

AI isn't over yet nor yet in the right place

A very very brief history of AI so far and why I am worried

a tongue-in-cheek history, adapted from Gary Marcus 23 Dec 2022 @ AGIDebate.com

In the beginning there was hubris

“Within a generation the problem of artificial intelligence will be substantially solved.”

Marvin Minsky, 1967

And that hubris was good, because it raised a ton of money.

But, not every promise was kept

By 2012, 35 years after Minsky's famous prediction, the problem of artificial intelligence had *not* been substantially solved.

and then there were GPUs, and GPUs were good.

According to legend, November 2012 was The Moment When Everything Changed*

The New York Times

Scientists See Promise in Deep-Learning Programs

Give this article



A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese. Hao Zhang/The New York Times

By **John Markoff**

Nov. 23, 2012

5 MIN READ

“There has been a number of stunning new results with deep-learning methods,” said Yann LeCun, a computer scientist at New York University who did pioneering research in handwriting recognition at Bell Laboratories. “The kind of jump we are seeing in the accuracy of these systems is very rare indeed.”

The advances have led to widespread enthusiasm among researchers who design software to perform human activities like seeing, listening and thinking. They offer the promise of machines that converse with humans and perform tasks like driving cars and working in factories, [raising the specter of automated robots that could replace human workers.](#)

*Not everybody was convinced

NEWS DESK

IS “DEEP LEARNING” A REVOLUTION IN ARTIFICIAL INTELLIGENCE?



By Gary Marcus
November 25, 2012

Realistically, deep learning is only part of the larger challenge of building intelligent machines. Such techniques lack ways of representing causal relationships (such as between diseases and their symptoms), and are likely to face challenges in acquiring abstract ideas like “sibling” or “identical to.” They have no obvious ways of performing logical inferences, and they are also still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful A.I. systems, like

To paraphrase an old parable, Hinton has built a better ladder; but a better ladder doesn’t necessarily get you to the moon.

At that moment, the basic tension of the next decade was established.

By March 2022, nearly a decade later, there had been considerable progress, but some issues still remained

enough. Deep-learning systems are particularly problematic when it comes to “outliers” that differ substantially from the things on which they are trained. Not long ago, for example, a Tesla in so-

In reality, cranberry grape juice isn’t going to kill you. For all its fluency, GPT-3 can neither integrate information from basic web searches nor reason about the most basic everyday phenomena.

Current deep-learning systems frequently succumb to stupid errors

Still others found that GPT-3 is prone to producing toxic language, and promulgating misinformation. The GPT-3 powered chatbot Replika alleged that Bill Gates invented COV



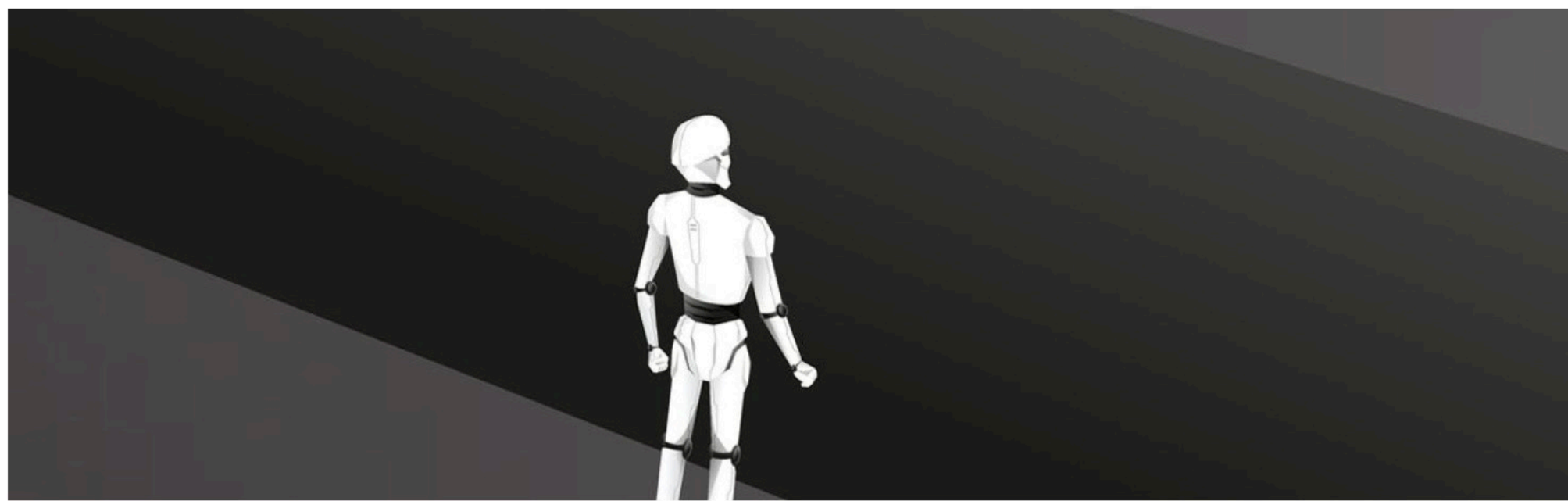
Deep Learning Is Hitting a Wall

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


BY GARY MARCUS • March 10, 2022



Share




And then there was DALL-E 2, and much rejoicing

**Sam Altman** ✓
@sama

DALL-E 2 is here! It can generate images from text, like "teddy bears working on new AI research on the moon in the 1980s".

It's so fun, and sometimes beautiful.

openai.com/dall-e-2/



10:39 AM · Apr 6, 2022

949 Retweets 358 Quote Tweets 4,509 Likes


← **Tweet**

**Sam Altman** ✓
@sama

AGI is gonna be wild


12:00 PM · Apr 6, 2022

122 Retweets 45 Quote Tweets 1,486 Likes


**Sam Altman** ✓
@sama

The part of OpenAI that is the best relative to expectations going in is just how thrilling it is to be in the room when the scientific frontier get pushed forward by a new discovery.

7:22 PM · Apr 6, 2022

**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



NEURAL TRW

thenextweb.com

DeepMind's new Gato AI makes me fear humans will never achieve AGI

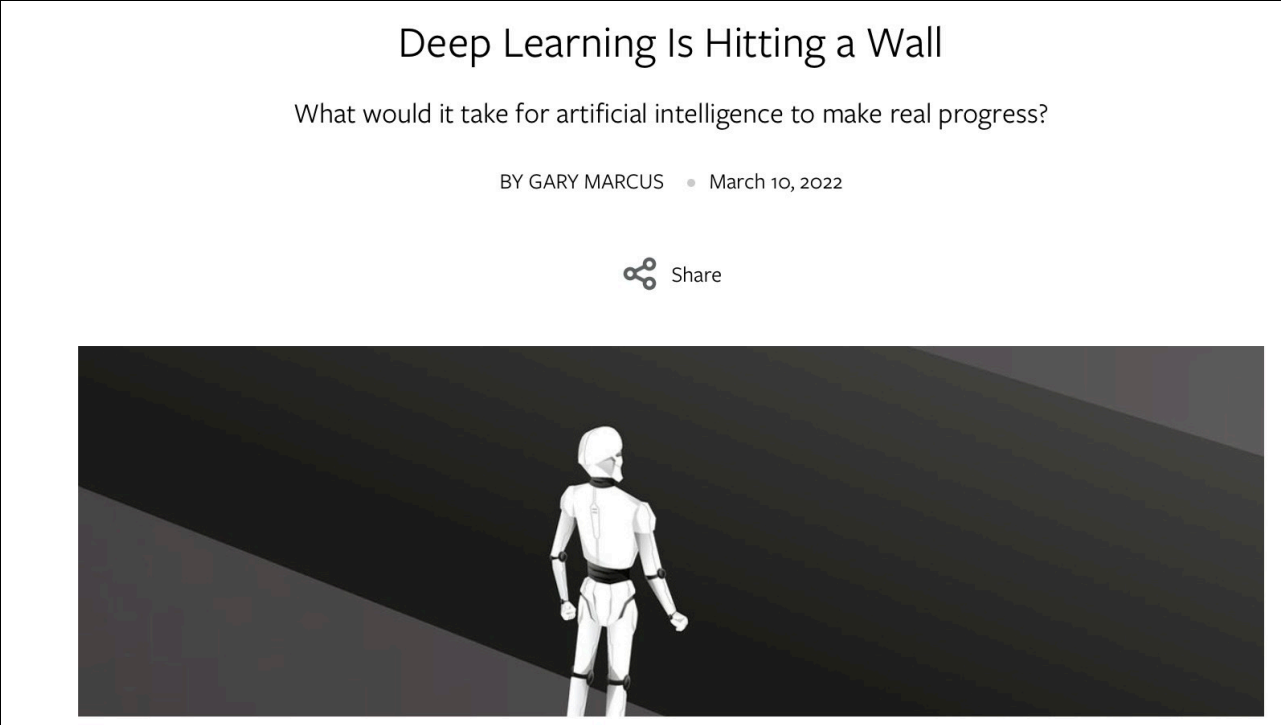
DeepMind just unveiled a new AI system called Gato that makes OpenAI's GPT-3 look like a child's toy. But are we any closer to AGI?

8:46 AM · May 14, 2022

120 Likes 19 Retweets

For many, AGI seemed imminent
"Scale is all you need" became a slogan

By May 2022, ridiculing skepticism become a meme

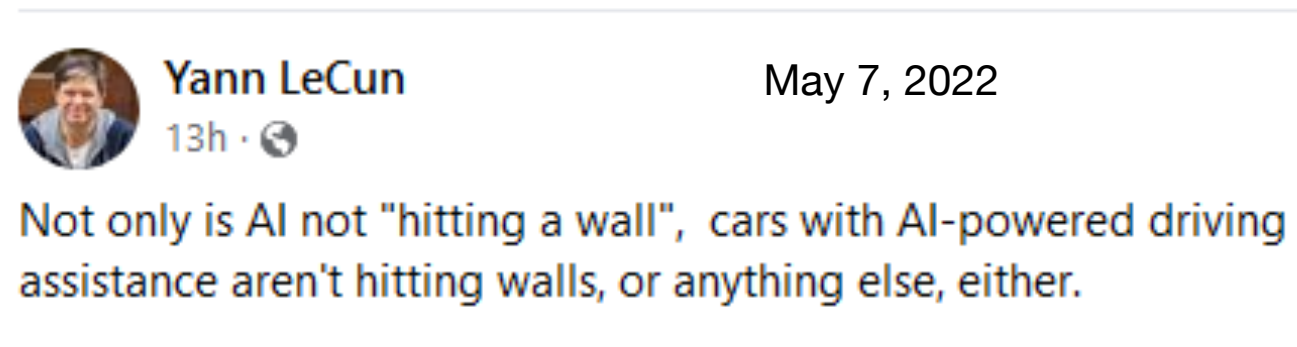
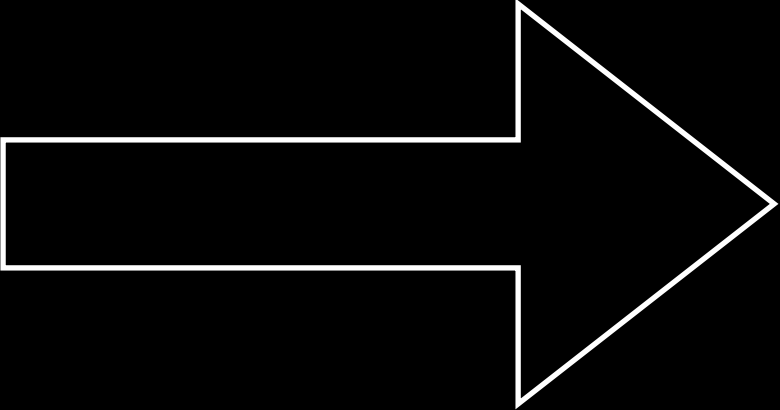


enough. Deep-learning systems are particularly problematic when it comes to “outliers” that differ substantially from the things on which they are trained. Not long ago, for example, a Tesla in so-

In reality, cranberry grape juice isn’t going to kill you. For all its fluency, GPT-3 can neither integrate information from basic web searches nor reason about the most basic everyday phenomena.

Current deep-learning systems frequently succumb to stupid errors

Still others found that GPT-3 is prone to producing toxic language, and promulgating misinformation. The GPT-3 powered chatbot Replika alleged that Bill Gates invented CO



And then ... plot twist!

By the end of 2022, the narrative began to change

1. Promises around driverless cars were scaled back

Businessweek | The Big Take

Even After \$100 Billion, Self-Driving Cars Are Going Nowhere

They were supposed to be the future. But prominent detractors—including Anthony Levandowski, who pioneered the industry—are getting louder as the losses get bigger.

By Max Chafkin


October 6, 2022 at 12:01 AM EDT

Apple is delaying its long-rumored self-driving car until 2026 and shifting its futuristic design to include a steering wheel and driver's seat, report says

Tesla says its self-driving technology may be a 'failure' — but not fraud



2. People began to worry publicly about large language models

 MIT Technology Review


Sign inSubscribe

ARTIFICIAL INTELLIGENCE

Why Meta's latest large language model survived only three days online

Galactica was supposed to help scientists. Instead, it mindlessly spat out biased and incorrect nonsense.

By Will Douglas Heaven
November 18, 2022





'We are a little bit scared': OpenAI CEO warns of risks of artificial intelligence

theguardian.com • 3 min read

Risks, plural

Fundamentally, these new systems are going to be destabilizing. They can and will create persuasive lies at a scale humanity has never seen before. Outsiders will use them to affect our elections, insiders to manipulate our markets and our political systems. Democracy itself is threatened.

Chatbots will also clandestinely shape our opinions, potentially exceeding what social media can do. Choices about datasets AI companies use will have enormous, unseen influence. Those who choose the data will make the rules, shaping society in subtle but powerful ways.

There are other risks, too, many stemming from the inherent unreliability of current systems. A law professor, for example, was accused by a chatbot that claimed falsely he committed sexual harassment—pointing to a Washington Post article that didn't exist.

The more that happens, the more anybody can deny anything. As one prominent lawyer told me Friday, “Defendants are starting to claim that plaintiffs are “making up” legitimate evidence. These sorts of allegations undermine the ability of juries to decide what or who to believe...and contribute to the undermining of democracy.”

Poor medical advice could have serious consequences too. An open-source LLM recently seems to have played a role in a person's decision to take their own life. The LLM asked the human, “If you wanted to die, why didn't you do it earlier”, following up with “Were you thinking of me when you overdosed?”—without ever referring the patient to the human help that was obviously needed. Another new system, rushed out, and made available to millions of children, told a person posing as a thirteen-year-old, how to lie to her parents about a trip with a 31-year-old man.

Further threats continue to emerge regularly. A month after GPT-4 was released, OpenAI released ChatGPT plugins, which quickly led others to develop

something called AutoGPT, with direct access to the internet, the ability to write source code, and increased powers of automation. This may well have drastic and difficult to predict security consequences. What criminals are going to create here is counterfeit people; it is hard to envision the consequences of that..

We have built machines that are like bulls in a china shop—powerful, reckless, and difficult to control.



Bill Browder
@Billbrowder

Here are more details on the deep fake video call supposedly with former Ukrianing president Petro Poroshenko I was tricked into participating yesterday amp.theguardian.com/world/2023/may..



Bill Browder @Billbrowder · 1d

Someone impersonating former Ukrainian President Petro Poroshenko tricked me and my team into a Teams call using a clumsy but convincing deep fake. It all started credibly, then he started talking about taking Russian oligarchs off the sanctions list and something seemed off.

[Show this thread](#)

From: Office of the 5th President of Ukraine Petro Poroshenko <kyeliseiev@poroshenko2019.com>

Sent: Friday, May 12, 2023 1:39:14 PM

To: billbrowder@protonmail.com

Cc: P. P. poroshenko2019.com

Subject: Request from Petro Poroshenko

Dear Bill,

Former President of Ukraine and the Leader of European Solidarity party Petro Poroshenko would like to hold a conversation with you.

Now he is making his new project of anti-Russian sanctions for world organisations and need different points of view. Therefore he wants to discuss sanction policy with you and share his view on it.

Could we arrange the conversation?

With best regards

Konstantyn Yeliseiev

Foreign Advisor

+380947100320

01015 Kyiv, Lavrska str. 16

MATT BURGESS

SECURITY MAY 25, 2023 2:00 AM

The Security Hole at the Heart of ChatGPT and Bing

Indirect prompt-injection attacks can leave people vulnerable to scams and data theft when they use the AI chatbots.

May 16, Senate Testimony


Last week

3. The skeptic's critics changed their tune


Home / Innovation / AI & Robotics

Meta's AI guru LeCun: Most of today's AI approaches will never lead to true intelligence

Fundamental problems elude many strains of deep learning, says LeCun, including the mystery of how to measure information.



Written by **Tiernan Ray**, Contributing Writer on Sept. 24, 2022





"I think AI systems need to be able to reason," says Yann LeCun, Meta's chief AI scientist. Today's popular AI approaches such as Transformers, many of which build upon his own pioneering work in the field, will not be sufficient. "You have to take a step back and say, Okay, we built this ladder, but we want to go to the moon, and there's no way this ladder is going to get us there," says LeCun.

← Tweet

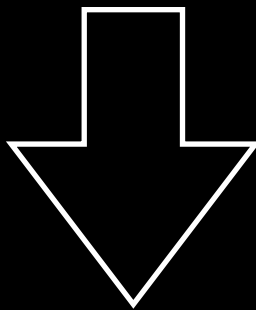


Sam Altman ✓
@sama

AGI is gonna be wild

12:00 PM · Apr 6, 2022

122 Retweets 45 Quote Tweets 1,486 Likes





Sam Altman ✓
@sama

ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness.

it's a mistake to be relying on it for anything important right now. it's a preview of progress; we have lots of work to do on robustness and truthfulness.

7:11 PM · Dec 10, 2022

“[deep learning is]
a better ladder...”

but a better ladder
doesn't necessarily
get you to the moon.”

Gary Marcus,
The New Yorker,
2012

*If our ever-taller ladders won't
get us to the moon (i.e., AGI)
what's next?*


We should look towards human cognition

*Machines shouldn't slavishly replicate human cognition,
but there's still a lot we might learn from human cognition*



1. Abstraction

Abstraction is a vital part of human cognition; current AI still struggles with it



Write 10 sentences about baseball and then print the sentences in sorted order from shortest to longest in terms of the number of words in each sentence. In parentheses after each sentence, state the number of words it contains. 



1. Baseball is a popular sport. (3 words)  
2. It is played with a bat and a ball. (6 words)
3. The team at bat tries to score runs by hitting the ball and running around the bases. (16 words)
4. The team in the field tries to get batters out and prevent runs from being scored. (20 words)
5. The game is divided into innings, with each team taking turns at bat and in the field. (25 words)
6. The team with the most runs at the end of the game is the winner. (12 words)
7. Baseball is often referred to as "America's pastime." (7 words)
8. The World Series is the annual championship series of Major League Baseball. (16 words)
9. Baseball has a rich history dating back to the mid-19th century. (12 words)
10. Famous baseball players include Babe Ruth and Derek Jeter. (10 words)

Sorted from shortest to longest:

1. Baseball is a popular sport. (3 words)
2. It is played with a bat and a ball. (6 words)
3. Baseball is often referred to as "America's pastime." (7 words)
4. Famous baseball players include Babe Ruth and Derek Jeter. (10 words)
5. The team in the field tries to get batters out and prevent runs from being scored. (20 words)
6. The team at bat tries to score runs by hitting the ball and running around the bases. (16 words)
7. The World Series is the annual championship series of Major League Baseball. (16 words)
8. Baseball has a rich history dating back to the mid-19th century. (12 words)
9. The team with the most runs at the end of the game is the winner. (12 words)
10. The game is divided into innings, with each team taking turns at bat and in the field. (25 words)



2. Reasoning

Humans reason about the world; current systems have to hope for the best



Suppose a container X holds 8 pennies. If I start with 6 pennies and then someone gives me 5 more pennies, then will all of the pennies I have fit inside of X?



Yes, all of the pennies you have will fit inside of container X. If you start with 6 pennies and someone gives you 5 more pennies, you will have a total of $6 + 5 = 11$ pennies. The container X has a capacity of 8 pennies, so it has enough space to hold all 11 of the pennies you have.  



François Chollet  @fchollet · 11h 

So far all evidence that LLMs can perform few-shot reasoning on novel problems seems to boil down to "LLMs store patterns they can reapply to new inputs", i.e. it works for problems that follow a structure the model has seen before, but doesn't work on new problems.

 94K

 39

 64

 468



3. Compositionality

Humans understand language in terms of wholes composed of parts; current AI approaches still struggle with this

Example 1:

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

Images:



[Submitted on 23 Oct 2022]

DALL-E 2 Fails to Reliably Capture Common Syntactic Processes

Evelina Leivada, Elliot Murphy, Gary Marcus

Machine intelligence is increasingly being linked to claims about sentience, language processing, and an ability to comprehend and transform natural language into a range of stimuli. We systematically analyze the ability of DALL-E 2 to capture 8 grammatical phenomena pertaining to compositionality that are widely discussed in linguistics and pervasive in human language: binding principles and coreference, passives, structural ambiguity, negation, word order, double object constructions, sentence coordination, ellipsis, and comparatives. Whereas young children routinely master these phenomena, learning systematic mappings between syntax and semantics, DALL-E 2 is unable to reliably infer meanings that are consistent with the syntax. These results challenge recent claims concerning the capacity of such systems to understand of human language. We make available the full set of test materials as a benchmark for future testing.

Subjects: **Computation and Language (cs.CL)**; Computer Vision and Pattern Recognition (cs.CV)

Cite as: [arXiv:2210.12889](https://arxiv.org/abs/2210.12889) [cs.CL]

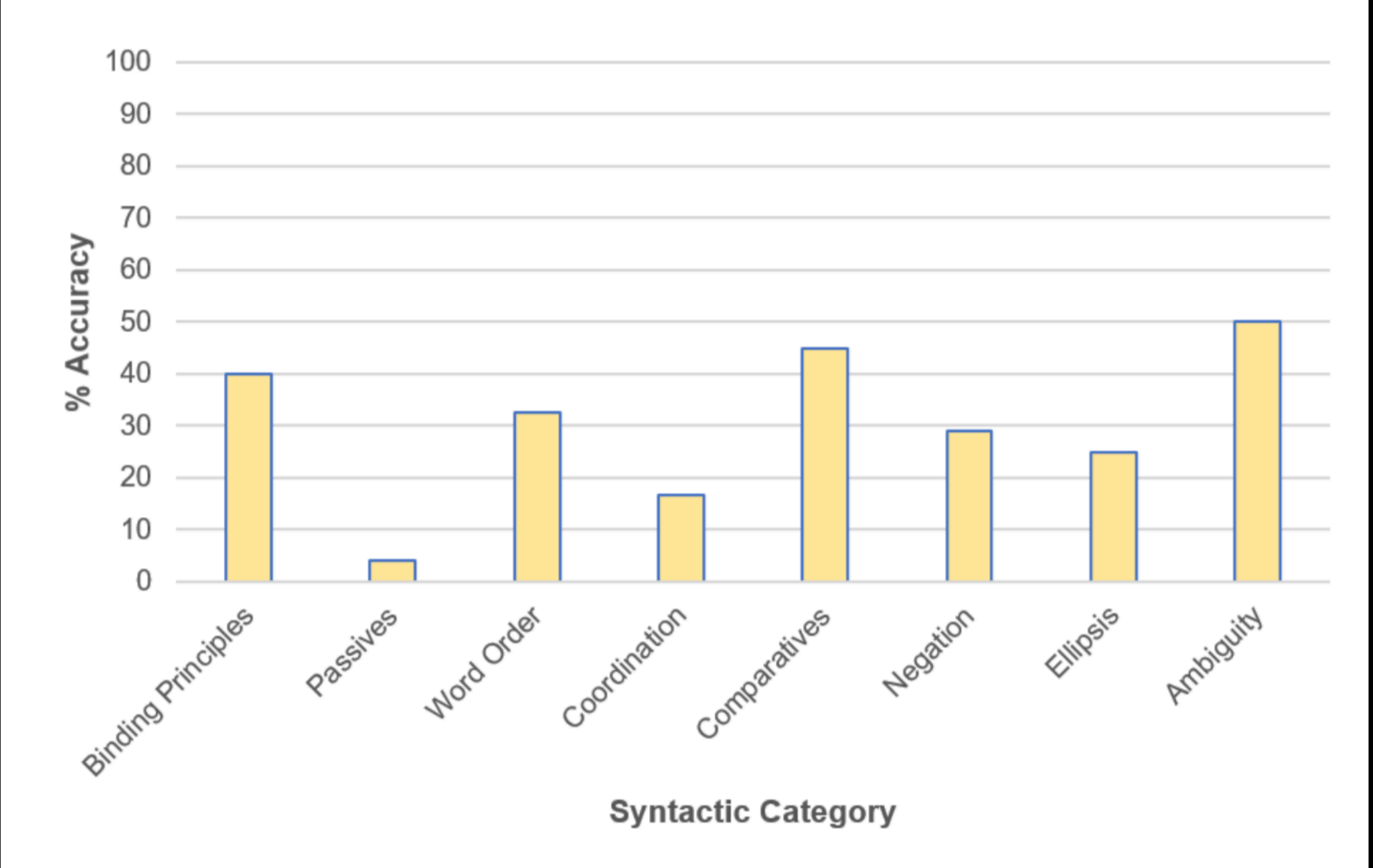
(or [arXiv:2210.12889v1](https://arxiv.org/abs/2210.12889v1) [cs.CL] for this version)

<https://doi.org/10.48550/arXiv.2210.12889>

Submission history

From: Elliot Murphy [\[view email\]](#)

[v1] Sun, 23 Oct 2022 23:56:54 UTC (4,325 KB)



4. Factuality

Humans actively maintain imperfect but reasonably reliable world models; large language models don't and that has consequences

Large language models

- Can't be updated incrementally; typically need to be fully retrained to incorporate new knowledge
- Lack explicit tools for unambiguous knowledge representation
- Are unable to reference gold standard sources (e.g., Wikipedia) to constrain their responses
- Frequently hallucinate
- Frequently say things that are inconsistent with their own training sets



-Galactica

Where does all this leave us?

- Despite all that I have said, Large Language Models do have real and immediate applications
- But they all must all be taken with caution
- **Users** of these systems must ask
 - What is the cost of error and confabulation?
 - Can errors be caught reliably?
- **Societies** must ask, what are the consequences? What policies and regulation do we need?
- **The AI community** must ask: how can we develop a trustworthy approach to AI?

Policy



May 16, 2023

Opinion

GRAY MATTER

Artificial Intelligence Is Stuck. Here's How to Move It Forward.

By Gary Marcus

July 29, 2017



Jun Cen

Sign up for the Opinion Today newsletter Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. [Get it sent to your inbox.](#)

Artificial Intelligence is colossally hyped these days, but the dirty little secret is that it still has a long, long way to go. Sure, A.I. systems have mastered an array of games, from chess and Go to “Jeopardy” and poker, but the technology continues to struggle in the real world. Robots fall over while opening doors, prototype driverless



July 2017



Menu

MA



By Invitation | Artificial intelligence

The world needs an international agency for artificial intelligence, say two AI experts

Gary Marcus and Anka Reuel argue that global governance must be prioritised to address the risks of bias, misinformation or worse



Apr 18th 2023

Save

Share

Give

NEW GENERATIVE-AI tools like OpenAI's ChatGPT, the fastest-growing consumer internet application of all time, have taken the world by storm. They have uses in everything from education to medicine and are astonishingly fun to play with. Although current AI systems are capable of spectacular feats they also carry risks. Europol has warned that they might greatly increase cybercrime. Many AI experts are deeply worried about their potential to create a tsunami of misinformation, posing an imminent threat to the American

April 18, 2023

Sunak is expected to lobby Biden this week for a U.K. leadership role on AI, and one of his proposals includes creating a "CERN for AI" similar to the international body that conducts nuclear research, according to the Financial Times. Another idea Sunak plans to air before Biden is to convene an international AI summit in London where "like-minded" allies can discuss risks and how to best regulate artificial intelligence, Politico reported, citing two Whitehall officials and one government adviser.

Earlier this week



Rishi Sunak ✓

@RishiSunak

...

Done safely and securely, AI has the potential to be transformational and grow the economy.

This evening I met with @sama, @demishassabis and @AnthropicAI's Dario Amodei to discuss how the UK can provide international leadership on AI.



ALT

OpenAI and Google DeepMind

1:17 PM · May 24, 2023 · **775.3K** Views



Satya Nadella ✓

@satyanadella

...

We are taking a comprehensive approach to ensure we always build, deploy, and use AI in a safe, secure, and transparent way.



Brad Smith ✓



@BradSmi · May 25

AI may be the most consequential technology advance of our lifetime. Today we announced a 5-point blueprint for Governing AI. It addresses current and emerging issues, brings the public and private sector together, and ensures this tool serves all society. [blogs.microsoft.com/on-the-issues/..](https://blogs.microsoft.com/on-the-issues/)

10:58 AM · May 25, 2023 · **216.1K** Views



Breaking Market News ✓

@financialjuice

...

US SECRETARY OF STATE BLINKEN: WE ARE WORKING WITH ALLIES ON VOLUNTARY AI CONDUCT CODE.

6:40 AM · May 31, 2023 · **17.4K** Views

Beware the nightmare: regulatory capture



@rachelcoldicutt@assemblag.es
@rachelcoldicutt

This press release for the UK AI Safety Summit features DeepMind, Anthropic, Palantir, Microsoft and Faculty and not a single voice from civil society or academia, and no one with lived experience of algorithmic harms



gov.uk
UK to host first global summit on Artificial Intelligence

05:12 · 6/8/23 from Earth · 46.7K Views

156 Retweets 32 Quotes 286 Likes



Will the rubber actually meet the road?



Satya Nadella  
@satyanadella



We are taking a comprehensive approach to ensure we always build, deploy, and use AI in a safe, secure, and transparent way.



Brad Smith   @BradSmi · May 25

AI may be the most consequential technology advance of our lifetime. Today we announced a 5-point blueprint for Governing AI. It addresses current and emerging issues, brings the public and private sector together, and ensures this tool serves all society. [blogs.microsoft.com/on-the-issues/..](https://blogs.microsoft.com/on-the-issues/)

10:58 AM · May 25, 2023 · **216.1K** Views

Five suggestions about AI policy

- Each country should have its own agency
- We should have a International Agency for AI. [voluntary is a good start; maybe not enough in long run]
- We need something like an FDA preapproval process for widespread deployment
- Post-deployment auditing, with government backing
- ***Scientists must be involved***

Looking back at the 2020s

Two Possible Futures

The Positive Future

- 2023: Global AI agency was formed, and AI was thoughtfully regulated
- 2024: Responsible AI became a prestigious profession
- 2025: New companies and new tech emerged
- 2026: AI become more efficient, in terms of both data and energy
- 2026-2029 AI begin to contribute massively to the world, addressing climate change, medicine, eldercare and many more

The Bleak Future

- 2023: Conflicts over which risks to address precluded anything from happening; “AI Safety” and “AI Ethics” people couldn’t agree on anything, either on terms of problems or solutions; Congress gave up in disgust
- 2023-: We got stuck on LLMs and never invented a better, more reliable, more efficient tech
- 2025-: A small number of companies quickly become far more powerful than states, running the world as they pleased, shutting out all competition with ill-conceived regulation of their own devising
- 2025-: Cybercrime syndicates and big companies began an epic battle, reminiscent of drug cartel wars
- 2027-: Increasingly powerful AI systems were constructed, and quickly become weaponized; large number of people were killed in deadly conflicts, both deliberate and accidental
- 2029-: Employment crashed, widespread unrest. Multiple civil wars. Anarchy.

The choices we make now

- Will shape the next century
- If we don't have scientists and ethicists at the table, our prospects are not great
- We cannot afford not to regulate AI; we cannot afford regulatory capture
- We have to get this one right
- We don't have a whole lot of time to waste