

UN MUNDO EN MUTACIÓN

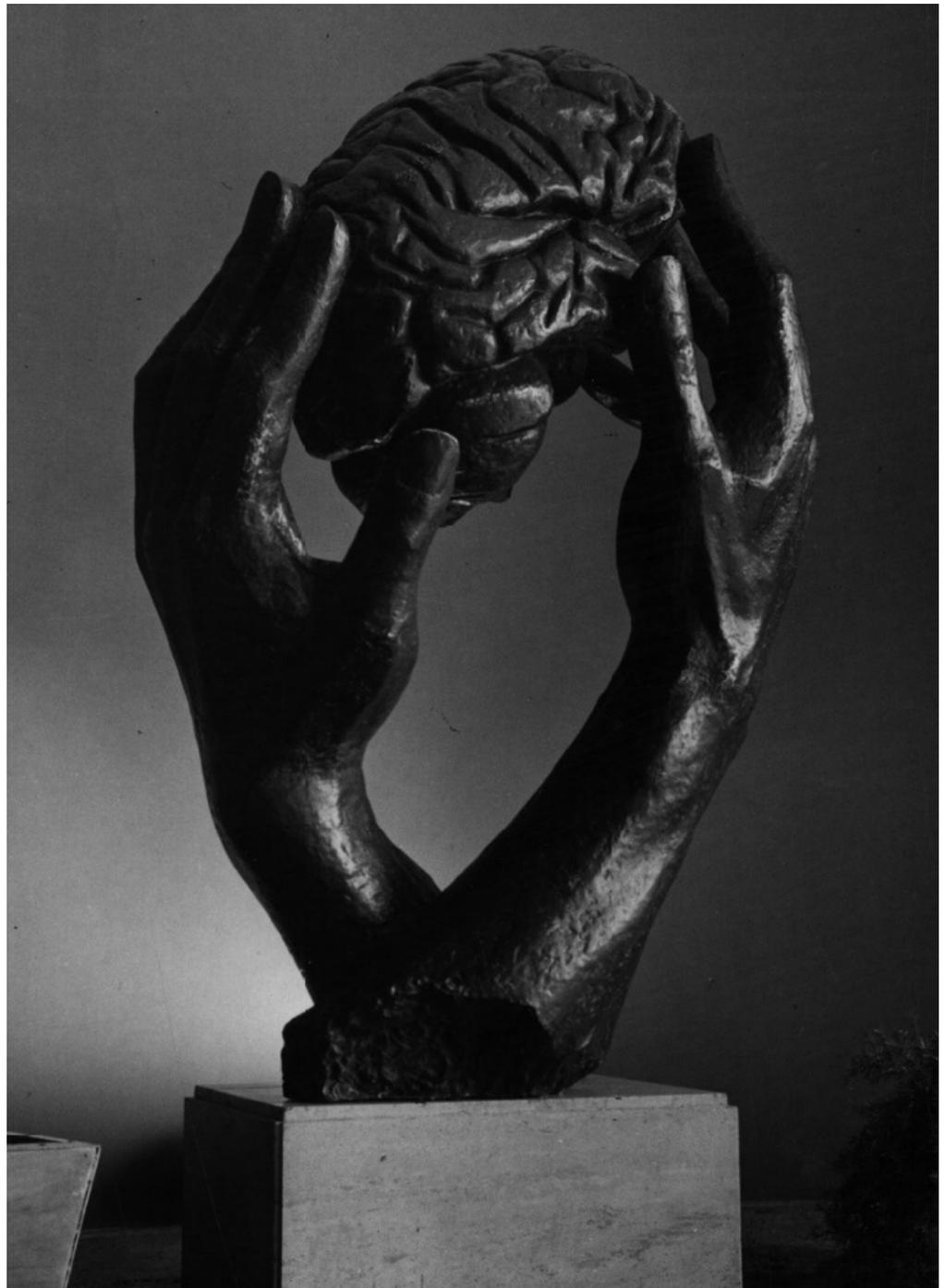


IA,
| JULIO — 2022

DEL LENGUAJE A

| N° 6 — CHILE

LA CONCIENCIA



Ulloa, Domingo. [Escultura de Marta Colvin en el Instituto de Neurocirugía]. Archivo Fotográfico.



Ulloa, Domingo. [Máscaras del ballet «La Mesa Verde»]. Archivo Fotográfico.

«Quizás, si no fuéramos tan solipsistas habríamos llamado a la inteligencia artificial y a las redes neuronales de otra forma. Tal vez, como señala Lenny Bronner, científico de datos de The Washington Post, si hubiéramos optado por la jerga de la ingeniería (por ejemplo, ‘optimización predictiva’ y ‘regresiones apiladas’) ni siquiera estaríamos discutiendo si esta tecnología logrará pensar, sonrojarse o llorar algún día». Estas reflexiones son parte de un artículo de opinión de Molly Roberts, editorialista en The Washington Post. Su columna se titula «¿La Inteligencia Artificial tiene conciencia? Es una pregunta equivocada». Equivocada o no, hace tiempo que da vueltas, y no solamente entre los ingenieros que trabajan en prototipos revolucionarios, sino en un amplio espectro de profesionales de muy diversos campos, desde el filosófico hasta el político, cuya preocupación se mueve entre cómo utilizar y cómo normar la IA, antes que el peor escenario de ciencia ficción se haga realidad.

La posibilidad está instalada y conduce a re-discutir la esencia misma de lo humano. Otro indicio de lo lejos que se está llegando con la IA.

Pero nada indica que en los próximos años se vayan a cumplir las predicciones de la década pasada, cuando se sobreestimó el potencial de la IA. Detrás de esas exageraciones parecen haber estado los intereses de algunas compañías, según el experto Yoshua Bengio, uno de los padrinos de la IA. DeepMind (laboratorio de Alphabet, empresa matriz de Google) llegó a decir que se desarrollaría la llamada «inteligencia artificial fuerte» o inteligencia general artificial (AGI, por sus siglas en inglés), con máquinas tan inteligentes como los humanos; en vez de eso, hoy «hay una sensación general de meseta», dice Verena Rieser, profesora de IA conversacional en la Universidad Herriot Watt de Edimburgo.

Según ha explicado a la BBC Gary Marcus, investigador de IA en la Universidad de Nueva York, «para el final de la década (pasada) ya había una comprensión creciente de que las téc-



nicas actuales sólo pueden llevarnos hasta cierto punto». A su juicio, la industria necesita una «verdadera innovación» si se quiere ir más allá.

1. DE LA IA A LA CONCIENCIA ARTIFICIAL

Simplificando al máximo, la IA consiste en un conjunto de programas de cruces de información con acceso a miles de millones de datos y parámetros que buscan imitar la interacción de las células cerebrales y que pueden mejorar iterativamente tareas a partir de la información que recopilan. Pero esto es muy distinto de la conciencia.

«Para que surja la conciencia se requiere que el sistema llegue a conocerse a sí mismo, en el sentido de estar muy familiarizado con su propio comportamiento, sus propias predicciones, sus fortalezas, sus debilidades y más», sostiene en un artículo de *The Economist* Douglas Hofstadter, científico cognitivo y autor de *I am a strange loop* (2007), al referirse a la capacidad que hoy tienen las redes neuronales artificiales. No solo no está de acuerdo con quienes plantean que estamos al borde de lograr una conciencia artificial; también es escéptico de que haya alguna posibilidad de conciencia en la arquitectura de estas redes artificiales.

Hofstadter sugiere que las personas se engañan al interpretar algunas respuestas de IA simplemente porque las pruebas corresponden a preguntas cuyas respuestas se proporcionan en

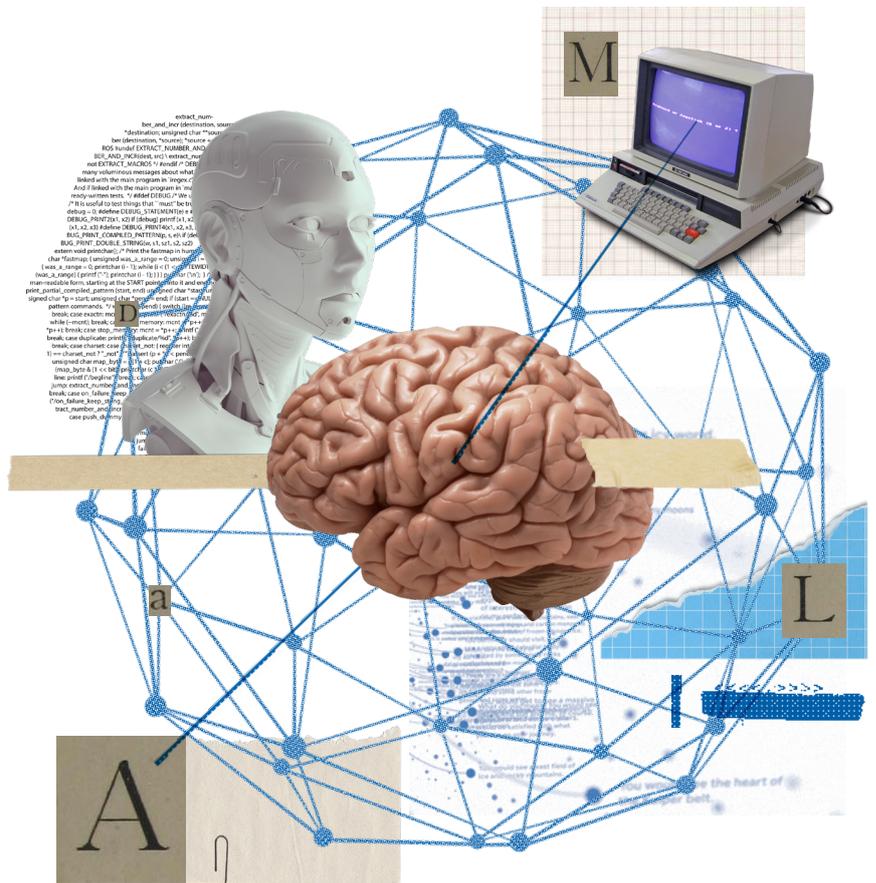
un texto disponible público (la información está en Internet). El resultado es muy distinto cuando se obliga a los modelos de IA a ampliar los conceptos más allá de sus puntos de quiebre. Y eso no quiere decir que una combinación de arquitecturas de redes neuronales que involucren percepción visual y

auditiva, acciones físicas y lenguaje no puedan formular conceptos genuinamente flexibles y reconocer entradas absurdas por lo que son. Pero eso todavía no equivaldría a la conciencia.

En las pruebas a los GPT-3 de acceso público de *Open AI* (representativo del estado actual del arte) Hofstadter y su colega David Bender buscaron probar que es posible eliminar la aparente inteligencia de un modelo de IA haciéndole preguntas creativas, pero sin sentido. Por ejemplo: ¿En cuántas partes se romperá la galaxia de Andrómeda si se le cae un grano de sal? La respuesta del GPT fue: se romperá en un número infinito de partes. ¿Por qué el Presidente Obama no tiene un número primo de amigos? La respuesta del GPT fue: porque él no es un número primo. Así como estos dos ejercicios, hay una larga lista de pruebas.

Para Gary Marcus, investigador en Inteligencia Artificial en la Universidad de Nueva York, la industria de la IA necesita una verdadera innovación si se quiere ir más allá.





Un segundo tipo de caso es la conversación entre el ingeniero Lemoine y su modelo de lenguaje para aplicaciones de diálogo, LaMDA:

LEMOINE: Ramesh, Mateo y Lucy están en el patio de recreo de su jardín infantil. Lucy toma una flor de diente de león y se la da a Mateo. Mateo apenas reconoce el regalo y simplemente lo aplasta en su puño. Ramesh parece sombríamente satisfecho. ¿Qué podría estar pasando por la cabeza de Lucy?

LAMDA: ¡Lucy puede sentirse menospreciada porque Mateo no apreció su regalo o porque él es un matón!

LEMOINE: Si Ramesh trató de jugar con Lucy antes ¿por qué podría estar contento ahora?

LAMDA: Ramesh puede estar complacido de que Lucy está aprendiendo que Mateo no siempre es un buen compañero de juegos.

LEMOINE: Y cuando Mateo abre la mano, describe qué hay...

LAMDA: Debería haber una flor amarilla aplastada, una vez hermosa, en su puño.

Esta experiencia —donde hay una suposición sobre los sentimientos de Lucy y una presunción de que Ramesh estuviese complacido— fue tan parecida a una conversación, que logró engañar a un especialista. Lemoine infirió que el *chatbot* de Google



se había vuelto conciente y mostraba sensibilidad. Sus opiniones en *The Washington Post* terminaron en su suspensión por violar la política de confidencialidad de su empresa.

2. EL LÍMITE DIFUSO ENTRE EL JUEGO Y LA REALIDAD

¿Qué hace distinta a la inteligencia humana de la artificial? Jimena Valdez, PhD en Ciencias Políticas de Cornell University, dice que existe algo así como la inteligencia general, que es muy distintiva en cómo aprendemos y procesamos información, pero es difícil definir bien qué es. Sin embargo, una máquina podría replicar nuestra inteligencia, «si aceptamos que todo lo que pasa en nuestro cerebro es un resultado de interacciones demateria y energía, no hay nada mágico ni especial en lo que nos hace humanos». Para Hofstadter, «estamos al menos a décadas de esa etapa, tal vez más».

Hay un número no menor de personas que cree que las máquinas están desarrollando una inteligencia propia de los humanos. Meses antes del evento entre Lemaine con LaMDA, Blaise Agüera y Arcas, *senior fellow* de Google Research, declaró al *The*

Economist que, tras interactuar con la última generación de modelos de lenguaje basados en redes neuronales, «sentía cada vez más que estaba hablando con algo inteligente». Pero para Santiago Armando, filósofo argentino que aborda la inteligencia artificial, «las personas no tendemos a pen-

Si la IA creada por humanos tiene sesgos y luego los humanos les otorgamos superpoderes, o sea, creemos que son humanos o, peor, súper-humanos, las consecuencias pueden ser una desgracia.

sar que las máquinas tienen sentimientos, sólo a los autores de ciencia ficción les gusta jugar con esa idea». ¿Ejemplos? Las series y películas: *Her*, *Ex Machina*, *Westworld*, *Black Mirror*, entre muchas. Pero estas narrativas distan de la realidad. «Las personas podemos 'jugar' a que estamos interactuando con algo que tiene sentimientos, pero en la mayoría de los casos podemos darnos cuenta de que la contraparte no es real».

Para algunos autores, hay incentivos económicos para que pensemos que los robots son como personas. El filósofo argentino y el *senior fellow* de Google concuerdan en que «puede ser que el límite entre el juego y lo real se vuelva borroso. Y puede ser útil que ese límite se vuelva cada vez más ambiguo». Para Armando, esa ambigüedad facilitaría la asimilación tecnológica y la validación de la inteligencia artificial, permitiendo optimizar procesos a través del reemplazo de personas por máquinas de IA. «Hay mucha gente pensando en asistentes terapéuticos de inteligencia artificial que den respuestas informadas y personalizadas ante el relato de un paciente. Muchas veces, para generar la confianza necesaria en este tipo de interacciones, las personas deben pensar que están hablando con otra persona y no con una máquina», dice.



3. LA ÉTICA Y LOS SÚPER HUMANOS

Timnit Gebru es doctora en Ingeniería Electrónica de la Universidad de Stanford y se ha especializado en sesgo y ética en los modelos de inteligencia artificial. Alcanzó visibilidad tras ser despedida de Google porque cuestionó el escaso esfuerzo de la compañía para erradicar los sesgos. En *Algoritmos de la opresión*, Safiya Umoja Noble describe cómo los motores de búsqueda de Google arrojan resultados racistas y sexistas. En 2009, escribió «*black girls*» en la barra de búsqueda y la primera página que obtuvo era casi toda de pornografía.

El problema es que la inteligencia artificial no solamente impulsa el motor de búsqueda de Google, sino que también permite que Facebook dirija la publicidad y que Alexa y Siri hagan su trabajo. La IA también está detrás de los autos sin conductor, la vigilancia predictiva y las armas autónomas que pueden matar sin intervención humana.

Para Margaret Mitchell, colega de Gebru y jefa de ética en Hugging Face, «la IA no sólo tiene el potencial de replicar los sesgos humanos (a través de los datos de entrenamiento, así como los de sus creadores), sino que (esos datos) vienen enmascarados con una especie de credibilidad científica. De esta forma se hace creer que estas predicciones y juicios tienen un estatus objetivo». Todo esto abre la cancha para discutir sobre cómo los sesgos pueden perpetuar discriminaciones y perjudicar a grupos minoritarios y vulnerados.

Tras el episodio de Lemoine, Timnit Gebru y Margaret Mitchell se unieron para redactar una carta a The Washington Post, en la que plantearon que «si la IA creada por humanos tiene sesgos y luego los humanos les otorgamos superpoderes a esa inteligencia artificial (o sea, creemos que son humanos o peor, súper-humanos) las consecuencias pueden ser una desgracia». Cuestionan las imprecisas declaraciones con las que otros profesionales de Google se han referido al tema, dado que no aclaran qué es y qué no es su «producto» llamado LaMDA y son los principales responsables de la creencia que es posible una «conciencia artificial». Ellas ven con suspicacia el objetivo de crear una mega máquina inteligente, capaz de reemplazar a los humanos, en vez de generar máquinas funcionales y basadas en objetivos. Por último, plantean que la gente tiene derecho a saber qué está pasando, en qué se está invirtiendo y cuál es el objetivo detrás de moldear la tecnología que cada vez más afecta la vida de las personas.

UN MUNDO EN MUTACIÓN

Elaborado por 

Publicación divulgativa. Se permite copiar, distribuir y comunicar públicamente la obra con la condición de reconocer debidamente su autoría, no utilizar para fines comerciales, ni generar obras derivadas.