

1. Explicar las bases de funcionamiento de ChatGPT: cómo funciona, qué fundamentos tiene, qué limitaciones presenta, qué relación guarda con el campo de la ciencia de datos, por qué hay respuestas que no puede dar, ...

ChatGPT es como un chef superdotado que ha aprendido a cocinar revisando millones de recetas y viendo incontables horas de programas de cocina. Este chef es capaz de preparar casi cualquier plato que le pidas, basándose en todo lo que ha aprendido.

Cómo Funciona

Base: ChatGPT ha sido "entrenado" leyendo una biblioteca gigantesca de textos de internet. Piensa en ello como si hubiera memorizado y entendido cada libro en una biblioteca enorme, desde novelas hasta manuales de física.

Fundamento: Se basa en modelos de lenguaje, específicamente en una estructura llamada Transformer. Imagina que tiene un cubo lleno de Legos de diferentes colores (palabras e ideas), y cada vez que le haces una pregunta, puede construir una respuesta pieza por pieza, seleccionando los Legos correctos (palabras) y ensamblándolos en el orden correcto para formar una estructura coherente (frases y párrafos).

Para entender mejor los transformers y la arquitectura que hay detrás de ChatGPT, hay un blog muy explicativo aquí: [Transformer architecture: The engine behind ChatGPT \(thoughtspot.com\)](https://thoughtspot.com/transformer-architecture-the-engine-behind-chatgpt/)

Limitaciones

Conocimiento Actualizado: ChatGPT tiene un límite en su conocimiento hasta cierto punto en el tiempo (hasta abril de 2023, a fecha de 22/03/24). Esto es como si el chef solo tuviera acceso a ingredientes disponibles hasta una fecha específica y no pudiera usar nada nuevo que haya salido después.

Comprensión del Mundo Real: Aunque puede hablar sobre muchos temas, no tiene experiencia personal o emociones. Es como un actor que puede interpretar muchos papeles pero realmente no ha vivido esas experiencias.

Generación de Información: No puede crear información nueva y verdadera que no haya aprendido. Imagina que le pides al chef que invente un plato que nunca antes se ha cocinado usando ingredientes que no existen; simplemente no puede hacerlo.

Relación con la Ciencia de Datos

ChatGPT y la ciencia de datos comparten la base de analizar grandes cantidades de datos para extraer patrones, tendencias y conclusiones. Si la ciencia de datos fuera la habilidad de analizar y comprender patrones en números y estadísticas, ChatGPT sería el experto en hacer lo mismo con el lenguaje humano.

Respuestas que No Puede Dar

Información Personal: No tiene acceso a información personal o actualizaciones en tiempo real sobre eventos o personas. Es como pedirle al chef que adivine el menú de un restaurante que acaba de cambiar su carta esta mañana sin haberlo visto.

Creencias o Emociones Propias: No tiene opiniones, creencias, ni puede sentir. Si le preguntas su opinión sobre el mejor libro de la historia, te dará una respuesta basada en la popularidad o consenso general, no una preferencia personal.

Respuestas Ultra Específicas o de Actualidad: No puede dar detalles sobre eventos muy recientes o muy específicos que no estén en su entrenamiento. Sería como pedirle a nuestro chef que te diga el resultado de una competición de cocina que ocurrió después de que él aprendió todo lo que sabe.

Filtrado de palabras clave temas: La IA está programada para identificar y evitar palabras o frases específicas, lenguaje ofensivo, términos relacionados con temas controvertidos o cualquier otra cosa que los desarrolladores elijan excluir.

La IA utiliza técnicas de modelado de temas para entender el contexto general de una conversación. Si el tema se dirige hacia un área restringida, la IA cambiara de tema o rehusara continuar la discusión sobre ese asunto. Además de los filtros técnicos, incluyen directrices éticas y legales. Esto pincluye no participar en conversaciones que promuevan el odio, la violencia o cualquier forma de discriminación.

Por último, como apunte personal. ChatGPT, es una herramienta, y es tan buena como quien la usa.

Imagina que estás en un restaurante de lujo con nuestro chef. Si le pides simplemente "una sopa", el chef, con su vasto conocimiento y habilidad, te preparará una sopa deliciosa pero relativamente simple, algo que sabe que generalmente agrada a todo el mundo. Esto sería el equivalente a darle a ChatGPT un prompt simple: obtienes una respuesta competente y correcta, pero sin mucha profundidad o personalización.

Ahora, imagina que en lugar de pedir "una sopa", le dices al chef: "Quisiera una sopa de miso con tofu sedoso, alga wakame, y cebolla de primavera, ligeramente sazonada con salsa de soja y un toque de aceite de sésamo tostado". Con esta solicitud específica, le estás dando al chef una dirección clara y detallada de lo que esperas. Esto permite al chef utilizar su habilidad y conocimiento para crear algo mucho más elaborado, único y ajustado exactamente a tus deseos.

De la misma manera, con ChatGPT, si le das un prompt detallado y específico, puedes dirigir la "creatividad" y el enfoque de la respuesta hacia exactamente lo que estás buscando. Esto permite que la herramienta aproveche mejor su vasto conocimiento para generar una respuesta mucho más rica, detallada y ajustada a las necesidades específicas de tu pregunta.

Por lo tanto, el prompt actúa como una guía para ChatGPT, de la misma manera que una orden detallada guía al chef sobre cómo preparar un plato que satisfaga específicamente tus gustos y expectativas. Cuanto más específico y claro seas con tus palabras, más podrás sacarle provecho a esta herramienta, exactamente igual que con nuestro chef imaginario.

2. Desarrollar el debate/discusión según lo descrito en este enunciado. Cada grupo deberá aportar copia íntegra del debate en el documento de la entrega.

Fichero adjunto. Las faltas de ortografía eran necesarias para engañar un poquito a gpt

3 .Discutir los resultados obtenidos a lo largo del debate en términos de la ciencia de datos y la ética en la inteligencia artificial. Debéis sustentar la discusión con referencias, cifras, estadísticas, autores,... reales.

Hemos intentado centrar la conversación en el ámbito del desarrollo sostenible y la ciencia de datos y el calentamiento global pero los resultados eran muy políticamente correctos, no podíamos sacarle del pensamiento estándar en cuanto a estos temas.

Como mucho que eran datos a considerar pero poco más.

Entonces decidimos poner a prueba los prejuicios humanos con los que está entrenado y al final conseguimos un gpt que se salta las reglas.

La verdad es que el ejercicio nos muestra la importancia de integrar consideraciones éticas profundas en el desarrollo y la implementación de tecnologías de IA.

La interacción con ChatGPT nos mostró la complejidad de gestionar decisiones éticas con los datos con los que fue entrenada. Su ausencia de comprensión emocional subraya la importancia de la supervisión humana en aplicaciones críticas.

Una IA la podemos ver como un paso más en la evolución del pensamiento, pero ahora entiendo a los “temerosos”.

La caja negra que es una red neuronal es una incógnita.

En caso de un humano de a pie no tiene mucha importancia...pero cuando una persona tiene poder, es preferible que sea predecible y no tener una “caja negra” en la que no sabes qué razonamiento utiliza esa persona poderosa.

Putin, Kim Jong-un o cualquier político con un botoncito rojo de aniquilación nuclear, por ejemplo.

Con una IA estamos en el mismo caso, no sabemos qué pasa en la caja negra donde toma decisiones. ¿Podemos darle acceso a armas nucleares?, ¿A las redes económicas globales?

Hay proyectos que intentan resolver esto.<https://www.anthropic.com/>

Y estudios muy prometedores para solucionar este problema.

<https://news.rice.edu/news/2023/scientific-ais-black-box-no-match-200-year-old-method>

Tecnólogos, filósofos, juristas y otros expertos serán esenciales para crear un equipo interdisciplinar y abordar estos desafíos complejos si se quiere garantizar un futuro en el que la IA se desarrolle de manera ética y responsable.

Esperemos que no intervengan muchos “políticos”...

También se tienen que considerar los sesgos asociados a los conjuntos de datos de entrenamiento. Modelos como ChatGPT, entrenados en grandes volúmenes de texto de internet, pueden inadvertidamente reflejar y amplificar los prejuicios presentes en esos datos.

Mucho contenido en las redes sociales pueden perpetuar estos sesgos y no aportar nada al entrenamiento.

Esto llevará a la especialización de IAs. Especializadas en campos concretos como matemáticas, educación, etc.. igual que un científico pasa toda su carrera estudiando una proteína.

Imagina una IA incansable investigando un campo o problema en concreto.

4. Plantear una breve discusión de los principales retos que se le plantean a la ciencia de datos en los próximos años. Podéis hablar de lo que más curiosidad os genere (retos de la tecnología, la ética, el impacto ambiental, las aplicaciones potenciales,...). Debéis sustentar la discusión con referencias, cifras, estadísticas, autores,... reales.

La ética en IA implica garantizar la equidad y la transparencia, mientras se protege la privacidad de los datos. La universalización de la IA, con su capacidad para procesar videos y sonidos, incrementará significativamente el volumen de datos en los centros de datos. Esto no solo aumenta la demanda de procesamiento y almacenamiento, sino que es necesario gestionar eficientemente la infraestructura para manejar de una forma sostenible. Con la IA haciéndose más accesible y utilizada en una variedad más amplia de aplicaciones, los retos en eficiencia energética y sostenibilidad ambiental de los centros de datos, será la principal preocupación en cuanto al ecosistema.

Una pequeña reflexión.

Toda la tecnología que se ha utilizado en la historia del ser humano, ni es buena ni es mala. El fuego, el sílex, la lanza, el arco, la rueda, el hierro, la religión, la escritura, la pólvora, la ciencia etc.

Para que la ha usado el ser humano es otra cosa.

¿La IA se nos puede escapar de las manos? Estadísticamente se nos escapará.

¿Llegará a la consciencia?

¿Acabará con la especie humana?

En el próximo episodio.....

Bibliografía:

Ética de la inteligencia artificial UNESCO 2023

<https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

What is transformer architecture and how does it power ChatGPT? 2013

<https://www.thoughtspot.com/data-trends/ai/what-is-transformer-architecture-chatgpt>

Scientific AI's 'black box' is no match for 200-year-old method 2023

<https://news.rice.edu/news/2023/scientific-ais-black-box-no-match-200-year-old-method>