

LLM Models Explored

Serve AI

Overview

- A **Large Language Model (LLM)** is a type of artificial intelligence algorithm that uses deep learning techniques and massively large datasets to understand, summarize, generate and predict new content (Kerner, 2024).
- This exploration aims to provide a brief description on the top LLMs in the AI industry and evaluate them based on their accuracy and ethicality of the information presented to the general public.
- This set of slides will mainly focus on the top and most popular LLMs as of July, 2024 so please note that the list is subject to change.



ChatGPT

- Created by **OpenAI** in November 2022
- The information that is produced by ChatGPT is publicly available on the internet from the Common Crawl dataset and sites like Wikipedia, licensed information from third-parties, and human-provided information
- The basic version of GPT is free, but ChatGPT Plus/4.0 is \$20/mo



ChatGPT

Pros	Cons
Highly efficient in automating repetitive tasks, such as answering frequently asked questions, reducing workload and human error.	Handling sensitive data raises privacy and security concerns, requiring robust data protection measures.
Can handle numerous queries simultaneously, making it ideal for fluctuating workloads.	Provides incorrect information, especially when dealing with ambiguous or unfamiliar queries.
Can communicate in multiple languages, broadening its reach to a global audience.	Lacks emotional understanding, making it less effective in handling emotionally charged or complex situations.

Gemini (formerly known as Bard)

- Created by **Google** in March 2023
- The information that is produced by Gemini is comprised of web pages available on the internet from the Common Crawl dataset and sites like Wikipedia, scientific papers, mathematical equations, and human and programming languages
- The basic version of Gemini is free, but Gemini Advanced is \$20/mo



Gemini (formerly known as Bard)

Pros	Cons
Better accessibility features such as a speech-to-text feature and vice-versa wherein Gemini can read your AI-generated response back to you.	As part of Google, it may raise concerns about data privacy and how user data is utilized.
Can search the internet and cite its sources. Gemini can refer you to various pieces of data that it used in creating its response.	Heavy dependence on Google's ecosystem might be a drawback for users seeking independence from Google's infrastructure and access may be restricted or limited based on specific Google services.
Benefits from Google's advanced search capabilities, offering more accurate and relevant information retrieval.	While innovative, the speed of new feature rollouts may be slower due to the scale and complexity of Google's infrastructure.

Claude

- Created by **Anthropic AI** in March 2023
- The information that is produced by is publicly available on the internet from the Common Crawl dataset and sites like Wikipedia
- The basic version of is free, but is \$20/mo



Claude

Pros	Cons
Claude can assist with coding, provide data summaries, and even summarize user-provided PDF files, making it a versatile tool for various tasks.	Claude is not connected to the internet, which limits its ability to provide real-time information or perform web-based tasks.
The summaries of large amounts of data are better than other LLMs, providing users with reliable information.	There is a limit on how many responses can be obtained at one time, especially in the free version, requiring users to be mindful of their usage.
Claude ranks as one of the best LLMs in terms of maintaining security and privacy of information.	Customization options are limited compared to some open-source models.

Pi AI

- Created by **Inflexion AI** in May 2023
- The information that is produced by is publicly available on the internet from the Common Crawl dataset and sites like Wikipedia, other language models, and user feedback.
- Currently Pi AI is completely free to use, but it is possible that in the future, Inflexion AI may introduce additional features or tiers that could come with a cost



Pi AI

Pros	Cons
Pi is the most conversational of the AI chatbots, using a friendly and engaging tone, often including emojis, which can make interactions more pleasant and enjoyable.	Pi is not as proficient when it comes to assisting with coding or summarizing data in various formats, limiting its usefulness for technical tasks.
With a character limit increased from 1000 to 4000, users can provide more detailed prompts, allowing for more comprehensive conversations within a single response.	Despite the increased character limit, Pi still cannot analyze large amounts of data at once, which can be a significant drawback for users needing in-depth data analysis.
Pi excels in discussions about life and career topics, offering a sense of empathy and understanding that can be comforting for users seeking emotional support.	Pi's friendly demeanor might lead some users to rely on it for emotional support, but it is important to remember that it is not a substitute for professional help in serious matters.

Conclusion

- It is important to note that even though ChatGPT stands out for its accessibility, user-friendliness, and strong community support, other LLMs may excel in specific areas such as accuracy, contextual understanding, and domain specialization.
- All four chatbots have their merits and their limitations, so it largely depends on the task at hand. Choosing the right LLM depends on the specific needs and context of the application which is important to keep in mind moving forward.



Links

<https://www.techtarget.com/whatis/definition/large-language-model-LLM>

<https://www.techtarget.com/whatis/feature/12-of-the-best-large-language-models>

<https://chatgpt.com/>

<https://medium.com/@jerryngoma6/5-advantages-and-disadvantages-of-using-chat-gpt-b5fdefd5f234>

<https://www.coursera.org/articles/bard-vs-chatgpt>

<https://www.linkedin.com/pulse/which-generative-ai-best-chatgpt-vs-gemini-vs-claude-2-alena-gorb-z6nhe/>