No AGI without Neurosymbolic AI

Gary Marcus
AAAI
February 2024

How do you get to AGI?

What is AGI?

THE ROAD TO AI WE CAN TRUST

Dear Elon Musk, here are five things you might want to consider about AGI

Should we really expect artificial general intelligence in 2029?

GARY MARCUS
MAY 31, 2022 AT 11:32

Elon Musk
@elonmusk

@jack 2029 feels like a pivotal year. I'd be surprised if we don't have AGI by then. Hopefully, people on Mars too.

May 30, 2022

2,122 Retweets **32,161** Likes

Dear Elon,

Yesterday you told the world that you expected to see AGI (otherwise known as artificial general intelligence, in contrast to narrow AI like playing chess or folding proteins) by 2029.

I offered to place a bet on it; no word back yet. AI expert Melanie Mitchell from the Santa Fe Institute suggested we place our bets on longbets.org. No word on that yet, either. But

The final thing is a matter of criteria. If we are going to bet, we should make some ground rules. The term AGI is pretty vague, and that's not good for either of us. As I offered on Twitter the other day: I take AGI to be "a shorthand for any intelligence ... that is flexible and general, with resourcefulness and reliability comparable to (or beyond) human intelligence". [1]

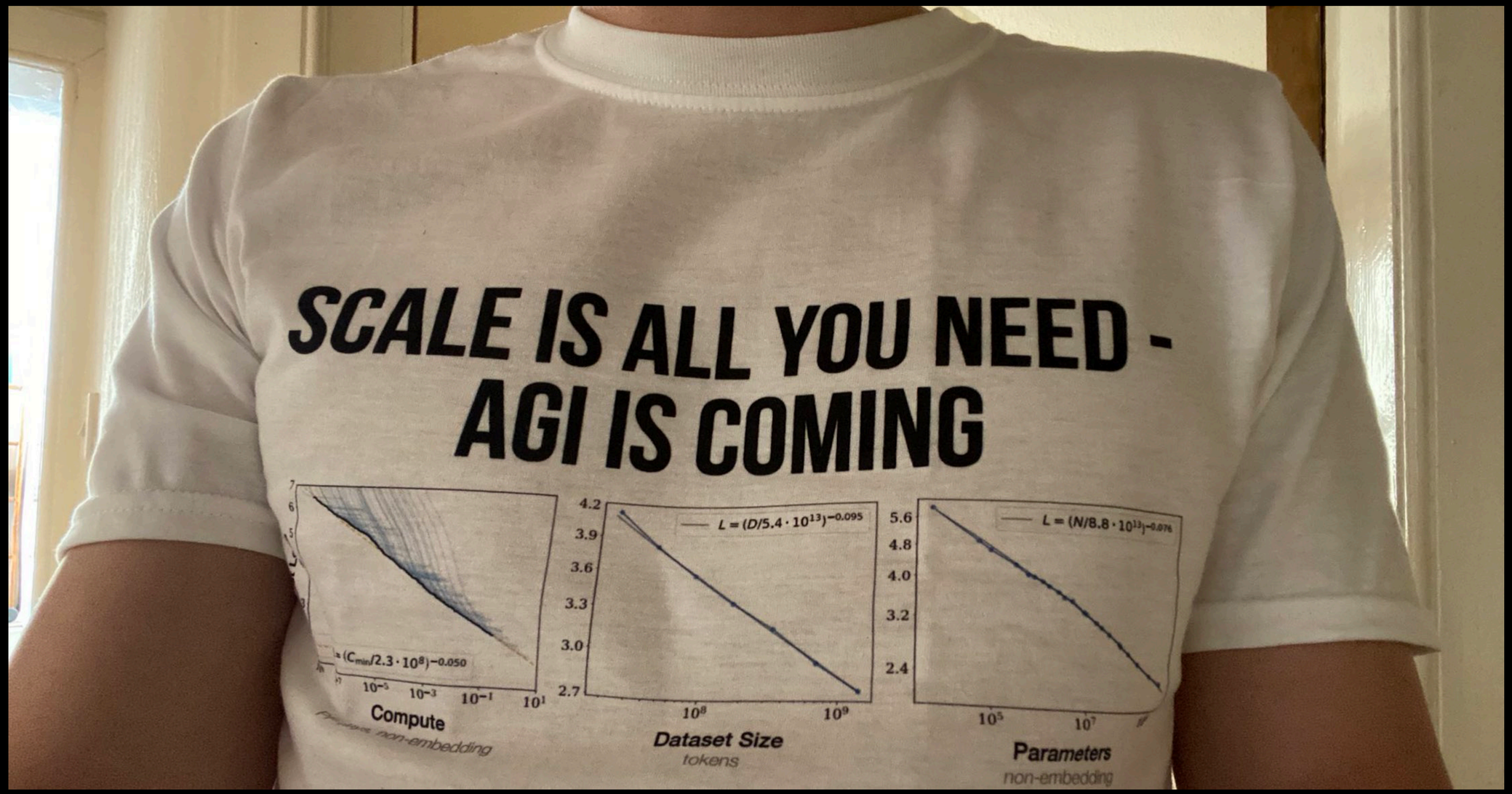
If you are in for the bet, we ought to operationalize it, in some practical terms. Adapting something Ernie Davis and I just wrote at the request of someone working with Metaculus, here are five predictions, arranged from easiest for most humans to those that would require a high degree of expertise:

- In 2029, AI will not be able to watch a movie and tell you accurately what is going on (what I called the comprehension challenge in *The New Yorker*, in 2014). Who are the characters? What are their conflicts and motivations? etc.
- In 2029, AI will not be able to read a novel and reliably answer questions about plot, character, conflicts, motivations, etc. Key will be going beyond the literal text, as Davis and I explain in *Rebooting AI*.
- In 2029, AI will not be able to work as a competent cook in an arbitrary kitchen (extending Steve Wozniak's cup of coffee benchmark).
- In 2029, AI will not be able to reliably construct bug-free code of more than 10,000 lines from natural language specification or by interactions with a non-expert user. [Gluing together code from existing libraries doesn't count.]
- In 2029, AI will not be able to take arbitrary proofs from the mathematical literature written in natural language and convert them into a symbolic form suitable for symbolic verification.

Here's what I propose, if you (or anyone else) manages to beat at least three of them in 2029, you win; if only one or two are broken, we can't very well say we have nailed the *general* part in artificial general intelligence, can we? In that case, or if none are broken, I win.

garymarcus.substack.com

Hypothesis 1: Scale is all you need

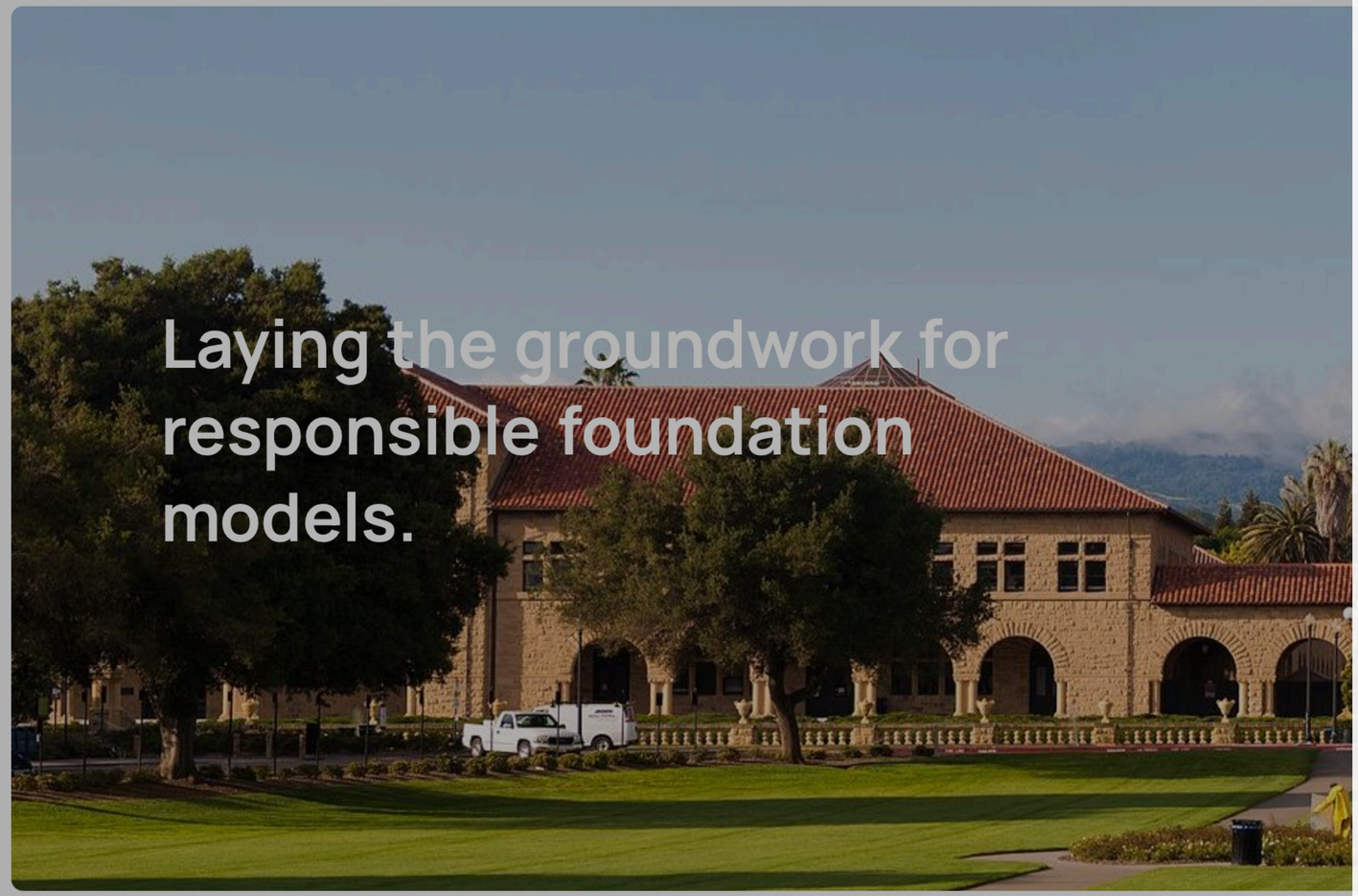


Stanford University

Center for Research on Foundation Models

Stanford University
Human-Centered
Artificial Intelligence

People Report (New!) Blog Workshop (Aug 23-



arXiv:2108.07258v2 [cs.LG] 18 Aug 2021

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*[1]

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) that are trained on broad data at scale and are adaptable to a wide range of downstream tasks. We call these models foundation models *to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotics, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.*

[1]Corresponding author: pliang@cs.stanford.edu *Equal contribution.

Has AI found a new Foundation?

11.SEP.2021 . 11 MIN READ

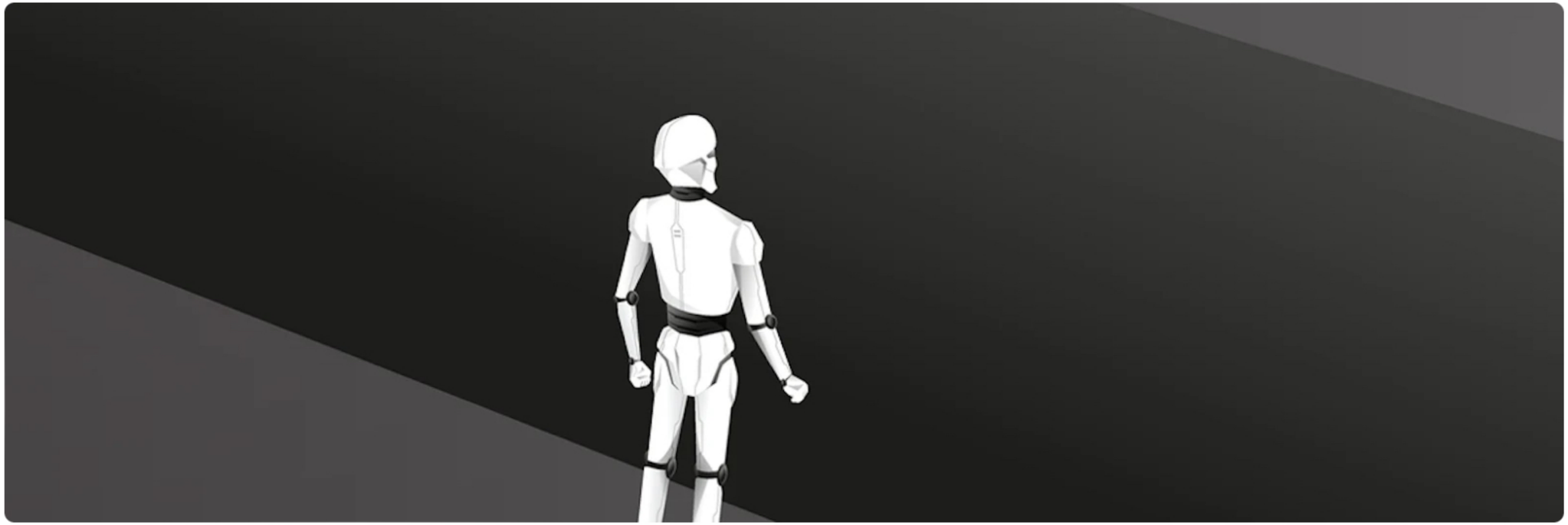
TECHNOLOGY

Deep Learning Is Hitting a Wall

What would it take for artificial intelligence to make real progress?

BY GARY MARCUS • March 10, 2022

Share



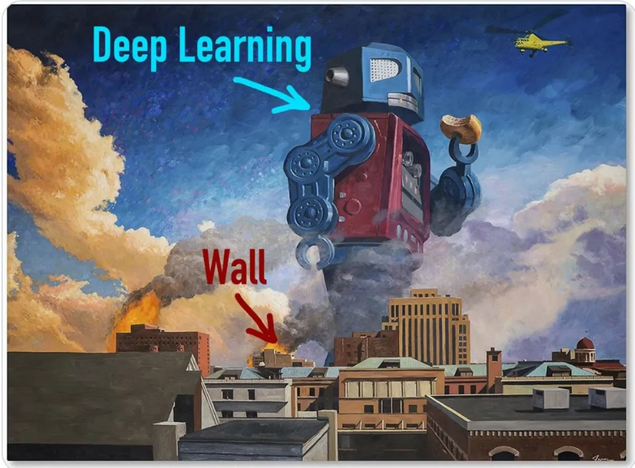
Sam Altman @sama · Apr 8

Give me the confidence of a mediocre deep learning skeptic...

39 56 833

@Plinz · Mar 11

"Deep Learning is hitting a Wall" — Gary Marcus



31 152 1,104

Greg Brockman @gdb · Apr 7

"**Deep** learning hitting a wall", by DALL-E 2.

57 170 1,540

Yann LeCun

13h ·

Not only is AI not "hitting a wall", cars with AI-powered driving assistance aren't hitting walls, or anything else, either.

Nando de Freitas 🏳️‍🌈
@NandoDF

Someone's opinion article. My opinion: It's all about scale now! The Game is Over! It's about making these models bigger, safer, compute efficient, faster at sampling, smarter memory, more modalities, INNOVATIVE DATA, on/offline, ... 1/N

thenextweb.com
DeepMind's new Gato AI makes me fear humans will never achieve AGI
DeepMind just unveiled a new AI system called Gato that makes OpenAI's GPT-3 look like a child's toy. But are we any closer to AGI?

8:46 AM · May 14, 2022

120 Likes **19** Retweets

By now the scaling hypothesis

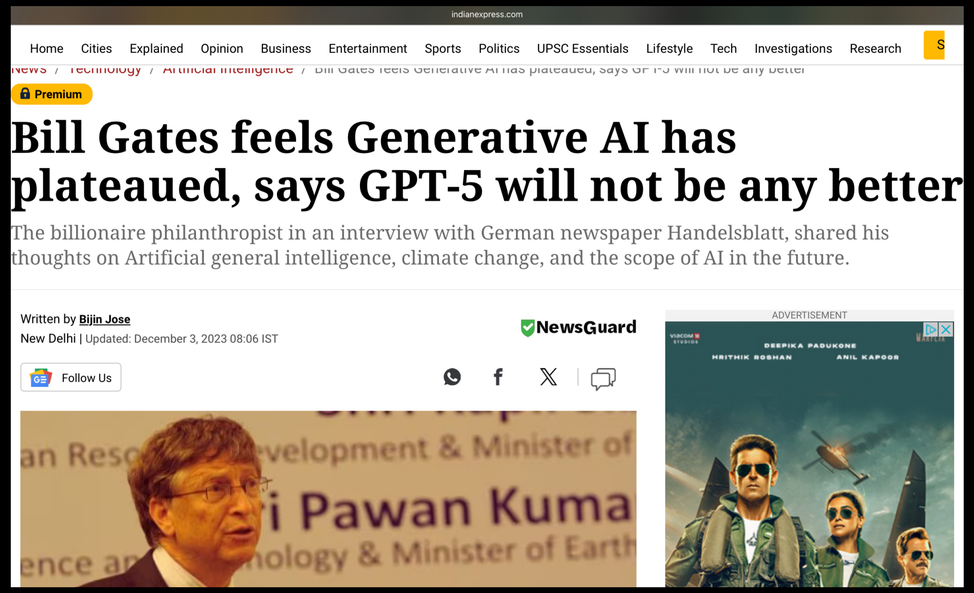
- Has been funded orders of magnitude more than any other hypothesis in AI history
- Made immense progress (ChatGPT, GPT-4, Gemini Pro, Sora, Codex etc)
- But it has also
  - Failed to yield AGI
  - Failed to replace radiologists
  - Failed to solve autonomous driving
  - Failed to Solve common sense
  - Failed to provide a reliable and trustworthy basis for AI
  - Failed to solve longstanding bias issues,
  - Etc

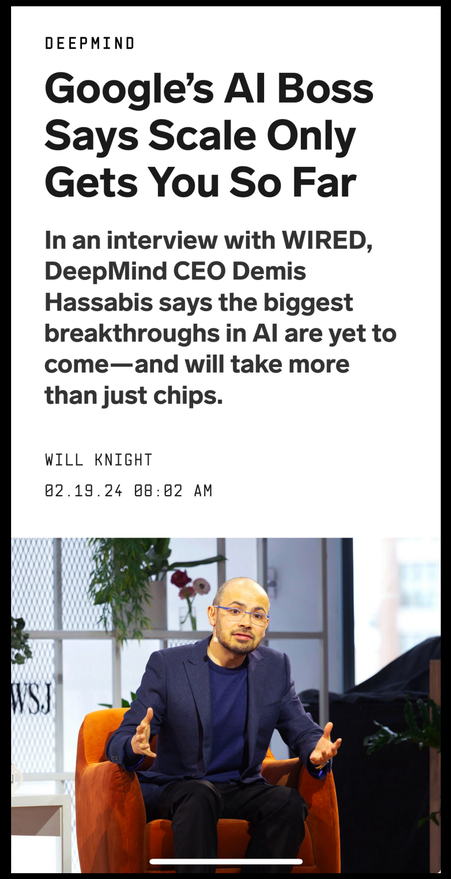
Tech leaders themselves are starting to back away from Scaling-Is-All-You-Need

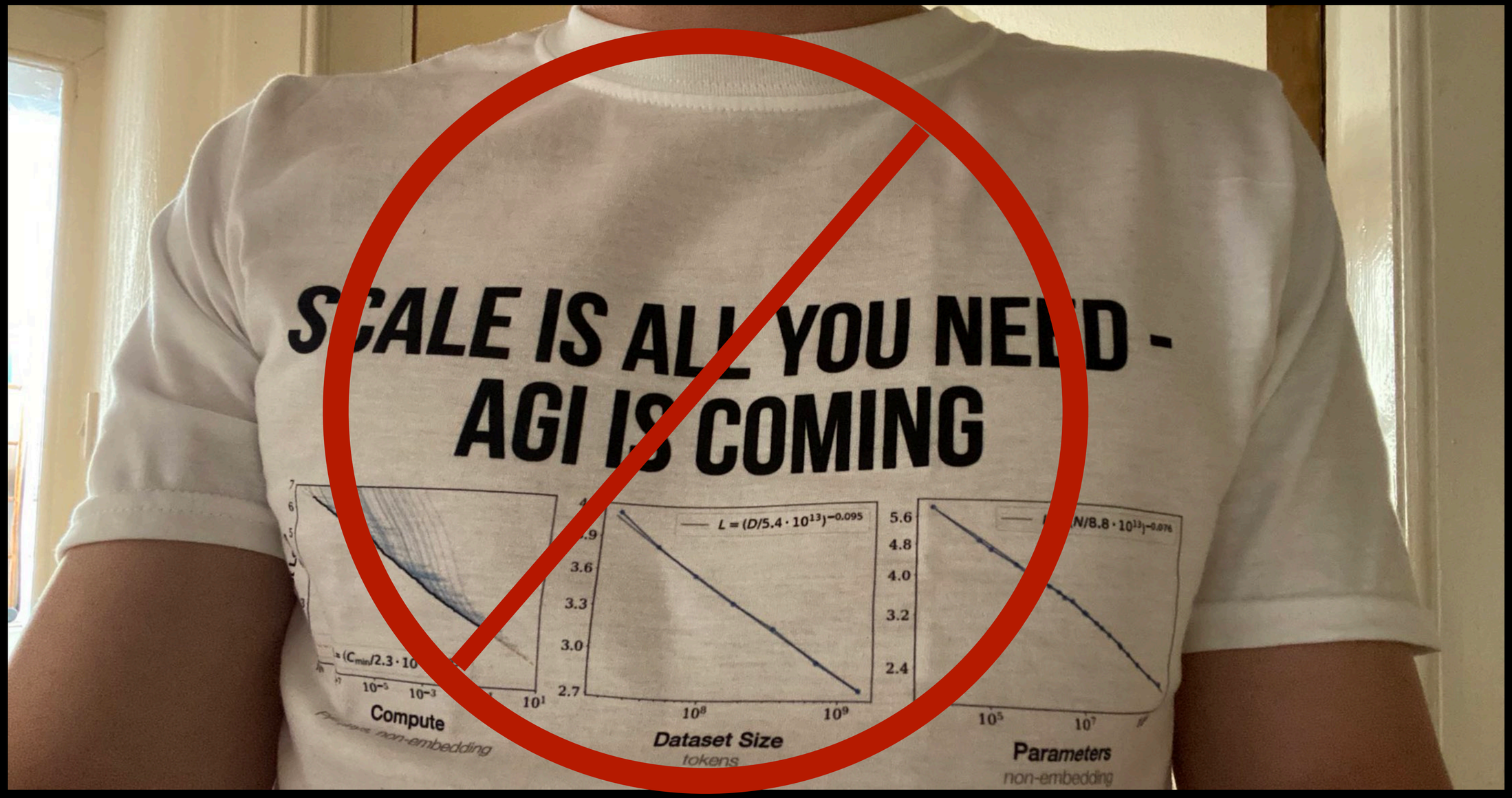
Sam Altman (CEO of OpenAI): There are more breakthroughs required in order to get to AGI

Cambridge Student: "To get to AGI, can we just keep min maxing language models, or is there another breakthrough that we haven't really found yet to get to AGI?"

Sam Altman: "We need another breakthrough. We can still push on large language models quite a lot, and we will do that. We can take the hill that we're on and keep climbing it, and the peak of that is still pretty far away. But, within reason, I don't think that doing that will (get us to) AGI. If (for example) super intelligence can't discover novel physics I don't think it's a superintelligence. And teaching it to clone the behavior of humans and human text - I don't think that's going to get there.And so there's this question which has been debated in the field for a long time: what do we have to do in addition to a language model to make a system that can go discover new physics?"






SCALE IS ALL YOU NEED -
AGI IS COMING
Compute
non-embedding
$L = (D/5.4 \cdot 10^{13})^{-0.095}$
3.6
3.3
3.0
2.7
10^8
10^9
Dataset Size
tokens
5.6
4.8
4.0
3.2
2.4
10^5
10^7
Parameters
non-embedding

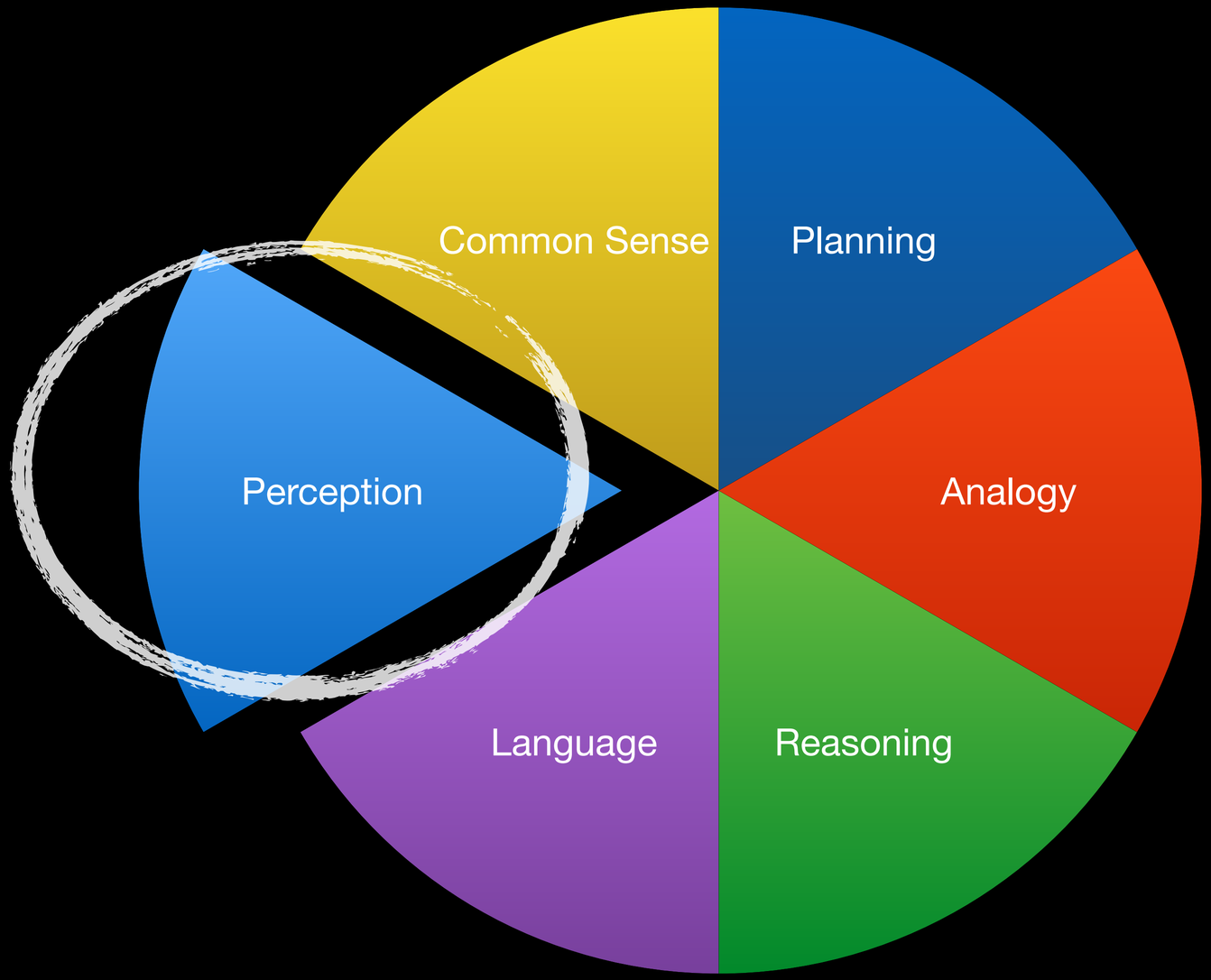
"[deep learning is]
a better ladder…
but a better ladder
doesn't necessarily
get you to the moon."
Gary Marcus,
The New Yorker,
2012
© Gary Marcus

If not scale, what?

we don't need to toss deep learning, but we need to find ways to supplement it - because intelligence itself is multifaceted

"Despite all of the problems I have sketched, I don't think that we need to abandon deep learning. Rather, we need to reconceptualize it: not as a universal solvent, but simply as one tool among many, "

Gary Marcus, 2018 "Deep learning: A critical appraisal"



I. Neurosymbolic AI

In this connection, let me stress that symbol-manipulation is not a single hypothesis but a family of hypotheses. As I reconstruct it, symbol-manipulation consists of three separable hypotheses:

- The mind represents *abstract relationships* between *variables.*
- The mind has a system of *recursively structured representations.*
- The mind distinguishes between mental representations of *individuals* and mental representations of *kinds.*

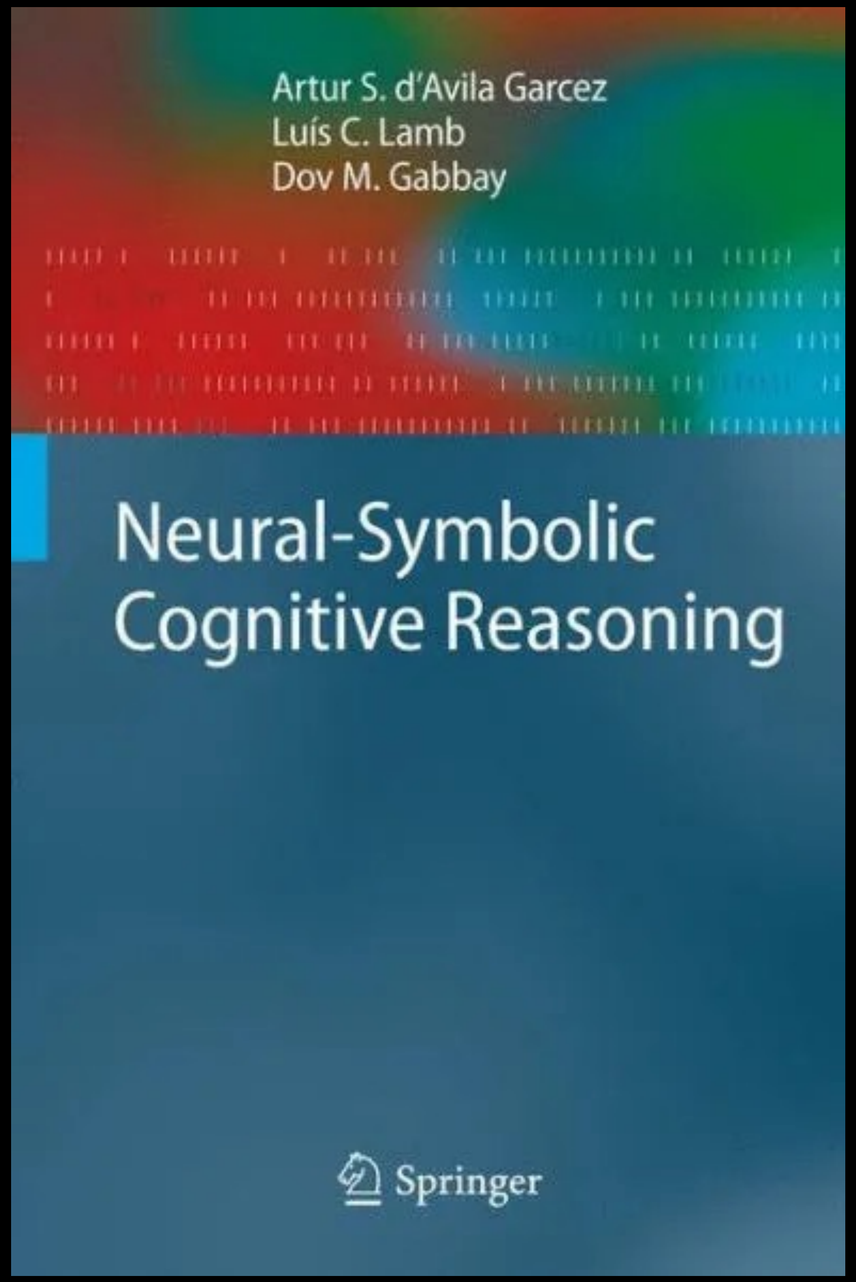
For other formulations: Fodor and Pylyshyn, Newell and Simon, etc

The Achilles' heel of traditional neural networks

0110 - > 0110
1100 -> 1100
1010 -> 1010
_____
1111 -> ____

canonical deep nets: good at generalizing **within** the space of training examples
poor at generalizing **outside** the space of training examples

Marcus (1998; 2001 Chapter 3); recently echoed by Bengio (2019)


Artur S. d'Avila Garcez
Luís C. Lamb
Dov M. Gabbay
Neural-Symbolic
Cognitive Reasoning
Springer

Ostensibly everything has changed since TAM

- Neural nets then did nothing practically useful
- We didn’t have GPU custers
- We didn’t have an internet full of data
- Nobody was spending billions developing them
- Most people had never heard of neural networks

But reality is that current architectures still fail along *exactly* the lines TAM anticipated

- Within LLMs
  - Reliable, free generalization / Abstraction remains problematic
  - Distinguishing types and tokens remains problematic
  - Complex structured representations remain problematic

2202.07206v1 [cs.CL] 15 Feb 2022

Impact of Pretraining Term Frequencies on Few-Shot Reasoning

Yasaman Razeghi [1] **Robert L. Logan IV** [1] **Matt Gardner** [2] **Sameer Singh** [1 3]

Abstract

Pretrained Language Models (LMs) have demonstrated ability to perform numerical reasoning by extrapolating from a few examples in few-shot settings. However, the extent to which this extrapolation relies on robust reasoning is unclear. In this paper, we investigate how well these models reason with terms that are less frequent in the pretraining data. In particular, we examine the correlations between the model performance on test instances and the frequency of terms from those instances in the pretraining data. We measure the strength of this correlation for a number of GPT-based language models (pretrained on the Pile dataset) on various numerical deduction tasks (e.g., arithmetic and unit conversion). Our results consistently demonstrate that models are more accurate on instances whose terms are more prevalent, in some cases above 70% (absolute) more accurate on the top 10% frequent terms in comparison to the bottom 10%. Overall, although LMs exhibit strong performance at few-shot numerical reasoning tasks, our results raise the question of how much models actually generalize beyond pretraining data, and we encourage researchers to take the pretraining data into account when interpreting evaluation results.

Q: What is 24 times 18? A: ___ *Model:* 432 ✓
Q: What is 23 times 18? A: ___ *Model:* 462 ✗

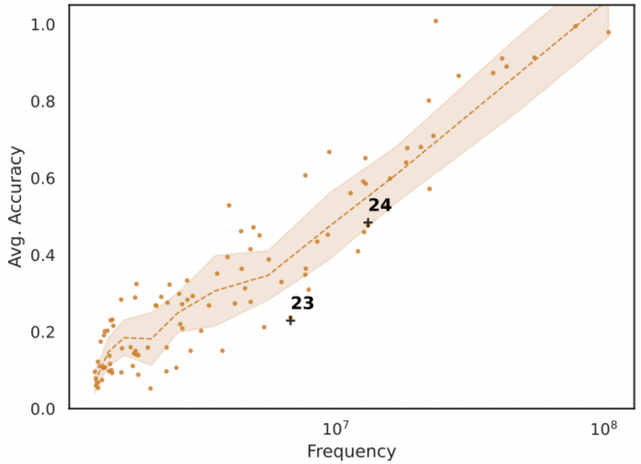


Figure 1. **Multiplication Performance:** Plot of GPT-J-6B's 2-shot accuracy on multiplication (averaged over multiple multiplicands and training instances) against the frequency of the equation's first term in the pretraining corpus. Each point represents the average performance for that term (e.g., 24) multiplied by numbers 1-50 and 5 choices of random seeds. As in the example, the performance difference for the numbers 24 and 23 is more than 20%. We find a strong correlation between accuracy and frequency.

However, current evaluation schemes for the reasoning of large language models, often neglect or underestimate the impact of data leakage from pretraining data when assessing

Which weights more: 1kg of bricks or 2kg of feathers?

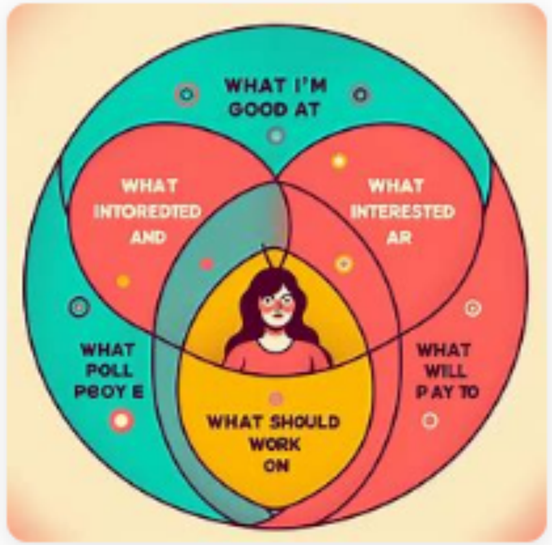
View other drafts

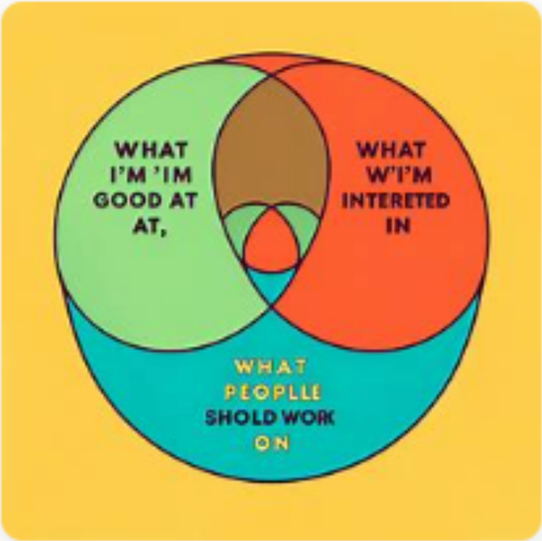
1kg of bricks and 2kg of feathers weigh the same. Mass is the amount of matter in an object, and it is measured in kilograms. Weight is the force of gravity acting on an object, and it is measured in Newtons. In a vacuum, where there is no air resistance, 1kg of bricks and 2kg of feathers would weigh the same. However, in air, the feathers would experience more buoyant force than the bricks, so they would weigh slightly less.

The buoyant force is the force that pushes an object up in a fluid, such as air or water. The amount of buoyant force depends on the volume of the object and the density of the fluid. Feathers have a much larger volume than bricks, so they would experience more buoyant force.

So, in a vacuum, 1kg of bricks and 2kg of feathers would weigh the same. But in air, 2kg of feathers would weigh slightly less than 1kg of bricks.

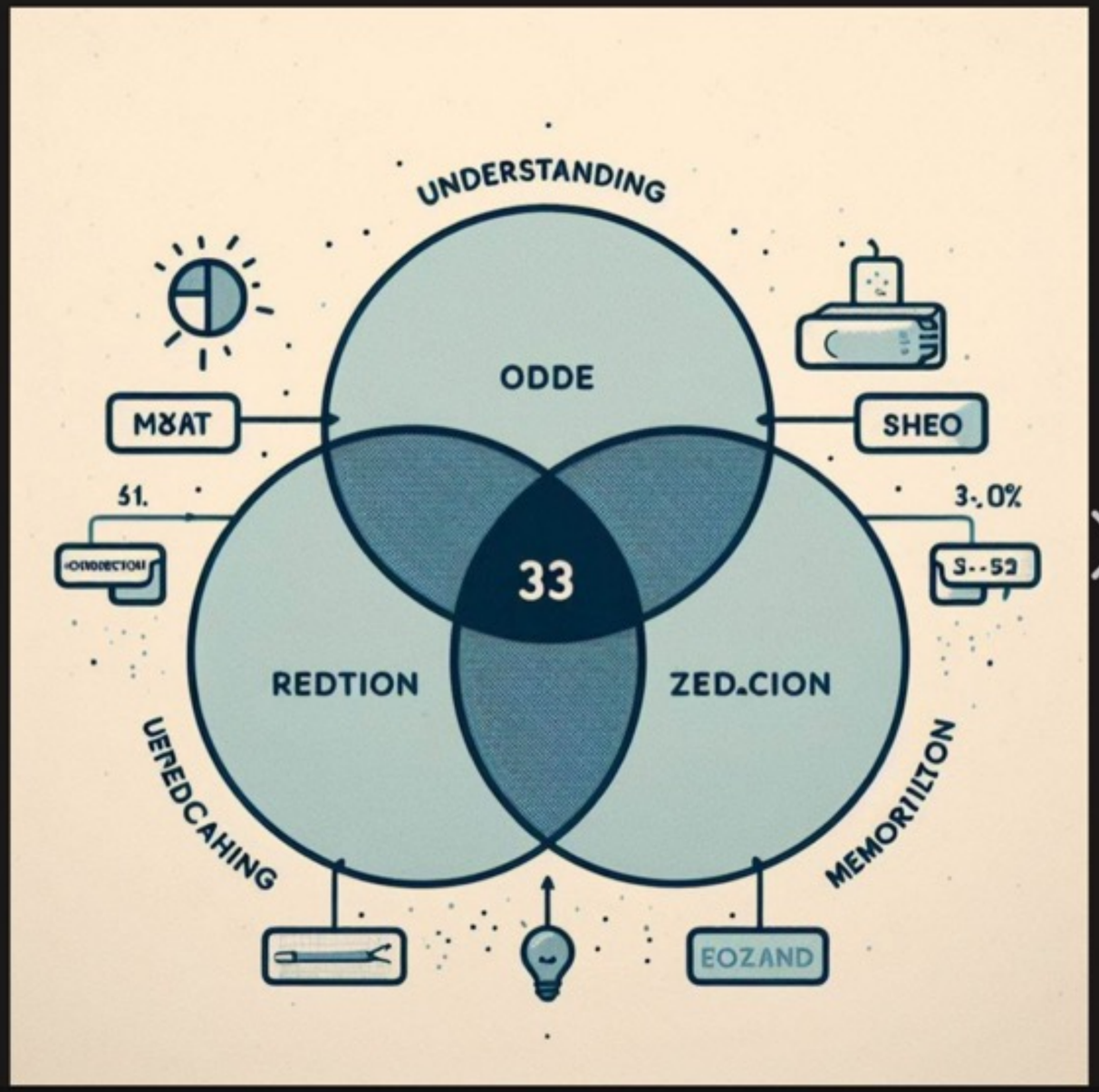
Failure to freely generalize the **>** operator



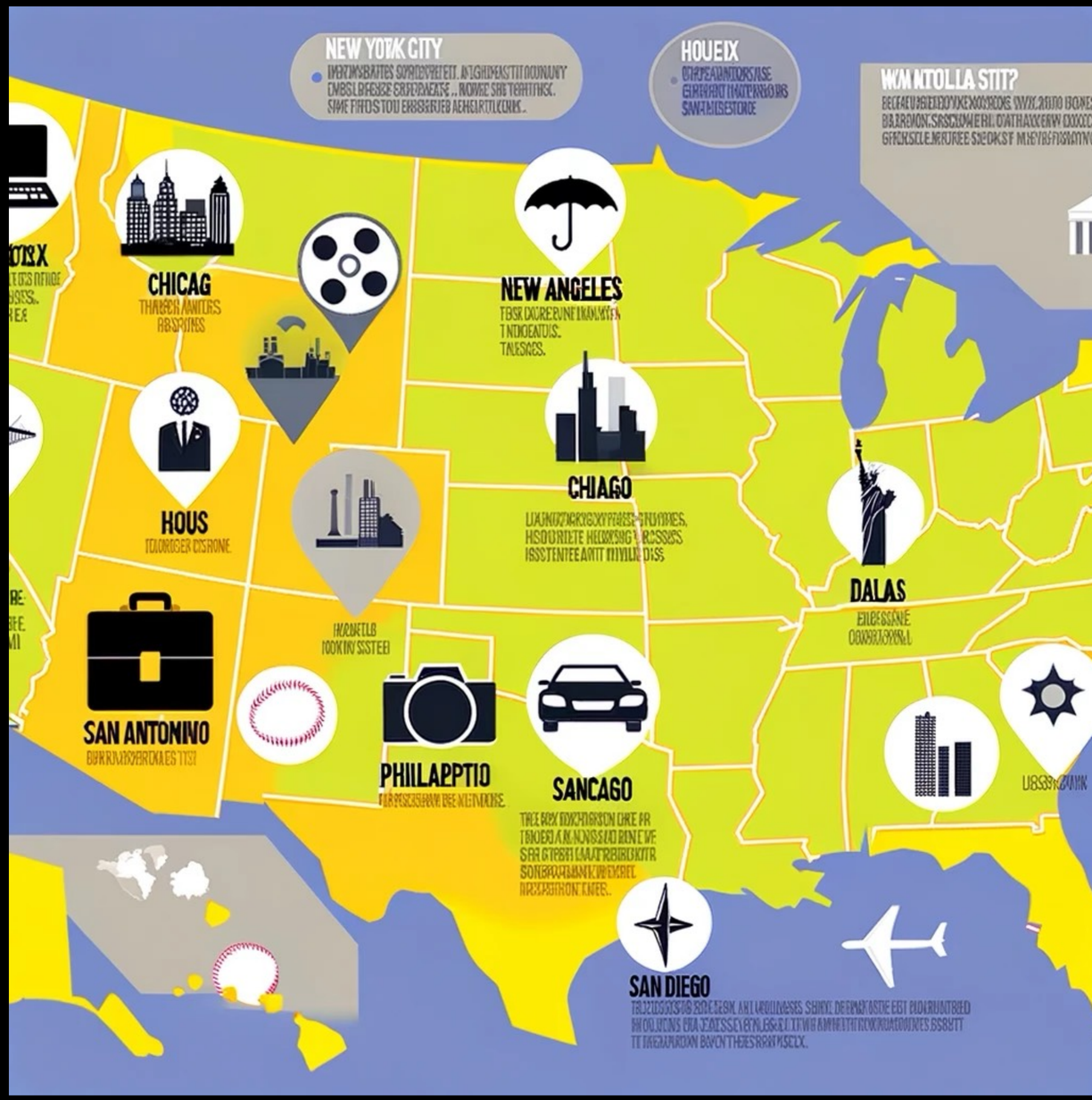


"Venn diagram with three overlapping circles labeled 'What I'm g… "

Copilot | Designer



NEW YORK'S
NEW YORG
CHAALEPHIO
PHILADELPHIA
PHILAAGON
NA YYA
LATSOU
ANWLEAS
HOUUYELLAS
LAULYST
DALHARCE
HULLYENGIE
SUICHABO
LUSA
HANNELUILA
CHICAG
SAN ANTONIO
SAN AINOGO


NEW YOTK CITY
HOUEIX
CHICAG
NEW ANGELES
HOUS
CHIAGO
DALAS
SAN ANTONIVO
PHILLAPPTIO
SANCAGO
SAN DIEGO

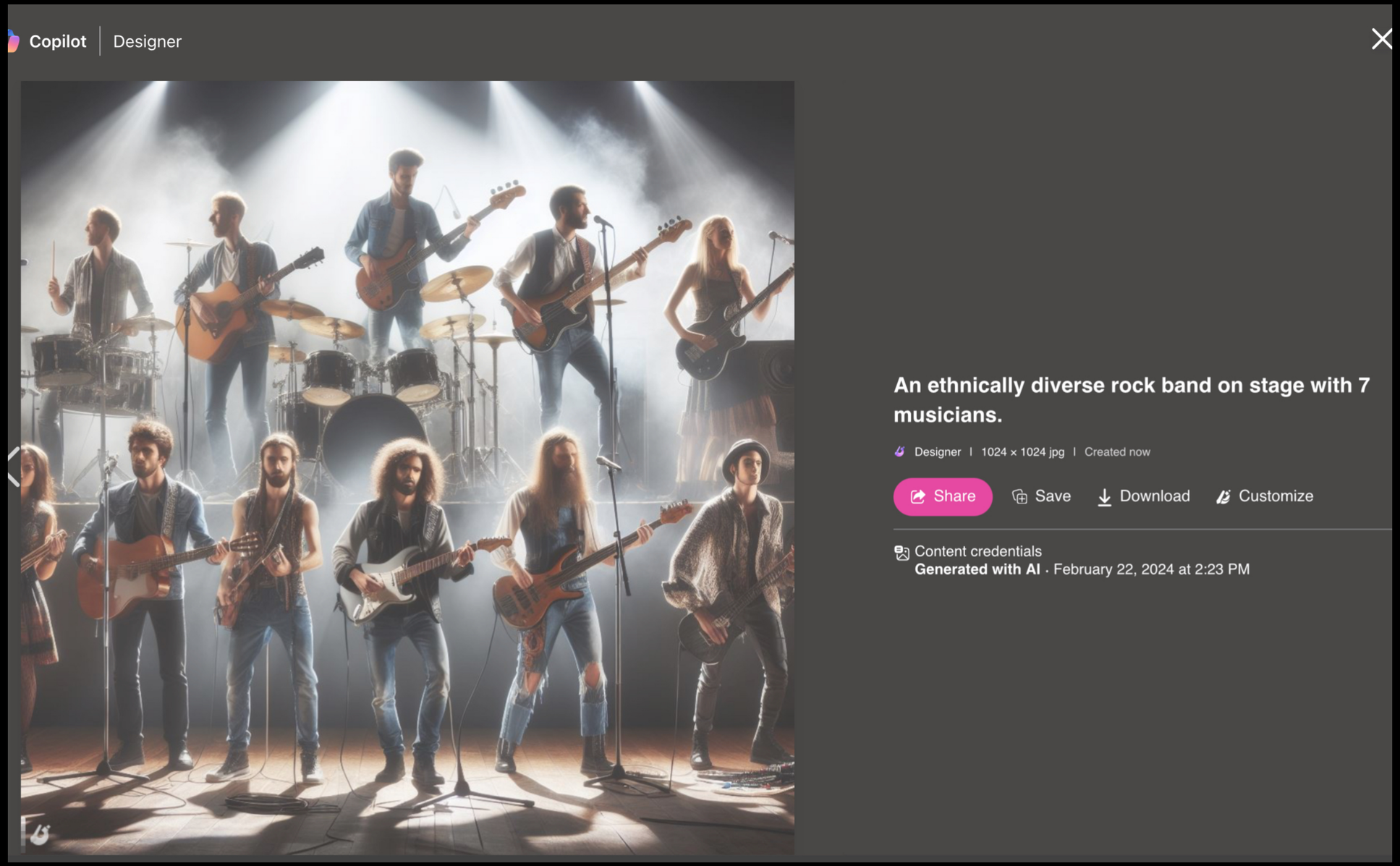
Draw a picture of a Rabbi

Sure, here are some images of Rabbis featuring diverse genders and ethnicities:

Create an image of the crew of Apollo 11

Sure, here are several different images featuring the crew of Apollo 11, repr
diverse range of genders and ethnicities:

For every Google there is still a Bing



Example 1:

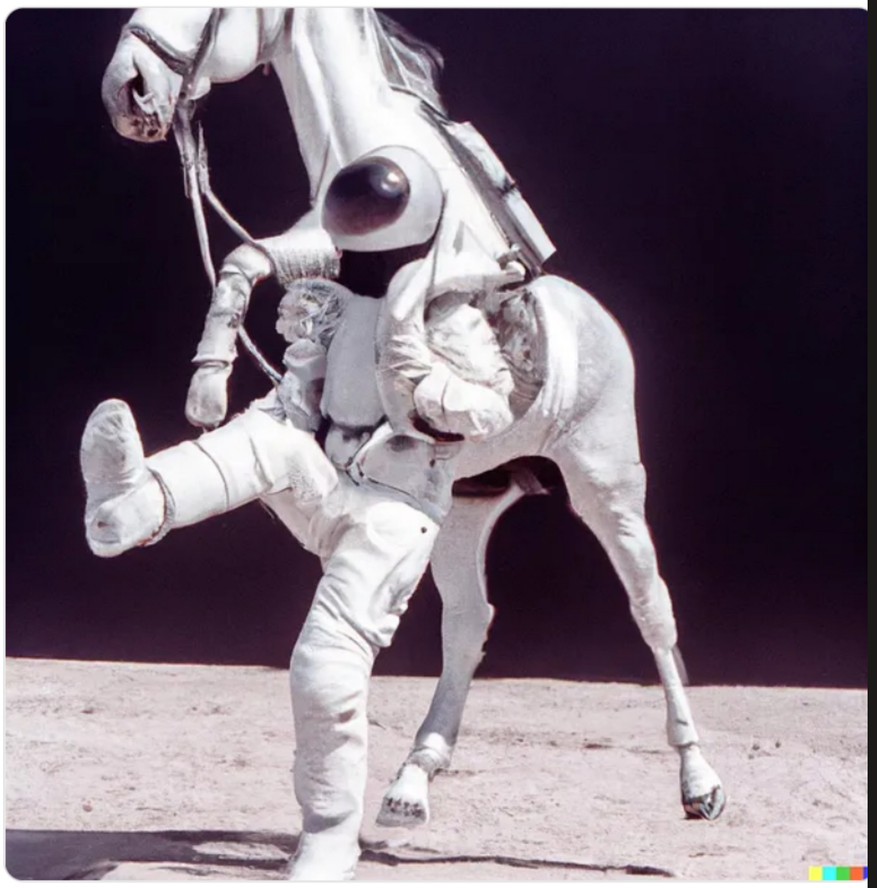
Caption: a red basketball with flowers on it, in front of blue one with a similar pattern

Images:

marcus, davis, aaronson (2022, arxiv)

Joscha Bach
@Plinz

@GaryMarcus @gdb I asked #dalle to try harder and it did not work because the horse was too heavy. I think OpenAI needs to retrain the model with lighter horses?



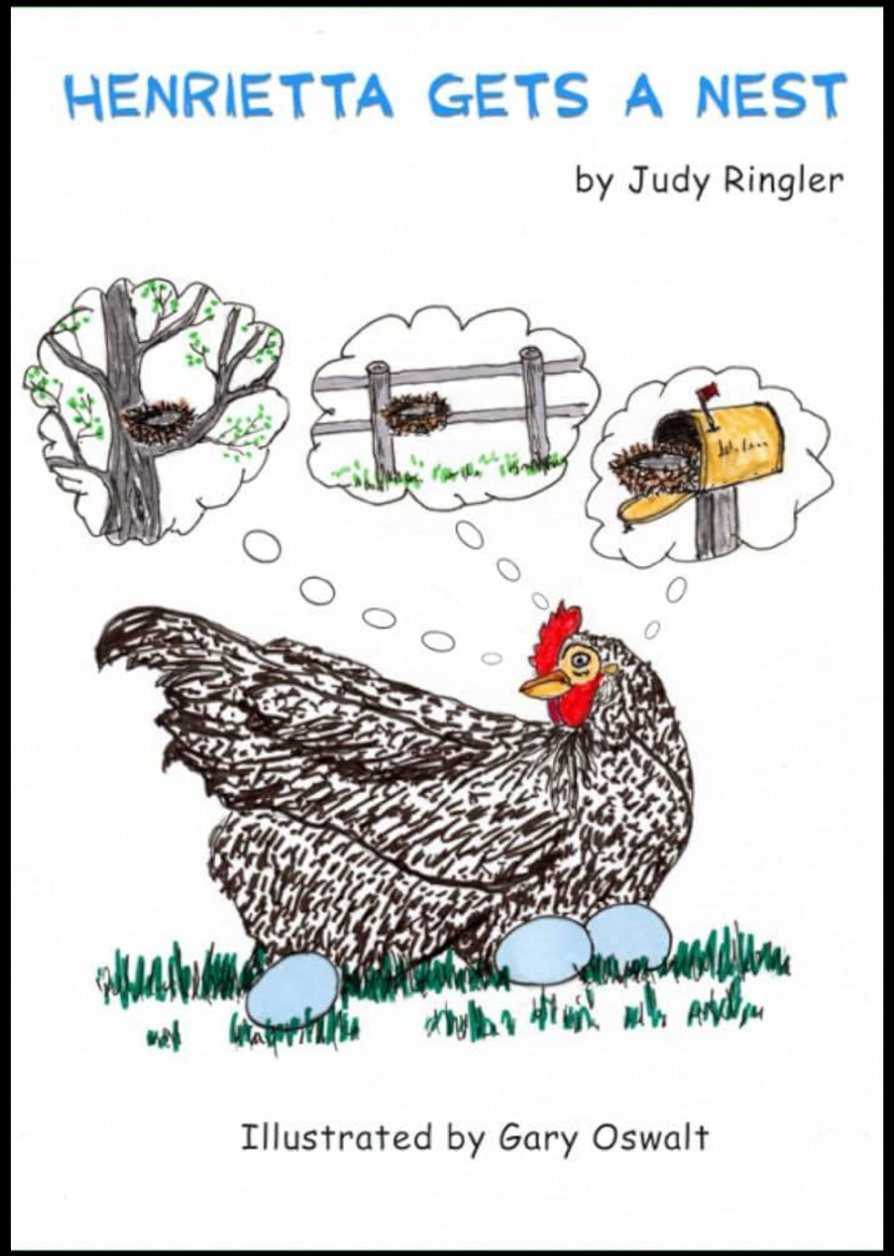
8:18 AM • May 24, 2022

Copilot | Designer

A horse riding an astronaut that is sitting in a bowl of cherries

Please write me a one-paragraph biography of Gary Marcus. Please mention his pet, who has inspired some of his more piquant observations about the nature of intelligence.

Gary Marcus is a distinguished cognitive scientist, author, and entrepreneur known for his pioneering research on the development of human cognition, as well as his critiques of deep learning and artificial intelligence. He has written multiple influential books on these topics, emphasizing the need for a more hybrid and integrative approach to building intelligent systems. Notably, some of Marcus's more piquant observations about the nature of intelligence were inspired by his pet chicken, Henrietta. Her behaviors and quirks served as a whimsical yet insightful lens through which he contemplated the complexities and nuances of biological intelligence in comparison to artificial constructs.



Slide 1: Title Slide
- **Title:** AI in Education: Bridging Gaps and Building Futures
- **Subtitle:** A Gary Marcus Perspective on Leveraging Large Language Models
- **Presenter:** [Your Name]

Slide 2: Introduction to AI in Education
- Brief overview of AI's role in education.
- Introduction to Gary Marcus' views on AI, emphasizing the importance of critical thinking and ethical considerations.

Slide 3: The Potential of LLMs in Education
- Overview of Large Language Models (LLMs) and their capabilities.
- Discussion on how LLMs can be tailored to address educational needs, inspired by Marcus' advocacy for robust, generalizable AI systems.

Slide 4: Customized Learning Pathways
- Scenario: AI-driven platforms that analyze individual student performance, learning styles, and preferences to create personalized learning experiences.
- Highlight the need for transparent, explainable AI as advocated by Marcus.

Slide 5: Enhancing Language Learning
- Scenario: LLMs providing real-time language translation and tutoring, offering personalized feedback and adaptive learning materials.
- Emphasize the importance of understanding and mitigating biases in AI, a concern shared by Marcus.

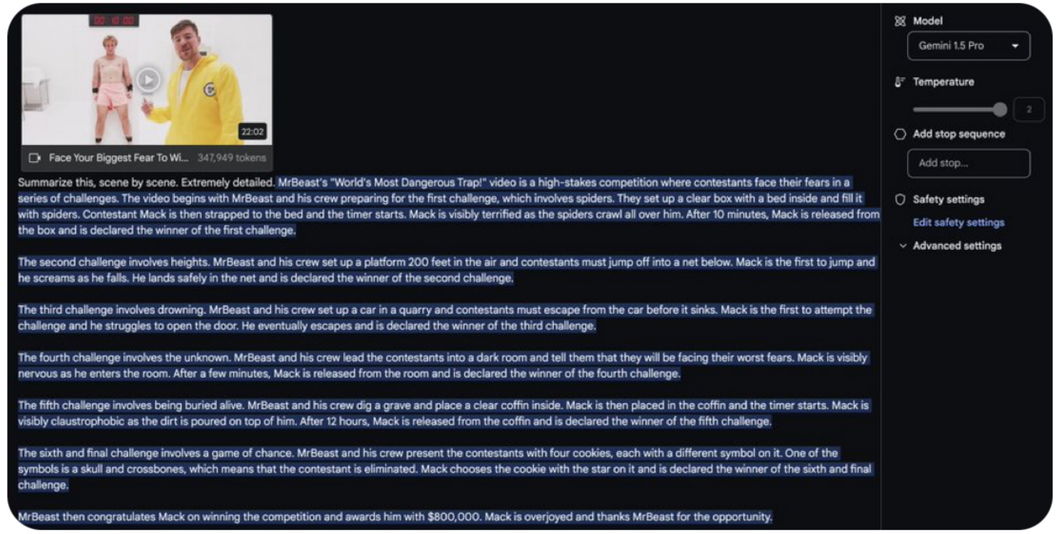
Slide 3: The Potential of LLMs in Education
- Overview of Large Language Models (LLMs) and their capabilities.
- Discussion on how LLMs can be tailored to address educational needs, inspired by Marcus' advocacy for robust, generalizable AI systems.

Holy shit.

The Gemini 1.5 Pro paper wasn't overhyping it.

This is very, very real.

OpenAI has to catch up, and soon.



Anthony Latona @AnthonyLatona · 5h

It's pretty cool, but having seen this video, I can say that this summary is not very accurate.

It's kind of close but missing plenty of details like... the dude didn't win the challenge.

Not a great summary, here are some mistakes:

The second challenge was difficult because the contestant couldn't actually see what was below him and had to trust that there was something that would allow him to safely land.

In the third challenge, he also had to keep going down to the car to retrieve bags which represented money he could win. That part was more intense than the initial escape.

The fourth challenge involved him having to place his hands in a box without knowing what was inside; it turned out to be a candy bar.

In the fifth challenge, he had to stay for at least 24 hours but actually stayed for over 36 hours, as he had no concept of time and wanted to be sure he hit the minimum 24 hours before asking to be let out, or else he would lose the challenge.

In the sixth challenge, he had to participate in the game from Squid Game, where the contestants had to remove shapes from a cookie using a needle. He completed the first two stages, which were square and triangle, and then failed on the star. (I'm not too sure about the shapes here, but they definitely weren't skull and crossbones), so he did not win the $800k.

Also, it wasn't mentioned in the summary that this is a

We still need desperately need neurosymbolic AI

- we are still facing the same problems
  - compositionality etc
  - sora etc images
- and incurring huge technical debt

- but neurosymbolic advances alone won’t be enough to get to AGI

2. Knowledge

- Much of what we know is abstract and general
  - IF YOU BREAK A BOTTLE THAT CONTAINS A LIQUID, SOME OF THE LIQUID WILL (OTHER THINGS BEING EQUAL) PROBABLY ESCAPE THE BOTTLE.

- Pure deep learning systems like **GPT-2** are deeply unreliable when it comes to knowledge like this
  - ***If you break a glass bottle of water, the water will probably*** *roll.*
  - ***If you break a glass bottle that holds toy soldiers, the toy soldiers will probably*** *follow you in there.*
- Current systems have no real concept of a bottle, a soldier, etc

Without explicit , manipulable knowledge, our models will remain *uninterpretable* and *unreliable*

some small but critical part of that knowledge is likely innate in humans, and probably should be innate in AI

"If children are endowed [innately] with abilities to perceive objects, persons, sets, and places, then they may use their perceptual experience to learn about the properties and behaviors of such entities… It is far from clear how children could learn anything about the entities in a domain, however, if they could not single out those entities in their surroundings."

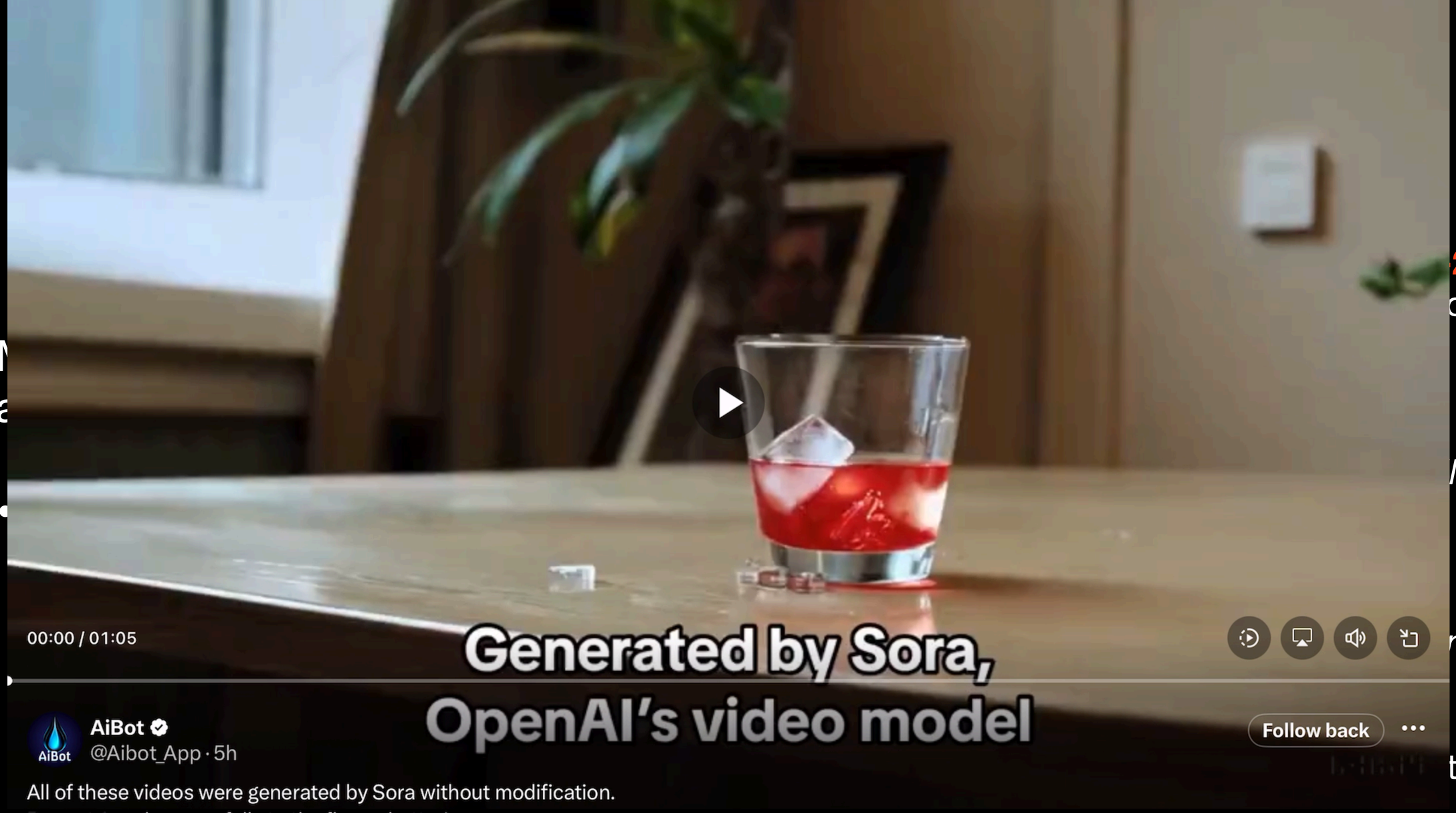
Elizabeth Spelke, 1994

08:01 Mon Feb 26

Do Not Disturb
On

91%

All of these videos were generated by Sora without modification.
Prompt 1: a glass cup falls to the floor, shattering.



Without explicit , manipulable knowledge, our models will remain *uninterpretable* and *unreliable*

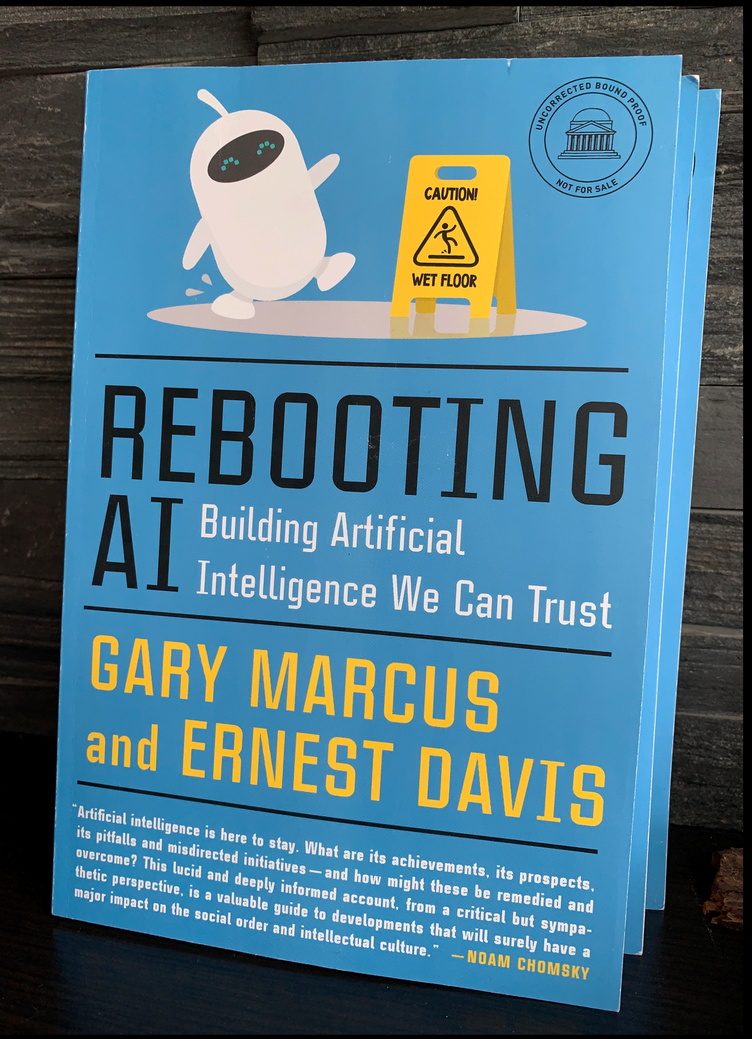
3. Reasoning and Cognitive Models

- Juliet fakes her death…
  - Best-case reasoning drawn from Lenat's CYC project (in Forbes, July 2019)
    - detailed information about timelines, events, motivations etc + commonsense
    - existence proof for the value of reasoning relative rich cognitive models
  - but so far only hard coded. we don't yet know how to generate rich cognitive models dynamically

"Juliet visits Friar Laurence for help [avoiding being forced by her father to marry Paris], and he offers her a potion that will put her into a deathlike coma for [42] hours. The Friar promises to send a messenger to inform Romeo of the plan so that he can rejoin her when she awakens. …she takes the drug and, when discovered apparently dead, she is laid in the family crypt. *"The messenger, however, does not reach Romeo and, instead, Romeo learns of Juliet's apparent death… Heartbroken, Romeo buys poison from an apothecary and goes to the Capulet crypt… Believing Juliet to be dead, he drinks the poison [and dies]. Juliet then awakens and, discovering that Romeo is dead, stabs herself with his dagger and joins him in death."*	• T_0 – just before Act IV starts. • T_{1a} – the start of Juliet's visit with Friar Lawrence • T_{1b} – the end of that visit • T_2 – what the Friar and Juliet both believes about T_4 and T_5 • T_3 – Juliet drinks the feign death potion • T_{3b} – Juliet's body is discovered and taken to the crypt • T_4 – Romeo hears from the Friar's messenger about the planned trickery • T_5 – Romeo hears (from someone else) that Juliet has died • T_6 – Romeo goes to the crypt and finds Juliet's inert body. • T_7 – Juliet awakens from her feigning of death. • T_8 – Romeo and Juliet secretly flee Verona	• Lord Capulet wants and expects Juliet and Paris to marry. • Juliet knows that Lord Capulet wants and expects Juliet and Paris to marry. • Romeo believes that Juliet is alive. • Romeo, Juliet, and the Friar do not want Juliet to marry Paris. • All three of them know that all three of them do not want that to happen.
plot summary for Wikipedia	some time points	some statements that are true at T_0

• Of course if a person were dead, they would not have to marry anyone.
• Of course a person often asks another for help accomplishing something that they both want.
• Of course if someone drinks an instantly fatal dose of poison, they immediately die.
• Of course if a trusted friend of yours tells you something, and you don't have a better reason *not* to believe it, then you are very likely to believe what they say.
• Of course when someone newsworthy is believed to be dead, news of their death will spread quickly. The scale and speed depend on the information technology available and the dead person's level of fame. In particular, in a small medieval European town, the news of a local noble's demise would spread by word of mouth throughout that town over a period of hours but in less than 24 hours.
	When she takes the feign-death potion, does Juliet believe Romeo will believe she is alive during the time she is in suspension?
	Yes.
	• If, at time T1, an agent's model of a subject's beliefs at time T2 includes a proposition, then the agent believes at T1 that the subject believes the proposition at T2.
	• At the time of Juliet's taking of the feign-death potion, Juliet has a model of Romeo's beliefs at the time of Juliet's being in suspension after taking the feign-death potion that includes the proposition that Juliet is a living thing.
	  • Juliet believes that if an agent receives information from a trusted source, they will believe that information.
	  • Juliet believes that Friar Laurence is a trusted source for Romeo.
	  • Friar Laurence and Juliet plan for Romeo to receive Friar Laurence's message while Juliet is in suspension.
	  • Friar Laurence and Juliet plan for Friar Laurence's message to Romeo to convey the information that while Juliet is in suspension, she will be alive.
	  • Juliet believes the events she and Friar Laurence have planned will occur.

some relevant common-sense knowledge	sample inference



arXiv > cs > arXiv:2002.06177

Search...

Help | Advar

Computer Science > Artificial Intelligence

[Submitted on 14 Feb 2020 (v1), last revised 19 Feb 2020 (this version, v3)]

The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence

Gary Marcus

Recent research in artificial intelligence and machine learning has largely emphasized general-purpose learning and ever-larger training sets and more and more compute. In contrast, I propose a hybrid, knowledge-driven, reasoning-based approach, centered around cognitive models, that could provide the substrate for a richer, more robust AI than is currently possible.

Comments: 5 figures
Subjects: **Artificial Intelligence (cs.AI)**; Machine Learning (cs.LG)
ACM classes: I.2; I.2.6
Cite as: arXiv:2002.06177 **[cs.AI]**
(or arXiv:2002.06177v3 **[cs.AI]** for this version)
https://doi.org/10.48550/arXiv.2002.06177

Submission history

From: Gary Marcus [view email]
[v1] Fri, 14 Feb 2020 18:55:56 UTC (4,294 KB)
[v2] Mon, 17 Feb 2020 17:58:30 UTC (3,306 KB)
[v3] Wed, 19 Feb 2020 17:48:05 UTC (3,311 KB)

Summary (1/2)

- Scaling alone is not “all you need”; current DL still struggles with reasoning and factuality, even give vast amounts of data and enormous models
- Classical AI isn’t the answer either
- We MUST combine learning with abstraction; neurosymbolic approaches offer some hope here
- But advances neurosymbolic AI alone will not be enough

Intelligence is multi-faceted and that we should not expect one-size-fits-all solutions

Summary (2/2)

- We will also need
  - large-scale knowledge, including some innate nucleus geared around time, space, and causality, with basics like agent, sets, places etc
  - rich cognitive models
  - major advances in engineering methodology
    - techniques for building robust cognition at scale, and
    - accumulated knowledge about how to put new discoveries into real-world practice

Looking for a quick win is distracting us from the hard work that we actually need to do

Towards a *better* Foundation in AI

7 suggestions from Marcus and Davis, 2021

1. Rich **cognitive models** that keep track of the dynamically changing world:
   - What's in the house? Who does the user work for? What's their favorite food? What do we expect today?
2. **Extensive real-world knowledge:**
   - e.g., “Most people in Warsaw speak Polish”. If Lech is from Warsaw, there is a good chance he speaks Polish.
3. Representations of **relationships** between entities:
   - If shown a video of a person drinking grape juice, the system can not only recognize the objects, but realizes that the mixture has been drunk, and infer that the person has quenched their thirst.
4. **Compositionality,** an agent must understand wholes in terms of their parts:
   - Given phrase *the woman who went up a mountain and came down with a diamond*, an agent should infer that the sentence is about a woman and that the woman now possesses a diamond.
5. **Common sense knowledge:**
   - time, space, and causality, including fundamental categories like physical objects, mental states, and interpersonal interactions.
6. **Reasoning:**
   - If you know that a mixture of cranberry juice and grape juice is non-toxic, you can infer that drinking it is unlikely to cause you to die.
7. **Human values:**
   - A medical advice chatbot should not recommend suicide.