Large Language Models

1. GPT-4

GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. It is a large multimodal model (accepting image and text inputs, emitting text outputs) that can solve difficult problems with greater accuracy than any of their previous models, thanks to its broader general knowledge and advanced reasoning capabilities.

Use Cases
GPT-4 is more creative and collaborative than ever before. It can generate, edit, and iterate with users on creative and technical writing tasks, such as composing songs, writing screenplays, or learning a user's writing style.

Safety Measures
OpenAI spent 6 months iteratively aligning GPT-4 using lessons from their adversarial testing program as well as ChatGPT, resulting in their best-ever results (though far from perfect) on factuality, steerability, and refusing to go outside of guardrails.

API Availability
Yes, GPT-4 can be accessed via an API.

2. AlphaCode

AlphaCode is the first AI code generation system that has "reached a competitive level of performance in programming competitions." It uses a large language model to build code in response to natural language descriptions of a problem1.

Use Cases
AlphaCode is capable of generating code at an unprecedented scale, and then smartly filters to a small set of promising programs.

3. GitHub Copilot

GitHub Copilot is an AI pair programmer that offers autocomplete-style suggestions as you code. It is powered by OpenAI Codex, a new AI system created by OpenAI.

Cost

GitHub Copilot is free to use for verified students, teachers, and maintainers of popular open source projects. If you are not a student, teacher, or maintainer of a popular open source project, you can try GitHub Copilot for free with a one-time 30-day trial. After the free trial, you will need a paid subscription for continued use. For individuals, it costs $10 per month or $100 per year. For businesses, it costs $19 per user per month.

Training Data

GitHub Copilot is trained on all languages that appear in public repositories. For each language, the quality of suggestions you receive may depend on the volume and diversity of training data for that language.

Use Cases

GitHub Copilot can help you code by offering autocomplete-style suggestions. You can receive suggestions from GitHub Copilot either by starting to write the code you want to use or by writing a natural language comment describing what you want the code to do.

Safety Measures

GitHub Copilot includes a filter that detects code suggestions matching public code on GitHub. You can choose to enable or disable the filter3. The training set for GitHub Copilot may contain insecure coding patterns, bugs, or references to outdated APIs or idioms. When GitHub Copilot produces suggestions based on this training data, those suggestions may also contain undesirable patterns.

4. Claude

Claude is a next-generation AI assistant based on Anthropic's research into training helpful, honest, and harmless AI systems. It is accessible through chat interface and API in their developer console.

Token Limit

Users can input up to 100K tokens in each prompt, which means that Claude can work over hundreds of pages of technical documentation or even a book.

Use Cases

Claude is capable of a wide variety of conversational and text processing tasks while maintaining a high degree of reliability and predictability. It can help with use cases including summarization, search, creative and collaborative writing, Q&A, coding, and more.

Safety Measures

Anthropic has been iterating to improve the underlying safety of Claude 2, so that it is more harmless and harder to prompt to produce offensive or dangerous output. They have an internal red-teaming evaluation that scores their models on a large representative set of harmful prompts, using an automated test while they also regularly check the results manually.

API Availability

Yes, Claude can be accessed via an API.

5. Cohere

Cohere Generate is a feature of Cohere that allows you to generate responses based on the data stored in your Weaviate instance. It harnesses the power of Cohere's API to access pre-trained language models for AI-driven text generation.

Use Cases

Cohere Generate produces unique content for emails, landing pages, product descriptions, and more.

API Availability

Yes, Cohere Generate can be accessed via an API.

6. PaLM

PaLM stands for "Pathways Language Model" and is a next-generation large language model that builds on Google's legacy of breakthrough research in machine learning and responsible AI. It excels at advanced reasoning tasks, including code and math, classification and question answering, translation and multilingual proficiency, and natural language generation better than their previous state-of-the-art LLMs.

Model Size

PaLM 2 is a 540-billion parameter, dense decoder-only Transformer model trained with the Pathways system.

Training Data

PaLM was trained using a combination of English and multilingual datasets that include high-quality web documents, books, Wikipedia, conversations, and GitHub code.

Use Cases

PaLM excels at advanced reasoning tasks, including code and math, classification and question answering, translation and multilingual proficiency, and natural language generation.

API Availability

Yes, PaLM can be accessed via an API.