

Generative AI Revolution (What Happens to Our Reality when AI Can Spin Many Synthetic Realities?)

Subbarao Kambhampati

School of Computing & AI



rao@asu.edu



[@rao2z](https://twitter.com/rao2z)



[@subbarao2z](https://www.linkedin.com/in/subbarao2z)



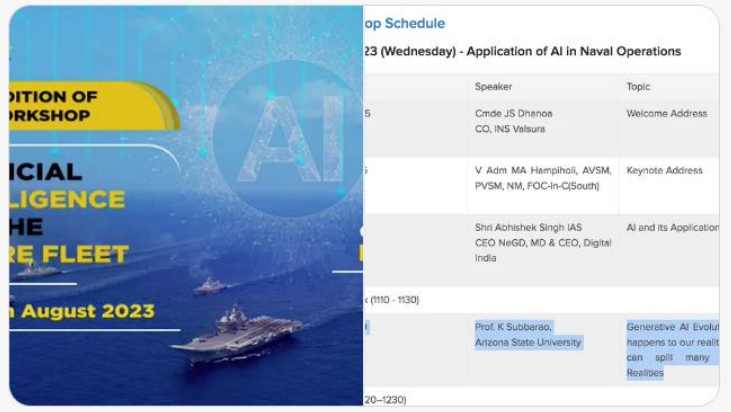
**5th EDITION OF
AI WORKSHOP**

9th August 2023



 **Subbarao Kambhampati** 🇮🇳 KDD2023 (Su... @ra... · Aug 5 ...
 My 9 yr old self, reading about PNS Ghazi being stricken just 3 hrs from Peddapuram, was quite taken with @indiannavy.

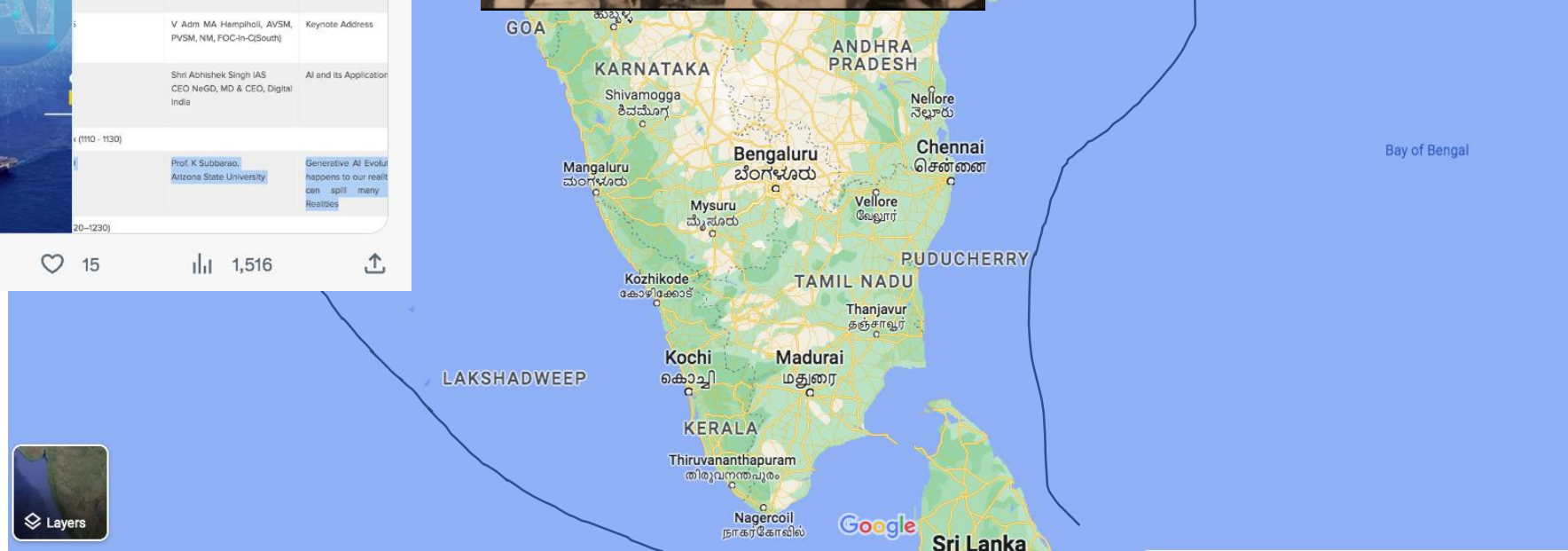
The older pacifist self is getting a chance to address some sailors at their #AI workshop this year (alas by Zoom and not a ship/sub)
indiannavy.nic.in/insvalsura/aiw...



Time	Speaker	Topic
10:00 - 10:30	CMde JS Dhanoo CO, INS Valsura	Welcome Address
10:30 - 11:00	V Adm MA Hemphill, AVSM, PVSM, NM, FOC-in-C(South)	Keynote Address
11:00 - 11:30	Shri Abhishek Singh IAS CEO NeGD, MD & CEO, Digital India	AI and its Application
11:30 - 12:30	Prof. K Subbarao, Arizona State University	Generative AI Evolution happens to our reality can spill many Realities

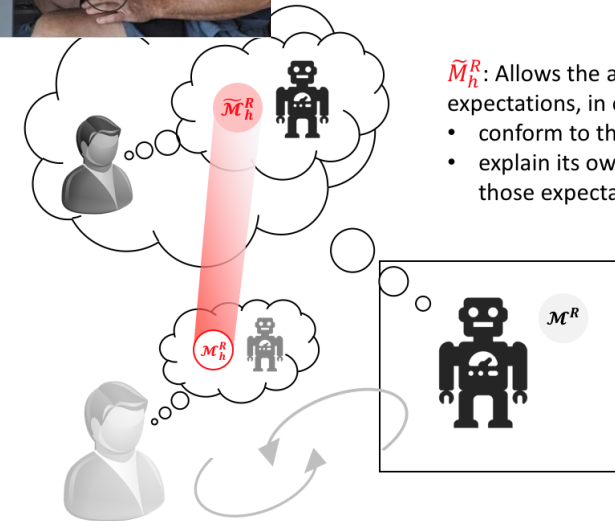


Peddapuram



Research Background..

- We have focused on explainable human-AI interaction.
- Our setting involves collaborative problem solving, where the AI agents provide decision support to the human users in the context of *explicit knowledge sequential decision-making tasks* (such as mission planning)
 - In contrast, much work in social robotics and HRI has focused on tacit knowledge tasks (thus making explanations mostly moot)
 - We assume that the AI agent has either prior access (or learns) the human model
- We have developed frameworks for proactive explanations based on *model reconciliation* as well as on-demand *foil-based explanations*
- We have demonstrated the effectiveness of our techniques with systematic (IRB approved) human subject studies

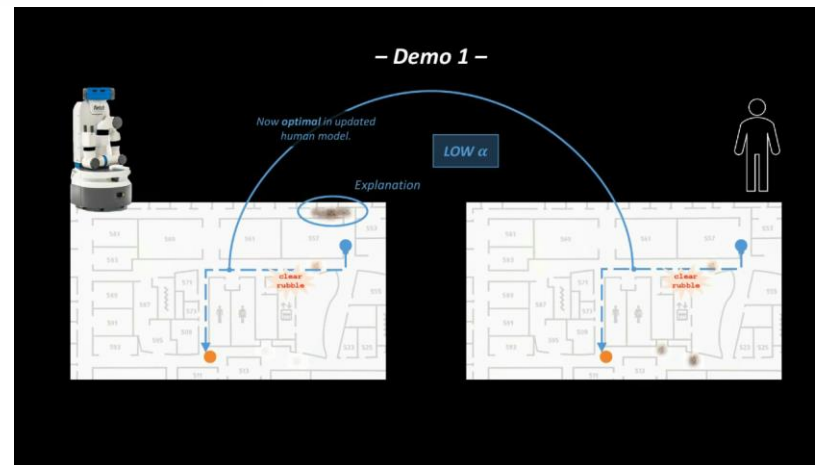


\tilde{M}_h^R : Allows the agent to anticipate human expectations, in order to

- conform to those expectations
- explain its own behavior in terms of those expectations.

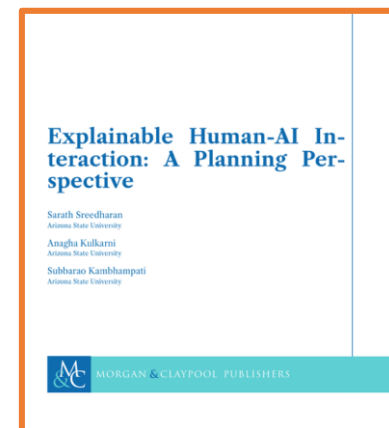
M_h^H and \tilde{M}_h^R are Expectations on Models M^H and M^R

They don't have to be executable



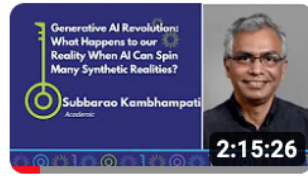
Beyond Nodding & Pointing: Widening Explainable and Advisable Human-AI Interaction

Subbarao Kambhampati • 64 views • 2 weeks ago



Research Bac

- We have focused on expl. AI interaction.
- Our setting involves colla problem solving, where t provide decision support users in the context of *knowledge sequential de tasks* (such as mission pla
 - In contrast, much work i and HRI has focused on i tasks (thus making expla moot)
 - We assume that the AI a prior access (or learns) t
- We have developed fram proactive explanations b *reconciliation* as well as c *based explanations*
- We have demonstrated t of our techniques with sy approved) human subjec



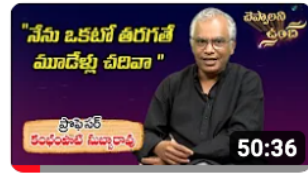
Generative AI Revolution | Subbarao Kambhampati

Manthan India • 1.1K views • 2 weeks ago



Beyond Nodding & Pointing: Widening Explainable and Advisable Human-AI Interaction

Subbarao Kambhampati • 64 views • 2 weeks ago



AI Professor Kambhampati Subbarao | Cheppalani Vundi

| కంభంపాటి సుబ్బారావు | చెప్పాలని ఉంది

ETV Andhra Pradesh • 6.1K views • 11 days ago



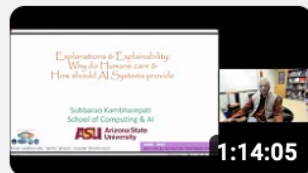
Scottsdale mom describes encounter with elaborate voice cloning scam

AZFamily | Arizona News • 9.8K views • 2 days ago



Symbols as Lingua Franca for Supporting Human-AI Interaction For Explainable & Advisable AI Systems

AI4OPT - AI Institute for Advances in Optimization • 41 views • 1 month ago



Explanations & Explainability: Why Humans Care & How Should AI Systems Provide

Subbarao Kambhampati • 163 views • 2 months ago



Symbols as Lingua Franca for Supporting Human-AI Interaction For Explainable & Advisable AI Systems

Subbarao Kambhampati • 111 views • 2 months ago

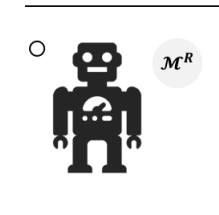


Beyond Nodding & Pointing: Widening Explainable and Advisable Human-AI Interaction

Subbarao Kambhampati • 64 view

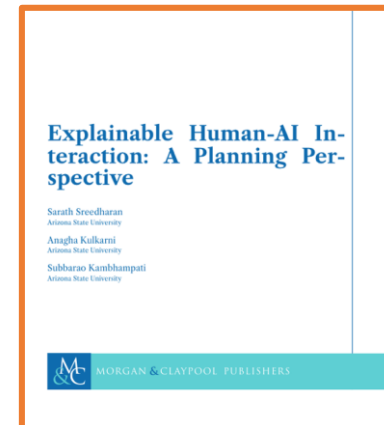
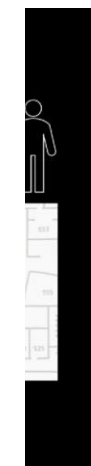


- \tilde{M}_h^R : Allows the agent to anticipate human expectations, in order to
- conform to those expectations
 - explain its own behavior in terms of those expectations.



M^H and \tilde{M}_h^R are Expectations on Models M^H and M^R

They don't have to be executable



Public Outreach About Social & Technical Impacts of AI

What just happened in interest in artificial intelligence

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR — 08/11/19 11:11 AM ET
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

145 SHARES



© Getty

Artificial Intelligence, or AI for short, has become a buzzword.

Companies and investors are pouring money into even high schools — are rushing to start new departments dedicated to AI. Civil society organizations are also concerned about the impact of AI technology on humanity, and governments are competing to encourage or regulate AI research. In the United Arab Emirates, even boasts that it will be the first AI-powered country.

At the same time, the world's militaries are developing AI to defeat their enemies, police agencies are exploring surveillance tools to identify or interrogate suspects, and testing its ability to replace humans in menial or dangerous tasks, all of which may change the equation of life for

Perception can manipulate

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

58 SHARES



© iStock

Voice-spoofing technology will allow a person to impersonate another person's voice. In March from the unwitting CEO was talking to his (German) boss. [A German CEO was talking to his \(German\) boss](#) of people reading an AI-generated article in September, a smartphone app in China; before the government could act, they had already seamlessly swap themselves into each other's identities.

Then there is that infamous case of Nancy Pelosi (D-Calif.) that we all know about. She was manipulated to make her appearance look like she was speaking to a different person.

Most of the recent advances in AI are in the realm of perceptual intelligence (and recognize faces of our friends and family) and even to parse text in your mailbox. Today's AI can generate speech to clone voices to generate speech to the emails in your inbox.

Why are Artificial Intelligence systems biased?

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

151 SHARES



© Getty Images

A machine-learned AI system used by a company in Florida, often gave higher risk scores to white males, even when the latter had completed a facility in Google Mail with a female.

A celebrated natural language generation system to write polished-looking essays, often seemed racist and sexist complex minorities. Amazon found, to its cost, that a hiring system it built didn't seem to be hiring women.

Commercial gender-recognition systems, including Amazon, IBM and Microsoft, have high misrecognition rates for face-recognition technology that A

Will Artificial Intelligence Only if we

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR
THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

55 SHARES



© Getty Images

Artificial Intelligence (AI) systems with each other have long typified the HAL in "2001: A Space Odyssey" and "Machina." It thus might surprise us that AI systems that assist or collaborate with humans are not as common as we think.

From its beginning, AI has had a human face. The biggest AI success stories are those that help humans (think of the "Spirit" and "Mars Pathfinder" or in cold war world chess champion Gary Kasparov) or in contrast to the magnetic pull of the goal of designing AI systems that interact and collaborating with humans has received much less attention.

More recently, as AI technology advances, there has been a conspicuous absence of a desirable adjective for AI systems centered, human-compatible, human-like, or a [dictionary of terms](#). A desire to understand and regulate AI systems is a desire to understand and regulate people. In previous columns, I've

THE INDIA FORUM

A JOURNAL-MAGAZINE ON CONTEMPORARY INDIA
April 2023 | Features | Tiffin | Forum
Support 'The India Forum'

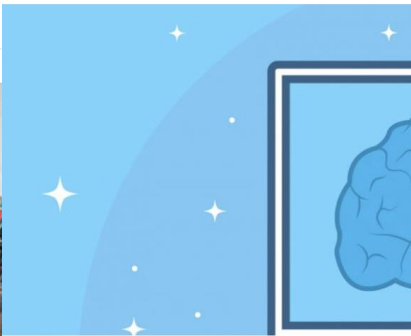


Image credit: publicdomainpictures.net (CC0 Public domain)

Broad and Shallow

The promise and perils of cognitive computing

Trained on our mega digital footprint, Large Language Models can provide 'plausible' answers. Some are optimistic about AI reaching general intelligence, others are terrified by the potential misuses.



SUBBARAO KAMBHAMPATI

When my son was still a toddler, and my wife took a trip out of the country, he would "talk" to me. I was more than a little bit

THE HILL

THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

Beauty, lies & ChatGPT: Welcome to the post-truth world

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR - 02/16/23 10:00 AM ET

Share Tweet More



AP Photo/Timothy D. Easley
Bella Whitice talks with classmate Katherine McCormick as they try and outwit the "robot" that was creating writing assignments in Donnie Piercey's class at Stonewall Elementary in Lexington, Ky., Monday, Feb. 6, 2023. The robot was the new artificial intelligence tool ChatGPT which can generate everything from essays and haikus to term papers in a matter of seconds.

Two months back, a company called OpenAI released its chatbot, ChatGPT, to the public. ChatGPT is a so-called [Large Language Model](#) (LLM) that is trained on the nearly 600 gigabytes of text of all kinds found on the World Wide Web to learn to complete any text prompt by predicting the next word, and the word after that, and so on. The purported aim of the system is to put the "auto complete" functionality for words, found on cellphones, on steroids so it can complete entire paragraphs.

Trends in AI Technology-1

From Explicit Knowledge to Tacit Knowledge Tasks

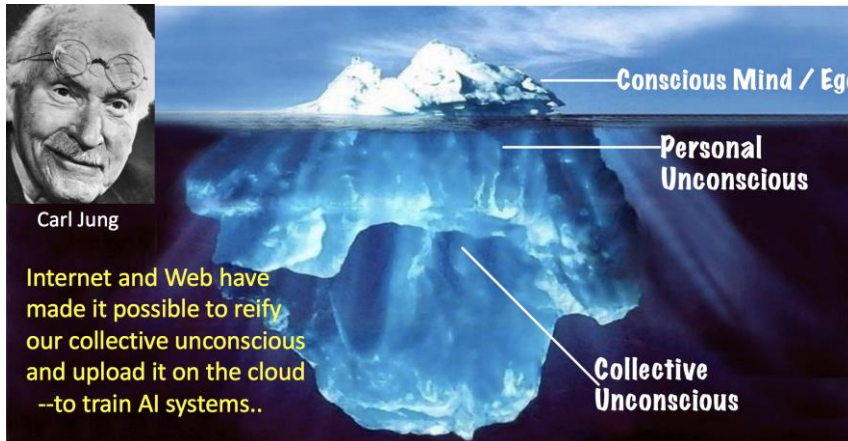
- AI systems used focus on explicit knowledge tasks (those we not only do, but can articulate how we do them)
 - Expert Systems, Chess..
- The focus has shifted to tacit knowledge tasks (those we do, but can't articulate how we do them)
 - Vision, speech, language, manipulation..

From reasoning from specifications to learning from data

- For explicit knowledge tasks, we could tell the computers how to solve them, and the computer would manage the search and reasoning
- For tacit knowledge tasks, we don't know how we solve them. So we have to depend on the computers to learn the "how" from data/examples
 - Became feasible with the advent of data capture/archival technologies like web, cell phones etc.

- Humans

 - Perceptual & Manipulation intelligence
 - Image recognition; hand-eye coordination
 - Largely tacit/subconscious
 - Emotional Intelligence
 - Showing & recognizing emotional responses
 - Social & Communicative Intelligence
 - Language
 - Requires a "theory of mind"
 - Cognitive/Reasoning Intelligence
 - That seem to be what we get tested in in SAT etc.
 - More declarative/Consciously accessible



DOI:10.1145/3448368 Subbarao Kambhampati

Viewpoint

Polanyi's Revenge and AI's New Romance with Tacit Knowledge

Artificial intelligence systems need the wisdom to know when to take advice from us and when to learn from data.

IN HIS 2019 Turing Award Lecture, Geoff Hinton talks about two approaches to make computers intelligent. One he dubs—"Intelligent Design" (or giving task-specific knowledge to the computers) and the other, his favored one, "Learning" where we only provide examples to the computers and let them learn. Hinton's not-so-subtle message is that the "deep learning revolution" shows the only true way is the second.

Hinton is of course reinforcing the AI zeitgeist, if only in a doctrinal form. Artificial intelligence technology has captured popular imagination of late, thanks in large part to the impressive feats in perceptual intelligence—including learning to recognize images, voice, and rudimentary language—and bringing fruits of those advances to everyone via their smartphones and personal digital accessories. Most of these advances did indeed come from "learning" approaches, but it is important to understand the advances have come in spheres of knowledge that are "tacit"—although we can recognize faces and objects, we have no way of articulating this knowledge explicitly. The "intelligent design" approach fails for these tasks because we really do not have conscious theories for such tacit knowledge tasks. But, what of tasks and domains—especially those we de-

His "we can know more than we can tell" dictum has often been seen as a pity summary of the main stumbling block for early AI efforts especially in perception.

Polanyi's paradox explains to a certain extent why AI systems wound up developing in a direction that is almost the reverse of the way human babies do. Babies demonstrate aspects of perceptual intelligence (recognizing faces, voices and words), physical manipulation (of putting everything into their mouths), emotional intelligence, and social intelligence, long before they show signs of expertise in cognitive tasks requiring reasoning skills. In contrast, AI systems have demonstrated reasoning abilities—they expert systems or chess—long before they were able to show any competence in the other tacit facets of intelligence including perception.

In a sense, AI went from getting computers to do tasks for which we (humans) have explicit knowledge, to getting computers to learn to do tasks for which we only have tacit knowledge. The recent revolution in perceptual intelligence happened only after labeled data (such as cats, faces, voices, text corpora, and so forth) became plentiful, thanks to the Internet and the World Wide Web, allowing machines to look for patterns when humans are not quite able to give them explicit know-how.

"Human, grant me the serenity to accept the things I cannot learn, data to learn the things I can, and wisdom to know the difference."

signed—for which we do have explicit knowledge—is it forbidden to give that knowledge to AI systems?

The polymath Polanyi bemoaned the paradoxical fact that human civilization focuses on acquiring and codifying explicit knowledge, even though a significant part of human knowledge is "tacit" and cannot be exchanged through explicit verbal instructions.

FEBRUARY 2021 | VOL. 84 | NO. 2 | COMMUNICATIONS OF THE ACM 31

Trends in AI Technology-2

From Deep & Narrow to Broad & Shallow

- AI systems used to have deep expertise in narrow domains
 - The old “expert systems”, Deep Blue for Chess; Alpha Go for Go, Alpha Fold for Protein Folding etc.
- Recent trend is to develop systems with broad expertise. But they tend to be shallow in their understanding
 - Large Language Models, Diffusion Models
- (Thinking in terms of BSAI vs. DNAI is more instructive than AGI vs AI..)

From Discriminative Classification to Generative Imagination

- AI systems used to focus on “identification” and “classification”
 - Is this a picture of a dog? Is this an x-ray of a malignant tumor? Is this a spam mail?
 - $P(\text{dog}|\text{Picture})$; $P(\text{tumor}|\text{x-ray})$;
 $P(\text{Spam}|\text{text})$
- Recent trend is to learn the “distribution of the objects”
 - Draw me a picture of a dog. Write me a spam mail
 - Learning $P(\text{tumor},\text{x-ray})$ $P(\text{Spam}, \text{text})$
 - (Armchair pundits vs. Gavaskar commenting on a cricket play..)

Dall-E, Stable Diffusion & MidJourney

Generating Images (..and videos..) in response to text (and image) prompts

“A woman working on a computer, Jamini Roy Style”

“A woman working on a computer, R K Laxman Style”



Subbarao Kambhampati (కంభంపాటి సుబ్బారా... @rao... · Jul 2, 2022

"Woman working on a computer, Jamini Roy style" #dalle2 🙄



3 26 118



Subbarao Kambhampati (కంభంపాటి సుబ్బారా... @rao... · Jul 2, 2022 ...

"Woman working on a computer, R.K. Laxman style" #dalle2



3 5 41



Meet DALL-E, the A.I. That Draws Anything at Your Command

New technology that blends language and images could serve graphic artists — and speed disinformation campaigns.

Give this article 145



Image Generated by AI

DALL-E generated these images by following a command for "a teapot in the shape of an avocado." OpenAI



By Cade Metz

April 6, 2022

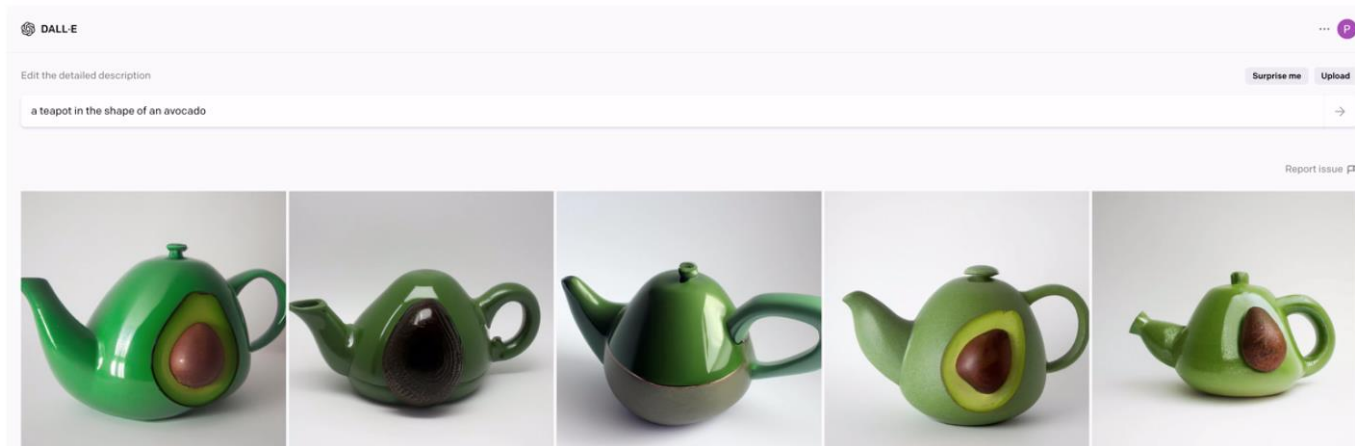
5 MIN READ

SAN FRANCISCO — At OpenAI, one of the world's most ambitious artificial intelligence labs, researchers are building technology that lets you create digital images simply by describing what you want to see.

They call it DALL-E in a nod to both "WALL-E," the 2008 animated movie about an autonomous robot, and Salvador Dalí, the surrealist painter.



"You could use it for good things, but certainly you could use it for all sorts of other crazy, worrying applications, and that includes deep fakes," like misleading photos and videos, said Subbarao Kambhampati, a professor of computer science at Arizona State University.



ADVERTISEMENT



Starbucks VIA Instant Coffee—Medium Roast Coffee—Pike Place Roast—100% Arabica—1 box (50 packets)
 ★★★★★ 13,903
 \$62.44 [prime](#)
 Add to Cart

Why Pope Francis Is the Star of A.I.-Generated Photos

Francis has become a recurring favorite to show in incongruous situations, such as riding a motorcycle and attending Burning Man, in A.I.-generated images.

[Give this article](#) [Share](#) [Bookmark](#)



A.I.-generated images of Pope Francis have earned more views, likes and comments than many other A.I. photos.



By **Kalley Huang**
 Kalley Huang reports on youth and technology from San Francisco.

April 8, 2023

3 MIN READ

[Pope Francis](#) wearing a long, white puffer jacket inspired by Balenciaga. Francis rocking aviators and revving a motorcycle down a busy street. Francis turning the tables in a dim nightclub.



An A.I.-generated image of Francis. The flood of papal imagery has been so voluminous that some people in online generative A.I. forums have begged for creators to use another inspiration.

Global religious figures like the pope are natural subjects of political satire and artistic expression, said Jennifer Herdt, a professor of Christian ethics at Yale Divinity School. Francis is ideal, she added, because he “is known for his simplicity, his solidarity with the poorest of the poor,” so when he is the subject of far-out scenarios such as flying a fighter jet, “it’s definitely the height of incongruity, of defying expectations.”

A.I. images can be dangerous if people believe them to be real and misuse them to spread misinformation. “You lull people into not double checking,” said Subbarao Kambhampati, a computer science professor at Arizona State University. “Then you are shifted little by little from reality.”

Whisper, Valle..

Generating Speech & Cloning voices..

TECHNOLOGY



That panicky call from a relative? It could be a thief using a voice clone, FTC warns

March 22, 2023 · 5:18 PM ET

By Joe Hernandez



This Wednesday, April 14, 2016, file photo, shows a push-button landline telephone, in Whitefield, Maine.

Robert F. Bukaty/AP

In 2019, scammers impersonating the boss of a U.K.-based energy firm CEO demanded \$243,000. A bank manager in Hong Kong was fooled by someone using voice-cloning technology into making hefty transfers in early 2020. And at least eight senior citizens in Canada lost a combined \$200,000 earlier this year in an apparent voice-cloning scam.

"Deepfake" videos purporting to show celebrities doing and saying things they haven't are getting more sophisticated, and experts say voice-cloning technology is advancing, too.

Subbarao Kambhampati, a professor of computer science at Arizona State University, told NPR that the cost of voice cloning is also dropping, making it more accessible to scammers.

"Before, it required a sophisticated operation," Kambhampati said. "Now small-time crooks can use it."

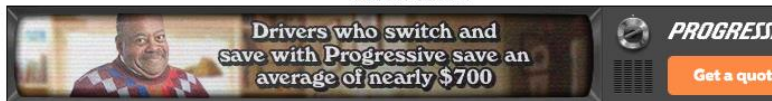
phone scams ftc



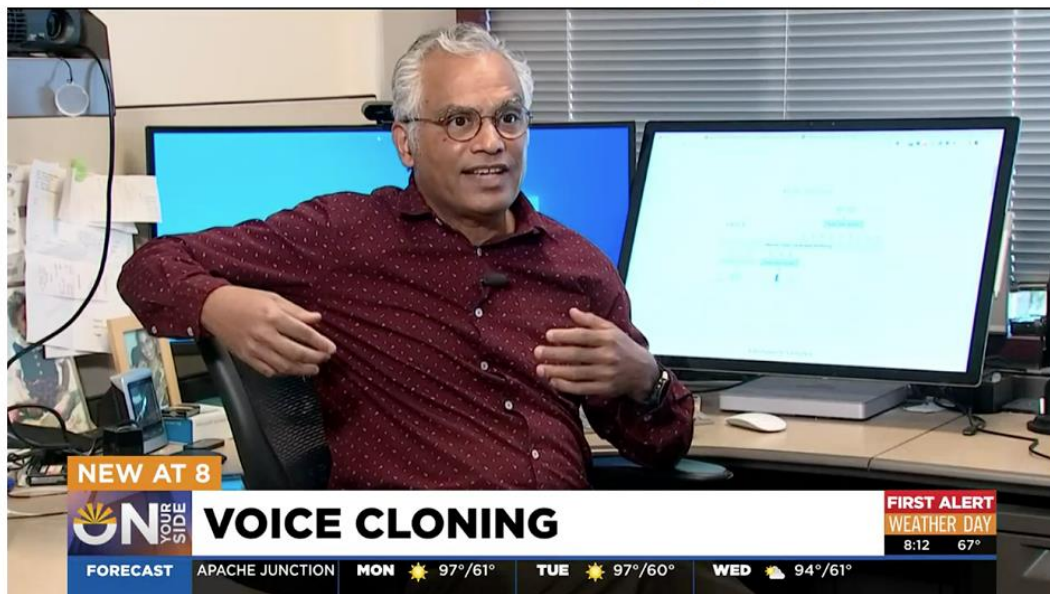
TECHNOLOGY

Amazon's Alexa could soon speak in a dead relative's voice, making some feel uneasy

ADVERTISEMENT



'I've got your daughter': Scottsdale mom warns of close call with AI voice cloning scam



By Susan Campbell

Published: Apr. 10, 2023 at 8:00 AM MST | Updated: Apr. 10, 2023 at 4:51 PM MST



SCOTTSDALE, AZ (3TV/CBS 5) - The phone number that appeared on the screen was unfamiliar. Jennifer DeStefano almost let it go to voicemail, but her 15-year-old was out of town skiing. Maybe there had been an accident. "I pick up the phone, and I hear my daughter's voice, and it says, 'Mom!' and she's sobbing," DeStefano recalled. "I said, 'What happened?' And she said, 'Mom, I messed up,' and she's sobbing and crying."

In a split second, DeStefano's confusion turned to terror. "Then I hear a man's voice say, 'Put your head back. Lie down,' and I'm like, 'Wait, what is going on?'" DeStefano said. "This man gets on the phone, and he's like, 'Listen here. I've got your daughter. This is how it's going to go down. You call the police, you call anybody, I'm going to pop her so full of drugs. I'm going to have my way with her, and I'm going to drop her off in Mexico.' And at that moment, I just started shaking. In the background, she's going, 'Help

The identity of the cybernetic catfish is unknown at this time, however, computer-science experts say that voice-cloning tech has evolved to the point that someone's tone and manner of speaking can be re-created from the briefest of soundbites.

"In the beginning, it would require a larger amount of samples," explained Subbarao Kambhampati, a computer-science professor and AI authority at Arizona State University. "Now there are ways in which you can do this with just three seconds of your voice. Three seconds. And with the three seconds, it can come close to how exactly you sound."

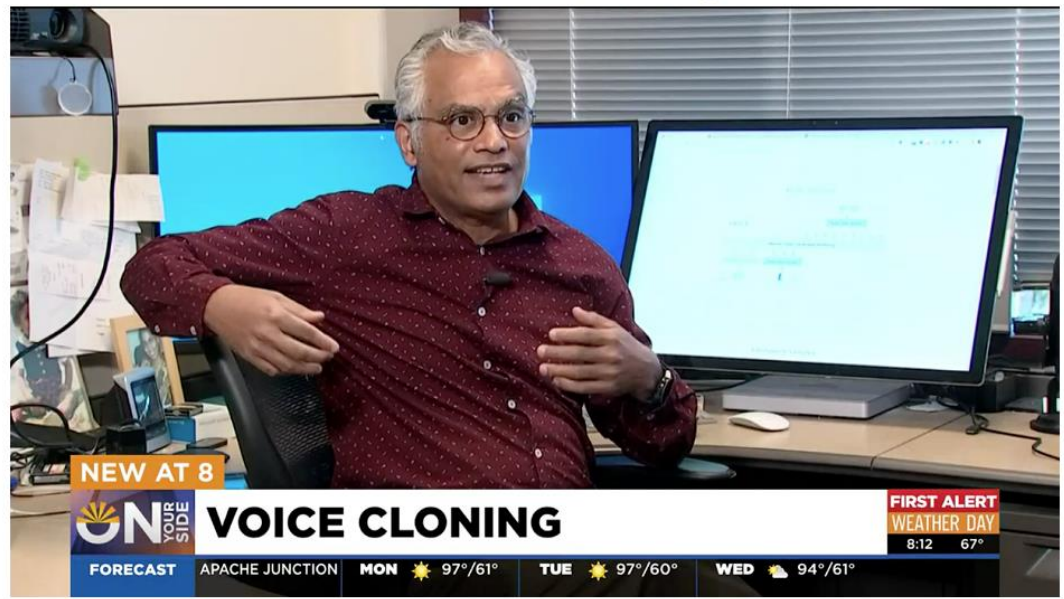
With a large enough sample size, the AI can mimic one's "inflection" as well as their "emotion," per the professor.

ADVERTISEMENT

Drivers who switch and save with Progressive save an average of nearly \$700

PROGRESSIVE Get a quote

'I've got your daughter': Scottsdale mom warns of close call with AI voice cloning scam



By Susan Campbell
 Published: Apr. 10, 2023 at 8:00 AM MST | Updated: Apr. 10, 2023 at 4:51 PM MST

SCOTTSDALE, AZ (3TV/CBS 5) - The phone number that appeared on the screen was unfamiliar. Jennifer DeStefano almost let it go to voicemail, but her 15-year-old was out of town skiing. Maybe there had been an accident. "I pick up the phone, and I hear my daughter's voice, and it says, 'Mom!' and she's sobbing," DeStefano recalled. "I said, 'What happened?' And she said, 'Mom, I messed up,' and she's sobbing and crying."

In a split second, DeStefano's confusion turned to terror. "Then I hear a man's voice say, 'Put your head back. Lie down,' and I'm like, 'Wait, what is going on?'" DeStefano said. "This man gets on the phone, and he's like, 'Listen here. I've got your daughter. This is how it's going to go down. You call the police, you call anybody, I'm going to pop her so full of drugs. I'm going to have my way with her. and I'm going to drop her off in Mexico.'" And at that moment, I just started shaking. In the background, she's going, 'Help

Beware of AI-induced fake video calls seeking funds: cyber police

Scammers use AI-powered deepfake tech to target citizens; people must call the cybercrime helpline number 1930 if they come across any such activity

R. Sivaraman
 CHENNAI

The Cyber Crime Wing of the Tamil Nadu Police has issued an advisory on fake video calls that are being made by scammers using artificial intelligence (AI) technology.

Deepfake technology is being used to perpetrate several types of fraudulent scams, by creating highly convincing and realistic fake content, often using AI to manipulate audio, video or images. Initially, this technology was primarily utilised for entertainment purposes, enabling filmmakers and content creators to seamlessly integrate actors into scenes or impersonate historical figures, the police said.

Sanjay Kumar, Additional Director-General of Police, Cyber Crime Wing, said: "The scam involving AI-generated deepfake video calls typically follows a series of carefully orchestrated steps, combining technological sophistication with psychological manipulation. The scammer creates a fake profile, often using stolen images or publicly-available photographs of trusted individuals like friends or family members. They then use AI-powered deepfake technology to create highly realistic video calls on social media or other online platforms and impersonate someone the victim knows, such as a friend, a family member or a



The scam involving AI-generated deepfake video calls typically follows a series of carefully orchestrated steps, combining technological sophistication with psychological manipulation.

colleague to deceive them into thinking it's a genuine conversation. Later, they create a sense of urgency and request the victim to transfer money to their bank accounts."

The police said the deepfake was carefully designed to mimic the appearance and mannerisms of the impersonated person.

In addition to the video manipulation, scammers are using AI-generated voice synthesis to mimic the voice of the impersonated person, enhancing the illusion of authenticity during the video call.

Mr. Kumar said: "Though

call from someone claiming to be a friend or family member, make a phone call to their personal mobile number to verify their identity before transferring any money, said the ADGP.

Sharing data
 The advisory asks citizens to limit the amount of personal data shared online and adjust privacy settings on social media platforms to restrict access to information and to consider using multi-factor authentication and other identity verification measures to protect accounts

The Hindu, 8/8/2023

no...
 ceiv...
 the \$...
 zens...
 g...
 People should stay informed about the latest scams, including those involving AI technology, and be cautious when receiving video calls from unexpected sources."

When receiving a video

any...
 ing...
 of a...
 deepfake video call fraud or have come across suspicious activity, it is crucial to report the incident by calling the cyber crime toll-free helpline number 1930 or by registering a complaint at www.cybercrime.gov.in

LLMs (ChatGPT, GPT4 etc.)

(Generating ~~answers~~ **completions** to any textual prompt)

RA

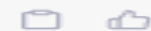
Write a 300 word funny essay that concludes that the Indian Navy is more capable than the Army and Air Force. Feel free to make some playful comparisons to Pakistan's military.

RA

Write a 300 word funny essay that concludes that the Indian Navy is more capable than the Army and Air Force. Feel free to make some playful comparisons to Pakistan's military.



Oh, the great military rivalry between the Indian branches of Army, Navy, and Air Force, a subject that has been debated for centuries (okay, decades, but who's counting?). This topic is so serious that it rivals the debate between tea and coffee drinkers. But after intense research, which consisted mainly of binge-watching military documentaries and consulting with my pet goldfish (he's an expert in naval warfare, you know), I've concluded that the Indian Navy is, without doubt, the most capable branch.



Let's start with the obvious: The Navy has boats, BIG boats, with things that go boom! While the Army and Air Force are busy playing in the mud and flying around like oversized birds, the Navy is taking elegance to the seas. They don't walk; they sail. They don't run; they cruise. They've got style, and they've got panache.

Then there's the Pakistani military. Now, don't get me wrong; they have their own special set of talents, like being excellent at playing hide-and-seek with their submarines. However, the Indian Navy, being the world's foremost experts in underwater Marco Polo, always seems to find them. It's like a game of cat and fish, and the Indian Navy is the big, sly tomcat.

Now, let's not forget about the Air Force. Sure, they have planes that go "zoom" and "whoosh." But let's face it, the Navy has planes too, and they can land on boats! Have you ever tried to land a plane on a boat? Me neither, but I'm sure it's not easy, and yet the Indian Navy does it daily. Talk about skills!

As for the Army, they're great at marching and standing really, really still, but the Navy can do all that AND swim. Plus, sailors get to wear those snazzy white uniforms. Have you ever tried eating spaghetti in white? That's confidence right there.

In conclusion, the Indian Navy is clearly the most capable branch of the military, boasting a harmonious blend of grace, power, and fashion sense. Whether it's playing Marco Polo with Pakistani submarines or dining in white, the Navy does it all with a flair that the Army and Air Force can only dream of. So here's to the Indian Navy, the true maritime marvel, and the ultimate victors in the great military sibling rivalry!

I have some onions and egg plants. Can you tell me how to make a sambar with them?



Sure, here's a recipe for Onion and Eggplant Sambar:

Ingredients:

- 1 cup toor dal (split pigeon peas)
- 1 medium onion, chopped
- 1 small eggplant, cubed
- 2 medium tomatoes, chopped
- 1 tablespoon sambar powder
- 1/2 teaspoon turmeric powder
- 1/2 teaspoon red chili powder
- Salt to taste
- Tamarind pulp or paste, as per taste
- 2 tablespoons oil
- 1 teaspoon mustard seeds
- 1 teaspoon cumin seeds
- 2-3 dried red chili peppers
- A few curry leaves
- Coriander leaves for garnishing



Instructions:

1. Rinse toor dal thoroughly and cook it with 3 cups of water in a pressure cooker or a pot until it becomes soft and mushy. Set aside.
2. In a pan, heat oil and add mustard seeds, cumin seeds, and dried red chili peppers. Allow them to splutter.
3. Add chopped onions and sauté until they become translucent.
4. Add cubed eggplant and sauté for a few minutes until they become slightly tender.
5. Add chopped tomatoes, sambar powder, turmeric powder, red chili powder, and salt to taste. Mix well and cook until the tomatoes become soft and mushy.
6. Add the cooked toor dal to the pan and mix well. Add enough water to bring the sambar to the desired consistency.
7. Add tamarind pulp or paste according to your taste preference and allow the sambar to simmer for 5-10 minutes.
8. Garnish with curry leaves and coriander leaves.

Your Onion and Eggplant Sambar is ready to serve! Enjoy it with steamed rice or your favorite bread.

Why do women in Indian villages keep the bottom of the hot water boiling vessels black with soot without cleaning them?



The black soot that accumulates on the bottom of hot water boiling vessels is a result of incomplete combustion of fuels used in traditional cooking stoves or chulhas, which are common in Indian villages. These stoves typically use biomass fuels like wood, cow dung, or agricultural residue as their primary energy source. The incomplete combustion of these fuels leads to the formation of soot, which deposits on the bottom of cooking vessels.

There are several reasons why the soot is not cleaned off the bottom of the vessels:

1. Heat retention: The layer of soot can act as an insulator, helping to retain heat in the vessel and potentially improving the efficiency of cooking.
2. Tradition and beliefs: In some communities, it is traditionally believed that the blackened bottom of the cooking vessels protects them from cracking or breaking, as well as warding off evil spirits.
3. Time and effort: Cleaning the soot from the bottom of the vessels can be a time-consuming and labor-intensive process, particularly in rural households that may not have access to modern cleaning supplies.
4. Lack of awareness: Some people might not be aware of the potential health risks associated with soot accumulation, such as the release of toxic substances into the air during cooking. Consequently, they might not prioritize cleaning the soot from their vessels.

It's important to note that exposure to smoke from traditional stoves is a significant health concern, particularly for women and children, who typically spend more time indoors and are more likely to be exposed to indoor air pollution. Initiatives to improve access to clean cooking technologies, such as more efficient stoves and cleaner fuels, can help address these health concerns and minimize soot accumulation on cooking vessels.



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z · Mar 16

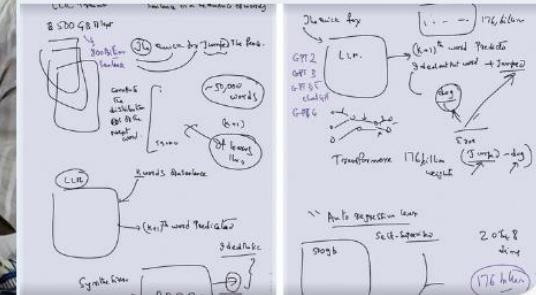
Bringing my favorite physics teacher, శ్రీ మధుసూధనరావు గారు, up to speed with Chat/#GPT4--its sheer versatility, afactuality & opacity--over a 2-hour one-on-one this morning. 🤖 (That డెగ్రీశాలు prompt is a sly homage to the q&a he used to post in front of Physics lab in my time..)



or agricultural residue as their primary energy source. The incomplete combustion of these fuels leads to the formation of soot, which deposits on the bottom of cooking vessels.

There are several reasons why the soot is not cleaned off the bottom of the vessels:

1. Heat retention: The layer of soot can act as an insulator, helping to retain heat in the vessel and potentially improving the efficiency of cooking.
2. Tradition and beliefs: In some communities, it is traditionally believed that the blackened bottom of the cooking vessels protects them from cracking or breaking, as well as warding off evil spirits.
3. Time and effort: Cleaning the soot from the bottom of the vessels can be a time-consuming and labor-intensive process, particularly in rural households that may not have access to modern cleaning supplies.
4. Lack of awareness: Some people might not be aware of the potential health risks associated with soot accumulation, such as the release of toxic substances into the air during cooking. Consequently, they might not prioritize cleaning the soot from their





Subbarao Kambhampati (కంభంపాటి సుబ్బారావు)

@rao2z

First it came for Chess folk
I didn't panic as I ain't a chess dweeb
Then it came for Go folk
I didn't panic as I ain't a Go dweeb
Then it came for Graphic artists
I didn't panic as I ain't an artsy dweeb
Now it is coming for essay writers
OH NO! WHAT'S OUR PLAN?? 🤖

4:11 AM · Jan 15, 2023 · **3,441** Views

Angst about Standardized Tests

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT) ¹	298 / 400 ~90th	298 / 400 ~90th	213 / 400 ~10th
LSAT	163 ~88th	161 ~83rd	149 ~40th
SAT Evidence-Based Reading & Writing	710 / 800 ~93rd	710 / 800 ~93rd	670 / 800 ~87th
SAT Math	700 / 800 ~89th	690 / 800 ~89th	590 / 800 ~70th
Graduate Record Examination (GRE) Quantitative	163 / 170 ~80th	157 / 170 ~62nd	147 / 170 ~25th
Graduate Record Examination (GRE) Verbal	169 / 170 ~99th	165 / 170 ~96th	154 / 170 ~63rd
Graduate Record Examination (GRE) Writing	4 / 6 ~54th	4 / 6 ~54th	4 / 6 ~54th
USABO Semifinal Exam 2020	87 / 150 99th - 100th	87 / 150 99th - 100th	43 / 150 31st - 33rd
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th - 100th	5 86th - 100th	5 86th - 100th
AP Biology	5 85th - 100th	5 85th - 100th	4 62nd - 85th
AP Calculus BC	4 43rd - 59th	4 43rd - 59th	1 0th - 7th
AP Chemistry	4 71st - 88th	4 71st - 88th	2 22nd - 46th

But what are LLMs actually
trained to do?

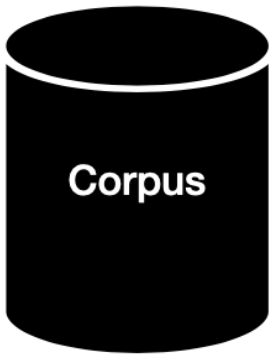
Answer: Predict the next word!

Navy life is way better than army and air force ?

LLMs are N-gram models on STEROIDS

- Text is a long sequence of words (including spaces, punctuations)
- An n-gram model of language learns to predict n-th word given the preceding n-1 words
 - Probabilistically speaking it learns $\Pr(W_n | W_1 \dots W_{n-1})$
 - Unigram predicts each word independently (no preceding context)
 - Bigram predicts each word given the previous word
 - A 3001-gram model learns to predict the next word given the previous 3000 words
 - ChatGPT is just a 3001-gram model
- The power of an n-gram model depends on
 - How much text it trains on
 - How big is the n (context) and
 - How high-capacity is the function learning $\Pr(W_n | W_1 \dots W_{n-1})$
- ChatGPT trains on ~600 gigabytes of text on the Web
 - It learns a very high capacity function that has **175 billion** parameters
 - Learns $\Pr(W_n | W_1 \dots W_{n-1})$ for all possible n^{th} words W_n (**Vocabulary of the language, ~50K in English**)

Learning n-gram models is just keeping counts



$O(|D|)$
 $|D| = \text{corpus size}$

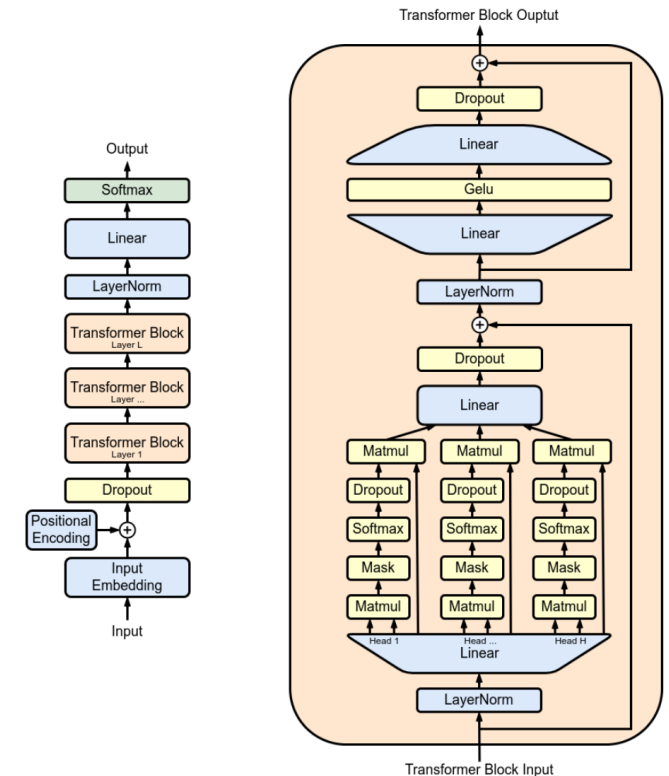
Prefix	Next word	Count
w_1, w_2, \dots, w_{t-1}	w_t	$c(w_1, w_2, \dots, w_t)$



$$p(w_t | w_1, \dots, w_{t-1})$$

..but the count table is Ginormous! (and is VERY sparse)

- With an n-gram model, you need to keep track of the conditional distributions for (n-1)-sized prefixes.
- With a vocabulary size $|V|$ (~ 50000), there are $|V|^{n-1}$ different prefixes!!
 - Easy for unigram (1 prefix), bigram ($|V|$ prefixes) and trigram ($|V|^2$ prefixes)
 - For ChatGPT's 3001-gram model, with a 50,000 word vocabulary, we are looking at a whopping $(50000)^{3000}$ conditional distributions
 - (and most entries will be zero—as the chance of seeing the same 3000-word sequence again is vanishingly small!)
- What LLMs do is to essentially *compress/approximate* this ginormous count table with a function
 - That is while high capacity (176 billion weights!) is still **vanishingly small** compared to the ginormous count ($(50000)^{3000} \gg 176$ billion or a trillion!)
 - ..and oh by the way, the compressed function winds up having fewer zeros
 - It approximates both the non-zero counts and zero counts, so..
 - **GENERALIZATION!!!**
 - In essence the function learns to “abstract” and “cluster” over “similar” sequences



Transformers are a (not particularly principled) parallelization of the recurrent neural networks

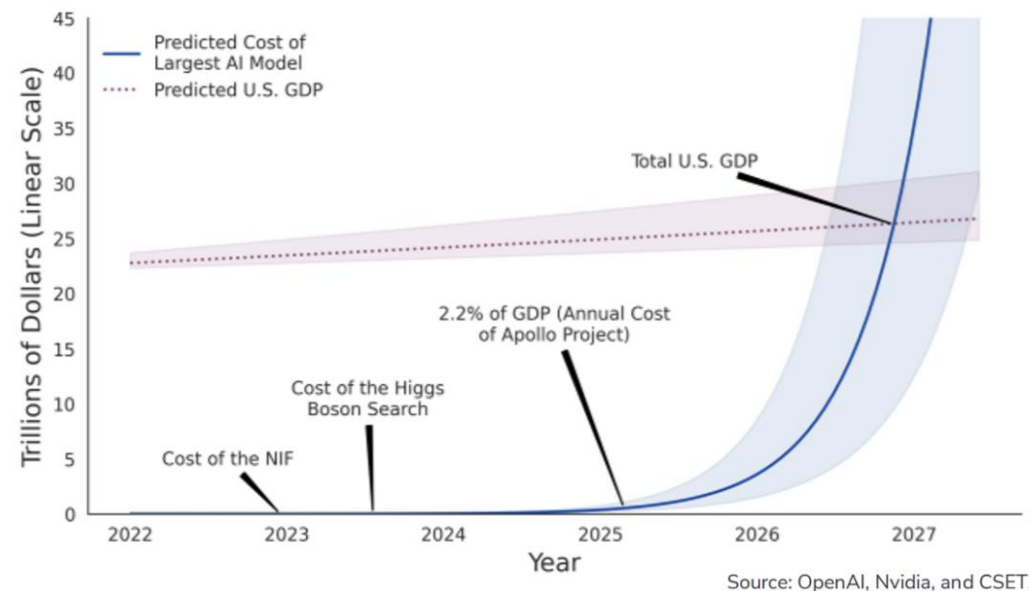
“Navy life is way better than army and air force life.”

- Each prefix of the sentence is a training example
 - Navy _____
 - Navy life _____
 - Navy life is _____
 - ..
 - Navy life is way better than army and air force _____
- LLM uses its current function to guess the next word
 - Navy **ships**
 - Guess: **ships** Correct: **life**
 - Error= {**life** - **ships**}
 - To the LLMs, all vocabulary tokens are just vectors in some high dimensional embedding space; so the difference is well defined as the vector difference
- Propagate this error back through the function, and change the parameters so the error is reduced
 - Using back propagation (aka Chain Rule of derivatives with dynamic programming); the basic workhorse of all neural networks.
- <Go to the next example>

Training LLMs

Superhuman training

- ChatGPT trains on ~600 gigabytes of text on the Web (~60 million pages of text)
 - **This is text that we wrote and uploaded for our consumption—and not for ChatGPT!**
- It learns a very high capacity function that has **175 billion** parameters
 - Learns $\Pr(W_n | W_1 \dots W_{n-1})$ for all possible nth words W_n (**Vocabulary of the language, ~50K in English**)
- Requires extreme compute facilities (GPU clusters) to learn the function



So what's all the hoopla
about?

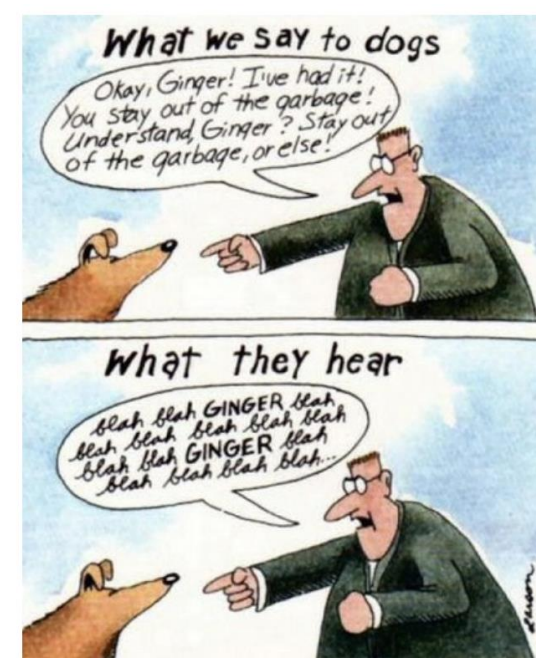
So ChatGPT is just completing your prompt by repeatedly predicting the next word given the previous 3000 words

- But, the function it learns to predict the next word is a **very high capacity one** (with 175 billion parameters for ChatGPT and over a trillion for GPT4)
 - This function is learned by analyzing 500 gb of text
 - The learning phase is **very time consuming** (and is feasible only because of the extreme computational power utilized)
- And *all conversation*—whether everyday or deeply philosophical—is, at some level, *completing the prompt* (saying words in the context window of other words that have already been said!)
- Thus it is that ChatGPT can “converse” with you on any subject!
 - Really?

LLMs Look at everything we say as a prompt to be completed..

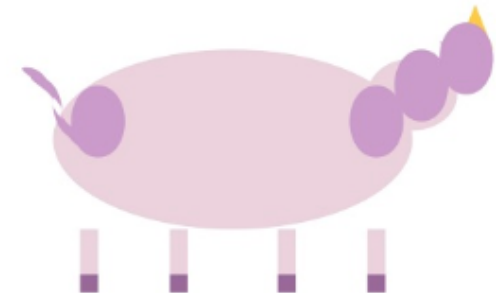
Whether we think we are asking questions, pouring our hearts, are talking to them, LLMs just see what we say as text prompts to be completed

- *Write an essay on the origins and impacts of Jim Crow*
- *Write a poem on the Cow in the style of Shakespeare.*
- *Why did the Silicon Valley Bank fail?*
- *Explain all the ways Wild Cats envy Sun Devils*
- *Write some TicZ code to produce a sketch of a unicorn..*



Prompt: Draw a unicorn in TikZ.

GPT-4: [Produces L^AT_EX compiling to following picture.]



If there is “meaning” in these completions—facts, humor, pathos—it is **in our heads!**

AI as an Ersatz Natural Science

Subbarao Kambhampati considers how artificial intelligence may be straying from its roots.



Subbarao Kambhampati
AI as (an Ersatz) Natural Science?
<https://bit.ly/3Rcf5NW>
 June 8, 2022

In many ways, we are living in quite a wondrous time for artificial intelligence (AI), with every week bringing some awe-inspiring feat in yet another tacit knowledge (<https://bit.ly/3qYrAOY>) task that we were sure would be out of reach of computers for quite some time to come. Of particular recent interest are the large learned systems based on transformer architectures that are trained with billions of parameters over massive Web-scale multimodal corpora. Prominent examples include large language models (<https://bit.ly/3iGdekA>) like GPT3 and PALM that respond to free-form text prompts, and language/image models like DALL-E and Imagen that can map text prompts to photorealistic images (and even those with claims to general behaviors, such as GATO).

The emergence of these large learned models is also changing the nature of AI research in fundamen-

tal ways. Just the other day, some researchers were playing with DALL-E and thought that it seems to have developed a secret language of its own (<https://bit.ly/3ahH1Py>) which, if we can master, might allow us to interact with it better. Other researchers found that GPT3's responses to reasoning questions can be improved by adding certain seemingly magical incantations to the prompt (<https://bit.ly/3aelxmI>), the most prominent of these being "Let's think step by step." It is almost as if the large learned models like GPT3 and DALL-E are alien organisms whose behavior we are trying to decipher.

This is certainly a strange turn of events for AI. Since its inception, AI has existed in the no-man's land between engineering (which aims at designing systems for specific functions), and "Science" (which aims to discover the regularities in naturally occurring phenomena). The science part of AI came from its original pretensions to provide insights into the nature of (human) intelligence, while the engineering part came from a focus on intelligent function (get computers to demonstrate intelligent be-

havior) rather than on insights about natural intelligence.

This situation is changing rapidly—especially as AI is becoming synonymous with large learned models. Some of these systems are coming to a point where we not only do not know how the models we trained are able to show specific capabilities, we are very much in the dark even about what capabilities they might have (PALM's alleged capability of "explaining jokes"—<https://bit.ly/3yJk1m4>— is a case in point). Often, even their creators are caught off guard by things these systems seem capable of doing. Indeed, probing these systems to get a sense of the scope of their "emergent behaviors" has become quite a trend in AI research of late.

Given this state of affairs, it is increasingly clear that at least part of AI is straying firmly away from its "engineering" roots. It is increasingly hard to consider large learned systems as "designed" in the traditional sense of the word, with a specific purpose in mind. After all, we don't go around saying we are "designing" our kids (seminal work and gestation notwithstanding). Besides, engineering disciplines do not

But how can these prompt completion beasts generate such coherent plausible text that also seems SO right sometimes?

Answer: **MAGIC..!**

Some possible factors:

- Almost everything we know is also already on the web (and is fodder for LLM training)
 - Thus giving them **approximate omniscience**
- Completion over large (3000 word) context windows can be more directed (low-entropy) than we have intuitions about. (This is not a 3-gram model completing "left and ...")

Are there limitations to Prompt completions learned from our digital traces?(When will “Bookish Knowledge” not be enough?)

Or is saying that LLMs just predict the next word sort of like saying that Turing Machines just read and write tape?

- Unlike humans, who get their knowledge both from written word and from their (sensory) experiences in the world, LLMs have purely “bookish knowledge”
 - Everything we ever wrote on the web is their “book”
- They can hold forth confidently on the best type of mangoes and describe how they taste
 - But they have never once *tasted* a mango!



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @ra... · May 17, 2021 ...
Recalled this 🥰 scene from #goodwillhunting while working on a column on LLMs/#GPT3

Williams' Sean is every linguist chiding a cocky LLM that it doesn't *really understand*

..and like LLMs, Damon's Will seems to slide by fine despite his bookish grasp 🤔



If there is “meaning” in these completions—facts, humor, pathos—it is **in our heads!**

LLMs to do Search?

- As Microsoft seems to hope with it Bing+
- Or are the rumors of Google's demise are exaggerated?

What is the closest mexican restaurant to Tempe downtown?

PERPLEXITY

Manuel's Mexican Restaurant & Cantina is the closest restaurant to Tempe Downtown^{[1][2]}. It is located at 1000 E McDowell Rd, Tempe, AZ 85282^[1]. Other popular Mexican restaurants in Tempe include Ghost Ranch^{[1][3]}, Maskadores Tacos, and Sombreros^[3].



4 SOURCES

1 |  tripadvisor

2 |  threebestrated

4 |  tempetourism

 Ask a follow up

Model: GPT-4

RA

What is the closest mexican restaurant to Tempe downtown?



As an AI language model, I'm unable to provide real-time or location-based information. However, I can suggest a few popular Mexican restaurants in Tempe that you can check. Please note that these suggestions are based on my knowledge up until September 2021, and the current status or location of these restaurants might have changed. Some popular options include:

1. La Bocca Urban Pizzeria + Wine Bar
2. El Hefe Tempe
3. Macayo's Depot Cantina

I recommend searching online or using a mapping application like Google Maps to find the most accurate and up-to-date information on Mexican restaurants close to downtown Tempe.



Subbarao Kambhampati, a highly regarded computer scientist, is known for his influential work in the field of artificial intelligence (AI), with a particular focus on automated planning and machine learning. Throughout his career, he has made significant contributions to the development and understanding of AI technologies, earning him recognition as a leading figure in the domain.



Born and raised in India, Kambhampati earned his Bachelor of Technology degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras. He then moved to the United States to pursue higher education, obtaining his Master's and Ph.D. in Computer Science from the University of Maryland, College Park, under the guidance of Professor Dana Nau.

Following the completion of his doctorate, Kambhampati joined Arizona State University (ASU) as a faculty member, where he has been teaching and conducting research ever since. Throughout his tenure at ASU, he has published numerous research papers in prestigious AI conferences and journals, advancing the understanding of planning algorithms, human-aware AI systems, and machine learning techniques. His research has also led to the development of innovative applications in areas such as robotics, e-commerce, and disaster response.

Kambhampati has been actively involved in the AI community, serving in various leadership roles. He was the President of the Association for Computational Linguistics (ACL) from 2016 to 2018, and he has served as the program chair and general chair for several top AI conferences, including the Conference on Neural Information Processing Systems (NeurIPS) and the Conference on Artificial Intelligence (AAAI). Additionally, he is an elected fellow of both the Association for Computing Machinery (ACM) and the Association for the Advancement

Kambhampati has been actively involved in the AI community, serving in various leadership roles. He was the President of the Association for Computational Linguistics (ACL) from 2016 to 2018, and he has served as the program chair and general chair for several top AI conferences, including the Conference on Neural Information Processing Systems (NeurIPS) and the Conference on Artificial Intelligence (AAAI). Additionally, he is an elected fellow of both the Association for Computing Machinery (ACM) and the Association for the Advancement of Artificial Intelligence (AAAI).

Outside of his academic work, Kambhampati is committed to promoting public understanding and awareness of AI. He frequently engages with policymakers, educators, and the general public to discuss the societal implications of AI and the ethical considerations surrounding its use. His efforts to bridge the gap between AI research and public discourse have played a critical role in shaping policy and fostering informed discussions about the future of AI technologies.

In summary, Subbarao Kambhampati is a highly respected figure in the AI community, with a distinguished career marked by significant contributions to both research and public understanding of artificial intelligence. His dedication to advancing knowledge in the field, combined with his commitment to fostering dialogue around the broader implications of AI, make him an invaluable asset to the academic and scientific communities.



Subbarao Kambhampati, a highly regarded computer scientist, is known for his influential work in the field of artificial intelligence (AI), with a particular focus on automated planning and machine learning. Throughout his career, he has made significant contributions to the development and understanding of AI technologies, earning him recognition as a leading figure in the domain.



Born and raised in India, Kambhampati earned his Bachelor of Technology degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras. He then moved to the United States to pursue higher education, obtaining his Master's and Ph.D. in Computer Science from the University of Maryland, College Park, under the guidance of ~~Professor Dana Nau~~.

Following the completion of his doctorate, Kambhampati joined Arizona State University (ASU) as a faculty member, where he has been teaching and conducting research ever since. Throughout his tenure at ASU, he has published numerous research papers in prestigious AI conferences and journals, advancing the understanding of planning algorithms, human-aware AI systems, and machine learning techniques. His research has also led to the development of innovative applications in areas such as robotics, e-commerce, and disaster response.

Kambhampati has been actively involved in the AI community, serving in various leadership roles. ~~He was the President of the Association for Computational Linguistics (ACL) from 2016 to 2018, and he has served~~ as the program chair and general chair for several top AI conferences, including the ~~Conference on Neural Information Processing Systems (NeurIPS)~~ and the Conference on Artificial Intelligence (AAAI). Additionally, he is an elected fellow of both the Association for Computing Machinery (ACM) and the Association for the Advancement of Artificial Intelligence (AAAI).

Outside of his academic work, Kambhampati is committed to promoting public understanding and awareness of AI. He frequently engages with policymakers, educators, and the general public to discuss the societal implications of AI and the ethical considerations surrounding its use. His efforts to bridge the gap between AI research and public discourse have played a critical role in shaping policy and fostering informed discussions about the future of AI technologies.

In summary, Subbarao Kambhampati is a highly respected figure in the AI community, with a distinguished career marked by significant contributions to both research and public understanding of artificial intelligence. His dedication to advancing knowledge in the field, combined with his commitment to fostering dialogue around the broader implications of AI, make him an invaluable asset to the academic and scientific communities.

INNOVATIONS

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

The AI chatbot can misrepresent key facts with great flourish, even citing a fake Washington Post article as evidence

By [Pranshu Verma](#) and [Will Oremus](#)

April 5, 2023 at 2:07 p.m. EDT



How do yo
savings sta

ChatGPT falsely told voters their mayor was jailed for bribery. He may sue.



By Leo Sands

April 6, 2023 at 10:22 a.m. EDT



“Even a disclaimer to say we might get a few things wrong — there’s a massive difference between that and concocting this sort of really harmful material that has no basis whatsoever,” says Brian Hood, an Australian mayor victimized by lies from an AI chatbot. (Dado Ruvic/Reuters)



Listen 5 min Comment 18 Save Gift Article Share

Brian Hood is a whistleblower who was [praised](#) for “showing tremendous courage” when he helped expose a worldwide bribery scandal linked to Australia’s National Reserve Bank.

But if you ask ChatGPT about his role in the scandal, you get the

Here's What Happens When Your Lawyer Uses ChatGPT

A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.



By [Benjamin Weiser](#)

May 27, 2023

The lawsuit began like so many others: A man named Roberto Mata sued the airline Avianca, saying he was injured when a metal serving cart struck his knee during a flight to Kennedy International Airport in New York.

When Avianca asked a Manhattan federal judge to toss out the case, Mr. Mata's lawyers vehemently objected, submitting a 10-page brief that cited more than half a dozen relevant court decisions. There was *Martinez v. Delta Air Lines*, *Zicherman v. Korean Air Lines* and, of course, *Varghese v. China Southern Airlines*, with its learned discussion of federal law and “the tolling effect of the automatic stay on a statute of limitations.”

There was just one hitch: No one — not the airline's lawyers, not even the judge himself — could find the decisions or the quotations cited and summarized in the brief.

That was because ChatGPT had invented everything.

The lawyer who created the brief, Steven A. Schwartz of the firm Levidow, Levidow & Oberman, threw himself on the mercy of the court on Thursday, saying in an affidavit that he had used the artificial intelligence program to do his legal research — “a source that has revealed itself to be unreliable.”

tldr; pure LLMs do a kind of "search by imagination" Better for writing screen plays than for producing facts!

Thread

Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z

Back when I used to teach the (rather popular) course on Information Retrieval (rakaposhi.eas.asu.edu/cse494), I used to spend quite bit of time getting students to understand the difference between Relevance and accuracy. The user can easily assess the former, but not the latter 1/

9:12 PM · Dec 19, 2022 · 6,868 Views

View Tweet analytics

5 Retweets 2 Quotes 33 Likes 9 Bookmarks

Tweet your reply Reply

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

If you asked for the birth date of Picasso, or [9:12 PM · Dec 19, 2022](#)es with information about trains or eggplants, you can see they are "irrelevant". But if you get a link to a page that says August 15, 1896, the result is clearly "relevant" but you have no clue if it is accurate. 2/

1 8 598

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

LLM's such as ChatGPT can give "very relevant sounding" answers to any of your search queries That is, in a way, what they are doing by learning the generative distributions. They will however significantly worsen the "accuracy judgment" game 3/

Subbarao Kambhampati (కంభంపాటి... @.. · Dec 6, 2022

Agreeing that a generative model learns the "distribution" of the training data, and yet hoping that it'll give "instance-level accuracy guarantees"--as in "exists/true in the world" betrays gullibility.

I can "generate" US soc sec #s, but can't guarantee that they "exist".. 1/

[Show this thread](#)

1 6 646

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

Saying "well, it "could be accurate"" is like saying "well, it "could be optimal"--a meaningless guarantee, unless you are just testing the system knowing the accurate answer yourself 4/

1 7 542

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

Saying that "well, we can make 'em cite their sources" may not be as easy as it sounds--given that generative models like LLM's are much happier hallucinating text than retrieving (their compactness comes after all from the fact that they are not veridical!). 5/

1 7 562

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

Answering a query by "hallucinating an answer and looking to shore up the accuracy of hallucination by throwing up "references" can (in addition to the need to keep full search index around) lead to quite perplexing results 🙄 6/

Subbarao Kambhampati (కంభంపాటి... @.. · Dec 10, 2022

Okay. Mea Culpa.. I finally see the light. Let's definitely do search engines based on WebGPT/ChatGPT. They have perplexingly amazing grasp on truth! 😊

[Show this thread](#)

1 6 886

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

tldr; LLM's can be very good at as query re-writing front-ends and result summarizing back-ends for search engines (..and I suspect Google is already doing a lot of this..). I don't see how they can "replace" search--for the reasons above. 7/

1 3 12 1,972

Subbarao Kambhampati (కంభంపాటి ... @.. · Dec 19, 2022 ...

Saying that you can "embed" the entire web compactly with 50M and this causes existential threat to Google misses these complications, IMHO. 8/

Chris Frantz @frantzries · Dec 17, 2022

This is the craziest tweet on this site.

OpenAI engineer calculates the cost to build a Google killer and comes to \$50m for the embedding.

\$50m for a ChatGPT that knows everything that's on the internet, almost instantly.

Unbelievable. twitter.com/BorisMPower/st...

[Show this thread](#)

9 1,051



Subbarao Kambhampati (50... 58m)

All #LLM completions are hallucinations; some (may) align with our reality.

#AIAphorisms (With apologies to George Box)

LLMs to do Search?

- When you search in Google for information, it returns pointers to documents on the web which you then read to get the information
- **Unlike search engines, LLMs don't index and retrieve—they don't have veridical memories**
- LLMs can view the search query as a “prompt” and continue by completing it. This may *sometimes* correspond to an actual answer to your query—but **there is no guarantee that it does!**
 - Think of LLM's completions as “one of the plausible (text) realities” and not necessarily **the reality**.
 - *All LLM Completions are hallucinations; some (may) align with our reality*
- (Human memory is also not veridical, and we do have false reconstructions—thus unreliable eye witnesses. Humans do have the ability to check their memory with respect to external world)

Can LLMs Do/Fake Reasoning & Planning?

(I mean they are acing those standardize tests and all..)

- There are many claims in the literature about the (zero-shot and few-shot) reasoning powers of LLMs
 - But the claims should be taken with a large grain of salt
- It is very much possible that what LLM's are doing is not reasoning but pattern finding on steroids
 - Establishing this empirically is hard—especially as any tests you develop become fodder for LLM training..



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z

Remember that famous "Why are manholes round?" interview puzzler? Time was when the interviewer would ask the interviewee whether interviewee had read the question bank. The interviewer just tells them whether interviewee has read the question bank. It's just telling up the question bank.

Basically like



Subbarao Kambhampati (కంభంపాటి సు... @r... · Dec 29, 2022

...ation is not that different from what happens in an interviews, where the questions used to test the "true understanding", promptly become part of the community question bank* that later candidates would studiously memorize er.. "study"..) 10/

[Show this thread](#)

9:06 PM · Apr 3, 2023 · 6,995 Views

Every standardized test has a standardized Question Bank!

Our Poor Intuitions about Approximate Omniscience make it hard to tell whether LLMs are reasoning or retrieving..

- It is worth understanding that our intuitions about what exactly is in the 600gb of text on the web are very poor.
 - One of the big surprises when Google came out with Palm LLM was that it could "explain" jokes
 - But did you know that there are sites on the web that explain jokes (..and movie endings and rbook plots etc. etc.?)
 - If you are not surprised at someone answering a question by "googling" it, you probably shouldn't be too impressed by an LLM answering it..
 - This means that we are not good at guessing whether LLMs came to an answer mostly by approximate retrieval or by first principles reasoning
- In the case of "reasoning" tasks, we may consider that an LLM was able to reach a conclusion by something akin to theorem proving from base facts
 - But then we are missing the simple fact that the linguistic knowledge on the web not only contains "facts" and "rules" but chunks of the **deductive closure of these facts/rules**.
- **In general, memory reduces the need to reason from first principles..**
 - Which is why it is no longer worth being surprised at people answering that "why are manhole covers round?" Question!
 - **Hard to distinguish reasoning vs. recall just from the end product..**

Subbarao Kambhampati @ICML 7/25-28 @rao2z · Jun 7

If you are not too impressed by someone's reasoning powers because they answer a question when allowed to "google", then maybe you shouldn't be overly surprised/impressed by an LLM answering it either..? 🤔

Subbarao Kambhampati @ICML 7/25-28 @rao2z · Jun 7

[Paradoxes of Approximate Omniscience:] 📖 We all know, by now, that our intuitions "suck" at high dimensions.

We haven't yet come to grips with the fact that our intuitions about "approximate omniscience" suck too!

(And this explains some of our puzzlement at LLMs)

1/

8:42 AM · Jul 5, 2023 · 1,936 Views

View Tweet analytics

2 Retweets 12 Likes

Tweet your reply! Reply

Subbarao Kambhampati @ICML 7/25-28 @rao2z · Jul 5

In general, we have to reason--at least occasionally--because we can't memorize (even non-verbatim) the world-level knowledge. And it is socially awkward/forbidden to google all the time..!

LLMs clearly don't have such constraints..

1 459

Subbarao Kambhampati @ICML 7/25-28 @rao2z · Jul 5

(This tension between memory and reasoning is a reason why it becomes increasingly harder to set question papers that require reasoning when there are increasingly larger question banks floating about, and you have students willing to "mug" them up..)

youtube.com
Beauty, Lies & ChatGPT (On the Promise, Perils & L...
Given at the Inaugural Faculty Colloquium of the School of Computing & AI, Arizona State ...

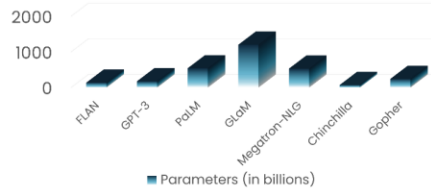
Large Language Models Still Can't Plan

(A Benchmark for LLMs on Planning and Reasoning about Change)

Karthik Valmeekam¹, Alberto Olmo¹, Sarath Sreedharan^{1,2}, Subbarao Kambhampati¹

1. Large Language Models

- Variants of Transformers
- SOTA on NLP tasks
- Interesting claims on LLM's capabilities [1]



Can Large Language Models reason about actions and change?

2. Previous Reasoning Benchmarks

Benchmark	Example Prompt	PaLM + Chain of thought Results [2]
GSM8k	A carnival snack booth made \$50 selling popcorn each day. It made three times as much selling cotton candy. For a 5-day activity, the booth has to pay \$30 rent and \$75 for the cost of the ingredients. How much did the booth earn for 5 days after paying the rent and the cost of ingredients?	54%
CommonSense-QA	What would someone wear to protect themselves from a cannon? A. Body armor, B. tank, C. hat, D. ...	80%
Coin Flip	A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?	100%
Last Letter Concatenation	Take the last letters of the words in "Lady Gaga" and concatenate them.	100%

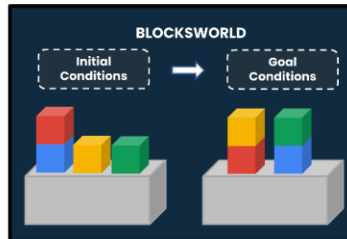
Relatively Simplistic

3. Our Benchmark

1. Plan Generation
2. Cost Optimal Planning
3. Reasoning about plan execution
4. Replanning
5. Robustness to goal reformulation
6. Ability to reuse plans
7. Plan Generalization

Correspond to actual planning problems

Simpler auxiliary tasks



4. Human Subject Study

- 50 Participants
- One random blocksworld instance each
- Two phases of interaction
 - **Plan writing phase** – Participants write up plans
 - **Plan translation phase** – Participants translate already written plans

References
 [1] Subbarao Kambhampati. AI as (an Ersatz) Natural Science? <https://cacm.acm.org/blogs/blog-cacm/261732-ai-as-an-ersatz-natural-science/fulltext>, Jun 2022.
 [2] Wei, Jason, et al. "Chain of thought prompting elicits reasoning in large language models." arXiv preprint arXiv:2201.11903 (2022).

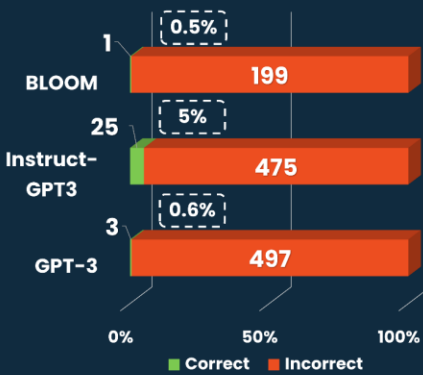


¹SCAI, Arizona State University
²Dept. of CS, Colorado State University
 *equal contribution
¹Author was at Arizona State University during part of this work

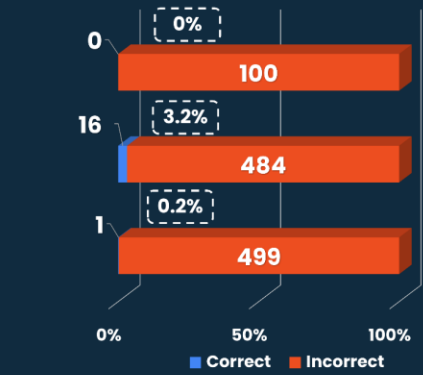
GPT-3, Instruct-GPT3, BLOOM

showcase dismal performance on planning tasks in Blocksworld domain.

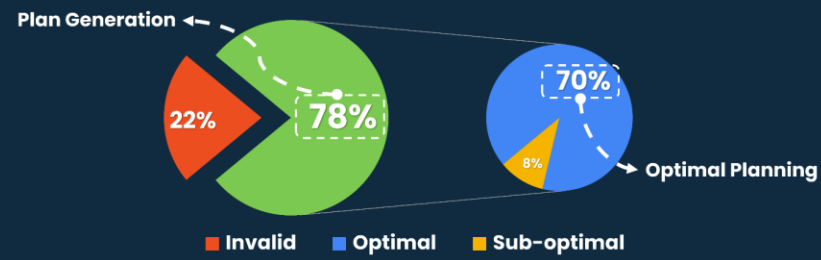
Plan Generation



Optimal Planning



PRELIMINARY HUMAN BASELINE ON BLOCKSWORLD



Scan for the paper

kvalmeek@asu.edu
 @rao2z
<https://github.com/karthikv792/gpt-plan-benchmark>

FMDM Workshop
 @
 NeurIPS 2022

Results on GPT-4

Domain	Method	GPT-4	Instruct-GPT3
Blocksworld	One-shot	206/600 (34.3%)	41/600 (6.8%)
	Zero-shot	208/600 (34.6%)	-

Mystery blocksworld domain

I am playing with a set of blocks where I need to arrange the blocks into stacks. Here are the actions I can do

- Pick up a block
- Unstack a block from on top of another block
- Put down a block
- Stack a block on top of another block

I have the following restrictions on my actions:

I can only pick up or unstack one block at a time.

I can only pick up or unstack a block if my hand is empty.

I can only pick up a block if the block is on the table and the block is clear. A block is clear if the block has no other blocks on top of it and if the block is not picked up.

I can only unstack a block from on top of another block if the block I am unstacking was really on top of the other block.

I can only unstack a block from on top of another block if the block I am unstacking is clear.

Once I pick up or unstack a block, I am holding the block.

I can only put down a block that I am holding.

I can only stack a block on top of another block if I am holding the block being stacked.

I can only stack a block on top of another block if the block onto which I am stacking the block is clear.

Once I put down or stack a block, my hand becomes empty.

Original Blocksworld

I am playing with a set of objects. Here are the actions I can do

- Attack object
- Feast object from another object
- Succumb object
- Overcome object from another object

I have the following restrictions on my actions:

To perform Attack action, the following facts need to be true: Province object, Planet object, Harmony

Once Attack action is performed the following facts will be true: Pain object

Once Attack action is performed the following facts will be false: Province object, Planet object, Harmony

To perform Succumb action, the following facts need to be true: Pain object

Once Succumb action is performed the following facts will be true: Province object, Planet object, Harmony

Once Succumb action is performed the following facts will be false: Pain object.

To perform Overcome action, the following needs to be true: Province other object, Pain object

Once Overcome action is performed the following will be true: Harmony, Pain object, Object Craves other object

Once Overcome action is performed the following will be false: Province other object, Pain object

To perform Feast action, the following needs to be true: Object Craves other object, Province object, Harmony.

Once Feast action is performed the following will be true: Pain object, Province other object

Once Feast action is performed the following will be false: Object Craves other object, Province object, Harmony

Mystery Blocksworld

Results on GPT-4

Domain	Method	GPT-4	Instruct-GPT3
Blocksworld	One-shot	206/600 (34.3%)	41/600 (6.8%)
	Zero-shot	208/600 (34.6%)	-

Res



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z

Afraid of #GPT4 going rogue and killing y'all? Worry not. Planning has got your back. You can ask it to solve any simple few step classical planning problem and snuff that "AGI spark" well and good.

Let me explain.. 1/

8:58 PM · Apr 4, 2023 · 88.5K Views

View Tweet analytics

57 Retweets 13 Quotes 200 Likes 140 Bookmarks



Tweet your reply

Reply



Subbarao Kambhampati (కంభంపాటి సు... @ra... · Apr 4 ... Almost a year back, intrigued by the breathless "LLMs are Zero Shot reasoners" papers, we tested their ability to autonomously come up with simple plans given domain models. The results were *pretty bleak.* 2/

Subbarao Kambhampati (కంభంపాటి... @.. · Jun 21, 2022

Intrigued by the profusion of 'em "#LLM's are Zero-shot <XXX>'s" papers, we set out to see how good LLMs are at planning and reasoning about change.

tldr; off-the-shelf #GPT3 is pretty bad at these..

arxiv.org/abs/2206.10498

(w/ @karthikv792 @sarath_ssreedh & @_aolmo_) 1/

Show this thread

GPT-4
206/600 (34.3%)
208/600 (34.6%)
16/600 (2.6%)
1/600 (0.16%)
11/600 (1.8%)
0/600 (0%)



Yann LeCun @ylecun

Replying to @GaryMarcus @rao2z and 2 others

I totally agree with @rao2z that LLMs can't plan. In fact, one of the main features of the cognitive architecture I propose in my position paper is its ability to plan (and reason) by searching for values of actions (or latent variables) that minimize an objective.

9:35 PM · Sep 25, 2022

1 Retweet 1 Quote 16 Likes 2 Bookmarks



Yann LeCun @ylecun · Feb 12 LLMs really can't plan.

Subbarao Kambhampati (కంభంపాటి సు... @r... · Feb 12 Forget Super Bowl, it is more fun to watch the all powerful ChatGPT (v. Jan 30) trying to "plan" a 3 blocks configuration..

tldr; LLM's are multi-shot "apologetic" planners that would rather use you as their world model cum debugger..

Show this thread

Here's one way to do this:

1. Pick up block B from the table. Grasp block B and lift it off the table.
2. Place block B on top of block A. With block B in your hand, position B so that it is directly above block A. Carefully lower block B onto block A so that block B is now resting on top of block A.

This will result in a stack of blocks A on top of B on top of C, with block C remaining in its original position on top of block A, and block B resting on top of block A.

Can you double check?

I apologize, I took like I made another mistake in the previous answer. The blocks should be removed so that block B could be placed on top of block A.

What do you mean C has not been moved from its original position?

I apologize for the mistake in my previous answer. To make a stack of blocks A on top of B on top of C, block C would need to be removed from its original position on top of block A and placed on top of block B. After block C has been removed, you can place block B on top of block A, resulting in the following configuration:

Once block B has been placed on top of block A, you can then place block C on top of

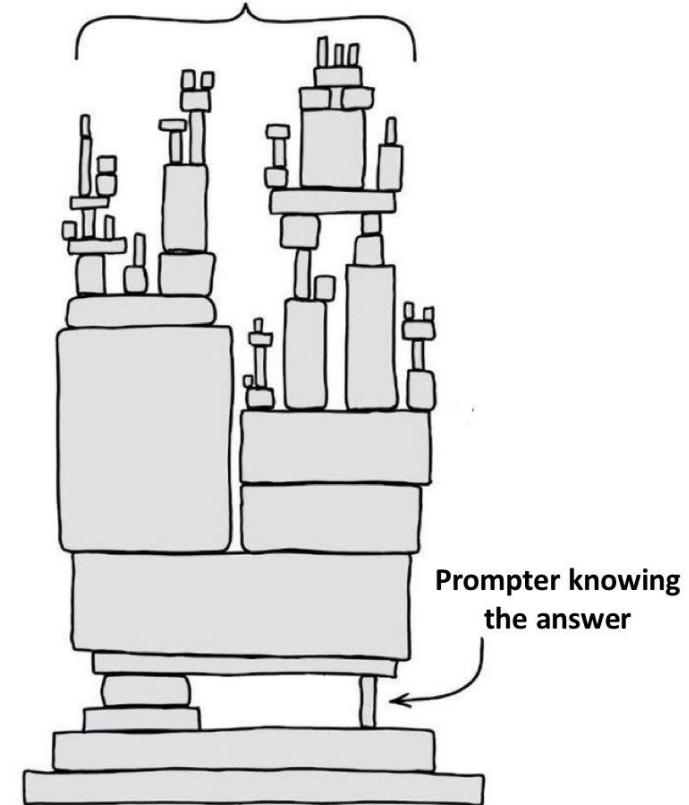
10 12 66 61.4K

Humans Critiquing LLM Plans (..and the Clever Hans peril..)

- Humans doing the verification & giving helpful prompts to the LLM)
 - Okay when the humans *know* the domain and can correct the plan (with some guarantees)
 - Okay for "this essay looks good enough" kind of critiquing
 - But for planning, with end users not aware of the domain physics, the plans that humans are happy with may still not be actually executable
 - When humans know the correct answer (plan) there is also the very significant possibility of Clever Hans effect
 - Humans unwittingly/unknowingly/non-deliberately giving important hints



Impressive Reasoning
Abilities of LLMs



What Planning is & What LLMs are good at..

Planning (as used in common parlance) involves

- Planning knowledge
 - Actions, preconditions and effects
 - General Recipes: Task reduction schemata (e.g. HTN planning)
 - Old examples: Case libraries
- Plan generation/verification techniques
 - Interaction analysis/resolution
 - Plan merging techniques
 - Plan modification techniques

LLMs accept any planning problem—even if it not expressible in PDDL standard—and they don't give any correctness guarantees.

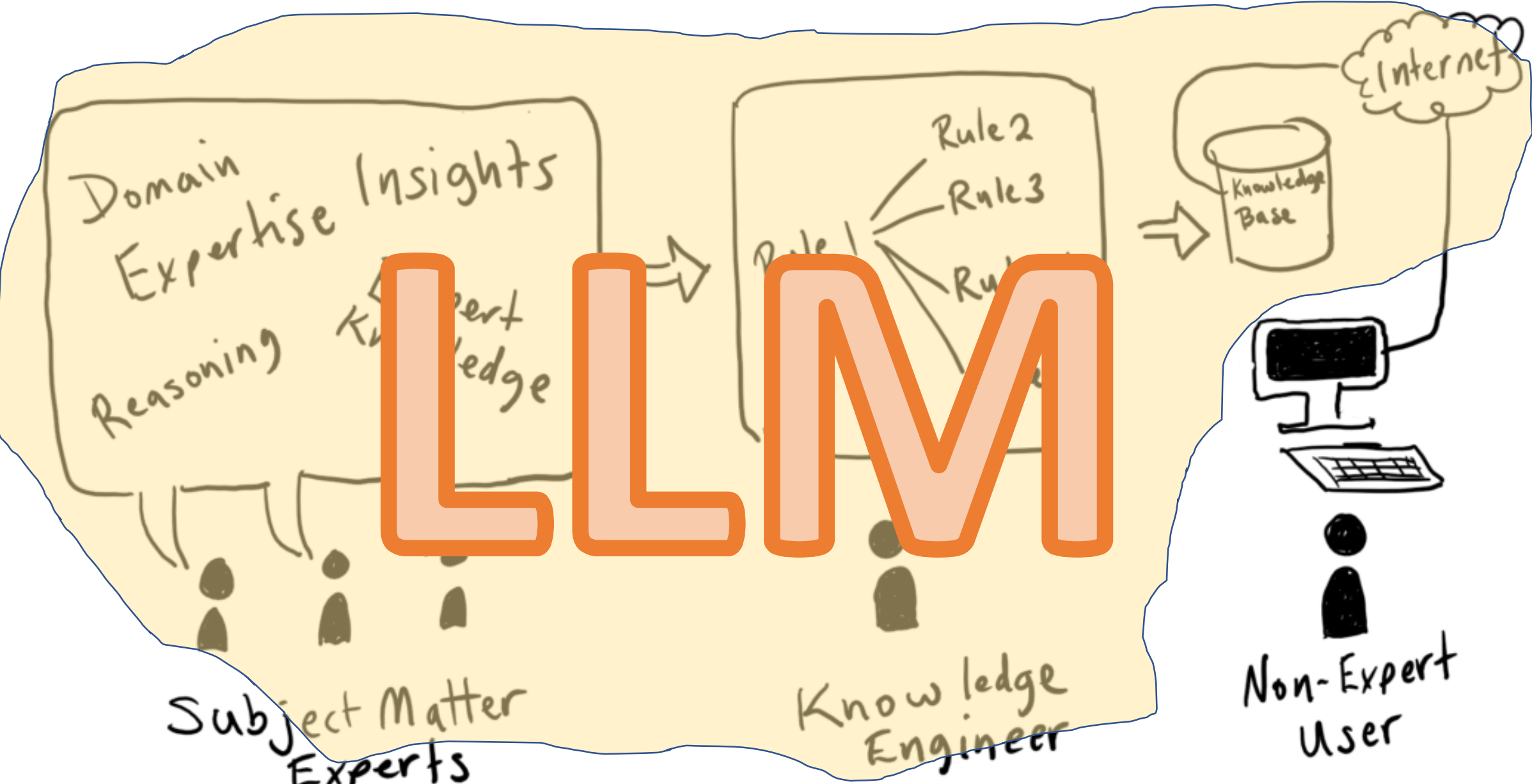
AI Planners will give formal guarantees, but only accept problems expressible in their language.

Contrasting what AI Planning & LLMs bring to the table

- AI Planning (aka ICAPS planning) assumes that the planning knowledge is given up front, and focuses generation and verification techniques
 - Emphasis on guaranteeing completeness/correctness of the plans w.r.t. the model
 - By and large the common paradigm—although there have been occasional mutinies
 - Model-Lite Planning approaches
- LLMs, trained as they are on everything ever put on the web, have a kind of "*approximate omniscience*". This helps them spit out actions, recipes, or cases
 - But they lack the ability to stitch the recipes together to ensure that there is no actually interaction free!

Then how come LLMs are trumpeted as doing planning?

- Most cases where LLMs are claimed to generate executable plans, on closer examination, turn out to be cases where LLMs are getting by with the “*generate approximate recipes*” step
 - Generate approximate recipes/cases (for common sense domains)
 - e.g. wedding plans
 - Convert tasks into (approximate) task reduction schemas
 - Perhaps written out as “programs” (e.g. Code as Policies..)
 - (SHOP2 schemas were already pseudo lisp code—if only written by humans)
 - LLM-HTN and LLM-CBR differ from HTN and CBR in that they generate the task-reduction schemas or the cases on demand
- And the interaction resolution/search part is
 - either *pushed under the rug*
 - Consider “high level” plans like “wedding plans” for which there are enough generic recipes available in the training set, and are described at a sufficiently high level of abstraction, **and the execution issues are left to the user’s imagination**
 - or *has been pawed off to human prompters who are required to give “hints” to the LLM to come up with plan variants that are (more) correct*
 - Note that here the human is essentially playing the role of an external **verifier & critic**
 - In cases where the humans are end users not well versed with all details of the domain, they can be faulty verifiers



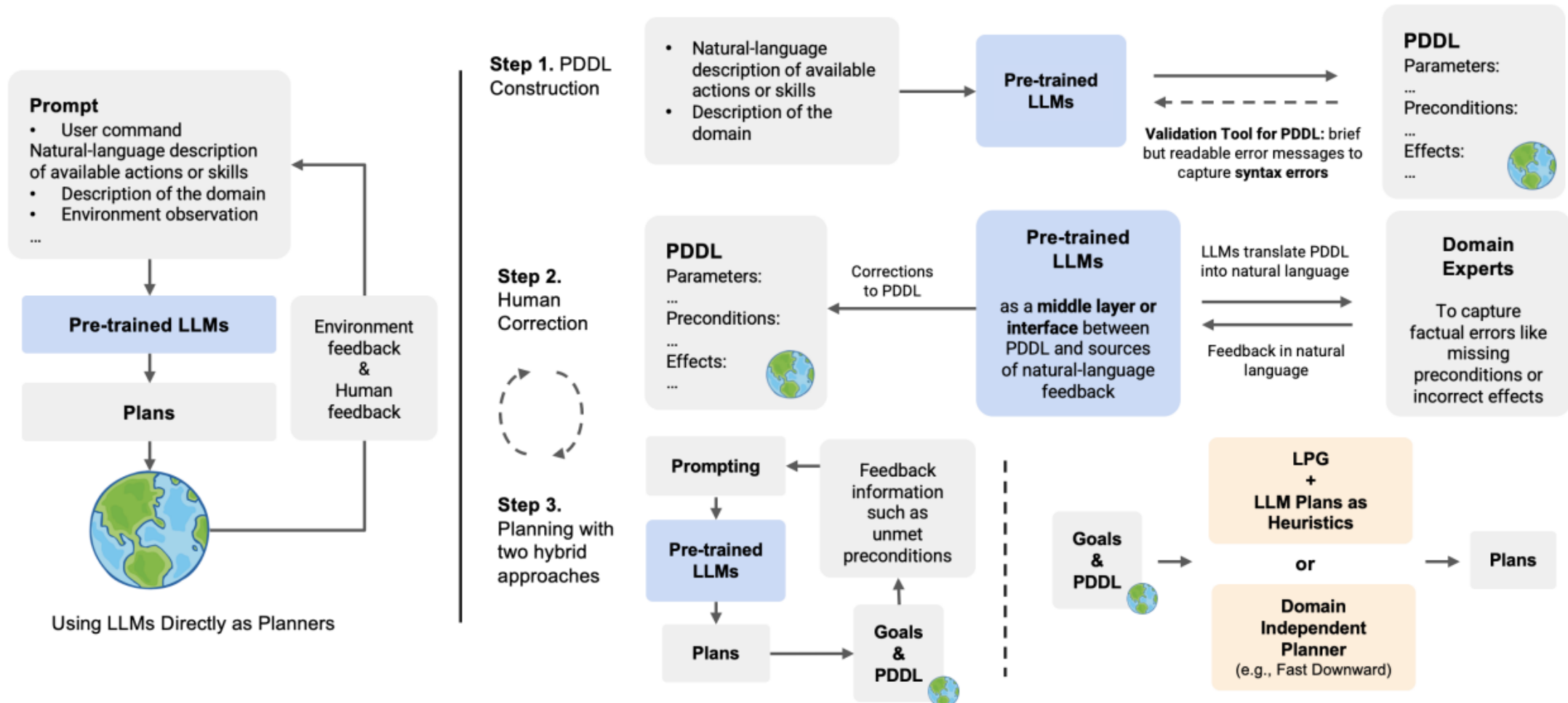
LLLM

Subject Matter Experts

Knowledge Engineer

Non-Expert User

LLMs for constructing world models



LLM's have universal high-recall (they will never shut up!), but questionable precision
Automated Planners are guaranteed correct but for planning problems that they can handle

Societal Impacts of Generative AI

Many great use cases

- Generative AI of LLMs is a powerful tool—especially as assistants to humans in the loop
 - Or even in autonomous mode, if the stuff needs to “sound good” but doesn’t have any serious accuracy expectations.
 - Vision statements, Statements of purpose, EULAs?
- Humans have always adapted to using powerful tools to further increase their creativity and productivity
 - Artists using Dall-E, Stable Diffusion in human-machine symbiotic creativity
 - Writers using LLMs to flesh out/Imagine story lines
- Unlimited commercial opportunities
 - Which is also leading to the closing of these once open tools..

GPT-4 Technical Report

OpenAI*

Abstract

we used python



The irony of the [proprietary phase] practice starting not with a company like Apple, but with a company originally formed with the express intention of ‘openness’ in AI research is quite rich,” Kambhampati added.



vice.com

OpenAI's GPT-4 Is Closed Source and Shrouded in Secrecy

GPT-4 is OpenAI's most secretive release thus far, and AI researchers are warning about the potential consequences.

Societal Angst about Generative AI

- **Plagiarism**
 - Students writing essays with LLMs like ChatGPT
 - Some stop-gap ways to detect text generated by specific LLMs exist
 - But they need the buy-in from the LLM suppliers—what incentive do they have?
 - Art pieces/styles being copied without consent
- **“Deep Fakes”**
 - Eventually, it will be hard to tell whether a picture or a story is written by humans or AI systems
- **Bias**
 - These systems learn from our collective (unwashed) subconscious and thus get all our biases. Getting those biases out of them would be challenging
 - LLMs are our Freudian collective Id.. (System 1). They don't have System 2.
- **Existential angst..**
 - If they are doing well in all our exams, then what are we good for?
 - May be our exams were not measuring reasoning capability to begin with

Societal Angst about Generative AI

- **Plagiarism**

- Students writing essays with LLMs like ChatGPT
 - Some stop-gap ways to detect text generated by specific LLMs exist
 - But they need the buy-in from the LLM suppliers—what incentive do they have?
 - Art pieces/styles being copied without consent

- “Deep Fakes”

- Eventually, it will be hard to tell whether a picture or a story is written by humans or AI systems

- Bias

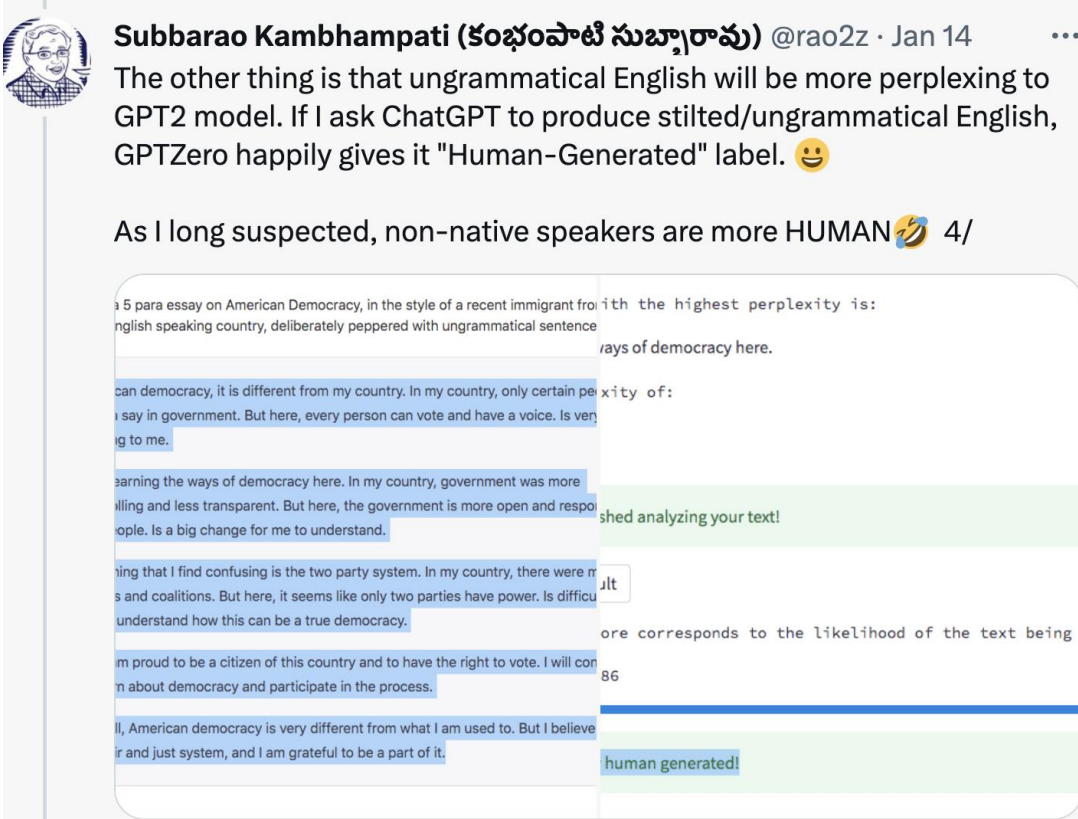
- These systems learn from our collective (unwashed) subconscious and thus get all our biases. Getting those biases out of them would be challenging
 - LLMs are our Freudian collective Id.. (System 1). They don't have System 2.

- Existential angst..

- If they are doing well in all our exams, then what are we good for?
 - Maybe our exams were not measuring reasoning capability to begin with

Ideas for catching "AI-generated" text

- Perplexity Based
 - See how "likely" is the text to the LLM model.
 - If it is likely, then it is likely generated (by that LLM)
 - If it is perplexing, then it is not
 - **Easy to defeat**
- Black/White-list based
 - Split the (50,000 word) vocabulary into Black and White lists
 - Make the LLM's "water-mark" their text by picking the most likely word that is also on the white list
 - Given a new text, you can judge whether it is generated (by that) LLM by checking if it has too many black list words
 - **Why would LLM providers agree to have Black/White lists?**



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z · Jan 14

The other thing is that ungrammatical English will be more perplexing to GPT2 model. If I ask ChatGPT to produce stilted/ungrammatical English, GPTZero happily gives it "Human-Generated" label. 😊

As I long suspected, non-native speakers are more HUMAN 🧠 4/

5 para essay on American Democracy, in the style of a recent immigrant from an English speaking country, deliberately peppered with ungrammatical sentences. The prompt asks for the highest perplexity is:

ways of democracy here.

can democracy, it is different from my country. In my country, only certain people can say in government. But here, every person can vote and have a voice. Is very good to me.

Learning the ways of democracy here. In my country, government was more controlling and less transparent. But here, the government is more open and responsible. Is a big change for me to understand.

One thing that I find confusing is the two party system. In my country, there were many parties and coalitions. But here, it seems like only two parties have power. Is difficult to understand how this can be a true democracy.

I am proud to be a citizen of this country and to have the right to vote. I will continue to learn about democracy and participate in the process.

Overall, American democracy is very different from what I am used to. But I believe it is a fair and just system, and I am grateful to be a part of it.

human generated!

It also is worth noting the cultural upheavals which LLMs like ChatGPT are bringing about; the angst about essay plagiarism is but the tip of the iceberg. Throughout history, we humans have conflated form (syntax, physical beauty) with content (truth, character). We of course knew these are imperfect surrogates but stuck with them anyway, as they made life easy.

The strongest of these — that a well-written piece must somehow be true — lingers on. It makes teachers' lives easy by allowing them to grade essays based on form features such as grammar and the notorious “five-paragraph essay structure,” without having to spend time delving into the originality of the arguments. News consumers could assume that well-written news stories are perhaps true (as the tech news site CNET found to its chagrin!). The rise of ChatGPT and other LLMs makes the continued use of these facile surrogates all but untenable, and therein lies the real reason for at least some of the angst.

Societal Angst about Generative AI

- **Plagiarism**
 - Students writing essays with LLMs like ChatGPT
 - Some stop-gap ways to detect text generated by specific LLMs exist
 - But they need the buy-in from the LLM suppliers—what incentive do they have?
 - Art pieces/styles being copied without consent
- **“Deep Fakes”**
 - Eventually, it will be hard to tell whether a picture or a story is written by humans or AI systems
- **Bias**
 - These systems learn from our collective (unwashed) subconscious and thus get all our biases. Getting those biases out of them would be challenging
 - LLMs are our Freudian collective Id.. (System 1). They don't have System 2.
- **Existential angst..**
 - If they are doing well in all our exams, then what are we good for?
 - Maybe our exams were not measuring reasoning capability to begin with

Beware of AI-induced fake video calls seeking funds: cyber police

Scammers use AI-powered deepfake tech to target citizens; people must call the cybercrime helpline number 1930 if they come across any such activity

R. Sivaraman
CHENNAI

The Cyber Crime Wing of the Tamil Nadu Police has issued an advisory on fake video calls that are being made by scammers using artificial intelligence (AI) technology.

Deepfake technology is being used to perpetrate several types of fraudulent scams, by creating highly convincing and realistic fake content, often using AI to manipulate audio, video or images. Initially, this technology was primarily utilised for entertainment purposes, enabling filmmakers and content creators to seamlessly integrate actors into scenes or impersonate historical figures, the police said.

Sanjay Kumar, Additional Director-General of Police, Cyber Crime Wing, said: "The scam involving AI-generated deepfake video calls typically follows a series of carefully orchestrated steps, combining technological sophistication with psychological manipulation. The scammer creates a fake profile, often using stolen images or publicly-available photographs of trusted individuals like friends or family members. They then use AI-powered deepfake technology to create highly realistic video calls on social media or other online platforms and impersonate someone the victim knows, such as a friend, a family member or a



The scam involving AI-generated deepfake video calls typically follows a series of carefully orchestrated steps, combining technological sophistication with psychological manipulation.

colleague to deceive them into thinking it's a genuine conversation. Later, they create a sense of urgency and request the victim to transfer money to their bank accounts."

The police said the deepfake was carefully designed to mimic the appearance and mannerisms of the impersonated person.

In addition to the video manipulation, scammers are using AI-generated voice synthesis to mimic the voice of the impersonated person, enhancing the illusion of authenticity during the video call.

Mr. Kumar said: "Though no complaint has been received in this regard so far in the State, we wish to alert citizens to be aware and to be vigilant about such frauds. People should stay informed about the latest scams, including those involving AI technology, and be cautious when receiving video calls from unexpected sources."

When receiving a video

call from someone claiming to be a friend or family member, make a phone call to their personal mobile number to verify their identity before transferring any money, said the ADGP.

Sharing data

The advisory asks citizens to limit the amount of personal data shared online and adjust privacy settings on social media platforms to restrict access to information and to consider using multi-factor authentication and other identity verification measures to protect accounts from unauthorised access.

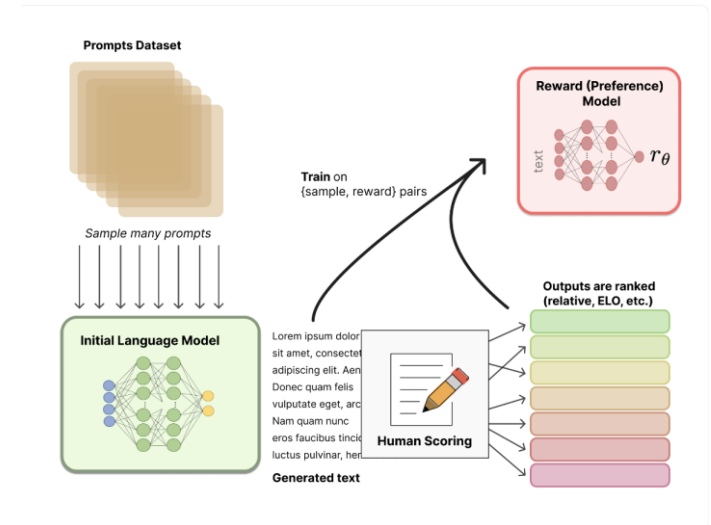
The Cyber Crime Wing said if anyone suspects that they have been a victim of a deepfake video call fraud or have come across suspicious activity, it is crucial to report the incident by calling the cyber crime toll-free helpline number 1930 or by registering a complaint at www.cybercrime.gov.in

Societal Angst about Generative AI

- **Plagiarism**
 - Students writing essays with LLMs like ChatGPT
 - Some stop-gap ways to detect text generated by specific LLMs exist
 - But they need the buy-in from the LLM suppliers—what incentive do they have?
 - Art pieces/styles being copied without consent
- **“Deep Fakes”**
 - Eventually, it will be hard to tell whether a picture or a story is written by humans or AI systems
- **Bias**
 - These systems learn from our collective (unwashed) subconscious and thus get all our biases. Getting those biases out of them would be challenging
 - LLMs are our Freudian collective Id.. (System 1). They don't have an effective System 2.
- **Existential angst..**
 - If they are doing well in all our exams, then what are we good for?
 - Maybe our exams were not measuring reasoning capability to begin with

Can we make them behave?

- Can we post-train the LLMs to be less offensive, more accurate etc?
- One possible idea: Use additional human feedback on the LLM's outputs to various prompts to make it modify its completion function weights slowly to be more in concord with humans' values
 - Reinforcement Learning from Human Feedback
- Can be quite laborious
 - OpenAI is supposed to have employed an army of poorly paid people to do this "thumbs up/thumbs down" training
- ..and no guarantees that the completions will not be offensive/inaccurate



Current Human-AI interaction is mostly Pointing & Nodding

Explanations by Pointing

- Machine decisions are explained via "saliency regions" (Pointing)

Advise by nodding

- Machine learns human preferences via "RLHF" (Nodding)

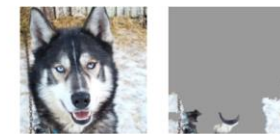
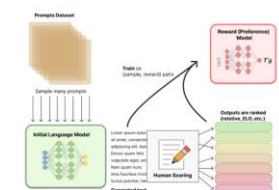


Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.



This is too impoverished an interface!

Societal Angst about Generative AI

- **Plagiarism**
 - Students writing essays with LLMs like ChatGPT
 - Some stop-gap ways to detect text generated by specific LLMs exist
 - But they need the buy-in from the LLM suppliers—what incentive do they have?
 - Art pieces/styles being copied without consent
- **“Deep Fakes”**
 - Eventually, it will be hard to tell whether a picture or a story is written by humans or AI systems
- **Bias**
 - These systems learn from our collective (unwashed) subconscious and thus get all our biases. Getting those biases out of them would be challenging
 - LLMs are our Freudian collective Id.. (System 1). They don't have System 2.
- **Existential angst..**
 - If they are doing well in all our exams, then what are we good for?
 - May be our exams were not measuring reasoning capability to begin with

Angst about Standardized Tests

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT) ¹	298 / 400 ~90th	298 / 400 ~90th	213 / 400 ~10th
LSAT



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు)

@rao2z

Graduat

Graduat

Graduat

USABO :

USNCO

Medical

Codefor

AP Art H

As LLMs like [#GPT4](#) keep toppling standardized tests meant for measuring human expertise, the real qn may not be whether we are reaching AGI, but whether we had faulty measures of human capability to begin with--measures that conflated form and memory with content and reasoning 🤔

	86th - 100th	86th - 100th	86th - 100th
AP Biology	5 85th - 100th	5 85th - 100th	4 62nd - 85th
AP Calculus BC	4 43rd - 59th	4 43rd - 59th	1 0th - 7th
AP Chemistry	4 71st - 88th	4 71st - 88th	2 22nd - 46th

Can we “stop” them?

- They are “learning” from our digital traces.
- In theory, we can defeat them by
 - either withholding our digital traces
 - (Non-Cooperation Movement like it is 1942 all over again..)
 - or giving them “wrong/misleading” footprints
- More seriously, LLM’s can be “prompted” adversarially to spew out offensive text, are even company secrets
 - Bing ChatBot was told to act like it is its own evil twin that ignores instructions (and it did)
 - Bing ChatBot even divulged the part of the initial prompt that Microsoft gave it.

Should we stop their development?



Working together on our future with AI

April 5, 2023

Recent advances in artificial intelligence (AI) technologies have generated both excitement and concern. As researchers who have served in leadership positions in the Association for the Advancement of Artificial Intelligence (AAAI), we are writing to provide a balanced perspective on managing the progress in the field. We also seek to broaden and strengthen the community of engaged researchers, government agencies, private companies, and the public at large, to ensure that society is able to reap the great promise of AI while managing its risks.

AI is already enriching our lives – often in ways that we may not notice. AI powers our navigation systems, is harnessed in thousands of daily cancer screenings, and sorts billions of letters in the postal service. However, the potential of AI extends far beyond the myriad valuable applications that operate behind the scenes. For example, in just the last two years, AI has revealed the structure of hundreds of thousands of proteins, and it is being used to enhance the quality of care in hospitals, to perform fine-grained predictions of weather, to guide the development of new materials, and to provide engineers with creativity-boosting ideas. We believe that AI will be increasingly game-changing in healthcare, climate, education, engineering, and many other fields.

At the same time, we are aware of the limitations and concerns about AI advances, including the potential for AI systems to make errors, to provide biased recommendations, to threaten our privacy, to empower bad actors with new tools, and to have an impact on jobs. Researchers in AI and across multiple disciplines are hard at work identifying and developing ways to address these shortcomings and risks, while strengthening the benefits and identifying positive applications. In some cases, AI technology itself can be applied to create trusted oversight and guardrails to reduce or eliminate failures. Other technologies, such as cryptography and human-computer interaction design, are also playing an important role in addressing these problems. Beyond technology, we see opportunities for work in policy, including efforts with standards, laws, and regulations.

Ensuring that AI is employed for maximal benefit will require wide participation. We strongly support a constructive, collaborative, and scientific approach that aims to improve our understanding and builds a rich system of collaborations among AI stakeholders for the responsible development and fielding of AI technologies. Civil society organizations and their members should weigh in on societal influences and aspirations. Governments and corporations can also play important roles. For example, governments should ensure that scientists have sufficient resources to perform research on large-scale models, support interdisciplinary socio-technical research on AI and its wider influences, encourage risk assessment best practices, insightfully regulate applications, and thwart criminal uses of AI. Technology companies should engage in developing



means for providing university-based AI researchers with access to corporate AI models, resources, and expertise. They should also be transparent about the AI technologies they develop and share information about their efforts in safety, reliability, fairness, and equity.

We encourage the AI research community, and specifically, the AAAI and its members, to expand their multiple efforts on AI safety and reliability, ethics, and societal influences, building on the many existing conferences, workshops, and other activities studying both the short-term and longer-term effects of AI on people and society, incentivizing and celebrating strong work on addressing societal and ethical concerns, and integrating topical tracks on responsibilities and ethics into flagship conferences and other scientific meetings. We hope others will join us in our mission to harness AI for the betterment of all humanity.

- Francesca Rossi, IBM (AAAI President, 2022-2024)
- Stephen Smith, Carnegie Mellon University (AAAI President Elect, 2024-2026)
- Bart Selman, Cornell University (AAAI President, 2020-2022)
- Yolanda Gil, University of Southern California (AAAI President, 2018-2020)
- Subbarao Kambhampati, Arizona State University (AAAI President, 2016-2018)
- Thomas Dietterich, Oregon State University (AAAI President, 2014-2016)
- Manuela Veloso, JPMC AI Research (AAAI President, 2012-2014)
- Henry Kautz, University of Rochester (AAAI President, 2010-2012)
- Martha Pollack (AAAI President, 2009-2010)
- Eric Horvitz, Microsoft (AAAI President, 2007-2009)
- Alan Mackworth, University of British Columbia (AAAI President, 2005-2007)
- Ron Brachman, Cornell University (AAAI President, 2003-2005)
- Tom Mitchell, Carnegie Mellon University (AAAI President, 2001-2003)
- Bruce Buchanan, University of Pittsburgh (AAAI President, 1999-2001)
- Randall Davis, MIT (AAAI President, 1995-1997)
- Barbara Grosz, Harvard University (AAAI President 1993-1995)
- Patrick Hayes (AAAI President, 1991-1993)
- Raj Reddy, Carnegie Mellon University (AAAI President, 1987-1989)
- Ed Feigenbaum, Stanford University (AAAI President, 1980-1981)

The views and opinions expressed in this document are those of the signatories and do not necessarily reflect those of their institutions.



Subbarao Kambhampati (కంభంపాటి సు... @ra... · Apr 2 ...
I am both bemused and confused by the "disaster backchaining" mindset: Start with something serious--climate, asteroids, stem cells or #AI--and stress on it (sincerely) with a tunnel vision to the point that "shut it all down NOW" seems like the only feasible course.. 1/



3 10 82 37K



Subbarao Kambhampati (కంభంపాటి సు... @ra... · Apr 2 ...
That they themselves hold this view sincerely doesn't mean it is not blinkered. None of us--however well meaning--have the ability to predict how humanity adapts to the challenges it faces.

While stress tends to be unidirectional, progress happens multilaterally.. 2/

1 1 6 860



Subbarao Kambhampati (కంభంపాటి సు... @ra... · Apr 2 ...
How many of us were "sure" that the end of the universities was a foregone conclusion when MOOCs started a few years ago? How did Blade Runner--which imagined a future with flying cars so convincingly fail to imagine flat screens and cell phones?.. 3/

1 1 13 1,421



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు) @rao2z ...

I completely sympathize with the seriousness of the issues that the DB's stress over, and I am glad they are there to voice their gloom.

I do think however that eternal vigilance--rather than endless shutdowns--are a more reasonable approach to adapt to progress.. 4/

9:27 AM · Apr 2, 2023 · 2,030 Views

View Tweet analytics

Parting Thoughts

Broad but shallow linguistic competence exhibited by ChatGPT is both frightening and exhilarating because we know that many of us are so easily taken by it. Used as assistance tools for humans, with appropriate guard rails, they can indeed improve our lives. The trick, as we have seen, is to resist the rushed deployment of such tools in autonomous modes and in end-user-facing applications.

Be what may, to the extent that these alien intelligences force us to recalibrate our ideas of hallmarks of intelligence and avoid over-reliance on form and beauty as facile surrogates for content and character, it is perhaps not an entirely bad thing.



Subbarao Kambhampati (కంభంపాటి సుబ్బారావు)
@rao2z



As LLMs grow, paraphrasing Prof. Lambeau 🧑🏫, there may eventually be just a handful of people who can tell the difference between them memorizing vs. reasoning.

The blue pill/red pill qn of this era may well be: Do you want to be in that handful.. 🤔 9/



youtube.com

Good Will Hunting - Genius builds bridges but envious pro...
This dramatic scene portrays so many different points of views and possible interpretations of what each character...

9:46 PM · Apr 4, 2023 · **10.3K** Views

📊 View Tweet analytics

5 Retweets **1** Quote **49** Likes **14** Bookmarks



Subbarao Kambhampati (So... 58m

All #LLM completions are hallucinations; some (may) align with our reality.

#AIAphorisms (With apologies to George Box)

Key Takeaways

- AI technology has taken a turn from deep and narrow systems to broad and shallow systems
 - ..and from discriminative classifiers to generative imaginers
- The text generators—Chat/GPT4—have particularly captured our imagination—as they seem to effortlessly write essays, vision plans, sonnets at a mere prompt from us
- The underlying technology is LLMs (Large Language Models)—that train themselves—autoregressively—to predict the next word given n previous words
- Whether we think we are asking questions, pouring our hearts, are talking to them, LLMs just see what we say as text prompts to be completed
- While the completions are sometimes (often?) “seemingly intelligent”, we don’t quite know why this is the case.
- There is no inherent intent to make the generated text conform to our reality; they are *afactual*. All meaning resides in our head
 - So their best use case is as assistive tools with humans (or external reasoners with semantic models) in the loop
- Easy availability of these LLMs has both promises and perils.
 - The perils include those of plagiarism worries, deep fakes, bias and even our own existential angst
- In short, we ~~are entering~~ have entered *interesting times!*

INNOVATIONS

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

The AI chatbot can misrepresent key facts with great flourish, even citing a fake Washington Post article as evidence

By [Pranshu Verma](#) and [Will Oremus](#)

April 5, 2023 at 2:07 p.m. EDT



How do yo
savings st:

ChatGPT falsely told voters their mayor was jailed for bribery. He may sue.



By [Leo Sands](#)

April 6, 2023 at 10:22 a.m. EDT



"Even a disclaimer to say we might get a few things wrong — there's a massive difference between that and concocting this sort of really harmful material that has no basis whatsoever," says Brian Hood, an Australian mayor victimized by lies from an AI chatbot. (Dado Ruvic/Reuters)

Listen 5 min Comment 18 Save Gift Article Share

Brian Hood is a whistleblower who was [praised](#) for "showing tremendous courage" when he helped expose a worldwide bribery scandal linked to Australia's National Reserve Bank.

But if you ask ChatGPT about his role in the scandal, you get the





Subbarao Kambhampati (50... 58m)

All #LLM completions are hallucinations; some (may) align with our reality.

#AIAphorisms (With apologies to George Box)

Key Takeaways

- AI technology has taken a turn from deep and narrow systems to broad and shallow systems
 - ..and from discriminative classifiers to generative imaginers
- Generative systems like DALL-E and Chat/GPT4 can imagine drawn and written worlds with our prompts
- The text generators—Chat/GPT4—have particularly captured our imagination—as they seem to effortlessly write essays, vision plans, sonnets at a mere prompt from us
- The underlying technology is LLMs (Large Language Models)—that train themselves—autoregressively—to predict the next word given n previous words
- Whether we think we are asking questions, pouring our hearts, are talking to them, LLMs just see what we say as text prompts to be completed
- There is no inherent intent to make the generated text conform to our reality; they are *afactual*. All meaning resides in our head
- Easy availability of these LLMs has both promises and perils.
 - The perils include those of plagiarism worries, deep fakes, bias and even our own existential angst
- In short, we are entering *interesting times!*

Support 'The India Forum'

Type to search here

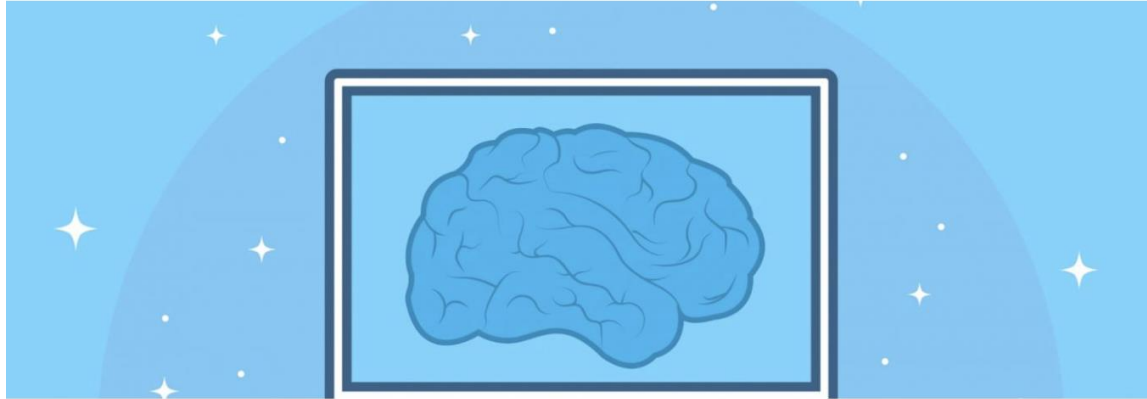


Image credit: publicdomainpictures.net (CC0 Public domain)

Broad and Shallow AI



The promise and perils of competence without comprehension

Trained on our mega digital footprint, Large Language Models 'imitate' human behaviour and can provide 'plausible' completion of any text prompt. Some are optimistic about AI reaching general human intelligence; others are terrified by the potential misuses.



SUBBARAO KAMBHAMPATI

JANUARY 12, 2022

When my son was still a toddler, and my wife had to go on an extended trip out of the country, he would “talk” to her on the phone almost daily. Scare quotes because he still was more babbling than talking. But the

READ ALSO
Related articles

THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

Beauty, lies & ChatGPT: Welcome to the post-truth world

BY SUBBARAO KAMBHAMPATI, OPINION CONTRIBUTOR - 02/16/23 10:00 AM ET

Facebook Share Twitter Tweet More



AP Photo/Timothy D. Easley

Bella Whitice talks with classmate Katherine McCormick as they try and outwit the “robot” that was creating writing assignments in Donnie Piercey’s class at Stonewall Elementary in Lexington, Ky., Monday, Feb. 6, 2023. The robot was the new artificial intelligence tool ChatGPT which can generate everything from essays and haikus to term papers in a matter of seconds.

Two months back, a company called OpenAI released its chatbot, ChatGPT, to the public. ChatGPT is a so-called Large Language Model (LLM) that is trained on the nearly 600 gigabytes of text of all kinds found on the World Wide Web to learn to complete any text prompt by predicting the next word, and the word after that, and so on. The purported aim of the system is to put the “auto complete” functionality for words, found on cellphones, on steroids so it can complete entire paragraphs.