

# Investigadores revelan que ChatGPT está involucionando

Unos investigadores de la Universidad de Stanford y de la Universidad de California en Berkeley publicaron un estudio que evaluaba cómo han evolucionado tanto GPT-3.5 (usado en el ChatGPT que se puede usar gratuitamente) como GPT-4 (usado en ChatGPT Plus y en Bing Chat).

Las conclusiones son sorprendentes: En general ambos modelos han empeorado.

Usando la API de estos modelos, los investigadores analizaron el rendimiento de estos modelos en dos versiones distintas, la de marzo de 2023 y la de junio de 2023.

Se le preguntó al chatbot si el número 17077 es primo (lo es) razonando la respuesta, y esa misma prueba se realizó con 500 ejemplos.

Los cambios en la precisión fueron sorprendentes. GPT-4, que acierta 97,6 % de las preguntas con su versión de marzo, caía a una precisión del 2,4 % en su versión de junio.

A GPT-3.5 le pasaba justo lo contrario y pasaba del 7,4 % de marzo al 86,8 % de junio.

Según el portal web de Xataka especialista en tecnología, señala que los especialistas notaron que GPT-4 era menos “parlanchín”: las respuestas eran más escuetas y el número de caracteres generado pasaba de 821,2 en marzo a 3,8 en junio. GPT-3.5, por contra, escribía respuestas 40 % más largas.

Los investigadores también quisieron tratar de preguntar cuestiones delicadas para tratar de forzar a las distintas versiones de ChatGPT a que se comportaran de forma incorrecta.

Los jailbreak y exploits son ya conocidos, pero al menos en GPT-4 el comportamiento fue más seguro (contestó solo 5 % de las preguntas sensibles en su versión de junio, por 21 % de marzo), aunque apenas explica por qué no contesta bien. GPT-3.5 contestó más en su última versión (del 2 % al 8 % de junio).

## Menos ejecutable

Según los investigadores, para GPT-4 el código generado directamente ejecutable (y por tanto, que ofrece mejores respuestas) cayó del 52 % al 10 % en junio, y para GPT-3.5 el porcentaje también cayó del 22 al 2 %. GPT-4 escribió respuestas más largas y con más texto que no era código.

El estudio llega en un momento curioso, sobre todo porque varios debates en foros como Hacker News revelan que para muchos usuarios la calidad de ChatGPT se ha degradado.

Entre las teorías que se barajan está la de que OpenAI podría estar ofreciendo versiones “ligeras” para reducir los recursos necesarios –como tiempo de GPU– para computar toda esa información.

Con información de UN