

Below is an example of a simple interaction with the LLM, ChatGPT, on August 30, 2024:

Mark Knepper said:

Is Wikipedia used to train ChatGPT?

ChatGPT said:

Yes, Wikipedia is one of the many publicly available data sources used in training models like ChatGPT. The training process involves using a diverse range of texts from books, websites, articles, and various forms of publicly accessible information, including Wikipedia, to help the model learn language patterns, facts, and general knowledge.

However, ChatGPT does not have direct access to Wikipedia or any specific data source after the training phase is complete. It doesn't retrieve real-time information from the web, including Wikipedia. Instead, it generates responses based on patterns and knowledge it has been trained on up until its knowledge cutoff date (in this case, September 2021).

Mark Knepper said:

When will the corpus for training ChatGPT be updated?

ChatGPT said:

OpenAI periodically updates its models, but specific dates for future updates or corpus revisions are not typically announced in advance. The timing of updates depends on various factors, including advancements in model architecture, improvements in training processes, and the availability of new data. Each new iteration of models, like GPT-4, often incorporates more recent data, enhanced training techniques, and feedback from users to improve performance.

While I don't have information on when the next training update will occur, OpenAI is committed to continually improving its models over time.