

Aplicaciones LLM

Qué son las alucinaciones y cómo se previenen

Las alucinaciones son respuestas falsas

- Una alucinación es una respuesta falsa de un LLM.
- Recuerda: un LLM no es un ser inteligente, sino un modelo estadístico entrenado para predecir la próxima palabra más probable en una frase.
- Recuerda: el nivel de acierto de un LLM depende de la calidad y cantidad de datos con los que se ha entrenado. Los Foundation LLMs tienen niveles de acierto altísimos cuando les preguntamos por las materias en las que han sido entrenados.
- El problema principal de las alucinaciones es que el LLM las dice de forma que parecen verdaderas.

¿Cómo pueden reducirse las alucinaciones?

- Regulando la “temperatura” del LLM.
 - La temperatura de los LLM regula el nivel de creatividad de sus respuestas.
 - Por ejemplo, la temperatura de chatGPT se puede regular entre 0 y 10.
 - A más temperatura, más creatividad y más alucinación. Esto puede ser bueno para que el LLM escriba una historia de ficción.
 - A menos temperatura, menos creatividad y menos alucinación. Esto es bueno para que el LLM responda a preguntas concretas.
- Afinando nuestros prompts.
 - Una forma muy eficaz es incluir ejemplos de preguntas y respuestas correctas en tu prompt para que el LLM sepa lo que debe y lo que no debe hacer.