

# Aprovechamiento de Grafos Para LLM

jzazooro

November 2023

## 1 Identificación y discusión de las imperfecciones de ChatGPT

ChatGPT, como cualquier sistema de inteligencia artificial, tiene varias imperfecciones y limitaciones. Aquí hay algunas de ellas:

- Falta de Conocimiento en Tiempo Real

Su conocimiento se basa en datos recopilados hasta enero de 2022. No tiene la capacidad de proporcionar información sobre eventos o desarrollos que hayan ocurrido después de esa fecha

- Sensibilidad al Contexto Limitada

Aunque trata de mantener el contexto en una conversación, a veces puede perder el hilo de la discusión, especialmente en conversaciones largas o cuando la información clave se encuentra lejos en el historial de la conversación.

- Generación de Contenido Inseguro o Inapropiado

Aunque hace un esfuerzo por moderar y filtrar contenido inapropiado, puede existir el riesgo de generar respuestas que no son seguras o socialmente aceptables. Siempre se debe tener precaución y revisar el contenido generado.

- Dependencia del Entrenamiento Inicial

Su conocimiento se basa en los datos con los que fue entrenado. Si bien se realizaron esfuerzos para garantizar una amplia variedad de información, puede haber sesgos inherentes a los datos de entrenamiento.

- No Discriminación de Fuentes

Durante el entrenamiento, no se discriminó entre fuentes confiables e información incorrecta. Por lo tanto, la información generada puede no ser siempre precisa o verificada.

- **Falta de Capacidad de Razonamiento Profundo**  
Aunque puede entender y generar respuestas basadas en patrones y contextos, carece de una verdadera comprensión o razonamiento profundo. No tiene la capacidad de procesar información de la misma manera que un ser humano.
- **Sensibilidad al Formato de Pregunta**  
La forma en que se hace una pregunta puede afectar a la calidad de la respuesta. Preguntas ambiguas o mal formuladas pueden generar respuestas imprecisas.
- **Manejo Limitado de Instrucciones Complejas**  
Aunque puede seguir instrucciones en lenguaje natural, su capacidad para manejar instrucciones complejas o tareas múltiples puede ser limitada.
- **Riesgo de sesgo y estereotipos**  
Puede reflejar sesgos presentes en los datos de entrenamiento, lo que podría dar lugar a respuestas que reflejen prejuicios o estereotipos existentes en la sociedad.
- **No hay Conciencia ni Emoción**  
No tiene conciencia, emociones ni intenciones. Sus respuestas se generan en función de patrones estadísticos aprendidos durante el entrenamiento.

Es importante tener en cuenta estas limitaciones al interactuar con sistemas de inteligencia artificial como ChatGPT. Aunque puede ser una herramienta útil, no sustituye la supervisión y el juicio humano.

De hecho se dió un caso en el que un abogado pregunto al chat gpt para realizar la demanda de un cliente, generando decisiones judiciales ficticias para respaldar sus argumentos.

## 2 Descripción clara del uso de grafos de conocimiento para mejorar ChatGPT

Los grafos de conocimiento son representaciones estructuradas de información que ayudan a modelar las relaciones entre diferentes entidades y conceptos. Integrar grafos de conocimiento en sistemas como ChatGPT puede mejorar significativamente la calidad de las respuestas y la comprensión del contexto. Aquí hay una descripción clara del uso de grafos de conocimiento para mejorar ChatGPT:

- **Representación Estructurada de Conocimiento:**  
Los grafos de conocimiento permiten representar la información de manera estructurada, con nodos que representan entidades y arcos que representan

relaciones entre ellas. Esto es beneficioso para modelar relaciones complejas entre conceptos, lo que puede ser útil para entender el contexto de una conversación.

- **Mejora de la Comprensión del Contexto:**

Integrar un grafo de conocimiento puede ayudar a ChatGPT a mantener un mejor seguimiento del contexto de una conversación. Puede recordar relaciones entre entidades y comprender cómo ciertos conceptos están relacionados entre sí, permitiendo respuestas más coherentes y contextualmente relevantes.

- **Resolución de Ambigüedades:**

Los grafos de conocimiento pueden ayudar a resolver ambigüedades en las preguntas proporcionando información adicional sobre las relaciones entre términos. Esto es especialmente útil en preguntas que podrían tener múltiples interpretaciones, ya que el grafo puede ayudar a refinar la comprensión y generar respuestas más precisas.

- **Ampliación del Conocimiento:**

Aunque ChatGPT tiene un amplio conocimiento previo, puede beneficiarse de la capacidad de acceder a bases de datos estructuradas a través de un grafo de conocimiento. Esto permite que el modelo responda preguntas que requieren información específica y actualizada que no estaba presente en los datos de entrenamiento originales.

- **Identificación de Conceptos Relacionados:**

Al utilizar grafos de conocimiento, ChatGPT puede identificar conceptos relacionados y ofrecer información adicional de manera más coherente. Esto mejora la capacidad del modelo para generar respuestas que aborden de manera más completa las necesidades del usuario.

- **Reducción de Sesgos mediante Contextualización:**

Integrar un grafo de conocimiento puede ayudar a contextualizar la información y reducir sesgos al proporcionar relaciones más completas entre entidades. Esto puede ayudar a evitar respuestas que reflejen sesgos presentes en los datos de entrenamiento

- **Adaptación a Preguntas Específicas:**

Los grafos de conocimiento permiten a ChatGPT adaptarse a preguntas específicas proporcionando información precisa sobre los temas abordados. Esto es particularmente útil en situaciones donde se requiere un conocimiento más profundo y detallado.

- **Validación y Verificación de Información:**

Al utilizar un grafo de conocimiento, ChatGPT puede verificar la información antes de generar respuestas, mejorando la precisión y confiabilidad de las respuestas proporcionadas.

En resumen, la incorporación de grafos de conocimiento en ChatGPT ofrece una estructura más rica y contextualizada para el conocimiento, lo que puede resultar en respuestas más informadas y precisas en una variedad de contextos de conversación.

### 3 Explicación detallada de cómo los grafos proporcionan contexto estructurado.

Los grafos proporcionan contexto estructurado al organizar la información en nodos y relaciones. Un grafo consta de nodos que representan entidades y relaciones que conectan esos nodos, ofreciendo un marco claro para modelar y entender la información. Aquí hay una explicación detallada de cómo los grafos proporcionan contexto estructurado:

- **Nodos y Entidades:**

**Nodos:** En un grafo, los nodos representan entidades o conceptos específicos. Por ejemplo, en un grafo de conocimiento sobre ciudades, puede haber nodos para ciudades como "Nueva York" o "París". **Entidades:** Cada nodo es una entidad individual. Puede contener información adicional sobre la entidad, como propiedades y atributos asociados. Por ejemplo, un nodo de ciudad puede incluir propiedades como población, ubicación geográfica, etc.

- **Relaciones y Conexiones:**

**Relaciones:** Los arcos o conexiones entre los nodos representan relaciones entre las entidades. Por ejemplo, un grafo sobre relaciones familiares puede tener relaciones como "padre de" o "hermano de" conectando nodos que representan personas. **Conexiones Dirigidas o No Dirigidas:** Las relaciones pueden ser dirigidas (con un inicio y un final) o no dirigidas, dependiendo de la naturaleza de la relación. Por ejemplo, "es hermano de" es una relación dirigida, mientras que "es amigo de" podría ser no dirigida.

- **Estructura Jerárquica y Redes Complejas:**

**Jerarquía:** Los grafos pueden tener una estructura jerárquica donde ciertos nodos actúan como nodos principales y otros como subnodos. Por ejemplo, un grafo de conocimiento sobre biología puede tener nodos principales para categorías generales como "animales" y subnodos para especies específicas. **Redes Complejas:** Además de la jerarquía, los grafos pueden formar redes complejas donde múltiples nodos están interconectados, permitiendo una representación más rica de las relaciones entre diversas entidades.

- **Contextualización y Relaciones Significativas:**

**Contexto Estructurado:** Al organizar la información en nodos y relaciones, los grafos ofrecen un contexto estructurado y significativo. Esto permite comprender cómo están conectados diferentes conceptos, lo que facilita la

interpretación del contexto en el que se encuentra una entidad. Identificación de Patrones y Tendencias: La estructura de grafo facilita la identificación de patrones y tendencias en los datos. Esto puede ser valioso para comprender las interconexiones y las dependencias entre diferentes elementos.

- **Facilitación de Consultas y Recuperación de Información:**

**Consultas Específicas:** Los grafos permiten realizar consultas específicas para obtener información sobre relaciones particulares o características de nodos. Esto facilita la recuperación de información precisa y relevante. **Eficiencia en la Búsqueda:** La estructura de grafo puede mejorar la eficiencia en la búsqueda y recuperación de información al limitar las consultas a partes específicas del grafo, evitando búsquedas exhaustivas.

- **Uso en Aplicaciones de Inteligencia Artificial:**

**Aplicaciones en Sistemas Conversacionales:** En el contexto de sistemas como ChatGPT, la incorporación de grafos de conocimiento puede permitir respuestas más informadas y contextualmente relevantes al comprender mejor las relaciones entre conceptos mencionados en la conversación.

En resumen, los grafos proporcionan un marco estructurado para organizar y entender la información, permitiendo una representación más rica y contextualizada de las relaciones entre diferentes entidades. Esto es esencial para proporcionar contexto estructurado en aplicaciones que requieren comprensión y generación de lenguaje natural.

## 4 Propuesta del ejemplo y creación del grafo de conocimiento

Esta sección se encuentra en los archivos .py del repositorio

## 5 Implementación de métricas para evaluar la precisión, relevancia y confiabilidad.

La evaluación de modelos de lenguaje, como ChatGPT, puede realizarse mediante diversas métricas que miden aspectos clave como precisión, relevancia y confiabilidad. Aquí te proporciono una descripción general de algunas métricas comunes que podrían ser útiles:

- **Precisión:**

**BLEU (Bilingual Evaluation Understudy):** Originalmente diseñada para evaluar traducciones automáticas, BLEU compara las similitudes entre las respuestas generadas y las respuestas de referencia. Cuanto más alto sea

el puntaje BLEU, más similar es la respuesta generada a las respuestas de referencia.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation): Diseñada para evaluar resúmenes automáticos, ROUGE compara la superposición de palabras y n-gramas entre las respuestas generadas y las respuestas de referencia.

- Relevancia:

F1 Score: Es una métrica que combina la precisión y el recall. En el contexto de evaluación de modelos de lenguaje, la precisión se refiere a la proporción de respuestas generadas que son relevantes, y el recall mide la proporción de respuestas relevantes que son generadas.

MRR (Mean Reciprocal Rank): Mide la calidad del ranking de las respuestas generadas. Cuanto más arriba en el ranking esté la respuesta correcta, mayor será el MRR.

- Confiabilidad:

Entropía de Respuesta: Mide la incertidumbre o la variabilidad en las respuestas generadas. Una baja entropía sugiere que el modelo es más consistente en sus respuestas.

Confianza del Modelo: Al implementar un modelo de clasificación de confianza, se puede evaluar cuánto confía el modelo en sus propias respuestas. Respuestas con alta confianza pueden considerarse más confiables.

- Evaluación Humana:

Evaluación por Pares (Human Rating): Solicitar a los evaluadores humanos que califiquen la calidad, relevancia y confiabilidad de las respuestas generadas. Esto puede proporcionar una perspectiva subjetiva y rica en detalles.

Calidad Subjetiva de Respuesta (Subjective Quality Assessment): Utilizar escalas de Likert u otros métodos para que los humanos califiquen la calidad subjetiva de las respuestas generadas.

Es importante adaptar las métricas a los objetivos específicos de tu aplicación y considerar el contexto en el que se utiliza ChatGPT. Además, ten en cuenta que la evaluación humana puede ser esencial para capturar aspectos subjetivos y complejos que las métricas automáticas pueden pasar por alto.

## 6 Reflexión y Recomendaciones

En la discusión de los resultados, se observa que ChatGPT ha demostrado un desempeño general sólido, destacando casos de éxito en respuestas relevantes y precisas. Sin embargo, se identificaron desafíos en la variabilidad de las respuestas y en la gestión de contextos complejos.

Como sugerencias para futuros desarrollos, se propone explorar la posibilidad de entrenamiento adicional y ajuste de hiperparámetros, así como la expansión y diversificación del conjunto de datos de entrenamiento. La integración de feedback de usuarios se plantea como una estrategia clave para mejorar continuamente el modelo, abordando áreas problemáticas específicas y refinando la estrategia de generación de respuestas.

Además, se sugiere la mejora de la calidad de los grafos de conocimiento, la implementación de estrategias para mitigar sesgos y mejorar la sensibilidad al contexto, y la investigación de técnicas que permitan un mayor control sobre el tipo y estilo de respuestas generadas. El desarrollo de una interfaz de usuario intuitiva y la planificación de actualizaciones periódicas del modelo son consideraciones importantes para una implementación efectiva y sostenible.

En resumen, el trabajo destaca la importancia de la mejora continua, la retroalimentación de usuarios y la adaptabilidad del modelo para mantener y elevar la calidad de las respuestas de ChatGPT en aplicaciones prácticas.