



Así responde Chat GTP a todo: millones de datos, matemáticas y adivinanzas

Inteligencia Artificial
En la semana que se ha celebrado la cumbre mundial, tres expertos explican cómo funcionan estas herramientas

JON GARAY



Cojan su teléfono móvil y empiecen a escribir un mensaje en Whatsapp. ¿Ven que cada vez que escriben una letra les sugiere varias palabras? Eso es justo lo que hacen, a gran escala, programas como ChatGPT o Bard, adivinar la siguiente palabra. «No hay una inteligencia detrás, ni conciencia, ni nada parecido. Son matemáticas y probabilidades. Términos como Inteligencia Artificial y redes neuronales son engañosos», asegura Eneko Agirre, el director de HiTZ, el Centro Vasco de Tecnología de la Lengua de la UPV/EHU.

Esta semana se ha celebrado en Bletchley Park –una histórica finca a 80 kilómetros de Londres donde se desentrañó el código secreto con el que se comunicaban los nazis en la II Segunda Guerra Mundial– una cumbre mundial para tratar de regular una tecno-

logía que es capaz de mejorar los diagnósticos médicos y de ayudar a luchar contra los incendios pero también de crear noticias falsas y poner en riesgo la propia democracia. Hay quien piensa que incluso nos podría superar en inteligencia.

ChatGPT es la más conocida de esta nueva generación de programas. Pertenece a lo que se llama Inteligencia Artificial Generativa. La Inteligencia Artificial es una rama de la computación que trata de imitar la cognición humana, es decir, «no solo cómo aprendemos, sino también cómo entendemos, cómo razonamos y cómo interactuamos», explica Mikel Díaz, director de Innovación de IBM en España.

La IA nació, como la Informática, hace setenta años. La Generativa se denomina así porque ‘generan’ un resultado a partir de una información inicial. Si traba-

jan con textos –otras, como Midjourney o Dall-E lo hacen con imágenes–, se llaman modelos de lenguaje. «Hasta hace cinco años, era la hermana pobre de la Inteligencia Artificial general. Se utilizaba en los traductores automáticos pero los resultados no eran demasiado buenos», explica Agirre. Ahora se les pregunta casi cualquier cosa y responden con sorprendente soltura. ¿Cómo lo hacen?

Todo empieza con una especie de juego de adivinanza. «Se le pide al sistema que adivine la siguiente palabra de una frase. Por ejemplo, ‘estamos comiendo un pincho en San ...’», ejemplifica el experto, que lleva trabajando en este campo desde 1988. La respuesta correcta sería San Sebastián. Para llegar a ella, hay que explicar el engranaje de estas tecnologías. Las claves son las redes neuronales, una base de datos gigantesca

y dos técnicas de aprendizaje automático (‘machine learning’ en inglés), el supervisado y por refuerzo.

Como los precios del ‘súper’

En el corazón de Chat GTP están las redes neuronales, un «modelo matemático inspirado en el funcionamiento del cerebro», explica Álex Rayón, doctor en Ciencias de la Computación y Tecnología. Como nuestro cerebro, tienen muchas ‘neuronas’ –llamados ‘nodos’– conectadas entre sí que pueden recibir y enviar información. Estas redes se organizan en capas relacionadas entre sí por unas conexiones llamadas ‘pesos’. Cada peso tiene un valor que se puede ajustar. La primera capa, la de entrada, recibe los datos iniciales, los transforma en números –vectores– y los envía a las capas intermedias. Los vectores recibidos se multiplican con unas matrices

–tablas de números organizadas en filas y columnas–, para llegar a la capa de salida, que produce el resultado final. A todo este proceso se le llama propagación hacia adelante. El resultado final se compara con el que se esperaba y si no es correcto, se le indica y la red neuronal hace el proceso inverso (propagación hacia atrás) para ajustar los pesos y mejorar en la siguiente palabra, en la siguiente adivinanza. «Es algo similar a como los supermercados establecen el precio óptimo de los precios de los productos, variándolos en función de si los clientes los compran o no», afirma Rayón.

ChatGPT cuenta además con una gigantesca base de datos. «Se ha ‘leído’ el equivalente a millones de libros, algo inalcanzable para nosotros. Su versión 3 contaba con 175.000 millones de parámetros –los datos que ha aprendido–. Los ejemplos se introducen uno a uno y se hace una propagación hacia adelante. La cuarta, la actual, no se sabe porque no lo han revelado, pero parece que ya ha alcanzado un límite». El 60% de los datos proceden de internet, otra importante porción de Reddit –un foro parecido a Twitter donde los usuarios intercambian noticias, opiniones...– y también cuenta con toda la Wikipedia. Para hacerse una idea del volumen que implica, esta última, en todos sus idiomas, solo supone un 3% de sus datos. Esto se traduce en un enorme gasto de energía y la necesidad de utilizar superordenadores para gestionarlos.

«Enseñarles a comportarse»

Sabido esto, volvemos a la frase de la adivinanza, ‘estamos comiendo unos pintxos en San Sebastián’. Se introduce en la capa de entrada, que la transforma en números (vectores), y se le pide que adivine la palabra que sigue. «El programa revisa todos sus datos y calcula las probabilidades para dar una respuesta correcta. Si responde San Sebastián, el programador se lo indica. Con la siguiente palabra se haría lo mismo y así sucesivamente. Solo es adivinar la siguiente palabra, no hay ningún proceso mental más complicado que este», enfatiza Agirre. A esto se le llama ‘aprendizaje supervisado’, ‘porque hay un humano que le indica si es correcto o no», añade Mikel Díaz.

Entre todos los millones de datos que el sistema ha leído hay también información poco deseable –como «cómo hacer una bomba o cómo matar a alguien»– o afirmaciones racistas y machistas. «Hay que enseñarles a comportarse, no tienen filtro», subraya Agirre. Entra aquí el ‘aprendizaje por refuerzo’. «Así aprenden también los niños. Si tocan una estufa y se queman, aprenden a no volver a hacerlo», pone como ejemplo Mikel Díaz. «También algunos robots de limpieza o los coches autónomos», continúa. Estos sistemas se relacionan con el entorno y toman decisiones que son recompensadas