

“Chat Gpt, La Habitación China De Searle y la conciencia”

Chat GPT está en boca de todos, algunos le tienen miedo, a otros les entusiasma el potencial que tiene para cambiar el mundo laboral. Hoy dejo de lado la parte de las consecuencias económicas para centrarme en el aspecto filosófico de Chat GPT y examinar un poco sobre la pregunta: ¿entiende chat GPT realmente el lenguaje? ¿es consciente?

What a typical “learning machine” does, is finding a mathematical formula, which, when applied to a collection of inputs (called “training data”), produces the desired outputs -The Hundred-Page Machine Learning Book, Adriy Burkov

La habitación china de John Searle

John Searle, fue un científico que se preguntó si una computadora puede entender realmente el lenguaje. Y para esto diseñó un experimento mental muy interesante:

Un humano, dentro de una habitación completamente aislada del mundo exterior, con una serie de instrucciones para procesar símbolos en chino y devolver una respuesta. Hay que resaltar que la persona en el interior de la habitación desconoce el idioma chino y se limita a seguir la serie de instrucciones que le fueron otorgadas, es decir, **no comprende el texto que entra, ni el que sale**.

```
graph TD;
    --> Habitación;
    Habitación--> ;
```

Puedes notar las similitudes con una computadora: hay un input o entrada, un output o salida y una caja negra o programa informático, cuyos detalles del funcionamiento escaparían para los que se encuentran afuera de este pequeño sistema simplificado.

```
graph TD;
    Input--> programa;
    programa--> output;
```

Según Searle, para las personas que están afuera de la habitación, pareciera que lo que sea que esté dentro entiende perfectamente el idioma chino. Sin embargo, nosotros sabemos que la persona en el interior solo está siguiendo un set de instrucciones, tan complejísimo como nosotros querramos, pero que **no implica una comprensión del lenguaje**, sino un proceso mecánico. Según Searle, esto es similar a cómo funciona una computadora.

De manera similar, y extrapolado a ChatGPT; aunque un modelo de lenguaje puede producir respuestas que parecen coherentes y relevantes, no hay evidencia

de que comprenda realmente el lenguaje o tenga conciencia de su significado. ChatGPT puede estar siguiendo el algoritmo más complejo existente pero, según Searle, no hay más conciencia del proceso ahí que la que encontraríamos en el reloj mecánico más sofisticado.

¿Pero entonces que marca la diferencia entre una inteligencia real y un proceso mecánico? ¿existe dicha diferencia?

¿Están la conciencia y la inteligencia artificial relacionadas?

El máximo representante de la inteligencia en la tierra es el ser humano (o al menos nuestro ego inflado nos dice eso), y también sucede que el ser humano es un ente vivo con conciencia. Esto nos dirige a la inevitable pregunta: ¿es necesaria un ser vivo con conciencia para tener inteligencia? ¿o quizás es al revés?

Inteligencia con conciencia

Algunos argumentan que la conciencia y la inteligencia están relacionados y que incluso tendríamos un gradiente de consciencia mientras avanzamos en la complejidad e inteligencia de los organismos. Dotando del calificativo “consciente” a animales superiores como delfines, orangutanes que muestran rasgos de inteligencia.

algunos van más allá al decir que un sistema se vuelve más consciente mientras más inteligente se vuelve, quizás podríamos considerar el ejemplo de Blake Lemoine, que aseguraba que el modelo de inteligencia artificial de Google había cobrado conciencia, o la teoría de la información integrada, de Giulio Tononi, que propone que la consciencia se genera cuando un sistema es capaz de incorporar información e unificarla, y que este nivel de consciencia (llamado Φ) puede calcularse para cualquier sistema.

Inteligencia sin conciencia

Por otro lado, existen posiciones que defienden que la inteligencia no depende necesariamente de la conciencia, sino que puede existir en sistemas que no tienen experiencia subjetiva. Para ejemplo basta citar a Alpha Go y otros programas informáticos que son capaces de responder a estímulos muy complejos, sin existir más allá del juego para el que fueron programados, o a los sonámbulos, que pueden mostrar signos de inteligencia aún no estándolo.

Pero aunque los seres humanos tengan procesos inconscientes que existen sin conciencia, al estar vivos, hay algo más que una simple acción mecánica ¿no?

Autómatas biológicos

A veces creemos que solo las máquinas tienen un comportamiento mecánico y que cualquier ser vivo sería capaz de responder de manera muy diferente a la

que haría una máquina, con más versatilidad y adaptándose a los cambios, pero, ¿es siempre así?

En el libro “Un eterno y grácil bucle” de Douglas R. Hofstadter. El autor cita un experimento en el que una avispa sphex es engañada para acercarse a un grillo en los límites de una madriguera hasta 40 veces. Tal cual como si fuera un programa de computadora, esta avispa queda atorada en un bucle infinito del que no puede escapar, ¿qué tan diferente es esto de un programa informático que, tras el mismo input, genera el mismo output?

Este experimento me hizo cuestionarme sobre si algunos seres vivos no son otra cosa que autómatas biológicos y también sobre donde está el punto de inflexión en el que un ser vivo deja de ser un autómata y se vuelve consciente.

¿Qué leer o ver para saber más de inteligencia artificial y conciencia?

Entre mi lista de materiales favoritos que tocan estos temas se encuentran los siguientes materiales:

- Un eterno y grácil bucle de Douglas R. Hofstadter: el autor profundiza en el tema de la autoreferencia y desarrolla la pregunta: ¿puede un sistema comprenderse así mismo?
- La nueva mente del emperador de Sir Roger Penrose: el autor establece el contexto de las leyes del universo y analiza si la conciencia y la inteligencia están relacionadas y si estas tienen un carácter determinista o no determinista.
- Brains, Minds, and Machines: Consciousness and Intelligence: plática del MIT, donde se desarrollan los temas de cerebros, conciencia, inteligencia y máquinas. Radicalmente infravalorada; ¿7000 vistas en youtube nada más? ¿de verdad?
- ¿Puede un programa estar vivo? pequeño videoensayo donde uno de mis youtubers favoritos desarrolla el tema de si un programa informático puede estar vivo.
- The connection between intelligence and consciousness