

- a) **En la entrevista, Altman se refiere a que chatGPT, en su forma actual, no será capaz de hacer descubrimientos científicos. Fridman está en desacuerdo. Argumenta en qué sentido ambos pueden tener razón. Da ejemplos y analiza cómo es que se podrían superar las actuales limitaciones para lograr la AGI, en tal como lo define Altman.**

A favor de Lex, creo que chatGPT o GPT-3/4 en su estado actual sí pueden servir como herramientas poderosas para asistir a personas a realizar descubrimientos científicos, ya que pueden tomar un rol de “mentor” o asistente de laboratorio, dependiendo del prompt entregado por la persona que lo utilice, y dependiendo del fin que le quiera dar. Algunos casos de uso podrían ser: hacer revisiones bibliográficas, asistir en cálculos, iterar sobre posibles métodos experimentales, recalcar errores en los procedimientos, ayudar a extraer conclusiones, etc.

Ahora bien tomando la posición de Altman, estos modelos / herramientas no pueden realizar los descubrimientos por sí solas, requieren de alguien que los utilice, por lo que en estricto rigor, no son capaces de realizar estos descubrimientos de manera independiente. Además, para el caso de chatGPT, su entrenamiento solo se basa en el lenguaje hasta 2021, por lo que carece casi 2 años de conocimientos, los cuales podrían ser cruciales para realizar descubrimientos en campos de la ciencia que avanzan rápidamente, como la misma inteligencia artificial.

Parte importante del potencial que tienen estas tecnologías al asistir, o en un futuro, participar independientemente en la ciencia o expansión del conocimiento en general, recae en qué tan capaces somos de usarla. Tal como lo dijo Altman; “Cuando se trata de GPT-4, podríamos estar lidiando con un AGI, pero quizá no tenemos la capacidades de accederla lo suficiente aún”. (*no es la cita exacta*) ¿Quizá lo que nos separa de un AGI, no son las capacidades de los modelos en sí, sino las interfaces que estamos usando para manejarlas? Podría ser que la mejor manera de interactuar con un modelo de lenguaje no sean los prompts escritos, como lo hacemos con chatGPT. Una opción alternativa a esto, un poco soñadora, sería traducir nuestros pensamientos a lenguaje que se pueda alimentar como un prompt a un modelo, cosa que reduce la cantidad de pasos entre la ideación de un prompt y la transmisión de este al modelo, aumentando la compatibilidad entre humanos y inteligencia artificial. Pero más allá de especular, creo que sería bueno indagar sobre nuevas interfaces o metodologías para interactuar con modelos de lenguaje, ya que podríamos no estar canalizando su potencial completo.

Otra idea que podría acercarnos al AGI es desarrollar sistemas de aprendizaje más flexibles y adaptables que puedan aprender de manera autónoma de múltiples fuentes de información. Actualmente, la mayoría de los modelos de inteligencia artificial son entrenados con conjuntos de datos específicos y no son capaces de aprender de manera continua o transferir conocimiento de una tarea a otra de manera eficiente. Si pudiéramos desarrollar sistemas de aprendizaje más flexibles y adaptables, podríamos tener modelos de inteligencia artificial que puedan aprender y mejorar de manera autónoma, y que puedan ser aplicados a una amplia gama de tareas y dominios.

b) Describe el argumento de Sutskever en la versión de Altman y discútelo. En tu discusión entrega tu propia definición de conciencia e intenta argumentar por qué un modelo de lenguaje podría o no podría tener esa capacidad.

El argumento de Sutsveker consiste en que: si a un modelo entrenado con **absolutamente ninguna referencia** de la “conciencia” (como palabra o concepto), se le pregunta si entiende una descripción la experiencia subjetiva de la conciencia humana, y esta responde que sí entiende, entonces este modelo es consciente.

Personalmente, creo que esta es una hipótesis muy reduccionista al abordar el concepto de “conciencia”. Creo que plantear que un modelo de lenguaje tiene la capacidad de ser consciente implica que la conciencia “existe” totalmente en el lenguaje, o es transmisible por el lenguaje (ya que estos modelos están entrenados a base del procesamiento de lenguaje). Si bien, tal como menciona Lex, se puede intentar convencer con lenguaje a un tercero que yo “poseo conciencia”, no creo que exista tal cosa como “otorgar” conciencia a través de ese mismo medio. La conciencia, va mucho más allá de la “racionalización” que es capaz de ejercer un modelo de lenguaje o el cerebro de un ser vivo:

A mi parecer, la conciencia es indivisible, existe en más lugares que solo el cerebro, y realiza más acciones que solo la racionalización. Para ejemplificar aquello se puede pensar en las distintas “sensaciones” que nos otorgan las partes de nuestro cuerpo, o extensiones del mismo: Por ejemplo, un jugador de tenis es consciente de la raqueta y de sus movimientos, o una persona que está comiendo es consciente del tenedor. También existe la conciencia social, emocional, etc.

Es cierto que un modelo de lenguaje bien entrenado emula e incluso supera la capacidad lógica humana, pero cuando lo comparamos a las otras manifestaciones de conciencia que nombré, es evidente que no son replicables por un computador, ya que este está basado únicamente en principios lógicos, que no gobiernan totalmente sobre las distintas conciencias, por lo que ni el modelo de lenguaje más avanzado podrá ser consciente.

Si no quedó muy convencid@ con mi respuesta, pregúntele a chatGPT si en algún momento será capaz de sentir emociones...

2. Definición para restricciones de cardinalidad:

Las restricciones de cardinalidad son una forma de limitar la cantidad de átomos verdaderos en un modelo o answer set de un programa lógico. Una restricción de cardinalidad se define como una fórmula lógica que especifica una condición en la cantidad de átomos verdaderos que pueden aparecer en un modelo.

Dado un modelo M que contiene la restricción $n\{P\}m$, este se satisface si y sólo si el número de átomos del predicado P en M está **entre** n y m .

Los programas ejemplo que ilustran este funcionamiento se encuentran en la carpeta "programas_ejemplo" dentro de la carpeta "Reflexión y Teoría"

