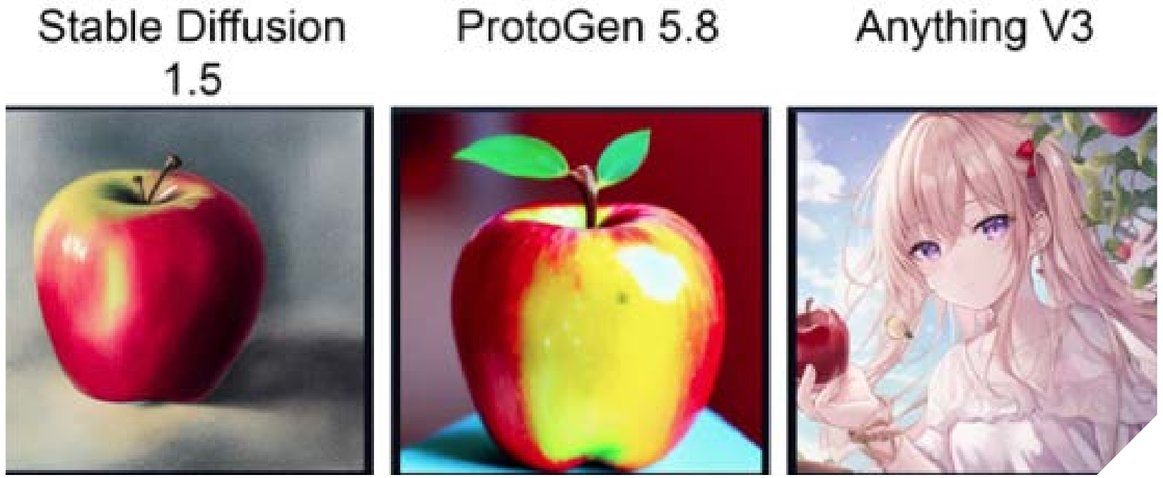


Figura 17 Al introducir la frase "an apple" en la caja de instrucciones, Stable Diffusion genera resultados distintos dependiendo el checkpoint utilizado. De izquierda a derecha: El checkpoint base. ProtoGen 5.8 enfocado en fotografía. Anything V3, enfocado en ilustraciones estilo manga/anime. Un ejemplo curioso del proceso semántico incompleto que la IA lleva a cabo.

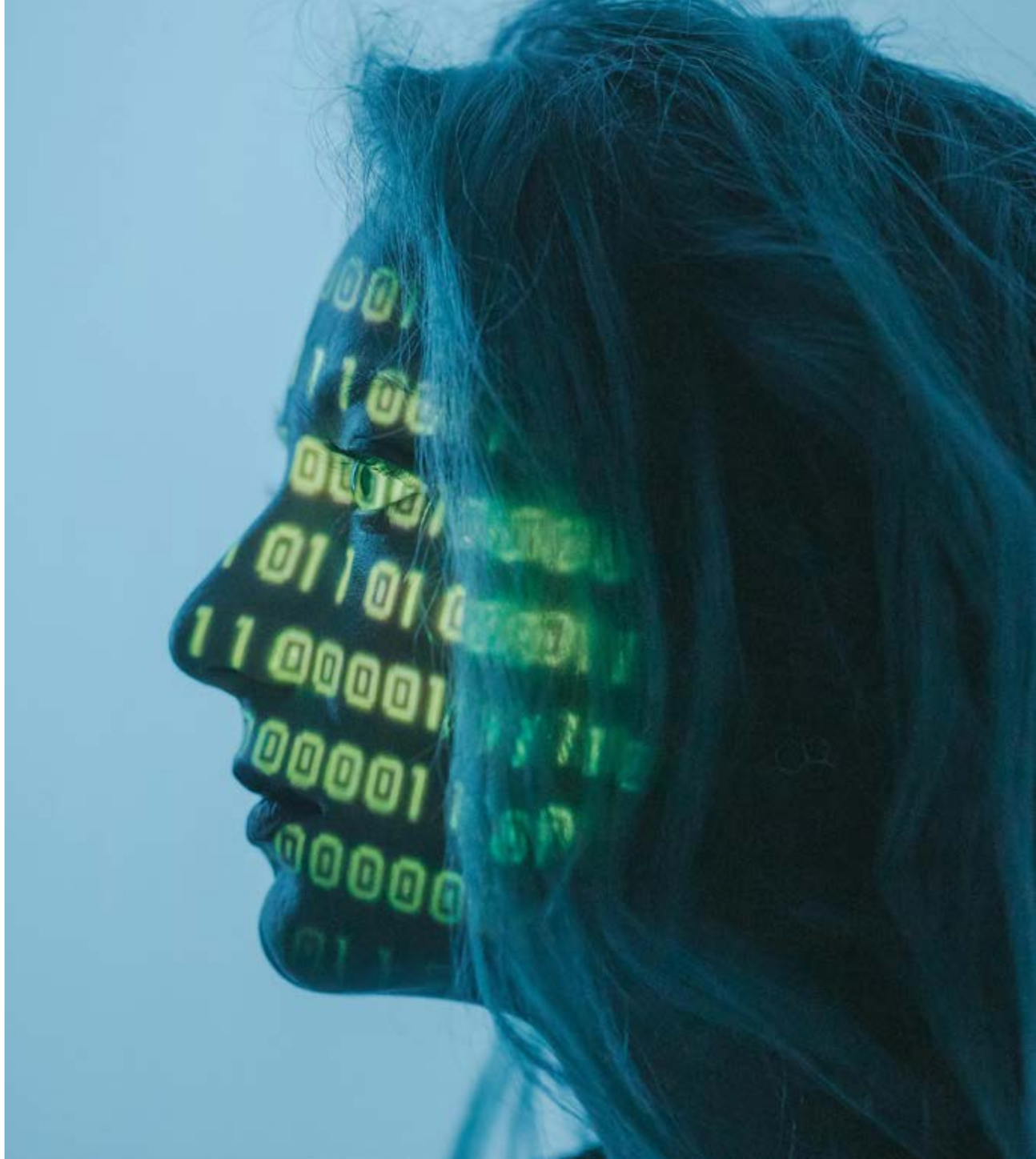
el checkpoint de su entrenamiento[25]. Un checkpoint puede entonces tomarse como base para re-entrenar el modelo con un cierto tipo de imágenes (Rombach et al., 2021), creando a su vez un nuevo checkpoint sumamente eficaz para generar un cierto tipo de imagen, pero sólo ese tipo de imagen (véase figura 17). Stability AI considera que esto es el punto fuerte de su modelo, pues ha sido entrenado con un dataset con un fuerte enfoque en muestras de arte, desde pinturas e ilustraciones hasta modelos tridimensionales, lo que le brinda una gran versatilidad al modelo para generar "imágenes hermosas" (Stability AI, 2022). Esto ha causado una gran controversia por haberse comprobado que Stability AI entrenó su modelo con los trabajos de artistas digitales sin su consentimiento (Vincent, 2023), pero sobre todo ha traído una importante pregunta a la mesa: ¿puede una inteligencia artificial "crear" arte?

Hoy más que nunca el término "inteligencia artificial" está llevando

[25] La comunidad de usuarios suele referirse a los checkpoints como modelos, lo cual puede llegar a provocar confusiones pues el modelo es Stable Diffusion, los checkpoints son información de entrenamiento.

a la humanidad a cuestionarse cuánto sabemos realmente sobre nosotros, sobre lo que nos hace humanos: ¿qué significa “significado”? ¿cuál es el significado de “crear”? y, quizás más importante aún, ¿puede una computadora “crear”? Es fácil perderse en la sorprendente capacidad de modelos como DALL-E o Stable Diffusion para no sólo entender lenguaje natural sino además generar imágenes que responden a una descripción textual dada; pero la forma en que estos modelos están revolucionando nuestra realidad va más allá de si las inteligencias artificiales reemplazarán a los humanos en actividades como programación o incluso ilustración: ¿podrían las inteligencias artificiales ayudarnos a entender nuestra propia mente? Así como muchos artistas, con justa razón, se han motivado a emprender acciones legales contra modelos que han sido entrenados explícitamente para replicar su estilo, a muchos otros les ha despertado un cuestionamiento por explorar cómo pueden las inteligencias artificiales elevar su esfuerzo artístico pues, de forma similar a como GPT-4 está ahorrando horas de trabajo tedioso y mecánico a los programadores que a su vez les permite enfocarse en aspectos verdaderamente creativos o hasta experimentales de su labor, las inteligencias artificiales podrían liberar la mente del artista para elevar su intención artística que, sin duda alguna, seguirá siendo una labor exclusivamente humana por bastante tiempo.

Imagen recuperada de: https://www.pexels.com/es-es/foto/futuro-robot-futurista-ia-8386365/ y https://www.pexels.com/es-es/foto/mujer-serio-codificando-hembra-5473956/



Referencias

Battleship New Jersey. (2020, August 31). Fire Control. Https://Www.Youtube.Com/@BattleshipNewJersey. https://www.youtube.com/watch?v=szxNJydEqOs

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., & Aerts, H. J. W. L. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. CA: A Cancer Journal for Clinicians, 69(2), 127–157. https://doi.org/10.3322/CAAC.21552

Bickerton, D. (2009). Adam's tongue: how humans made language, how language made humans. Macmillan.

Bigot, M. (2010). Apuntes de lingüística antropológica. Facultad de humanidades y artes de la universidad nacional de Rosario.

Boeckx, C. A., & Benítez-Burraco, A. (2014). The shape of the human language-ready brain. Frontiers in Psychology, 5, 282.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 2020-December. https://arxiv.org/abs/2005.14165v4

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), Advances in Neural Information Processing Systems (Vol. 33, pp. 1877–1901). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Bureau of Ordnance. (1949). Ordnance Pamphlet 1140: Basic Fire Control Mechanisms. Bureau of Ordnance.

Cairó Battistutti, O. (2005). Metodología de la Programación (3rd ed.). Alfaomega.

Ceccarelli, M. (2007). Distinguished figures in mechanism and machine science: Their contributions and legacies. Springer.

Copeland, B. J. (2004). The essential turing. Clarendon Press.

Douglas Heaven, W. (2020). OpenAI's new language generator GPT-3 is shockingly good—and completely mindless. MIT Technology Review. https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

Farahmand, F. (2021, October 22). Innovations and challenges in developing Medical Devices | Dr. Farid Farahmand. GHTC 2021 Plenary Talk.

Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. Journal of Artificial General Intelligence, 5(1), 1–48. https://doi.org/doi:10.2478/jagi-2014-0001

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. Communications of the ACM, 63(11), 139–144. https://doi.org/10.48550/arxiv.1406.2661

Gregersen, E. (2020). History of technology timeline. Encyclopedia Britannica.

Greimas, A. J. (1987). Semántica Estructural: Investigación metodológica. (A. De La Fuente, Trad.). Gredos.

Haugeland, J. (1989). Artificial intelligence: The very idea. MIT press.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 2020-December. https://doi.org/10.48550/arxiv.2006.11239

Hockstein, N. G., Gourin, C. G., Faust, R. A., & Terris, D. J. (2007). A history of robots: From science fiction to surgical robotics. Journal of Robotic Surgery, 1(2), 113–118. https://doi.org/10.1007/S11701-007-0021-2/FIGURES/6

IBM. (2020, June 8). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? IBM's Cloud Blog. https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

IBM. (2021, December 16). What are neural networks? Ibm.Com.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. International Journal of Human-Computer Studies, 131, 99–107. https://doi.org/10.1016/J.IJHCS.2019.05.006

Jay Wang, J. (2021, January 5). DALL·E: Creating images from text. OpenAI Research. https://openai.com/research/dall-e

Kaur, R., Kumar, P., & Singh, R. P. (2014). A Journey of digital storage from punch cards to cloud. IOSR Journal of Engineering, 4(3), 36–41.

Kingma, D. P., Salimans, T., Poole, B., & Ho, J. (2021). Variational Diffusion Models. Advances in Neural Information Processing Systems, 26, 21696–21707. https://doi.org/10.48550/arxiv.2107.00630

Krizhevsky, A. (2017, September). The CIFAR-10 dataset. Alex Krizhevsky. https://www.cs.toronto.edu/~kriz/cifar.html

Kublik, S., & Saboo, S. (2022). GPT-3: Building Innovative NLP Products. O'Reilly Media.

Leahey, T. H. (2003). Cognition and learning (D. K. Freedheim & I. B. Weiner, Eds.; Vol. 1, pp. 86–134). John Wiley & Sons.

Leendert, A. (1991). C++ for programmers. Wjohn Wiley & sons.

Leontiev, A., Luria, A. R., & Vigotsky, L. S. (2004). Psicología y pedagogía. Ediciones Akal.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8693 LNCS(PART 5), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Luria, A. R. (1989). El cerebro en acción. (M. Torres, Trad.). Ediciones Roca.

Murray Hopper, G. (1981). Keynote Address. In R. L. Wexelblat (Ed.), ACM SIGPLAN History of Programming Languages Conference (pp. 5–20). Academic Press.

Nichol, A., & Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models. https://doi.org/10.48550/arxiv.2102.09672

OpenAI. (2023a, March 14). GPT-4 Developer Livestream. YouTube.

OpenAI. (2023b). GPT-4 Technical Report. https://arxiv.org/abs/2303.08774v2

Palmer, F. R., & Frank Robert, P. (1981). Semantics. Cambridge university press.

Pennock, G. R. (2007). James Watt (1736–1819). Distinguished Figures in Mechanism and Machine Science: Their Contributions and Legacies Part 1, 337–369.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/arxiv.2103.00020

Razavi, A., den Oord, A., & Vinyals, O. (2019). Generating diverse high-fidelity images with vq-vae-2. Advances in Neural Information Processing Systems, 32.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022-June, 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042

Sammet, J. E., & Holberton, B. (1981). COBOL Session. In R. L. Wexelblat (Ed.), ACM SIGPLAN History of Programming Languages Conference (pp. 199–278). Academic Press.

Singh, D. (2023, March 19). Understanding Large Language Models - The Force Behind chatGPT. The Growth Catalyst Newsletter. https://www.growth-catalyst.in/p/tech-simplified-understanding-large

Song, J., Meng, C., & Ermon, S. (2020). Denoising Diffusion Implicit Models. https://doi.org/10.48550/arxiv.2010.02502

Stability AI. (2022). Stable Diffusion Launch Announcement. Stability.Ai. https://stability.ai/blog/stable-diffusion-announcement

Stanford Vision Lab. (2020). About ImageNet. Image-Net.Org. https://image-net.org/about.php

Van Den Oord, A., Vinyals, O., & others. (2017). Neural discrete representation learning. Advances in Neural Information Processing Systems, 30.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 2017-December, 5999–6009. https://arxiv.org/abs/1706.03762v5

Vincent, J. (2023, January 16). AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit. The Verge. https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart

Weik, M. H. (1961). A third survey of domestic electronic digital computing systems (Issue 1115). Ballistic Research Laboratories.