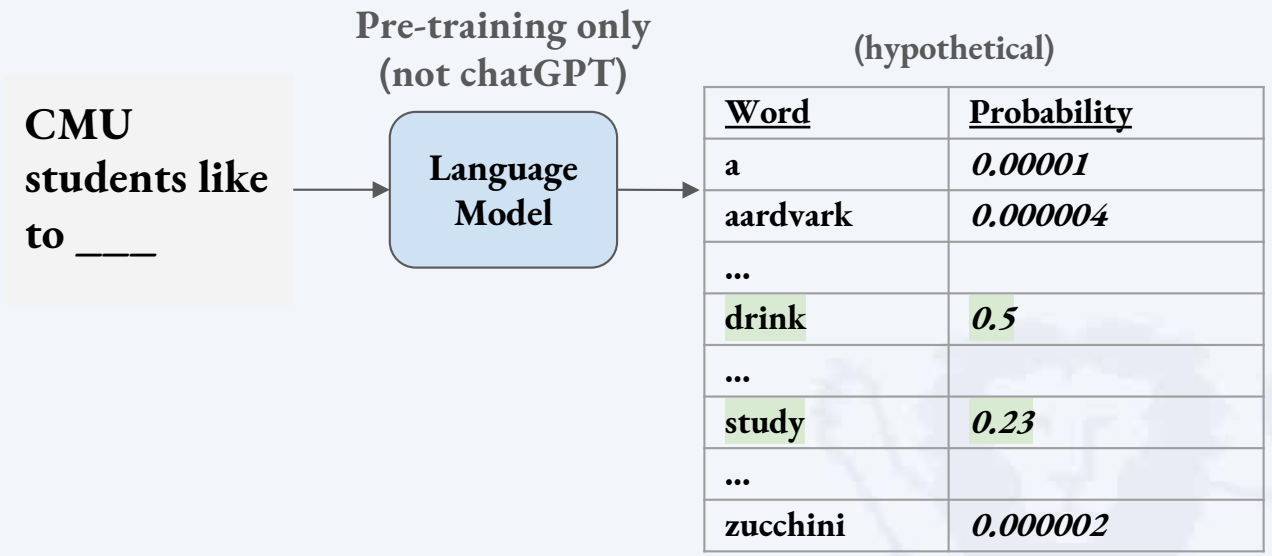
LLM, ChatGPT and Beyond

Speaker: Qizhe Xie

The speaker will elucidate the reasons behind the success of Large Language Models (LLMs) and provide an overview of the mechanisms that enable them to function effectively. The talk will cover topics such as the role of massive datasets and compute, while also discussing real-world applications and future prospects.

What are Language Models?



Word	Probability
a	*0.00001*
aardvark	*0.000004*
...	
drink	*0.5*
...	
study	*0.23*
...	
zucchini	*0.000002*

What do language models learn from next-word prediction?

Grammar	In my free time, I like to {run, banana}
Lexical semantics	I went to the zoo to see giraffes, lions, and {zebras, spoon}
World knowledge	The capital of Denmark is {Copenhagen, London}
Sentiment analysis	Movie review: I was engaged and on the edge of my seat the whole time. The movie was {good, bad}
Harder sentiment analysis	Movie review: Overall, the value I got from the two hours watching it was the sum total of the popcorn and the drink. The movie was {bad, good}
Translation	The word for "pretty" in Spanish is {bonita, hola}
Spatial reasoning	[...] Iroh went into the kitchen to make some tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the {kitchen, store}
Math question	First grade arithmetic exam: 3 + 8 + 4 = {15, 11}

[thousands (millions?) more]

Extreme multi-task learning!

What can't language models learn from next-word prediction?

Current world knowledge	The stock price of APPL on May 1st, 2023 is {???}
Arbitrarily long arithmetic	36382894730 + 238302849204 = {???}
Many-step reasoning	Take the nineteenth digit of Pi and multiply it by the e to the fourth power. The resulting ones-digit of the resulting number is {???}
Predict the future	The winner of the FIFA world cup in 2026 is {???}
Information not in the training data	Qizhe' s favorite color is {???}
Extremely long inputs	[2,000 page Harry Potter fan-fiction] What happened after Harry opened the chest for the second time? {???}

Emergence in science

General defn. in science

Emergence is *a qualitative change that arises from quantitative changes.*

Popularized by this 1972 piece by Nobel-Prize winning physicist P.W. Anderson.

SCIENCE

More Is Different

Broken symmetry and the nature of the hierarchical structure of science.

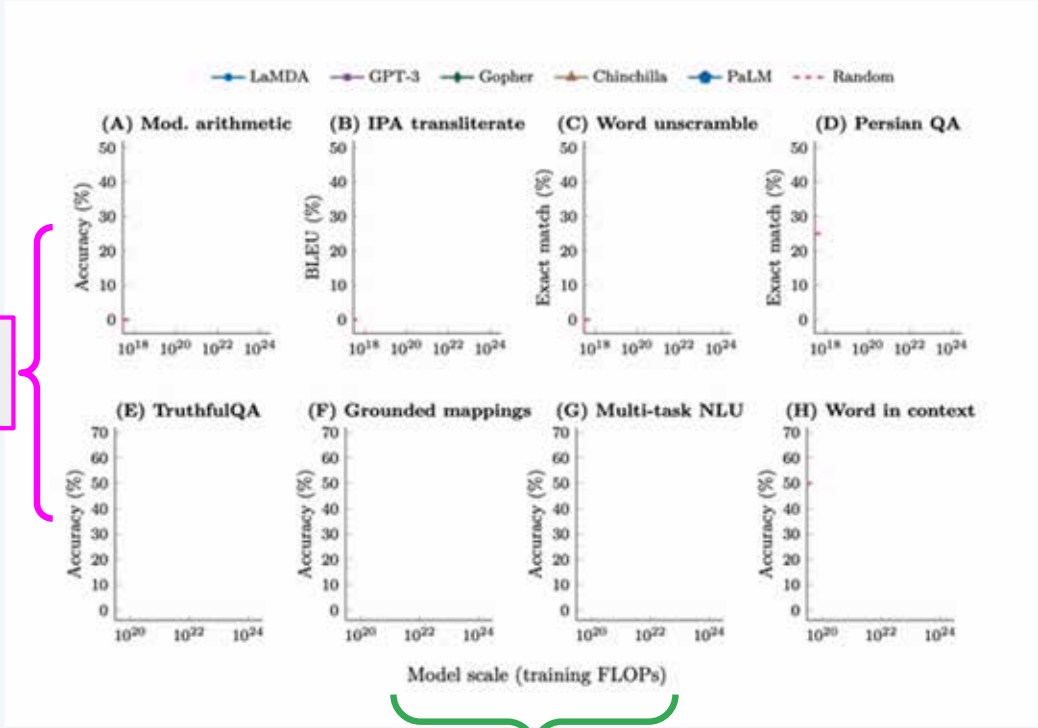
P. W. Anderson

With a bit of uranium, nothing special happens. With a large amount of uranium, you get a nuclear reaction.

Given only small molecules such as calcium, you can't meaningfully encode useful information. Given larger models such as DNA, you can encode a genome.

Suggested further reading:
Future ML Systems Will Be Qualitatively Different (2023).

Emergence in large language model (LLM)



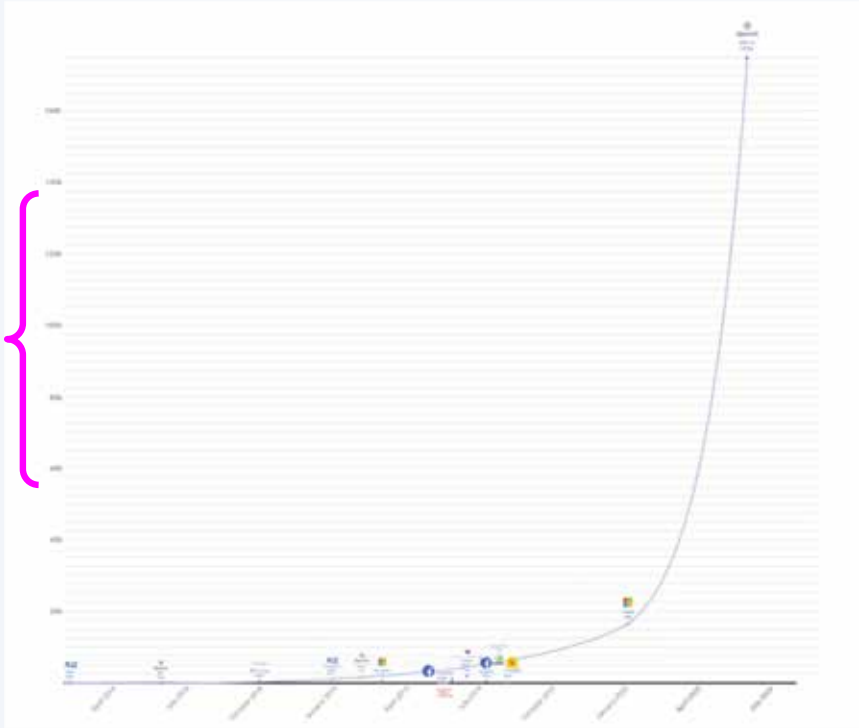
y-axis is performance on the task

x-axis is "scale"

Performance is flat for small models. Performance spikes to well above-random for large models.

Open research question: is it possible to predict emergence using only smaller model sizes?

Scaling LLMs



- GPT: The first LLM that uses the Transformer architecture
- GPT-2: LLM are general purpose models
- GPT-3: Scaling up GPT-2
- ChatGPT / InstructGPT: Aligning GPT to follow instructions
- GPT-4: ??

ChatGPT

GPT: supernaturally precocious child who learned from all human data

ChatGPT: the child who follow human instructions

Details covered by Rishabh

2018 → **Today (2023)** → **Future ...?**

…
Protein discovery
Clinical diagnosis
Play chess well
High-level planning
Abstract reasoning
Simple math
Commonsense reasoning
Know world knowledge
Translation
Sentiment analysis
Generate coherent text
Be grammatically correct

…
Protein discovery
Clinical diagnosis
Play chess well
High-level planning
Abstract reasoning
Simple math
Commonsense reasoning
Know world knowledge
Translation
Sentiment analysis
Generate coherent text
Be grammatically correct

…
(?) Protein discovery
(?) Clinical diagnosis
(?) Play chess well
(?) High-level planning
(?) Abstract reasoning
Simple math
Commonsense reasoning
Know world knowledge
Translation
Sentiment analysis
Generate coherent text
Be grammatically correct

Why did OpenAI succeed (my opinion)

(1) Clear vision: AGI

Aimed for Artificial General Intelligence (AGI) at inception (2015).

(2) Engineering + Research Culture

Member of the Technical Staff
OpenAI
Sep 2022 - Present · 8 mos

A top-down management approach, focus on engineering and research

All of these come with great cost!

(3) Product-centric mindset

DALL-E 2

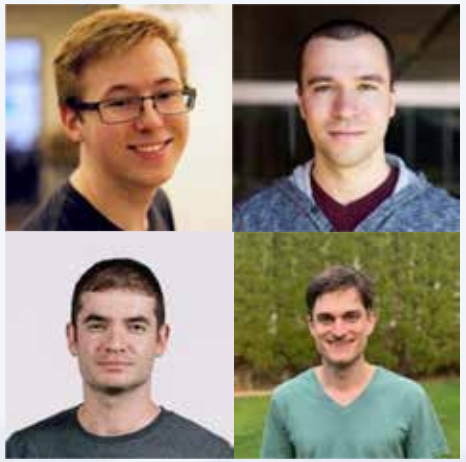
OpenAI Gym

Several orgs possessed both the technology know-how and insights. OpenAI built the right product.

AI is a collective endeavor

Researchers who laid the foundations of Deep Learning in the 80s

Pioneers in deep learning



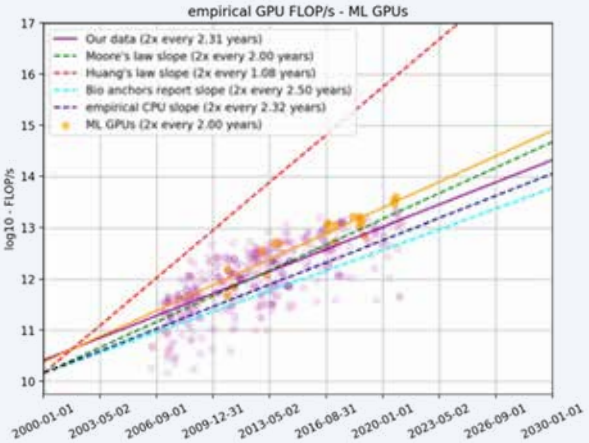
Great people from OpenAI

Google invested heavily in AI and tech in general. Google was the pioneer on scaling up model size and compute.
(e.g., A paper by Google used 10,000 GPUs in 2016)

Nvidia GPUs serve as the driving force behind the AI engine.

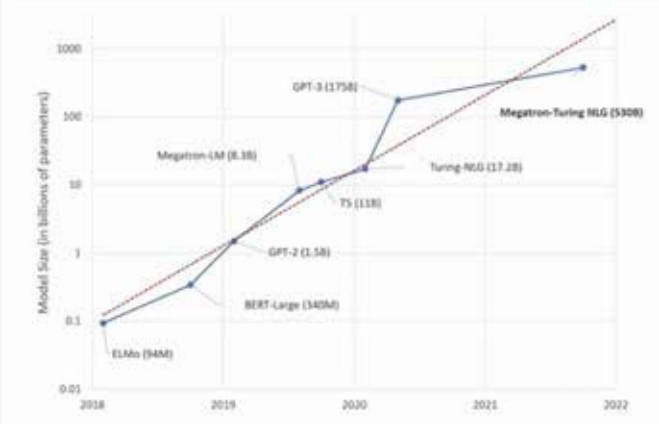
Exponential growth of AI Intelligence

(1) Hardware



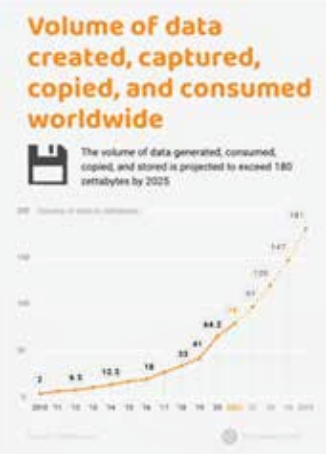
Exponential growth of computing power

(2) Model Size



Exponential growth of model size

(3) Data



Exponential growth of data

Scaling up AI = scaling up compute + model + data