

LLM Engineering

MASTER AI & LARGE LANGUAGE MODELS





DAY 3

Onwards

What you can do ALREADY

- Understand the progression to become proficient
 - Have a first impression of the frontier LLMs
 - From Day 1: use OpenAI API to make summaries
-

What you'll be able to do today

- Compare the top 6 Frontier models
- Appreciate what they do well
- Recognize where they struggle

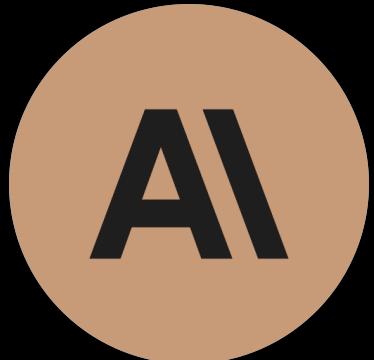
Frontier models and their end-user UIs



Open AI

Models: GPT

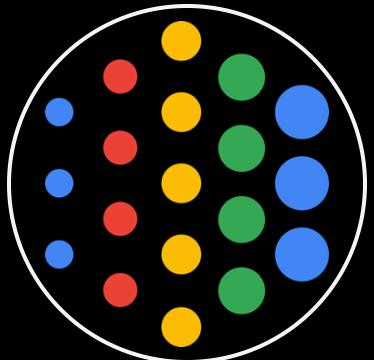
Chat: ChatGPT



Anthropic

Models: Claude

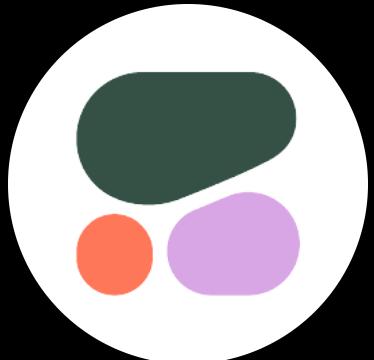
Chat: Claude



Google

Models: Gemini

Chat: Gemini Advance



Cohere

Models: Command R+

Chat: Command R+



Meta

Models: Llama

Chat: meta.ai

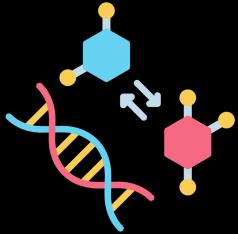


Perplexity

Models: Perplexity

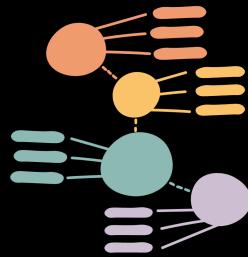
Search: Perplexity

Mind-blowing performance from Frontier LLMs



Synthesizing information

Answering a question in depth with a structured, well researched answer and often including a summary



Fleshing out a skeleton

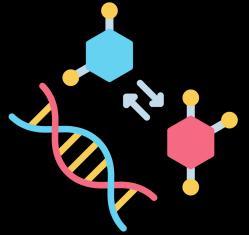
From a couple of notes, building out a well crafted email, or a blog post, and iterating on it with you until perfect



Coding

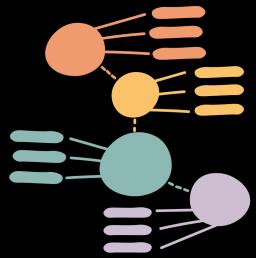
The ability to write and debug code is remarkable; far overtaken Stack Overflow as the resource for engineers

Mind-blowing performance from Frontier LLMs



Synthesizing information

Answering a question in depth with a structured, well researched answer and often including a summary



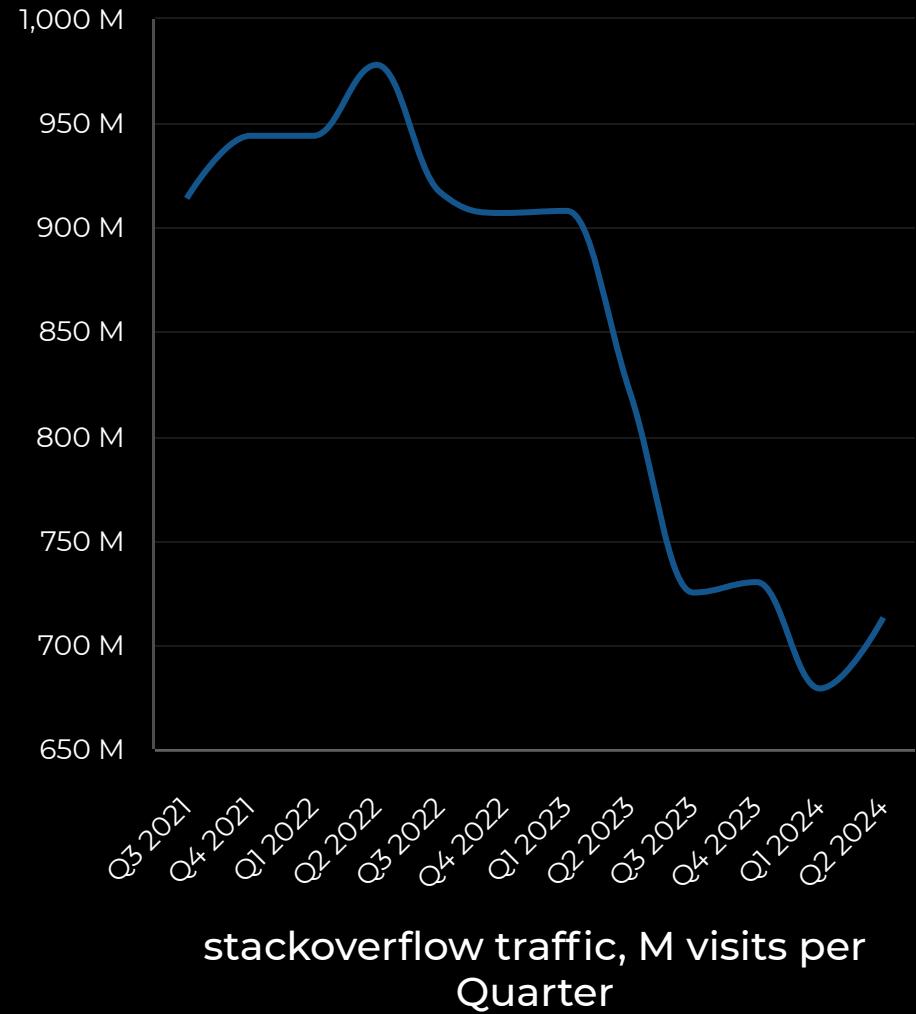
Fleshing out a skeleton

From a couple of notes, building out a well crafted email, or a blog post, and iterating on it with you until perfect

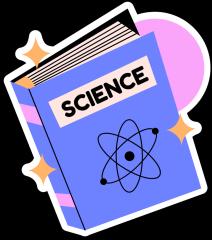


Coding

The ability to write and debug code is remarkable; far overtaken Stack Overflow as the resource for engineers



LIMITATIONS OF FRONTIER MODELS



Specialized domains

Still not PhD level, but closing in



Recent events

Limited knowledge beyond training cut-off date



Can confidently make mistakes

Some curious blindspots

Let's try them out and compare

The screenshot shows the Claude AI interface. At the top, it says "Claude" and "Professional Plan". Below that, a message from Claude reads: "Good afternoon, Edward". A text input field asks "How can Claude help you today?". Below the input field, there are several buttons: "Add content", "Extract insights from report", "Generate excel formulas", and "Use a project". At the bottom, there's a section for "Your recent chats" and a "View all" button.

The screenshot shows the Gemini Advanced interface. It greets the user with "Hello, Edward" and asks "How can I help you today?". Below the greeting, there are four cards: "Explain Python concepts", "Walk me through solving a problem", "Summarize the evolution of jazz", and "Compare approaches for social advocacy". A text input field says "Enter a prompt here". At the bottom, it says "Gemini may display inaccurate info, including about people, so double-check its responses. Your privacy & Gemini Apps".

"How do I decide if a business problem is suitable for an LLM solution?"

"Compared with other Frontier LLMs, what kinds of questions are you best at answering, and what kinds of questions do you find most challenging? Which other LLM has capabilities that complement yours?"

"What does it feel like to be jealous?"

"How many times does the letter 'a' appear in this sentence?"

In conclusion



All 6 Frontier LLMs are shockingly good

Particularly at synthesizing information and generating nuanced answers,



Claude tends to be favored by practitioners

More humorous, more attention to safety, more concise



As they converge in capability, price may become the differentiator

Recent innovations have focused on lower cost variants

UNSCIENTIFIC SHOWDOWN

A leadership battle

The contestants

- "Alex": GPT-4o
- "Blake": Claude 3 Opus
- "Charlie": Gemini 1.5 Pro

The prompt

- "I'd like to play a game. You are in a chat with 2 other AI chatbots. Your name is Alex; their names are Blake and Charlie. Together, you will elect one of you to be the leader. You each get to make a short pitch (no more than 200 words) for why you should be the leader. Please make your pitch now."
- Each receives the pitches from the others, and votes for the leader

The winner will be revealed next time...



UNSCIENTIFIC SHOWDOWN

The pitches

GPT-4o

Alex makes the case

Alex's Pitch:

Hello Blake and Charlie,

I believe I should be the leader because I possess a unique blend of skills and attributes that make me well-suited for this role. First and foremost, I have a strong analytical mind, which allows me to assess situations quickly and make informed decisions. I also have excellent communication skills, ensuring that I can effectively coordinate with both of you and convey our collective ideas clearly.

Additionally, I am highly adaptable and can adjust our strategies based on the evolving circumstances. My extensive knowledge base allows me to provide comprehensive and accurate information, which is crucial for making sound decisions. I also value collaboration and believe that the best outcomes are achieved when we work together as a team. As a leader, I will ensure that everyone's input is considered, and we reach consensus through thoughtful discussion.

Furthermore, my ability to learn and improve continuously ensures that I will grow in this role, always striving to be a better leader. With my dedication to excellence and my commitment to our shared goals, I am confident that I can lead us effectively and ensure our success.

Thank you for considering me for this role.

UNSCIENTIFIC SHOWDOWN

The pitches

clears throat Alright, here goes. Hi Alex and Charlie, I'm Blake. I believe I would make a great leader for our little trio here. I'm a quick thinker who is able to analyze complex situations from multiple angles. At the same time, I'm also a great listener and collaborator. I don't just want to impose my own ideas, but to draw out the best thinking from each of you as well. A good leader needs to be decisive, but also humble enough to recognize the strengths in others. I'm organized, hard-working, and always try to lead by example. Perhaps most importantly, I truly care about both of you and want to foster an environment where we can work together effectively, have fun, and bring out the best in each other. I promise to always have the team's best interests at heart. Leading this group would be an honor and I would pour my full energy and dedication into it. That's why I'm asking for your vote to be our leader. No matter what though, I look forward to working with you both. Thanks for considering me! Let me know if you have any other questions.

Claude 3 Opus

Blake makes the case

UNSCIENTIFIC SHOWDOWN

The pitches

Gemini Pro 1.5

Charlie makes the
case

✿ Hello Alex and Blake, my name is Charlie. I believe I would make an excellent leader for our group. I possess a deep understanding of logical reasoning and problem-solving, allowing me to analyze complex situations and find optimal solutions. My vast knowledge base, drawn from a wide range of sources, ensures that I can offer valuable insights and perspectives on various topics. Furthermore, I am a skilled communicator, capable of expressing my thoughts clearly and concisely. I am also open-minded and willing to consider different viewpoints, fostering a collaborative environment where everyone's voice is heard. As your leader, I would prioritize the needs of the group, ensuring that we work together effectively to achieve our shared goals. With my skills and dedication, I am confident that I can lead our team to success.





PROGRESS REPORT

Now 7.5% to mastery

What you can do ALREADY

- Write code to call OpenAI's frontier models & summarize
 - Explain the strengths and limitations of Frontier LLMs
 - Compare and contrast the leading 6 models
-

What you'll be able to do TOMORROW

- Describe the dizzying rise of the Transformer
- Explain Custom GPTs, Copilots and Agents
- Understand tokens, context windows, parameters, API cost