



redmarble.ai

AI RESEARCH FROM THE LAB



RESEARCH FOCUS:

GPT4all

Can you run GPT locally?

OVERVIEW

Our research team recently had the opportunity to investigate a new technology called **GPT4all**, which is a promising development in the field of language models. What sets **GPT4all** apart is its ability to be downloaded and run on your own computer. This means that you can use it **offline**, which is a great option for users who need to work without an internet connection.

KEY CONCEPTS

GPT4All is an open-source software ecosystem designed to enable individuals to train and utilise powerful and customised large language models on everyday hardware. It is optimised to run inference on CPUs of laptops, desktops, and servers, allowing for efficient deployment of language models with 7-13 billion parameters. The project is overseen by Nomic AI, ensuring quality, security, and maintainability of contributions to the ecosystem.

DEEP DIVE

The creators of GPT4all have developed the model by fine-tuning Meta's Llama 7B with 1 million question-answer pairs. To clean the data, they used a software called Atlas, which is a large text-database visualisation tool. They claim to have removed all malformed data from the set but this seems unlikely as there is no mention of any manual data cleaning in the report.



SETUP & TESTING

The installation of GPT4All was straightforward, and the model, occupying around 4GB, ran without any issues. The initial prompt produced a coherent response within 15 seconds. However, further testing revealed limitations. When queried for factual information, the model provided inaccurate answers and demonstrated outdated knowledge. It became evident that the model had limitations in terms of its currency and factual reliability.

Additionally, the model displayed a tendency to generate hallucinatory and nonsensical responses, similar to earlier versions of OpenAI's transformers. As more questions were asked, the model's responses became increasingly bizarre and detached from reality. Overall, due to its unpredictable behaviour and watered-down nature compared to existing OpenAI models, the GPT4All model was deemed unsuitable for practical use.



HITTING THE LIMITS

The limitations of the model must be taken into account. For example, the model is not up to date and cannot be trusted to always provide factual information. Additionally, the model tends to produce "hallucinations" and becomes confused with the tasks it is presented.



WHAT'S NEXT?

The testing of GPT4All revealed several limitations and pitfalls that need to be taken into consideration before proceeding further. Its performance fell significantly short of the latest models from OpenAI, making it unsuitable for practical use. Additionally, the data extraction process employed by GPT4All raised concerns about its reliability and compliance with OpenAI's terms of use.

Moving forward, it's crucial to consider alternative avenues for language model development and exploration. Exploring advancements made by OpenAI and other organisations in the field of large language models will provide valuable insights and inspiration. We will keep a close watch on emerging research, developments in natural language processing, and potential policy changes regarding the usage of AI models.

What's the verdict?

REAL TALK

Overall, while **GPT4all** has potential, it is not yet at a level where it can compete with the latest models from OpenAI. As the technology continues to develop, We are excited to see what new advancements will be made.

