

LLMs Unleashed: A New Chapter in Clinical Programming?



November 4, 2023

Phuse EU Connect

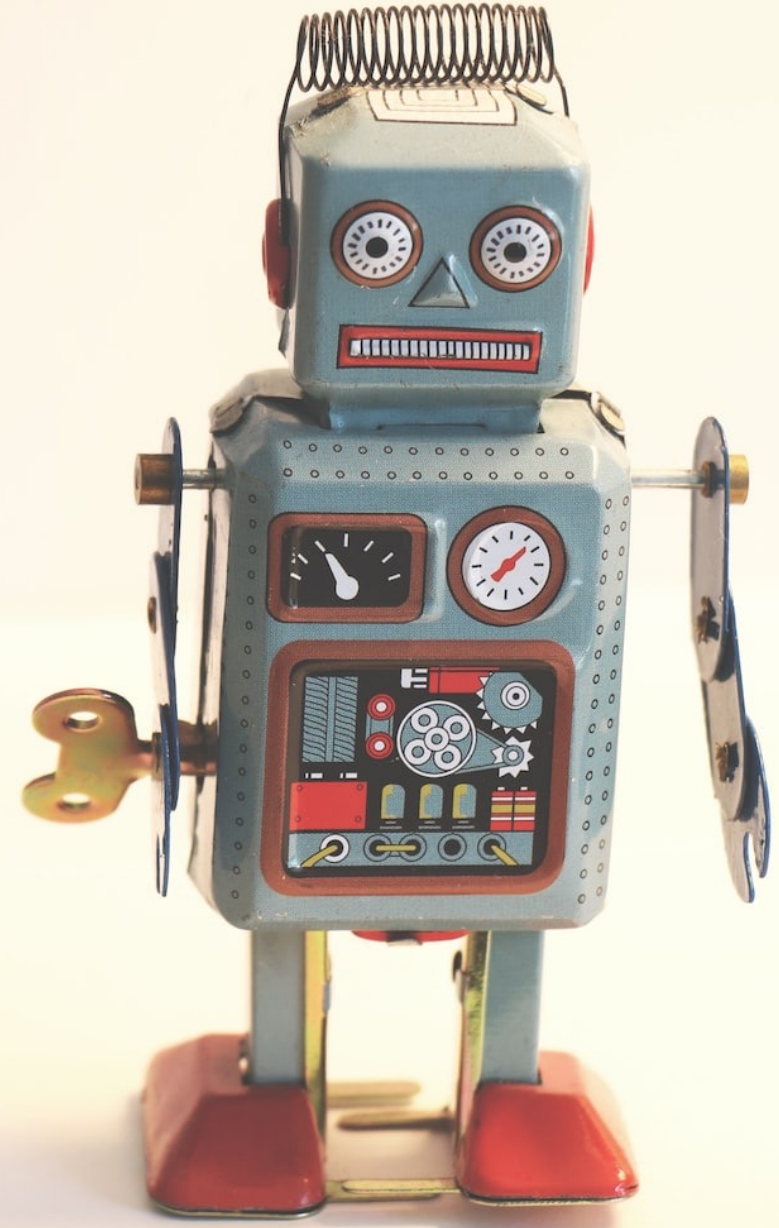
Karma Tarap

The Dawn of a New Era: Large Language Models

1. What Are Large Language Models?
2. How we interact with them
3. Applications & Considerations in Clinical Programming
4. A Glimpse Into the Future

What Are Language Models?

- Computational algorithms that understand, generate, or manipulate human language.
- Have evolved from simple n-gram models to complex architectures like RNNs, LSTMs, and Transformers.
- "Attention Is All You Need" paper revolutionized the field, giving rise to BERT, GPT-2, and GPT-3.
- Transformers enable parallel processing and more effective handling of long-term dependencies (The T in GPT)

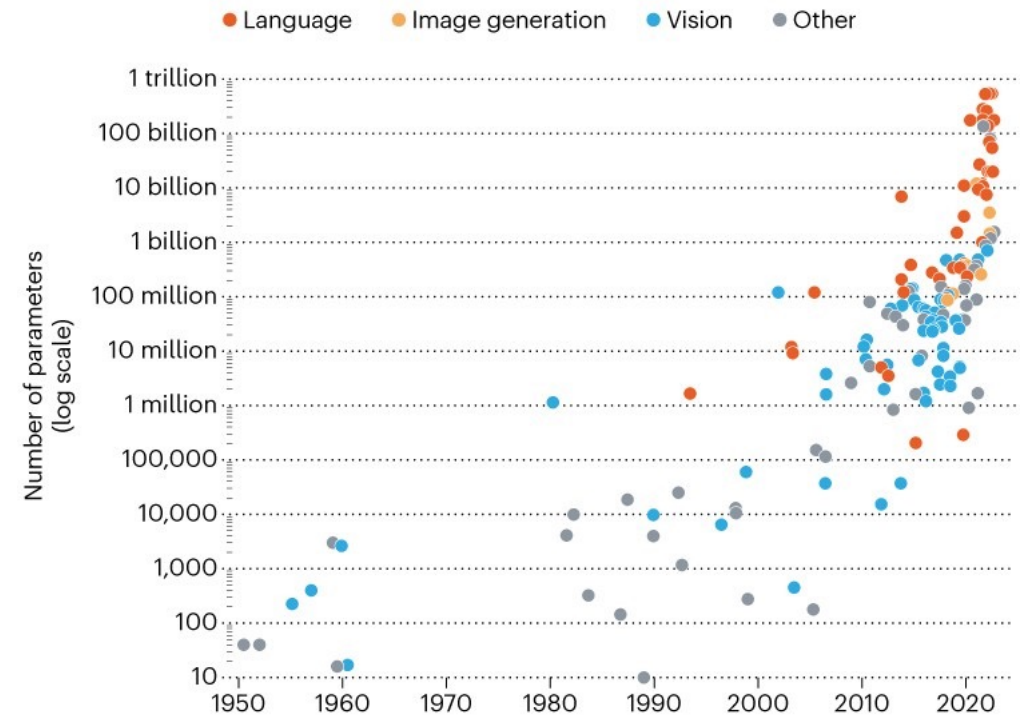


Rise of the Parameters

- Parameter count is a key metric for evaluating LLMs.
- GPT-3, with 175 billion parameters, demonstrates performance gains in natural language understanding.

THE DRIVE TO BIGGER AI MODELS

The scale of artificial-intelligence neural networks is growing exponentially, as measured by the models' parameters (roughly, the number of connections between their neurons)*.

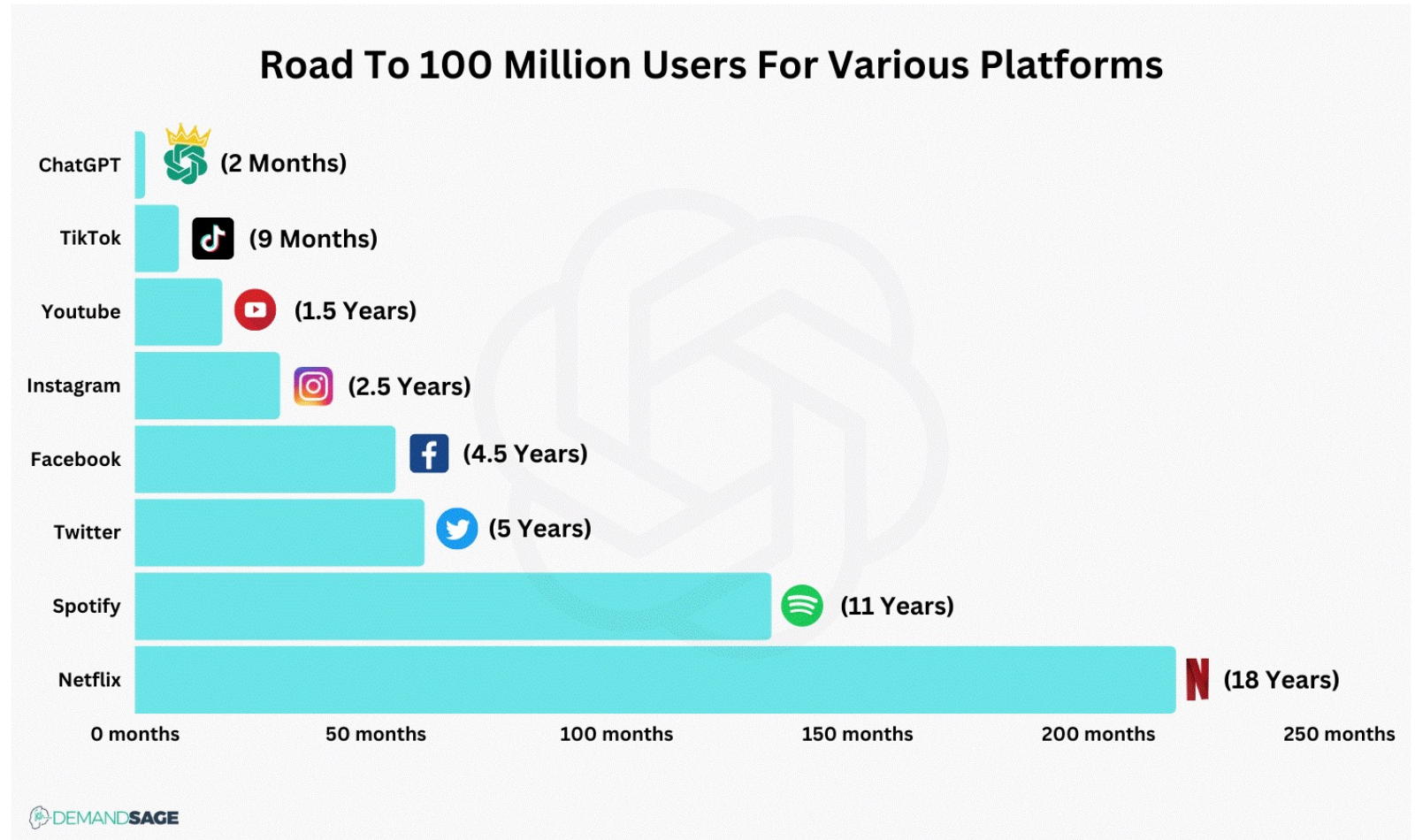


*'Sparse' models, which have more than one trillion parameters but use only a fraction of them in each computation, are not shown.

©nature

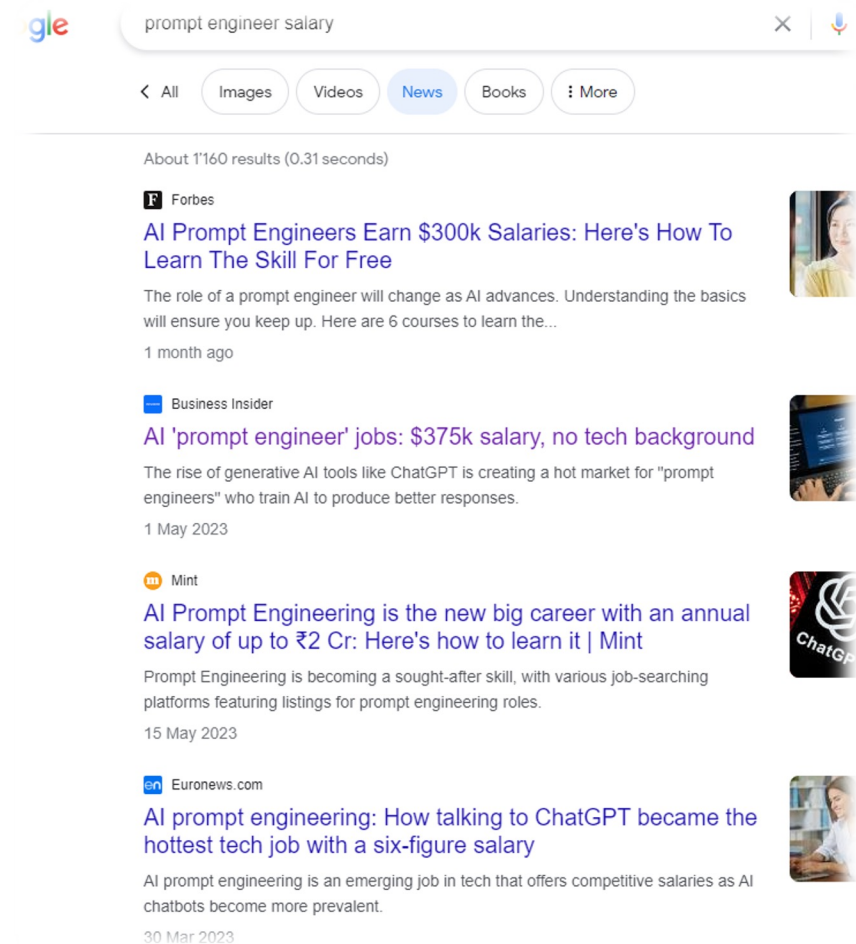
Public's Imagination: The ChatGPT Phenomenon

- ChatGPT gained 100M users within 2 months of its release.
- Its diverse knowledge base and human-like reasoning make it stand out.
- Built on the same architecture as GPT-3, but fine-tuned differently, making it more adaptable.
- Parameters alone are not the answer.



The New Conversation—Prompting

- Prompts are how we interact with LLMs
- Prompt engineering involves crafting questions or instructions for accurate and useful model outputs.
- This skill is becoming increasingly valuable, giving rise to roles like "Prompt Engineers."
- The type of prompt can dramatically affect the model's performance.



What's in a Prompt? The Art of Asking

TECHNIQUE	DESCRIPTION	EXAMPLE
Zero-shot	Straight-up query.	"What is the capital of France?"
Few-shot	Providing previous examples to guide the model.	"Question 1: What is the age of the patient? Answer: 45 Question 2: What are the symptoms? Answer: Fever and cough"
Reflexion	Forces self-reflection on the model's part. This technique improved GPT-4's accuracy from 80% to 91%.	"Review your response and propose any modifications that can improve your response."
Chain of Thought	Long-form, multi-step reasoning	The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2, 1. A: Adding all the odd numbers (9, 15, 1) gives 25. The answer is False. The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1. A: Imagine three different experts are answering this question. All experts will write down 1 step of their thinking, then share it with the group. Then all experts will go on to the next step, etc. If any expert realizes they're wrong at any point then they leave. The question is...
Tree of Thought	Branching paths of reasoning.	

A Real Challenge—Data Visualization

Goal: Allow users to rapidly analyze SDTM data without knowledge of programming or the data structures



Confidentiality - We cannot share patient data

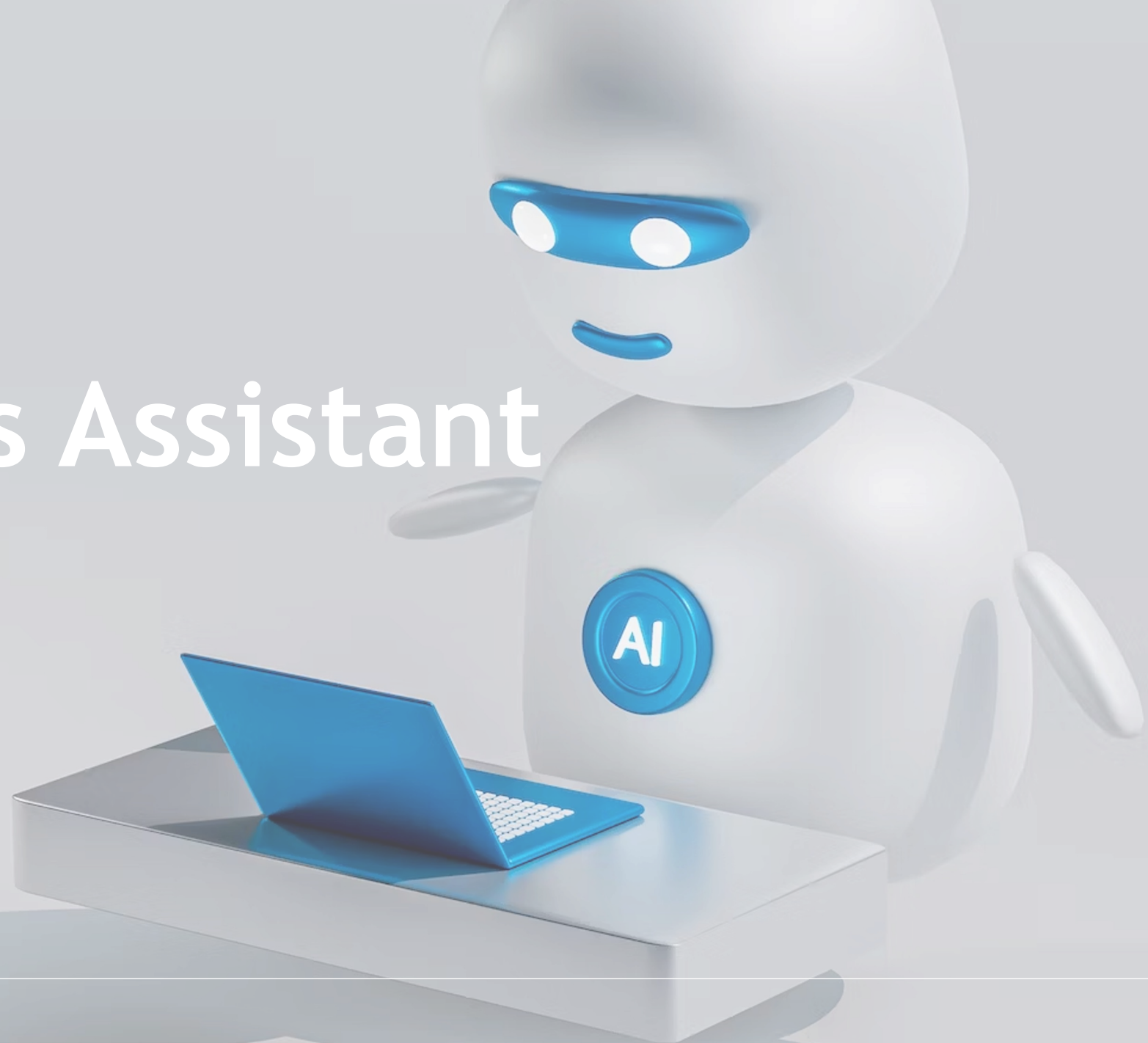


Accuracy - The generated outputs must match what is in our data

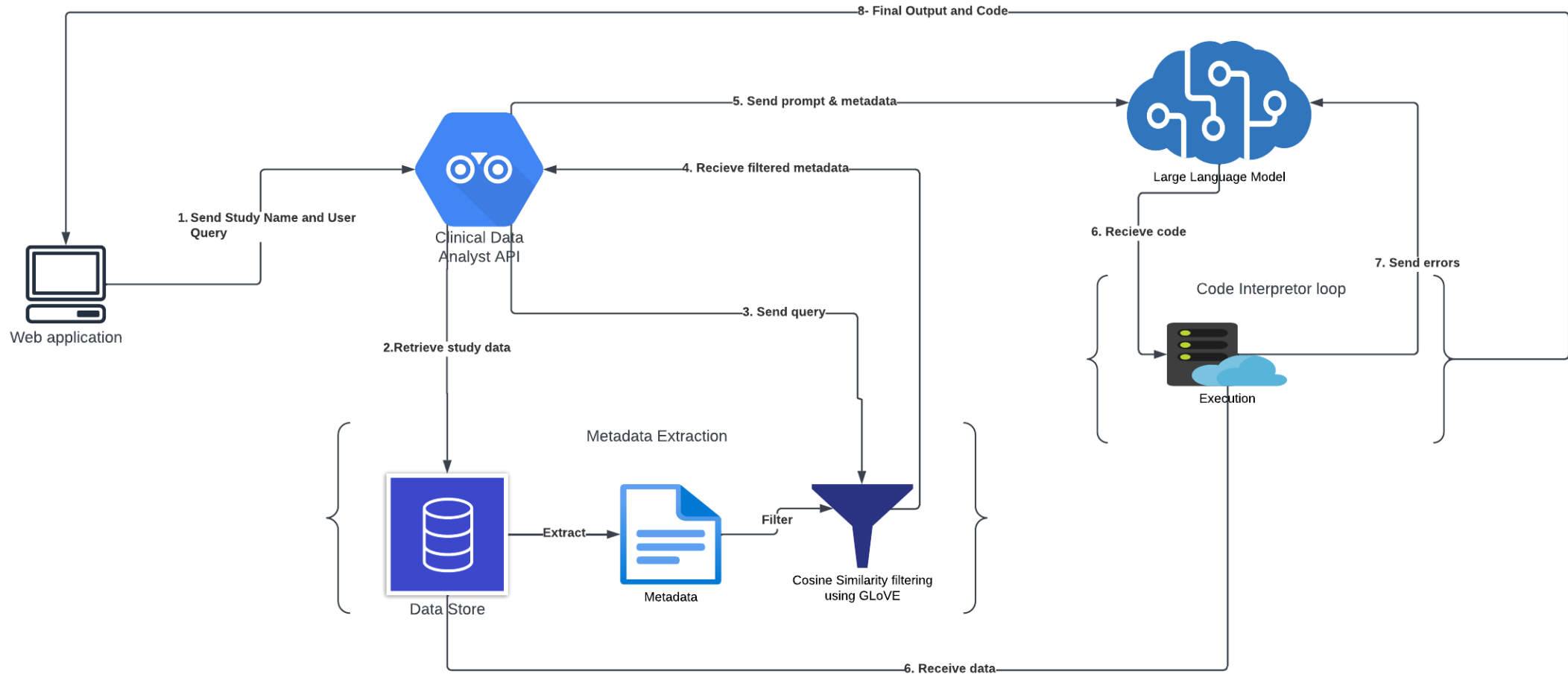


Reproducibility - The output should be reproducible within our SCE

Demo: Analytics Assistant

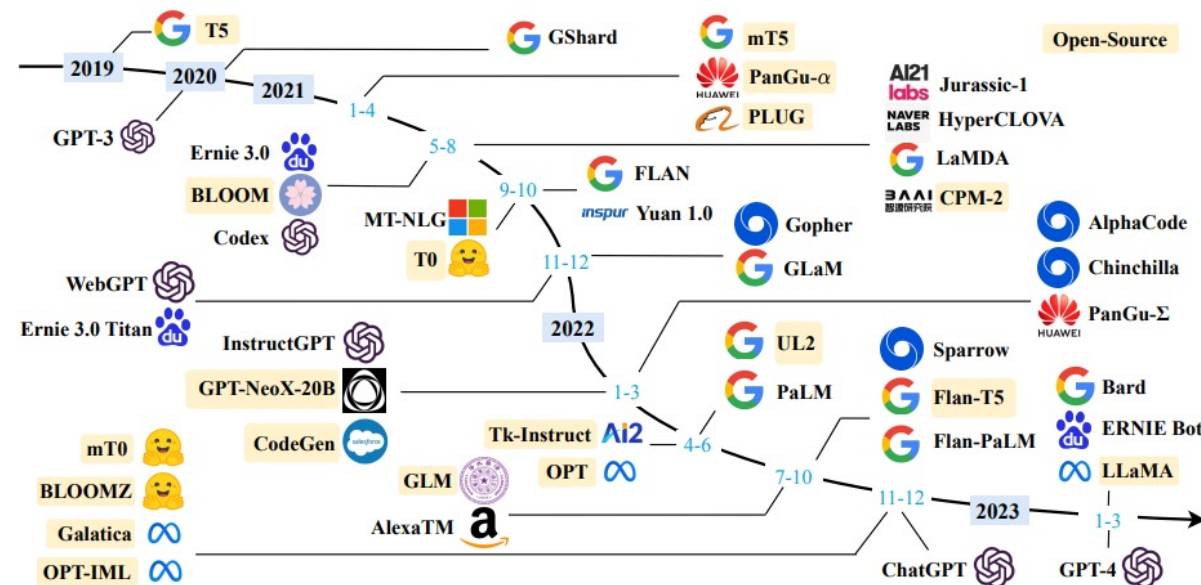


How does Analytics Assistant work

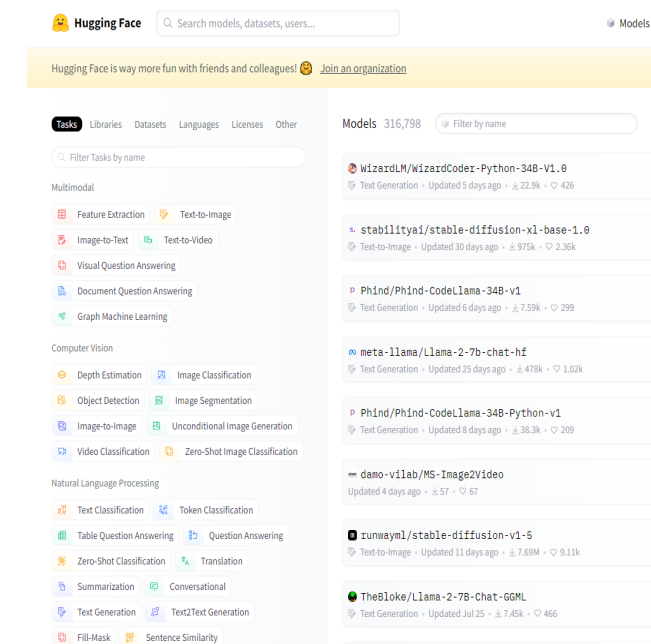


An Explosion[💣] of LLMs

- The landscape is continuously evolving with the rise of open-source models.
- Opensource models are critical for use cases within the pharmaceutical industry with confidential data.
- Opensource models are reaching or surpassing GPT4 on certain benchmarks.

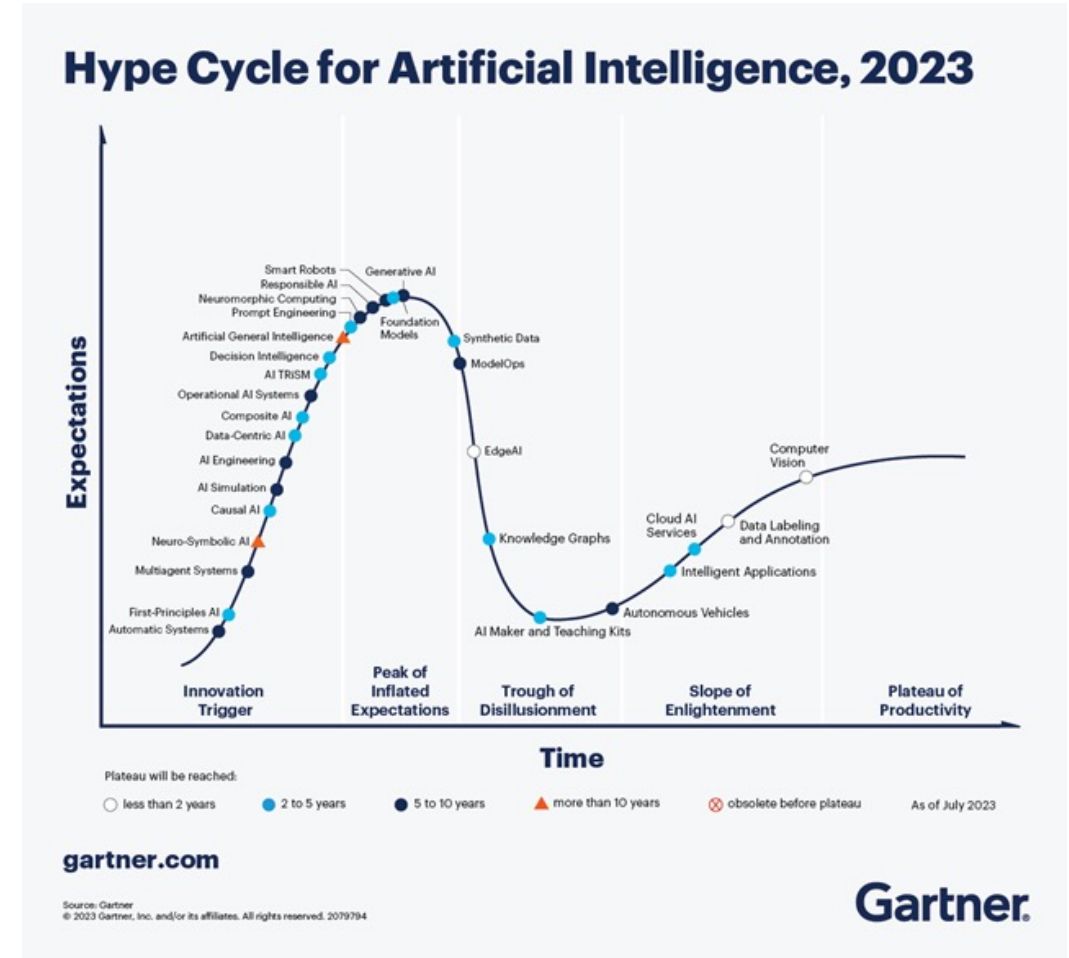


Model	Params	HumanEval
Closed-source models		
PaLM-Coder	540B	36.0
PaLM 2-S	?	37.6
Code-Davinci-002	-	47.0
PanguCoder2	15B	61.6
Unnatural Code Llama	34B	62.2
Claude 1.3	?	56.0
Claude 2	?	71.2
GPT-3.5 (2023/03/15)	?	48.1
GPT-4 (2023/03/15)	?	67.0
GPT-3.5 (2023/08/26)	?	72.5
GPT-4 (2023/08/26)	?	82.0
Open-source models		
StarCoder	15B	33.6
InstructCodeT5+	16B	35.0
CodeGen2.5-instruct	7B	36.2
Lemur	70B	38.0
Code Llama-Instruct	34B	41.5
OctoCoder	15B	46.2
Code Llama	34B	48.8
WizardLM v1.0	70B	50.6
Code Llama-Python	34B	53.7
WizardCoder V1.0	15B	59.8
LemurChat	70B	61.6
Phind-CodeLlama-V1	34B	67.6
Phind-CodeLlama-Python-V1	34B	69.5
SOTA Open-source model		
WizardCoder-Python V1.0	34B	73.2



Proceed with Caution: The Hype vs Reality

- Generative AI and prompt engineering are at the peak of Gartner's hype cycle.
- It's crucial to have balanced expectations and focus on solving concrete use cases.
- Key Challenges must be addressed!
 - Reliability (model hallucinations)
 - Better Prompting
 - Fine tuned models
 - Human checks and balances
 - Infrastructure
 - Company Specific
 - Model size
 - Llama.cpp



Bright Horizons: The New Era of Language Models

- **Rapid Technological Advancements:** We're at an unprecedented point in the capabilities of language models.
- **Accessibility and Ease:** Simple prompts can now achieve what once required extensive fine-tuning.
- **Imperfect but Impactful:** Even with limitations, these tools can be extremely useful.
- **Resources Abound:** Numerous guides, courses, and communities are available to help you start your projects.
- **Have Fun Hacking 🐱** : I look forward to seeing the amazing tools you will build!

Parting Shot: A Programmer's Joke from GPT4

Why did the SAS
programmer bring a
ladder to work?

Parting Shot: A Programmer's Joke from GPT4

Why did the SAS programmer bring a ladder to work?

Because he heard it was good for 'data step'ping!

Thank you for your Attention!

Useful Resources

- [A history of LLMs](#)
- [Prompt engineering breakthroughs](#)
- [Open source models](#)

