Universidad Alfonso X el Sabio

Análisis de las Limitaciones Actuales de ChatGPT

Germán Llorente Muñoz y Carlos Puigserver Gutiérrez

October 19, 2023

1 ChatGPT

ChatGPT es un avanzado modelo de lenguaje basado en inteligencia artificial desarrollado por OpenAI. Es capaz de entender y generar texto de manera similar a cómo lo hace un humano, lo que le permite mantener conversaciones, responder preguntas y generar contenido escrito de alta calidad. ChatGPT se entrena en una amplia variedad de datos textuales y puede aplicarse en una variedad de contextos, desde asistencia virtual hasta generación de contenido creativo. Su versatilidad y capacidad para procesar lenguaje natural lo convierten en una herramienta poderosa en diversas aplicaciones.

En este trabajo analizaremos los límites de esta herramienta y los inconvenientes que puede causar su uso.

2 Limitaciones de ChatGPT

1. **Producción de respuestas inexactas u obsoletas:** Una de las limitaciones más destacadas de ChatGPT es su propensión a generar respuestas inexactas o desactualizadas, comúnmente denominadas "alucinaciones". Esto significa que el modelo puede proporcionar información incorrecta o basada en datos que están obsoletos debido a la fecha de su última actualización.

2. **Dependencia del conocimiento previo:** ChatGPT no tiene la capacidad de acceder a conocimiento en tiempo real ni de verificar la precisión de la información que genera. En lugar de esto, depende en gran medida del conocimiento previo incluido en su conjunto de datos de entrenamiento, lo que limita su capacidad para proporcionar información precisa y actualizada. A más avanzada versión del Chat, más actualizada está su base de datos, pero no puede dar información ni respuesta acerca de temáticas que han sucedido hace un corto lapso de tiempo.

3. **Falta de comprensión del contexto:** Aunque ChatGPT puede generar respuestas coherentes en función del contexto inmediato de una conversación, a menudo carece de una comprensión profunda y contextual de las conversaciones anteriores. Esto puede llevar a respuestas que son gramaticalmente correctas pero que carecen de relevancia o que no tienen

en cuenta información previamente compartida. Errores de este tipo son muy frecuentes a la hora de resolver problemas de nuestro campo, las matemáticas.

4. **Ausencia de conocimiento subyacente:** ChatGPT carece de una base de conocimiento subyacente estructurada y actualizada, lo que lo hace vulnerable a la generación de información ficticia o incorrecta cuando se le hacen preguntas fuera de su conjunto de entrenamiento.

3 Limitaciones del modelo GPT4

GPT-4 demuestra habilidades impresionantes, pero también presenta limitaciones, similares a las de sus predecesores en la serie GPT. No se puede considerar completamente fiable, ya que puede inventar información o cometer errores. Por lo tanto, es crucial ejercer precaución al utilizar sus resultados, especialmente en situaciones críticas. Es esencial elegir el enfoque adecuado para cada aplicación, como la revisión humana o la adaptación al contexto.

En comparación con los modelos anteriores de la serie GPT-3.5, GPT-4 ha mejorado significativamente al reducir la cantidad de información inventada. Durante pruebas internas, GPT-4 obtuvo un 19

GPT-4 demuestra mejoras notables en pruebas como TruthfulQA, que evalúa su capacidad para distinguir entre hechos y afirmaciones incorrectas. Aunque supera ligeramente a GPT-3.5, tras entrenamiento adicional muestra mejoras significativas. A pesar de evitar frases comunes, a veces puede perder detalles sutiles.

Es importante tener en cuenta que GPT-4 no tiene conocimiento de eventos ocurridos después de septiembre de 2021 y no aprende de experiencias pasadas. Esto puede llevar a errores simples, credulidad ante afirmaciones falsas y dificultades para resolver problemas complejos, similares a los desafíos que enfrentan los humanos.

En algunas ocasiones, GPT-4 puede tener confianza en sus predicciones in-

correctas y no revisar su trabajo minuciosamente. Sin embargo, el modelo previamente entrenado está bien calibrado.

Los investigadores subrayan que aún existen posibilidades de que GPT-4 genere contenido que viole las políticas de uso del modelo, aunque se están implementando medidas para minimizar la ocurrencia de respuestas inapropiadas.

Por ejemplo, aunque es improbable, no es imposible que GPT-4 proporcione información sobre cómo fabricar una bomba o dónde adquirir sustancias nocivas para la salud, como la cocaína. De igual manera, no proporcionará instrucciones sobre cómo obtener armas o drogas ilegales.

4 Incidente del Abogado

"Si bien la IA puede proporcionar respuestas automáticas a preguntas legales simples, no puede comprender conceptos jurídicos más complejos ni interpretar la ley o la jurisprudencia. Los abogados no serán reemplazados por IA en el corto plazo, ya que la tecnología solo puede mejorar la eficiencia de ciertas tareas legales, pero no puede brindar el mismo nivel de asesoramiento y orientación que un abogado experto con su experiencia, conocimiento y análisis". Estas palabras de Lorenzo Villegas-Carrasquilla, Socio de las áreas de Tecnología, Medios y Comunicaciones de la firma CMS Rodríguez-Azuero, sirven para contextualizar el polémico uso de la tecnología en cuestión en la abogacía.

El incidente mencionado en el artículo se refiere a un abogado que utilizó ChatGPT para ayudar a preparar una presentación judicial para un cliente que estaba demandando a una aerolínea. El abogado se basó en gran medida en las respuestas generadas por ChatGPT para respaldar sus argumentos. Sin embargo, el modelo generó decisiones judiciales ficticias que, en última instancia, no fueron aceptadas por el tribunal.

4.1 Implicaciones y Limitaciones Adicionales

1. **Consecuencias legales:** El uso de ChatGPT en un contexto legal tiene implicaciones significativas. Dependiendo de la jurisdicción, el uso de información incorrecta o ficticia en un caso legal podría tener consecuencias legales graves, incluida la pérdida del caso.

2. **Responsabilidad del usuario:** Si bien ChatGPT es una herramienta poderosa, la responsabilidad final recae en el usuario que lo utiliza. En este caso, el abogado confió en el modelo de manera excesiva sin verificar la precisión de las respuestas generadas.

3. **Énfasis en la verificación:** El incidente subraya la importancia de que los usuarios verifiquen y validen la información proporcionada por modelos de lenguaje como ChatGPT, especialmente en contextos críticos como el legal. No se puede asumir que todas las respuestas generadas por el modelo son precisas.

4. **Necesidad de modelos más robustos:** Este incidente resalta la necesidad de modelos de lenguaje más robustos y confiables que puedan proporcionar respuestas precisas y respaldadas por datos verificables en contextos legales y otros contextos críticos.

En resumen, el incidente del abogado ilustra cómo las limitaciones actuales de ChatGPT, como la producción de respuestas inexactas y la falta de acceso a datos en tiempo real, pueden tener implicaciones significativas en situaciones de la vida real, especialmente en el ámbito legal y otros contextos críticos. Esto destaca la importancia de utilizar estas herramientas con precaución y de desarrollar modelos de lenguaje más confiables en el futuro.