

# About LLMs 2026

---

Origins, Latest Developments

# Administrative Details

- Attendance secret code!
  - <https://canvas.stanford.edu/courses/221239/quizzes/184660>
- Welcome to 420-040
- Continue forming teams! Due on Thursday 1/15
- More 2-3 minute project pitches
- Today we review LLMs



# Latest Developments

- Reasoning Models - Deepseek R1
  - Now reasoning levels integrated in APIs
- Multimodal workflows
  - Generate an Image, add a script, generate a clip, add music, a narration
- Agentic Ensembles
  - Orchestrator, specialized models and tools <https://sup.ai>
- Codebase Generation
  - Including basic devops functions
- Cutting Edge Maths Proofs
  - Erdős (AIR-dish) problem 728 solved by generating Lean proof
  - <https://mathstodon.xyz/@tao/115855840223258103>

# Back to Basics

Slide content due to:

- Monica Lam
- Eric Wallace
- Andrej Karpathy

# Language Modeling

$$p(x_1, x_2, \dots, x_L)$$

# Language Modeling

$$p(x_1, x_2, \dots, x_L)$$

$$p(\text{the}, \text{mouse}, \text{ate}, \text{the}, \text{cheese}) = 0.001$$

$$p(\text{mouse}, \text{the}, \text{cheese}, \text{ate}, \text{the}) = 10^{-20}$$

# Language Modeling

$$p(x_1, x_2, \dots, x_L)$$

=

$$\prod_{i=1}^L p(x_i \mid x_{1:i-1})$$

# Neural Language Models

---

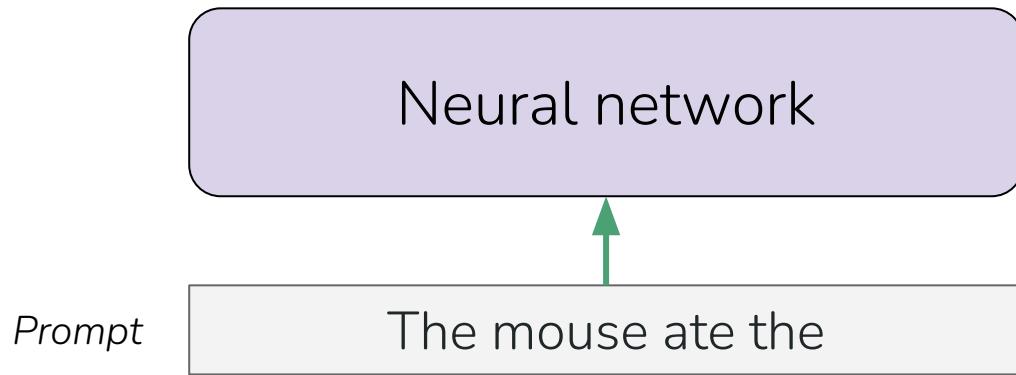
*Prompt*

The mouse ate the

$$\prod_{i=1}^L p(x_i \mid x_{1:i-1})$$

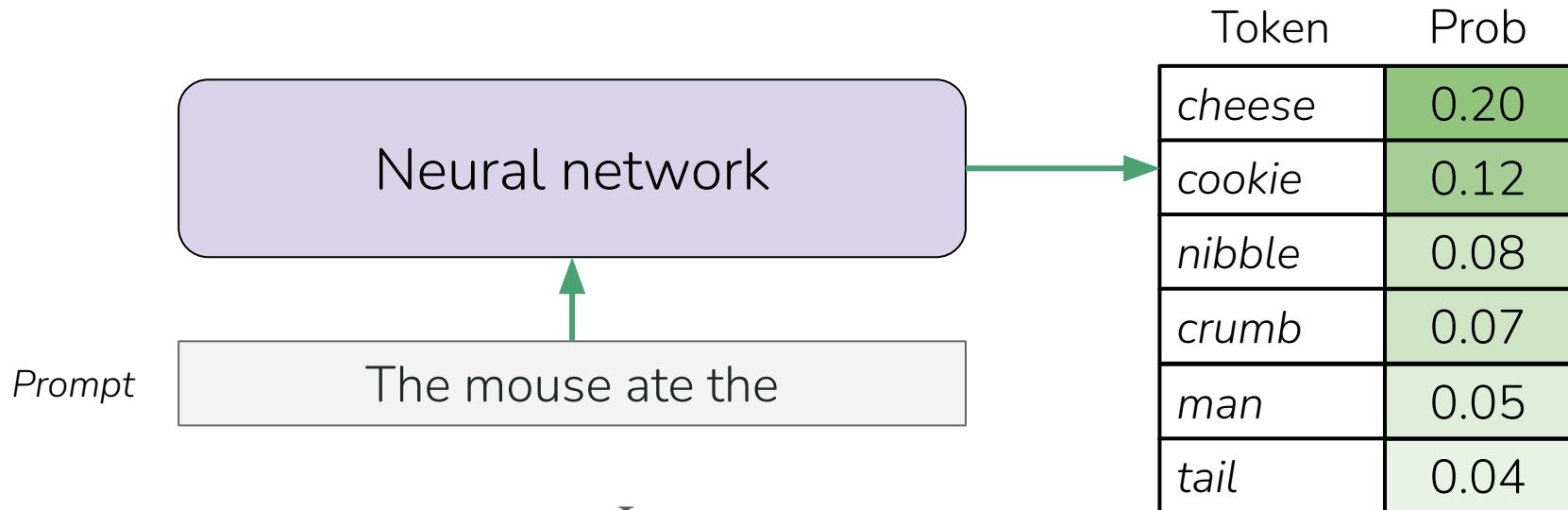
# Neural Language Models

---



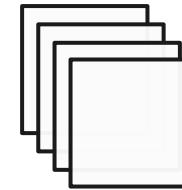
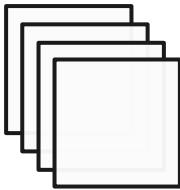
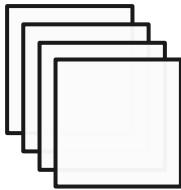
$$\prod_{i=1}^L p(x_i \mid x_{1:i-1})$$

# Neural Language Models



$$\prod_{i=1}^L p(x_i \mid x_{1:i-1})$$

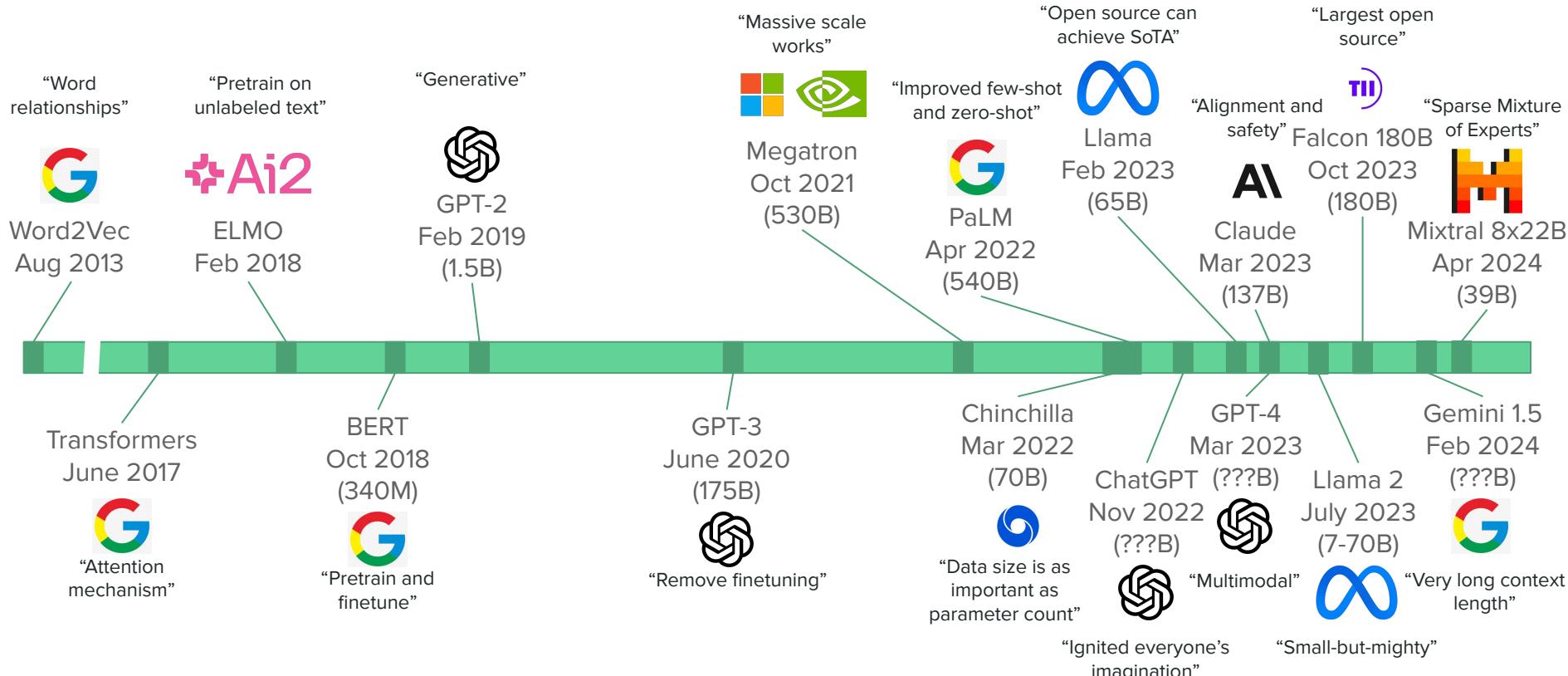
# Data for Pre-training



Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

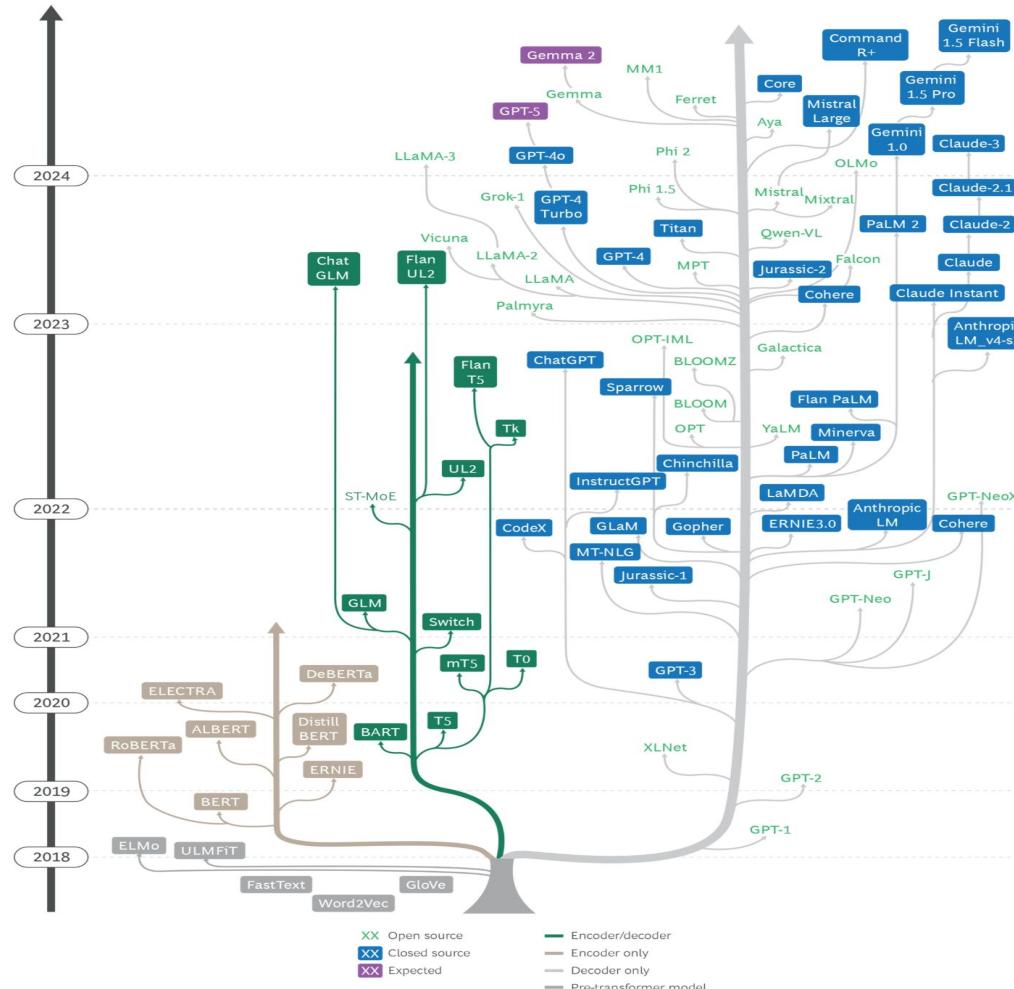
Table 1: **Pre-training data.** Data mixtures used for pre-training, for each subset we list the sampling proportion, number of epochs performed on the subset when training on 1.4T tokens, and disk size. The pre-training runs on 1T tokens have the same sampling proportion.

# Early LLMs



## Exhibit 1 - Evolutionary Tree of GenAI Models

# LLMs Today



John Whaley, Jan Jannink

# Failures of Pre-trained LMs

PROMPT *Explain the moon landing to a 6 year old in a few sentences.*

COMPLETION GPT-3

Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.

Language models are not *aligned* with user intent [[Ouyang et al., 2022](#)].  
Finetuning to the rescue!

# Failures of Pre-trained LMs

---

PROMPT *Explain the moon landing to a 6 year old in a few sentences.*

COMPLETION

**Human**

A giant rocket ship blasted off from Earth carrying astronauts to the moon. The astronauts landed their spaceship on the moon and walked around exploring the lunar surface. Then they returned safely back to Earth, bringing home moon rocks to show everyone.

Language models are not *aligned* with user intent [[Ouyang et al., 2022](#)].  
Finetuning to the rescue!

# LLM Accomplishments

ChatGPT: US Medical Licensing Exam

Pass!

GPT-4: Uniform Bar Examination

Pass!

<https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000198>

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4389233](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233)

# Is GPT Ready to Practice?

## Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions

**Molly Bohannon** Forbes Staff

*I cover breaking news.*

Follow

Jun 8, 2023, 02:06pm EDT

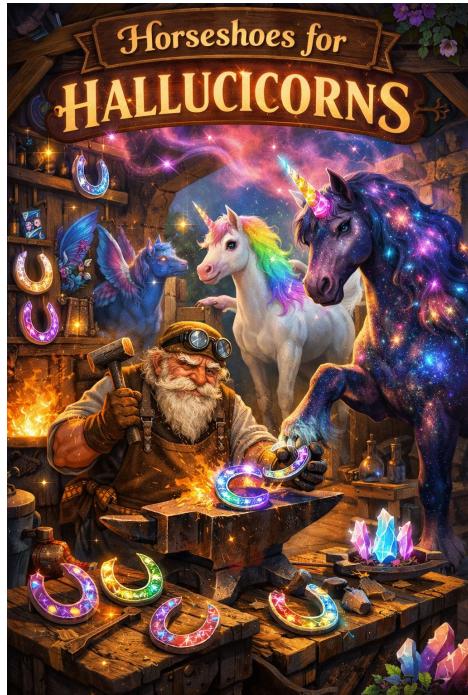
Exams are designed for humans, not AI

Humans passing the bar can practice

- They know when to look up info

But not AI bots

- don't know the meaning of their words
- make mistakes no humans will



# It Hallucinates!

Generative AI = Hallucination

# ChatGPT (7/2/2023)

+ Conversational

- Hallucinates on recent events!

Who is the monarch of UK?



As of September 2021, the monarch of the United Kingdom is Queen Elizabeth II. She has been the queen since February 6, 1952, following the death of her father, King George VI. Queen Elizabeth II is the longest-reigning monarch in British history and has played an important role in the country's political and cultural life over the past seven decades.

# ChatGPT (7/2/2023)

- + Conversational
- Hallucinates on long-tail info.  
With details and conviction

1978      Hinton, G. E.  
**Relaxation and its role in vision.**  
PhD Thesis, University of Edinburgh.

Where did Geoff Hinton get his PhD?



Geoffrey Hinton received his PhD in Artificial Intelligence from the University of Edinburgh in 1978. His doctoral thesis was titled "**Relational Learning of Pattern-Matchable Templates**". Hinton is widely regarded as one of the pioneers of deep learning and has made significant contributions to the field of artificial intelligence and machine learning throughout his career.

# Buying a Ticket from GPT-3

User: I'd like to buy a plane ticket to New York.

Agent: Sure, I can help you with that. What date and time would you like to depart?

User: July 7th, 9am.

Agent: Alright, I can book a flight for you departing on July 7th at 9am.

What airport would you like to depart from?

User: SFO

Agent: Great, I can book a flight for you departing from San Francisco on July 7th at 9am. Would you like to book this flight now?

User: What is the flight number?

Agent: The flight number is AA123.

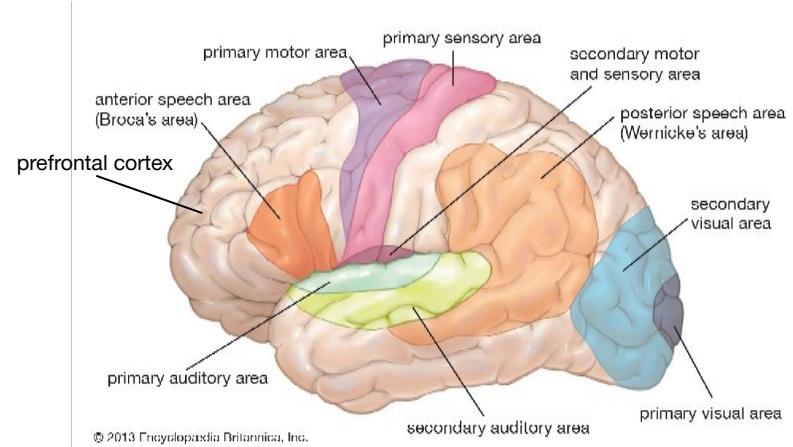
+ Perfect flow without training  
- Hallucinates: American AA123  
DFW  HNL  
- Needs to consult DB

# Examples of (Failed) Business Use Cases

- Google withdrew medical AI overviews
  - False information about blood tests
- EPIC medical use cases (with expert reviews)
- A large manufacturer
  - “Our pretrained agent hallucinates components that don’t exist”
- Microsoft travel article (with human reviewers)
  - “Visit Food Bank if you are hungry in Ottawa”

# We All Have an “LLM” in Our Brain!

- Our “LLM”: Speech area
- Prefrontal cortex
  - Inhibition and attention
  - Planning
  - Long-term memory
  - Monitoring external signals
- If prefrontal cortex is damaged, “Speak without thinking”



# Practical challenges

## Associated with training

- Bias
- Stereotyping
- Costs
- Access
- Legal issues

### OpenAI Sued for Using 'Stolen' Data, Violating Your Privacy With ChatGPT

The class-action lawsuit asks the court to temporarily shut down ChatGPT to rein in a 'powerful and wildly profitable AI' that was released 'without regard for the risks.'

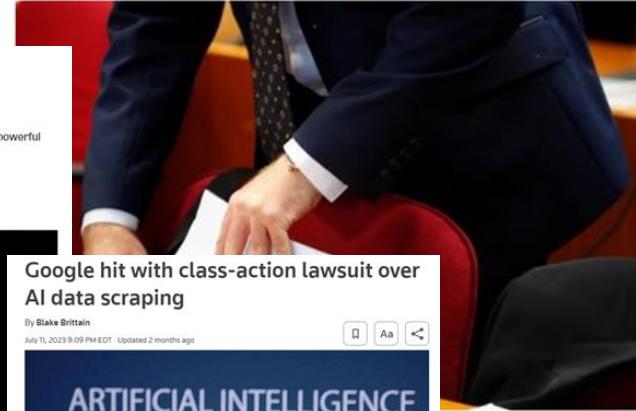
By Michael Kan June 29, 2023 f t



### Getty Images lawsuit says Stability AI misused photos to train AI

By Blake Brittain

February 6, 2023 12:32 PM EST · Updated 7 months ago



### Google hit with class-action lawsuit over AI data scraping

By Blake Brittain

July 11, 2023 9:09 PM EDT · Updated 2 months ago



# Practical challenges

Associated with the operation of the technology

- Hallucinations
- Catastrophic forgetting
- Centralization (versus open source)
- Legal issues
- Prompt injection
- Guardrails

ARTIFICIAL INTELLIGENCE / TECH / LAB

**AI-generated art cannot be copyrighted, rules a US federal judge**

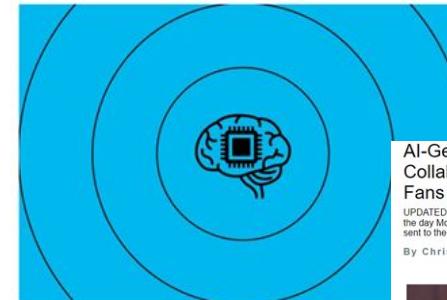


Illustration by Alex Castro / The Verge

/ DC District Court Judge Beryl A. Howell says human beings are an 'essential part of a valid copyright claim.'

AI-Generated Fake 'Drake'/'Weeknd' Collaboration, 'Heart on My Sleeve,' Delights Fans and Sets Off Industry Alarm Bells

UPDATED: The soundalike song in question began to disappear from streaming services throughout the day Monday. Universal Music was identified by YouTube as being behind the takedown notices sent to the DSPs.

By Chris Willman ▾



**OpenAI sued for defamation after ChatGPT fabricates legal accusations against radio host**