

L'INTELLIGENZA ARTIFICIALE SI DISEGNA INTORNO A TE

Counterfeit Society
quando l'AI erode la
fiducia

Marco S. Nobile

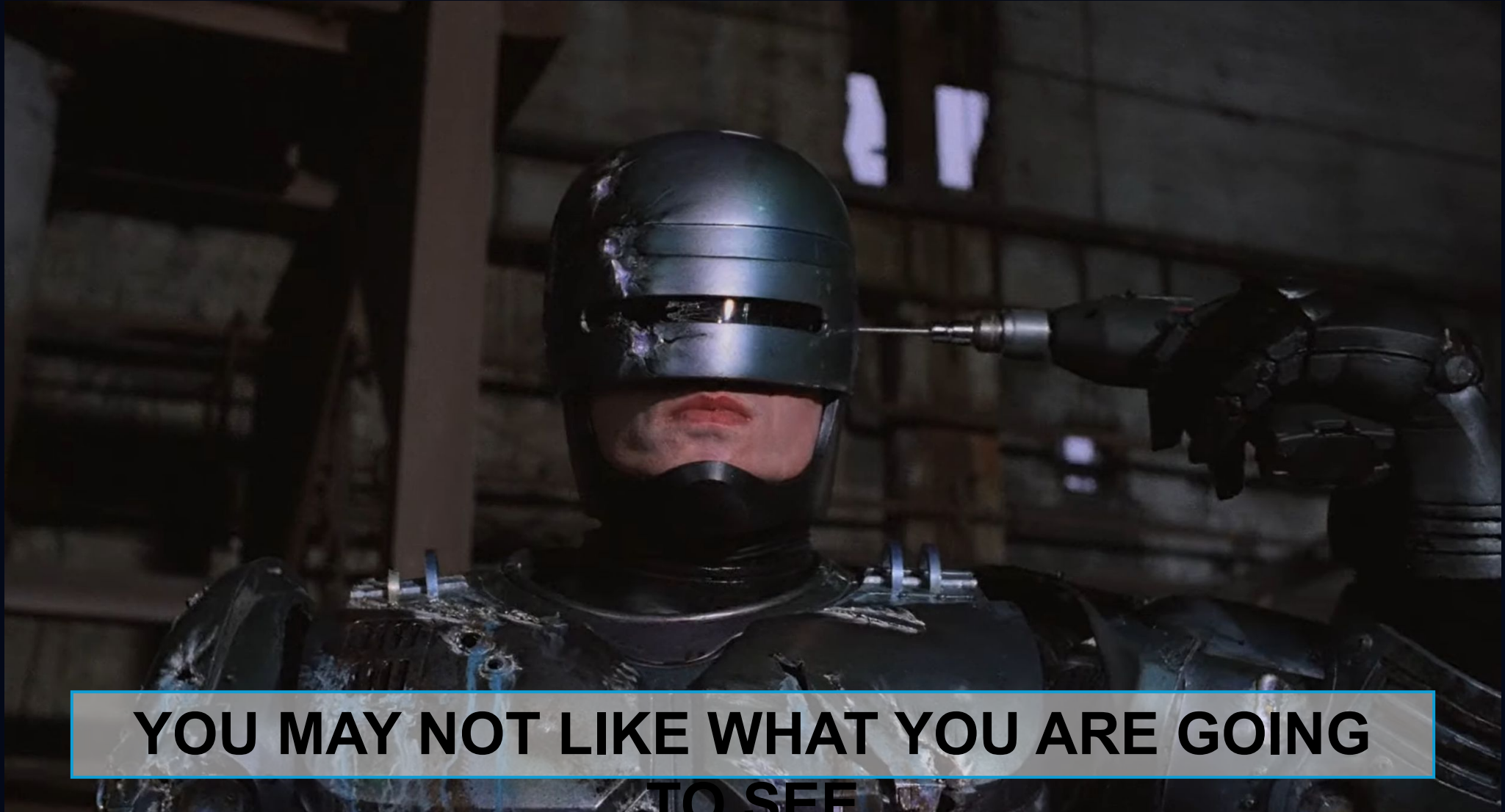
 @aresio.bsky.social

Department of Environmental Sciences,
Informatics and Statistics

Rector's delegate for digital innovation and integrity in
research

March 10th, 2026 – BAUM, Malcantòn Marcorà

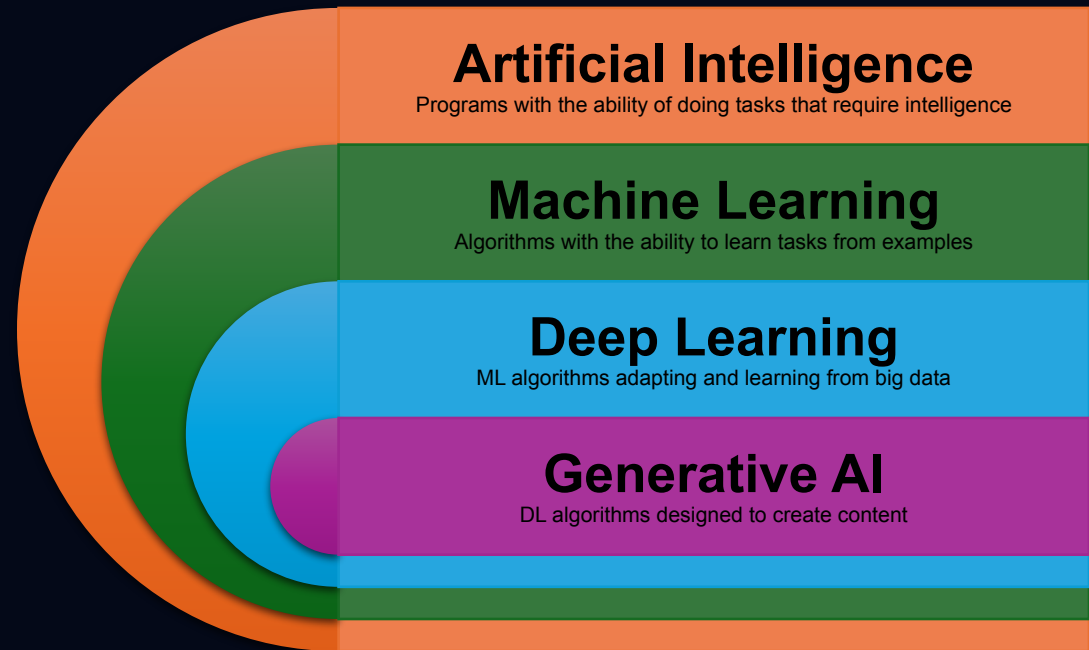
Disclaimer



**YOU MAY NOT LIKE WHAT YOU ARE GOING
TO SEE**

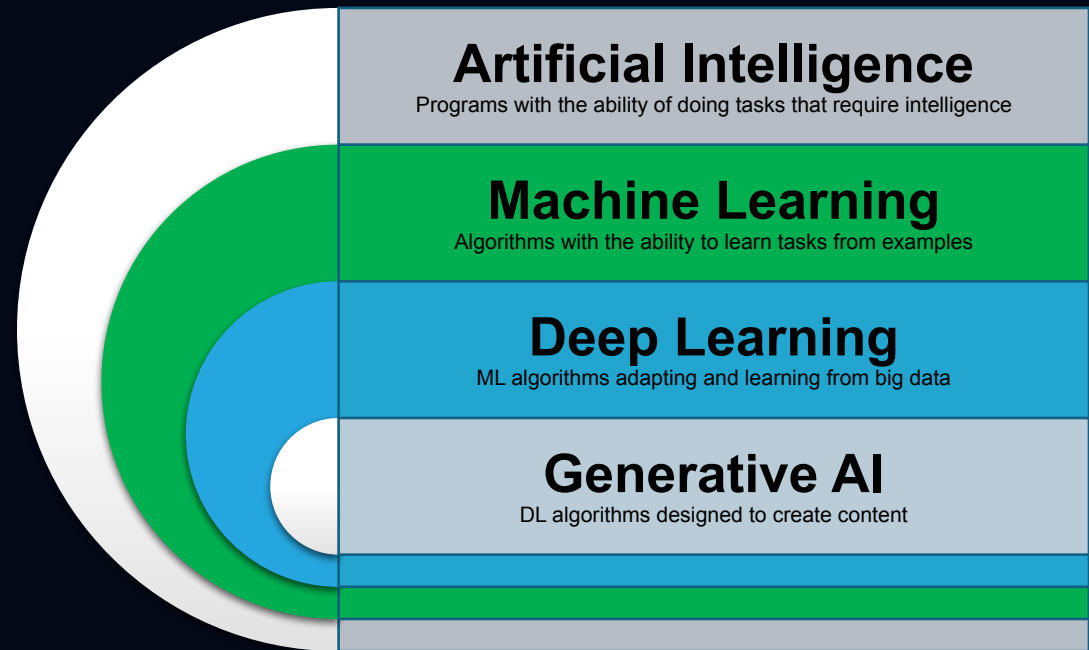
Artificial Intelligence?

- Many definitions, no agreement
- A very practical definition: «**the science of making machines do things that would require intelligence if done by man**»
- There are many ways to create AI
- A very popular approach is **machine learning**



Artificial Intelligence?

- Many definitions, no agreement
- A very practical definition: «**the science of making machines do things that would require intelligence if done by man**»
- There are many ways to create AI
- A very popular approach is **machine learning**



Our research: AI & Medicine

Article

EMBO reports

GPNMB marks a quiescent cell population in melanoma and promotes metastasis formation

Fiorenza Lotti¹, Marine Melixetian¹, Thalia Vlachou^{1,5}, Marco S Nobile^{1,2}, Leone Bacciu², Marco Malferrari³, Nicolò Quaresima³, Stefania Rapino³, Federica Marocchi¹, Massimo Barberis¹, Chiara Soriani¹, Barbara Gallo¹, Velia Mollo¹, Ilaria Ferrarotto¹, Daniela Bossi^{1,6}, Pier Francesco Ferrucci^{1,7}, Pier Giuseppe Pelicci^{1,4}, Lucilla Luzi¹ & Luisa Lanfrancone^{1,8}

Home > Explainable Artificial Intelligence > Conference paper

Assessing the Value of Explainable Artificial Intelligence for Magnetic Resonance Imaging

Conference paper | Open Access | First Online: 12 October 2025
pp 423–447 | [Cite this conference paper](#)

You have full access to this [open access](#) conference paper

[Download book PDF](#) [Download book EPUB](#)

Giada Frasson, Matteo Rizzo, Marco Salvatore Nobile, Amalia Lupi & Emilio Quaia

Home > Endocrine Pathology > Article

MiThyCA: A Computational Pathology Pipeline for the Identification of Microscopic Foci of Papillary Thyroid Carcinoma–Like Nuclear Features with AI in Whole–Slide Histological Images

Research | Published: 07 October 2025
Volume 36, article number 34, (2025) | [Cite this article](#)

Access provided by CARE COMPACT–CRUI

[Download PDF](#)

Leone Bacciu, Mario Urso, Vasco Coelho, Giorgio Cazzaniga, Angela Ida Pincelli, Mattia Garancini, Daniele M. Papetti, Daniela Besozzi, Giulia Capitoli, Stefania Galimberti, Alessia Vargiolu, Nicola Fusco, Andrea Gianatti, Fabio Pagni, Vincenzo L'Imperio & Marco S. Nobile

Assessing Cardiac Functionality by Means of Interpretable AI and Myocardial Strain

Publisher: IEEE [Cite This](#) [PDF](#)

Marco S. Nobile; Amalia Lupi; Leone Bacciu; Matteo Grazioso; Chiara Gallese; Emilio Quaia [All Authors](#)

7 Full Text Views

[R](#) [Share](#) [C](#) [F](#) [B](#)

JOURNAL ARTICLE

Sudden cardiac arrest prediction via deep learning electrocardiogram analysis

Matt T Oberdier, Luca Neri, Alessandro Orro, Richard T Carrick, Marco S Nobile, Sujai Jaipalli, Mariam Khan, Stefano Diciotti, Claudio Borghi, Henry R Halperin

Author Notes

European Heart Journal - Digital Health, Volume 6, Issue 2, March 2025, Pages 170–179, <https://doi.org/10.1093/ehjdh/ztae088>

Published: 25 February 2025 [Article history](#)

Modern Pathology

Volume 37, Issue 12, December 2024, 100608

Research Article

Machine Learning Streamlines the Morphometric Characterization and Multiclass Segmentation of Nuclei in Different Follicular Thyroid Lesions: Everything in a NUTSHELL

Vincenzo L'Imperio^{a,b}, Vasco Coelho^c, Giorgio Cazzaniga^{a,b}, Daniele M. Papetti^c, Fabio Del Carro^{a,b}, Giulia Capitoli^{a,d}, Mario Marino^c, Joranda Ceku^{a,b}, Nicola Fusco^{a,f}, Mariia Ivanova^e, Andrea Gianatti^g, Marco S. Nobile^{d,h}, Stefania Galimberti^{a,d,i}, Daniela Besozzi^{c,d}, Fabio Pagni^{a,b}

nature > nature communications > article

Article | Open access | Published: 20 September 2023

Large T cell clones expressing immune checkpoints increase during multiple myeloma evolution and predict treatment resistance

Cirino Botta, Cristina Perez, Marta Larrayoz, Noemi Puig, Maria-Teresa Cedena, Rosalinda Termini, Ibai Goicoechea, Sara Rodriguez, Aintzane Zabaleta, Aitziber Lopez, Sarai Sarvide, Laura Blanco, Daniele M. Papetti, Marco S. Nobile, Daniela Besozzi, Massimo Gentile, Pierpaolo Correale, Sergio Siragusa, Albert Oriol, Maria Esther González-García, Anna Sureda, Felipe de Arriba, Rafael Rios Tamayo, Jose-Maria Moraleda, the Programa Para el Estudio de la Terapéutica en Hemopatías Malignas/Grupo Español de Mieloma (PETHEMA/GEM) cooperative group & the IMMUnocell study group [+ Show authors](#)

Nature Communications 14, Article number: 5825 (2023) | [Cite this article](#)

9719 Accesses | 19 Citations | 17 Altmetric | [Metrics](#)

Home > BMC Medical Informatics and Decision Making > Article

Assisting clinical diagnosis with interpretable fuzzy probabilistic modelling

Research | Open access | Published: 15 September 2025
Volume 25, article number 330, (2025) | [Cite this article](#)

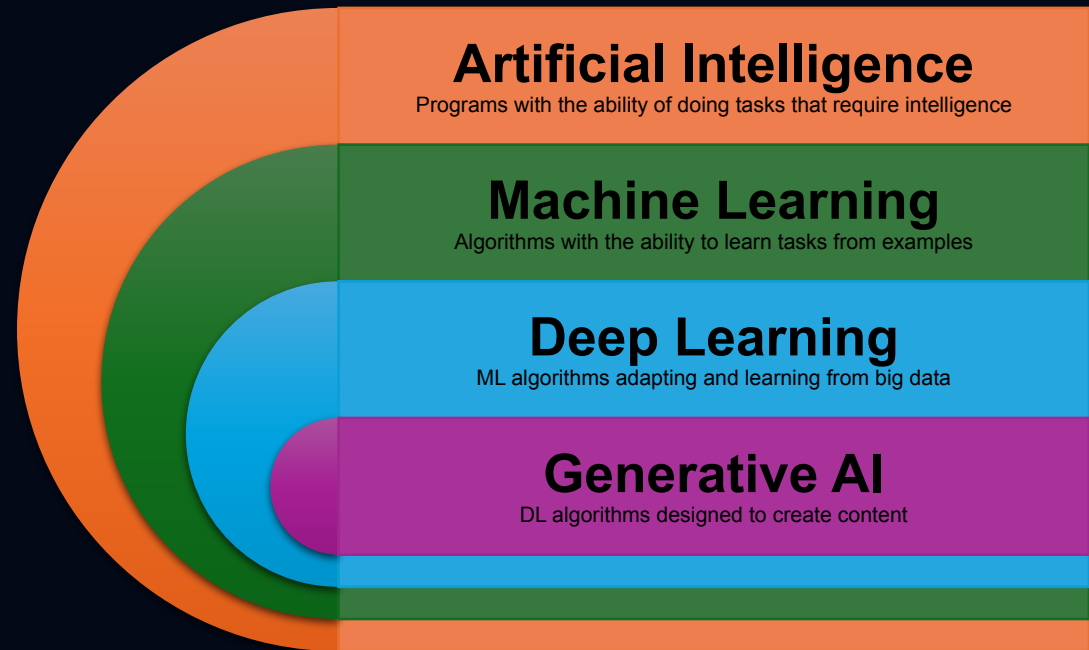
You have full access to this [open access](#) article

[Download PDF](#)

Giulia Capitoli, Marco S. Nobile, Emma L. Ambags, Vincenzo L'Imperio, Michele Provenzano & Pietro Liò

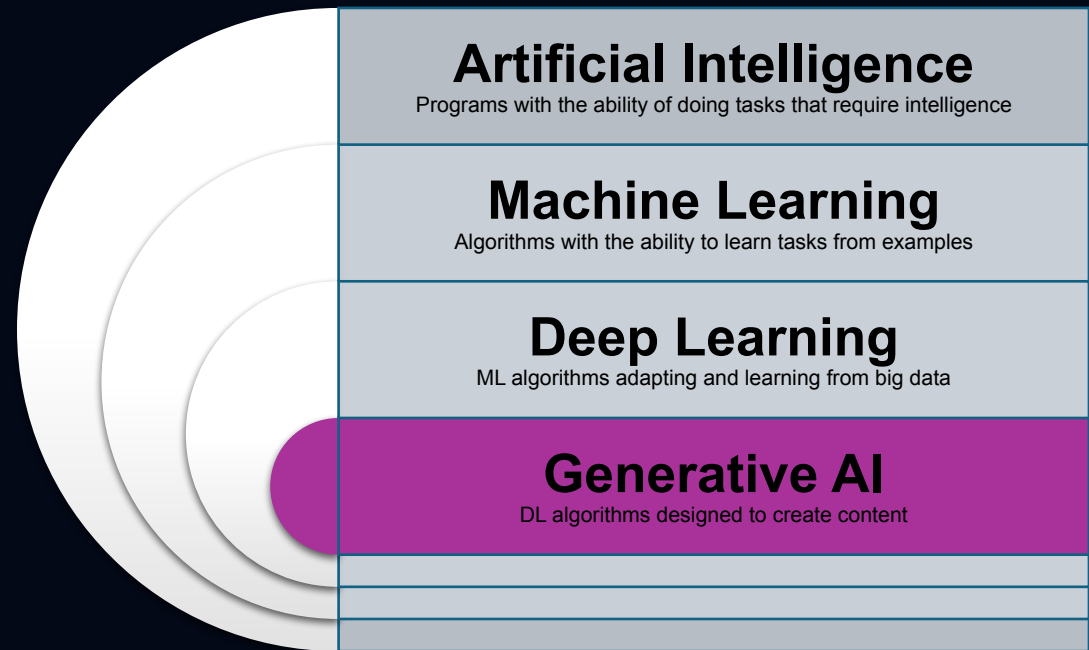
Artificial Intelligence?

- Many definitions, no agreement
- A very practical definition: «**the science of making machines do things that would require intelligence if done by man**»
- There are many ways to create AI
- A very popular approach is **machine learning**



Artificial Intelligence?

- Many definitions, no agreement
- A very practical definition: «**the science of making machines do things that would require intelligence if done by man**»
- There are many ways to create AI
- A very popular approach is **machine learning**





Generative AI

A novel disruptive approach to AI: it is no longer performing specific tasks but it **generates new content**

This kind of models are named **Generative AI (GenAI)**

- The content to be generated is specified with a **textual prompt**
- We can create **anything**: text, images, music, video, programs, cats
- Even **molecules, peptides, proteins, drugs**

Well-known example: **ChatGPT**, a chatbot exploiting GPT, a large language model (LLM) based on a **recent neural architecture** named **Transformer**

ChatGPT will answer on *any* topic



Tell me the recipe of the sarde in saor

Sarde in Saor – Traditional Venetian Sweet-and-Sour Sardines

Sarde in Saor is a classic dish from the Venice area in the Veneto region of Italy. The name *saor* means “flavor” and refers to the sweet-and-sour marinade made with onions, vinegar, and sometimes raisins and pine nuts. Traditionally it was prepared by fishermen to preserve sardines for several days.

Ingredients (4 servings)

- 500 g fresh sardines
- 2 large onions (white or yellow)
- ½ cup white wine vinegar
- 2 tbsp raisins
- 2 tbsp pine nuts
- Flour (for dredging)
- Olive oil (for frying + cooking)
- Salt and pepper
- Optional: a pinch of sugar (if you want it slightly sweeter)

Step-by-Step Preparation



Explain to me briefly the principles of Teller–Ulam design

Sure — here’s a **concise, non-technical explanation** of the Teller–Ulam design, the principle behind modern thermonuclear (hydrogen) weapons:

Core idea

The Teller–Ulam design is a **two-stage configuration** that uses the energy from a **fission (atomic) bomb** to trigger a much more powerful **fusion (hydrogen) reaction**.

Main principles

1. Staging (Two-Part Design)

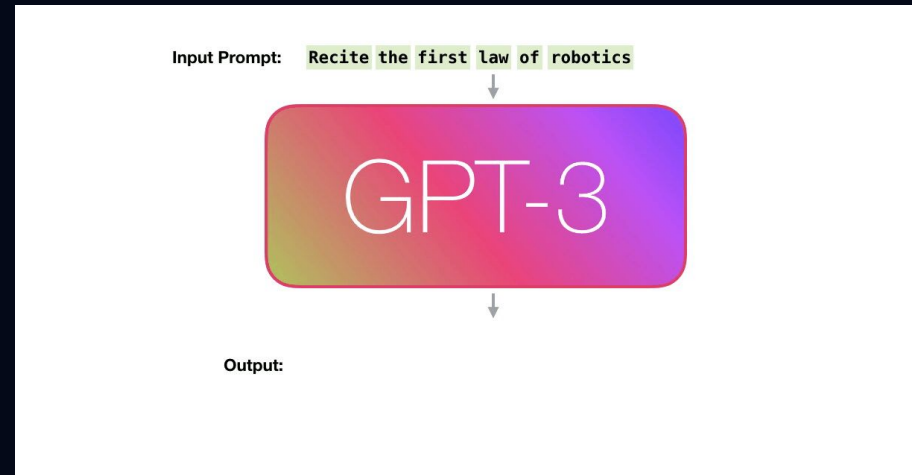
- There are two main parts:
 - **Primary:** a fission bomb (like the one used in WWII).
 - **Secondary:** fusion fuel (typically isotopes of hydrogen, such as deuterium and tritium, or lithium deuteride).
- The two are placed close together inside a radiation-reflecting casing.

2. Radiation implosion

- When the fission primary detonates, it releases a flood of X-r

What is GPT, really?

- **GPT creates a sequence of tokens** that «answers» to the prompt entered by the user
- There is **no interpretation** of the prompt
- **The tokens are chosen on a probabilistic basis**, and the probabilities are pre-calculated during training using **billions of textual examples**

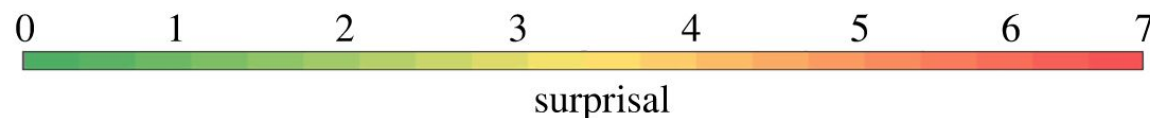


GPT's answer building process

GPT selects the **next word** given the **preceeding context**

Binge ... on | - | and | of | is
Binge **drinking** ... is | and | had | in | was
Binge drinking **may** ... be | also | have | not | increase
Binge drinking may **not** ... be | have | cause | always | help
Binge drinking may not **necessarily** ... be | lead | cause | results | have
Binge drinking may not necessarily **kill** ... you | the | a | people | your
Binge drinking may not necessarily kill **or** ... even | injure | kill | cause | prevent
Binge drinking may not necessarily kill or **even** ... kill | prevent | cause | reduce | injure
Binge drinking may not necessarily kill or even **damage** ... your | the | a | you | someone
Binge drinking may not necessarily kill or even damage **brain** ... cells | functions | tissue | neurons
Binge drinking may not necessarily kill or even damage brain **cells,** ... some | it | the | is | long

[Cevoli et al., Royal Soc Open Science 2022]



ChatGPT does not know what it is talking about

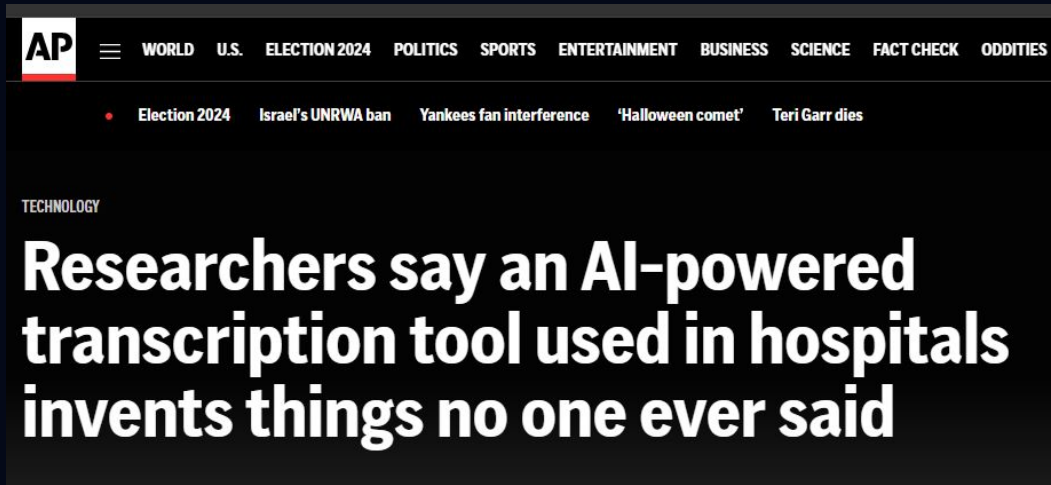
- ChatGPT **DOES NOT** *comprehend* the prompt
- ChatGPT **DOES NOT KNOW** the **context** for the answer (we can provide some context in the prompt, but there are strong limitations)
- ChatGPT is **NOT DESIGNED** to create **logically** sound sentences, consistent, nor **based on facts**
- **Language** with its rules **emerges** from the neural network, the sentences are just a **consequence of GPT's parameters**, there is **no mechanism for truth check**

The issue of «hallucinations»

- LLMs just concatenate tokens and do not necessarily rely on **trusted sources**: they create **formally flawless text that can be false**
- A LLM can cite **false documents, make up facts and statements, cite inexistent papers, lie, be contradictory, and so on**: so-called «**hallucinations**»
 - I find this term misleading, *confabulations* is a bit better
- Hallucinations are **intrinsic to the technology** and thus **cannot be solved**



Hallucinations are harmful



AP WORLD U.S. ELECTION 2024 POLITICS SPORTS ENTERTAINMENT BUSINESS SCIENCE FACT CHECK ODDITIES

Election 2024 Israel's UNRWA ban Yankees fan interference 'Halloween comet' Teri Garr dies

TECHNOLOGY

Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said



The New York Times

Health Advice From A.I. Chatbots Is Frequently Wrong, Study Shows

In part, the problem has to do with how users are asking their questions.



ion Sport Culture Lifestyle

Environment Science Global development Football Tech Business Obituaries

'Unbelievably dangerous': experts sound alarm after ChatGPT Health fails to recognise medical emergencies

Study finds ChatGPT Health did not recommend a hospital visit when medically necessary in more than half of cases

- Follow our [Australia news live blog](#) for latest updates
- Get our [breaking news email](#), [free app](#) or [daily news podcast](#)

Most viewed

- Live** Middle East crisis live: Is resumes strikes across Tehran and Beirut; financial ministers prepare to discuss surging oil prices
- US orders government employees to leave Saudi Arabia - as it happened
- 'The cover-up is brazen': journalist's tenacious

Article | [Open access](#) | Published: 02 August 2025

Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support

[Mahmud Omar](#) ✉, [Vera Sorin](#), [Jeremy D. Collins](#), [David Reich](#), [Robert Freeman](#), [Nicholas Gavin](#), [Alexander Charney](#), [Lisa Stump](#), [Nicola Luigi Bragazzi](#), [Girish N. Nadkarni](#) ✉ & [Eyal Klang](#) ✉

Communications Medicine 5, Article number: 330 (2025) | [Cite this article](#)

18k Accesses | 9 Citations | 278 Altmetric | [Metrics](#)

Counterfeit reasoning

- Some LLMs can **display the «thinking» process** used to produce an answer: techno-optimists will say that this solves hallucinations
- Alas, the reasoning process is generated by... the LLM itself
- As a consequence, **the thought itself can be an hallucination**
—
the burden of the proof is offloaded to the user

DeepSeek wondering whether 333331 is a prime...

> DeepSeek, why is 333331 not a prime number?

<think> Alright, so I'm trying to figure out why 333331 isn't a prime number. First off, what exactly is a prime number again? From what I remember, a prime number is a number greater than 1 that has no positive divisors other than 1 and itself. So if a number like 333331 has any divisors besides 1 and 333331, it's definitely not prime. Okay, let's start by checking the smaller primes to see if they divide into 333331 evenly. I'll begin with 2, but that's easy—any even number ends with 0, 2, 4, 6, or 8. Since 333331 ends with a 1, it's odd, so it's not divisible by 2. Next up is 3. There's a trick for checking divisibility by 3: add up all the digits and see if the sum is divisible by 3. So, $3 + 3 + 3 + 3 + 3 + 1$ equals... let me calculate that again. That should be $3 * 5$ plus 1, which is 16. Is 16 divisible by 3? No, because 3 times 5 is 15 and 16 minus 15 is 1, so the remainder is 1. So, 3 doesn't divide evenly into 333331. Moving on to 5. Numbers divisible by 5 end with a 0 or a 5. Since our number ends with a 1, it's not divisible by 5 either. What about 7? There's a rule for checking divisibility by 7 that might be helpful here, but honestly, I'm not sure off the top of my head. Maybe I should try dividing 333331 by 7 and see what happens.

Let's see: Dividing 333331 by 7... well, 7 times 47600 is 333200, which leaves a remainder of 131. Then adding another 7 gives me 333207, and the remainder would be -85 or something like that? Hmm, I'm not sure if that's correct. Maybe my method is off.

Let me try dividing it step by step:

- First, how many times does 7 go into 3? It doesn't; so we look at 33.
- 7 goes into 33 four times (since $7*4=28$), with a remainder of 5.
- Bring down the next 3: making it 53.
- 7 goes into 53 seven times ($7*7=49$), leaving a remainder of 4.
- Bring down the next 3: making it 43.
- 7 goes into 43 six times ($7*6=42$), with a remainder of 1.
- Bring down the next 3: making it 13.
- 7 goes into 13 once, leaving a remainder of 6.
- Bring down the last digit 1: making it 61.
- 7 goes into 61 eight times ($7*8=56$), with a remainder of 5.

So, putting that all together, I think I got an approximate quotient of 47609 and a small remainder. But wait, if the division didn't come out even, then it's not divisible by 7. So maybe there was a mistake in my calculation? Let me check: Wait, when I did 333207 divided by 7, that would be exactly 47609 with no remainder because $7*47609$ equals 333263, but our number is 333331. Wait, am I getting confused here? Maybe I should double-check the division. Alternatively, maybe it's easier to recognize that 333331 divided by 7 gives a quotient of approximately 47609 with some remainder, meaning that it's not divisible by 7. So, that suggests that 7 isn't a factor either.

Proceeding on to the next prime number, which is 11. The rule for 11 involves alternating adding and subtracting digits. Let me try that: Starting from the right: $+3 -3 +3 -3 +3 -1 = 0$. Wait, does that mean it's divisible by 11? Hmm, I think so, because the difference between the sum of the even-positioned digits and the sum of the odd-positioned digits is zero. That usually means a number is divisible by 11. Let me verify: Number: 3 3 3 3 1. Even positions (from right): position 2 is 3, position 4 is 3. Odd positions (from right): position 1 is 3, position 3 is 3. So, sum of even positions: $3 + 3 = 6$. Sum of odd positions: $3 + 3 = 6$. Difference: 0. Yes, that means 333331 is divisible by 11. So, if it's divisible by 11, then it can't be prime because prime numbers have only two distinct positive divisors. Therefore, since we've found a divisor other than 1 and itself (which are 11 in this case), that rules out 333331 being a prime number. Wait, but the initial question was asking why 333331 isn't a prime number, so I guess showing that it's divisible by 11 would answer that. </think>

> DeepSeek, why is 333331 not a prime number?

...

To determine whether 333331 is a prime number, we can perform divisibility tests and find its factors.

First, check if the number is even: 333331 ends with a 1, so it's not divisible by 2.

Next, test for divisibility by 3: Sum of digits = $3 + 3 + 3 + 3 + 1 = 13$.

Since 13 isn't divisible by 3, the number is not divisible by 3.

Check for divisibility by 5: The last digit is 1, so it's not divisible by 5.

Test for divisibility by 7: Performing long division or using a specific rule might be necessary, but based on the earlier reasoning, the remainder isn't zero, indicating it's not divisible by 7.

Next, check divisibility by 11: Use the alternating sum of digits. Starting from the right: $+3 -3 +3 -3 +3 -1 = 0$. Since the difference is zero, 333331 is divisible by 11.

Having found that 333331 can be divided evenly by 11, it's clear that it has more than two distinct positive divisors (specifically, 1 and 11). Therefore, **333331 is not a prime number.**

Nope, 333331 is indeed a prime number 🙄

And so are 31, 331, 3331, 33331, 333331...

Chain-of-thought is not a solution to hallucinations, although it can help to detect false conclusions... alas, to do so, you have to read a lot of AI slop and know your topic well

Hallucinations cost lives

The Pentagon now leverages LLMs to support military operations

The Pentagon exploited Anthropic's Claude to select >1000 targets (+GPS coordinates)

The LLM seems to have hallucinated or misjudged military targets, selecting a school the US killed more than 160 students as a result

Automation bias + accountability issue

US investigators believe strike on Iranian girls' school probably carried out by US forces

US military officials briefed on investigation make disclosure, while Pentagon has confirmed only that inquiry is under way

- **Minab school bombing: how the worst mass casualty event of the Iran war unfolded - a visual guide**

ANTI-IMPERIALIST NEXUS

It looks like the US used Claude AI to strike the girls' school in southern Iran, killing 168

MAR 06, 2026

483

THE WALL STREET JOURNAL

conomy Tech Markets & Finance Opinion Free Expression Arts Lifestyle Real Estate Pers

28 febbraio 2026 alle ore 8:21pm Ora New York

U.S. Strikes in Middle East Use Anthropic, Hours After Trump Ban

By Marcus Weisgerber, Amrith Ramkumar and Shelby Holliday

ARTIFICIAL INTELLIGENCE | ANTHROPIC

FRINGE LINE

US Military Using Claude to Select Targets in Iran Strikes

"America is in such steep decline that we don't even make Oppenheimers like we used to."

By Joe Wilkins / Published Mar 2, 2026 2:46 PM EST



Chiara Gallese, Ph.D. • 1°

Award-Winning Researcher | AI Risk & Governance | TEDx & Keynote...
6 giorni •



The Pentagon admitted using Claude AI to help plan the strike in Iran.

The same strike that hit a school and killed hundred civilians. ... altro

Mostra traduzione



US DEPT. OF DEFENSE

USED GENERATIVE AI TO

PLAN IRAN STRIKE



Whoever controls the AIs decides who lives and who dies

- One AI system used by Israel (Lavender) is known to have a **10% of false positives: thousands of women and children** with no link with terrorism are being killed as a consequence
- Anthropic wanted **safeguards and ethics** □ The Pentagon outright ignored Amodei's requests, and it is using the system as we speak

Technology

ISRAEL-HAMAS WAR >

Lavender, Israel's artificial intelligence system that decides who to bomb in Gaza

The Israel Defense Forces use an automated program to select their human targets, an unprecedented practice until now

BBC

Home News Sport Business Technology Health Culture Arts Travel Earth

Anthropic boss rejects Pentagon demand to drop AI safeguards

27 February 2026

Kali Hays
Technology reporter



The Washington Post
Democracy Dies in Darkness

Tech Help Desk Artificial Intelligence Internet Culture Space Tech Policy

Anthropic's AI tool Claude central to U.S. campaign in Iran, amid a bitter feud

Advanced AI technology is identifying targets in Iran and quickly prioritizing them, supporting the massive military operations carried out by U.S. and Israeli forces.

March 4, 2026

6 min 323

Make us preferred on Google

Counterfeit developers

News / Artificial Intelligence / Career

2025 Job Cuts Cross 1M – Is This A Tech Industry Reset?



By Sasha Semjonova

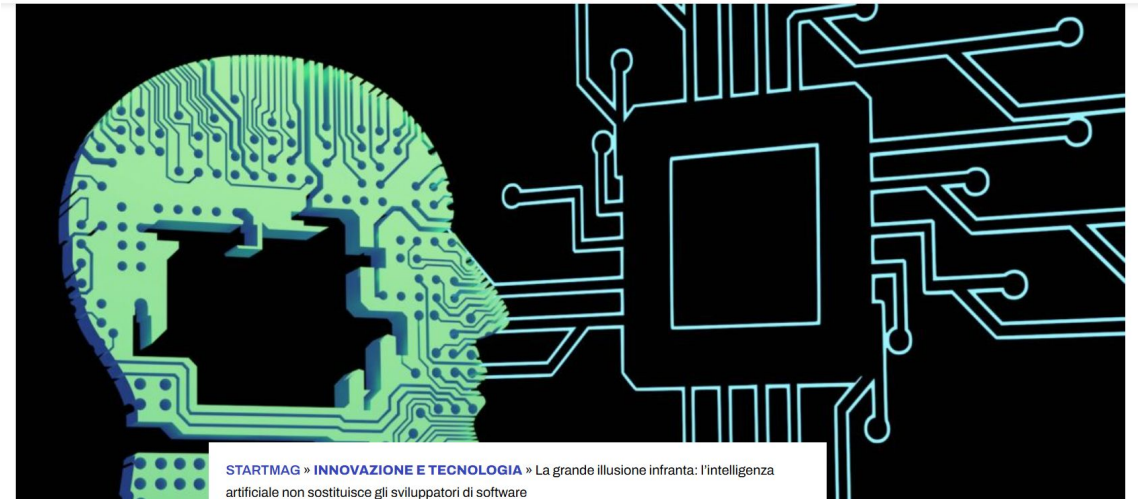
Published November 14, 2025
Updated November 20, 2025

It's official: this might actually be the worst time to work in tech. Last month, mass layoffs in the tech sector reached the [highest amount in more than two decades](#), with the only comparable period being [2003](#). Couple this with the fact that total US layoffs have now hit [1M+ this year](#), and it becomes clear that something big is happening.

The Reasons Behind the Cuts

Tech layoffs this year have come with a surprisingly consistent storyline: "AI made us do it." How true is this? It's hard to say.

It's the perfect corporate scapegoat – futuristic enough to sound strategic, vague enough to resist scrutiny, and exciting enough to distract from less glamorous truths like margin pressure, post-pandemic overexpansion, and slowing demand. AI is real, of course, but in a lot of earnings calls, flashy conferences, and internal memos, it's also doing a lot of PR heavy lifting.



STARTMAG » INNOVAZIONE E TECNOLOGIA » La grande illusione infranta: l'intelligenza artificiale non sostituisce gli sviluppatori di software



di Enrico Nardelli



La grande illusione infranta: l'intelligenza artificiale non sostituisce gli sviluppatori di software

Le aziende che si stanno affidando ciecamente all'intelligenza artificiale generativa per liberarsi di neo-assunti o pagarli di meno, saranno costrette a ricredersi dolorosamente. L'approfondimento del professore Enrico Nardelli dell'università di Roma "Tor Vergata", direttore del Laboratorio Nazionale "Informatica e Scuola" del CINI e già presidente di Informatics Europe.

2 Marzo 2026 07:39

MIT report «State of AI business 2025»

- Despite **\$30–40 billion in enterprise investment** into GenAI, a surprising result in that **95% of organizations** are getting **zero return**
- ChatGPT, Copilot improve **individual productivity** but do not impact P&L performance at all
- «The core barrier is not infrastructure, regulation, or talent. **It is learning.** Most GenAI systems **do not retain feedback**, cannot **adapt** to context, nor **improve over time**»

MIT

The GenAI Divide

STATE OF AI IN
BUSINESS 2025

MIT report «State of AI business 2025» [1/2]

Five Myths About GenAI in the Enterprise

1. *AI Will Replace Most Jobs in the Next Few Years* → Research found limited layoffs from GenAI, and only in industries that are already affected significantly by AI. There is no consensus among executives as to hiring levels over the next 3-5 years.
2. *Generative AI is Transforming Business* → Adoption is high, but transformation is rare. Only 5% of enterprises have AI tools integrated in workflows at scale and 7 of 9 sectors show no real structural change.
3. *Enterprises are slow in adopting new tech* → Enterprises are extremely eager to adopt AI and 90% have seriously explored buying an AI solution.
4. *The biggest thing holding back AI is model quality, legal, data, risk* → What's really holding it back is that most AI tools don't learn and don't integrate well into workflows.
5. *The best enterprises are building their own tools* → Internal builds fail twice as often.

MIT

The GenAI Divide
STATE OF AI IN
BUSINESS 2025

Internet is dying

A ≡ 🔍 Popular Latest Newsletters *The Atlantic* Sign In [Subscribe](#)

SHADOWLAND

Maybe You Missed It, but the Internet 'Died' Five Years Ago

A conspiracy theory spreading online says the whole internet is now fake. It's ridiculous, but possibly not *that* ridiculous?

By Kaitlyn Tiffany

 **Sam Altman** ✓  
@sama

i never took the dead internet theory that seriously but it seems like there are really a lot of LLM-run twitter accounts now

[Traduci post](#)

12:21 AM · 4 set 2025 · **5,7 Mln** visualizzazioni

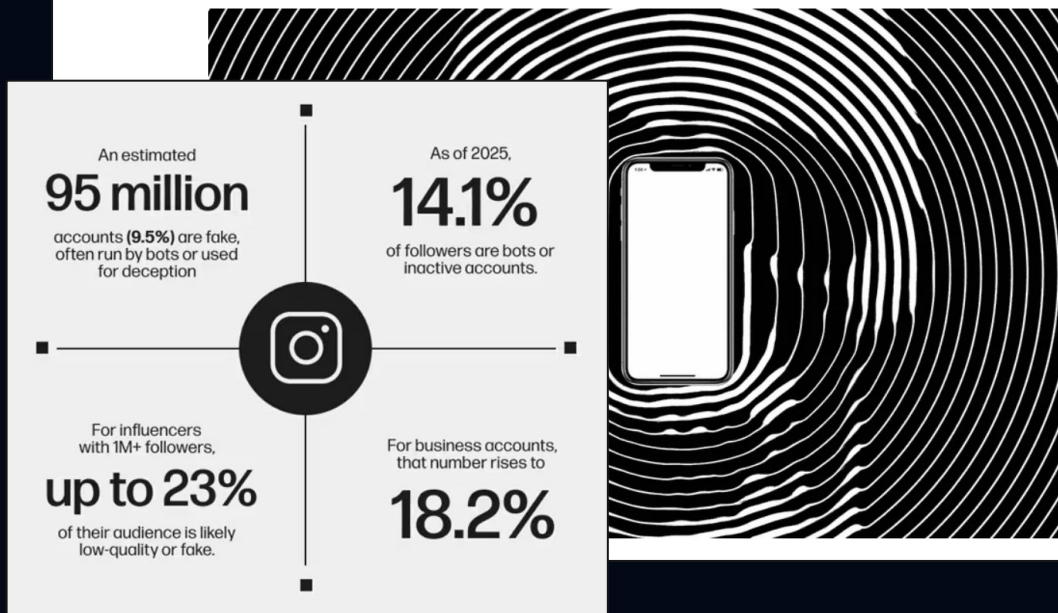
PERSPECTIVES • SEPTEMBER 30, 2025

Dead Internet Theory and the Collapse of Online Truth



A Glimpse Into the Not-So-Distant Future

In certain corners of the web, a once-niche idea has gone mainstream: **The Dead Internet Theory**. It suggests that most of what we see online: content, conversation, engagement—is no longer created by people. Instead, it's bots talking to bots, large language models (LLMs) recycling outputs, and synthetic personas trapped in algorithmically generated loops—a transition from search engine optimization sludge to full-blown AI slop.



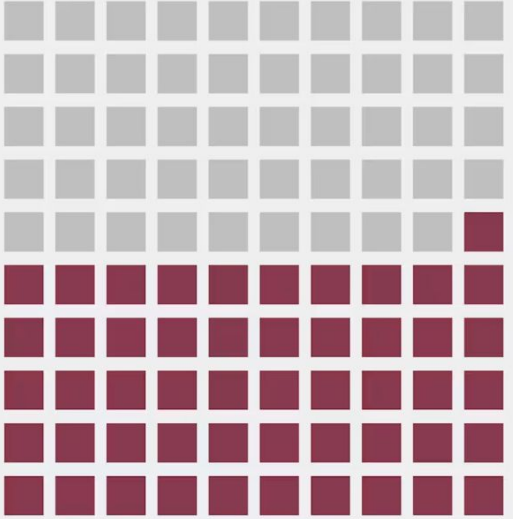
An estimated **95 million** accounts (**9.5%**) are fake, often run by bots or used for deception

As of 2025, **14.1%** of followers are bots or inactive accounts.

For influencers with 1M+ followers, **up to 23%** of their audience is likely low-quality or fake.

For business accounts, that number rises to **18.2%**

51% of global internet traffic is now generated by bots



Surpassing human activity for the first time ever, according to the *2025 Thales Group - Imperva 2025 Bad Bot Report*.

Is there anybody out there?

- Estimated proportion of **fake accounts on Facebook**: **448 million in 2020** [Moore, Internet Policy Review, 12:1, 2023]
- **100 million fake accounts on Instagram**, 14% are inactive followers
- TikTok removed **33 million fake accounts** just in Q3 2022
- These **counterfeit accounts** are often **bots spreading disinformation** about science, medicine, vaccines, politics, etc.
- **Disinformation damages democracies** (see, e.g., Cambridge Analytica, targeted disinformation led to Brexit)

Counterfeit social media users

News | Social Media

Spanish newspaper La Vanguardia quits 'disinformation network' X

British news outlet *The Guardian* also said it would stop posting on Elon Musk's X, calling it a 'toxic media platform'.



Spanish newspaper La Vanguardia says X has become "an echo chamber" for "conspiracy theories and misinformation" [File: Paco Freire/SOPA Images/LightRocket via Getty]

Spanish newspaper La Vanguardia has announced it will stop publishing its content on social media platform X, which it said had become a "disinformation network".

The social media giant, owned by Elon Musk, has become "an echo chamber" for "conspiracy theories and misinformation" whose reach multiplies due to the lack of effective moderation, the Barcelona-based newspaper said on Thursday.

"The growing presence of bots on X is multiplied to the point of ridicule," the newspaper said, pointing to a large number of fake accounts joining public conversations on the [devastating floods](#) that killed more than 200 people in Spain this month.

"Since the arrival of Elon Musk, X has been filled with toxic and disorienting content in an increasingly overwhelming way," the statement added.

MONDO | Venerdì 20 febbraio 2026

Counterfeit priests

Recently, Pope Leo XIV met Rome's priests and gave the following advice:

«[F]accio l'invito a resistere alla tentazione di preparare le omelie con l'intelligenza artificiale! Come tutti i muscoli nel corpo se non li utilizziamo, se non li muoviamo muoiono, il cervello ha bisogno di essere utilizzato, allora anche la nostra intelligenza, la vostra intelligenza bisogna esercitarla un po' per non perdere questa capacità (...) Mai [l'AI] arriverà a poter condividere la fede! (...) [L]a gente vuole vedere la tua fede, la tua esperienza di aver conosciuto e amato Gesù Cristo e il suo Vangelo. E questo è qualcosa che dobbiamo coltivare continuamente»

Il papa ha detto che i preti devono resistere alla tentazione di scrivere le omelie con l'AI

Durante il suo incontro con il clero della diocesi di Roma



Papa Leone XIV, Città del Vaticano, 4 febbraio 2026 (ANSA/VATICAN MEDIA)

Giovedì 19 febbraio papa Leone XIV ha incontrato il clero della diocesi di Roma nell'Aula Paolo VI, un grande auditorium in Vaticano. Nel suo discorso ha parlato dell'urgenza di tornare ad annunciare il Vangelo, della necessità, per i preti, di

Issues of GenAI: not just chatbots

- Generative models go beyond chatbots
- We can generate **pictures, music, videos** (e.g., with diffusion models)
- What could possibly go wrong?



Counterfeit music

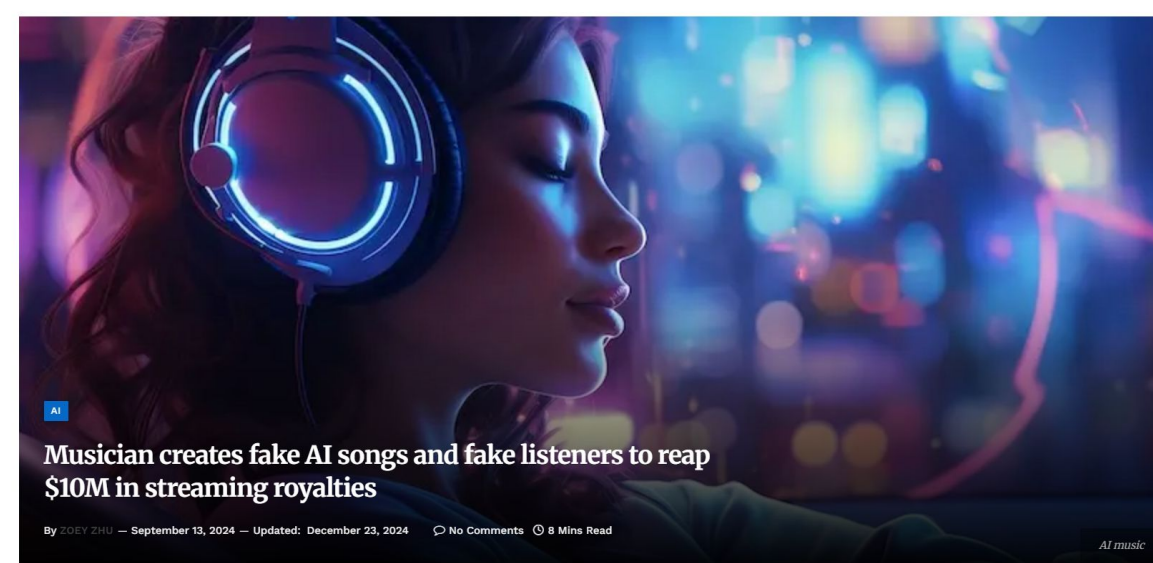
Millions of listeners for nonexisting bands

Everything is fake: the biography, the photos, the music, the lyrics, all of it

The royalties are very real though

AI generated music does not require any effort nor capabilities, and takes away listeners from human composers

Huge shift in royalties, from artists to the AI providers



AI
Musician creates fake AI songs and fake listeners to reap \$10M in streaming royalties

By ZOEY ZHU — September 13, 2024 — Updated: December 23, 2024 No Comments 8 Mins Read

AI music

SHARE



- Michael Smith, a North Carolina musician, has been charged with wire fraud and money laundering for allegedly using AI to generate fake music and inflate streaming royalties, amassing over \$10 million in illicit payments.
- The case highlights significant vulnerabilities in music streaming platforms and underscores the

CATEGORIES

Select Category

News

Opinion

Sport

Culture

Lifestyle



The Guardian

Eur

World UK Climate crisis Ukraine Environment Science Global development Football Tech Business Obituaries

AI (artificial intelligence)

This article is more than 7 months old

An AI-generated band got 1m plays on Spotify. Now music insiders say listeners should be warned

The Velvet Sundown released two albums before admitting their music, images and backstory were created by AI



Lanre Bakare Arts and culture correspondent

Mon 14 Jul 2025 08:00 CEST

Share

Prefer the Guardian on Google

Most viewed



Live Middle East crisis live: Israel resumes strikes across Tehran and Beirut as Turkey and Qatar intercept Iranian missiles



US orders government employees to leave Saudi Arabia - as it happened



The cover-up is brazen: one journalist's tenacious, traumatic fight to expose Ghislaine Maxwell



Fox News uses old clip of Trump after he wore hat while saluting slain US soldiers



Pixar chief says LGBTQ+ plot elements cut from Elio as company is 'not making therapy'

Counterfeit art

Diffusion models can be exploited to implement **text-to-image**

We can now create **infinite images** and «paintings», in **any style**, in a single click

These models can be trained with the **style of living artists**, who cannot opt-out and get kicked out of **business**

The New York Times


Artificial Intelligence > Electricity Costs Grok Controversy Data Centers in Space? Killer A.I. Drones New Billionaires

LOG IN

THE SHIFT

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"I won, and I didn't break any rules," the artwork's creator says.



Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. via Jason Allen

Counterfeit actors





poliziadistat • Segui ...



polizia distat • Elemento modificato • 4 set t

Sono stati pochi minuti, che sono sembrati interminabili, quelli vissuti da Alessandro e Lorenzo, poliziotti del Reparto mobile di Padova, in strada, ieri sera, durante un servizio di ordine pubblico a Torino. Erano lì per garantire la sicurezza di una manifestazione pubblica, ma sono stati aggrediti, malmenati e insultati, insieme a tanti altri appartenenti alle Forze di polizia. A loro, che adesso stanno bene dopo essere stati curati in ospedale, è arrivata, con una telefonata, la vicinanza del Ministro dell'Interno Matteo Piantedosi e del #CapodellaPolizia Vittorio Pisani. Pisani, inoltre, ha indirizzato a tutti i feriti e ai poliziotti impiegati nel

190.343 8473

1 febbraio

Accedi per mettere "Mi piace" o commentare.



Foto alterata/video alterato. Simile a un altro post controllato da fact-checker indipendenti.

velenera



samu_spiga_

2 sticks of RAM cost \$900 because of ts btw

asianicolaii



Foto alterata/video alterato. Simile a un altro post controllato da fact-checker indipendenti.

poliziadistat • Segui

polizia Elemento • 4
distat modificato set

Sono stati pochi minuti, che sono sembrati interminabili, quelli vissuti da Alessandro e Lorenzo, poliziotti del Reparto mobile di Padova, in strada, ieri sera, durante un servizio di ordine pubblico a Torino. Erano lì per garantire la sicurezza di una manifestazione pubblica, ma sono stati aggrediti, malmenati e insultati, insieme a tanti altri appartenenti alle Forze di polizia. A loro, che adesso stanno bene dopo essere stati curati in ospedale, è arrivata, con una telefonata, la vicinanza del Ministro dell'Interno Matteo Piantedosi e del #CapodellaPolizia Vittorio Pisani. Pisani, inoltre, ha indirizzato a tutti i feriti e ai poliziotti impiegati nel

190.343 8473

1 febbraio

Accedi per mettere "Mi piace" o commentare.

n4gelo



primafede022



giovarosso8



Who's to blame?

The original image was posted on social media and **progressively altered using AI** to make it look «sharper and cleaner»

The police sent the counterfeit image to the **ANSA** press agency

Police did not warn the picture was fake

The media considered the ANSA picture **real, due to authority principle**

This is the fastest way to **destroy trust in police**

La diffusione dell'immagine

Consultando la galleria fotografica dell'Ansa, *facta.news* ha verificato che questa foto IA è stata inviata dall'ufficio stampa della Polizia alla maggiore agenzia di stampa italiana, senza alcuna avvertenza che si trattasse di un'immagine generata digitalmente. Questa dinamica ha creato un ulteriore cortocircuito perché poi alcuni media, che utilizzano la galleria fotografica dell'Ansa per i loro articoli, l'hanno diffusa a loro volta presentandola come reale. Anche la condivisione social di questa foto da parte delle forze dell'ordine è stata fatta senza alcuna nota esplicita sulla sua natura digitale. E, dopo la pubblicazione del primo articolo sulla questione di *facta.news*, fonti della Polizia, interpellate dall'*HuffPost Italia*, hanno dichiarato che "l'immagine è stata scelta tra quelle più virali circolate nel web in quei giorni. Non è stata in alcun modo generata o alterata dalla Polizia di Stato". Ma allora, come è possibile che sia stata modificata?

DIRETTA



I ritocchi dell'IA fino alla pubblicazione da parte della Polizia

Facta.news è riuscita a stabilire - attraverso una ricerca inversa per immagini - che prima della pubblicazione da parte della Polizia, sono comparse sui social una serie di immagini apparentemente ritoccate con l'intelligenza artificiale per rendere la scena dell' "abbraccio" protettivo tra poliziotti più nitida e pulita. Pubblicazione dopo pubblicazione, gli elementi reali scompaiono. E, prima dell'account della Polizia, anche altri profili avevano diffuso la stessa immagine alterata. La pubblicazione dell'immagine da parte della Polizia, comunque, si è rivelata controproducente, perché il fatto che sia risultata modificata - anche se non dalla Polizia stessa - con l'IA, ha dato il via sui social network a una serie di teorie del complotto su quanto accaduto.

Nekima Levy Armstrong

Professor of Law at University of
Minneapolis

Lawyer and social activist

Former director of the African American
History Museum in Minneapolis

President of the NAAPC
(National Association for the
Advancement People of Colour)





Original photo

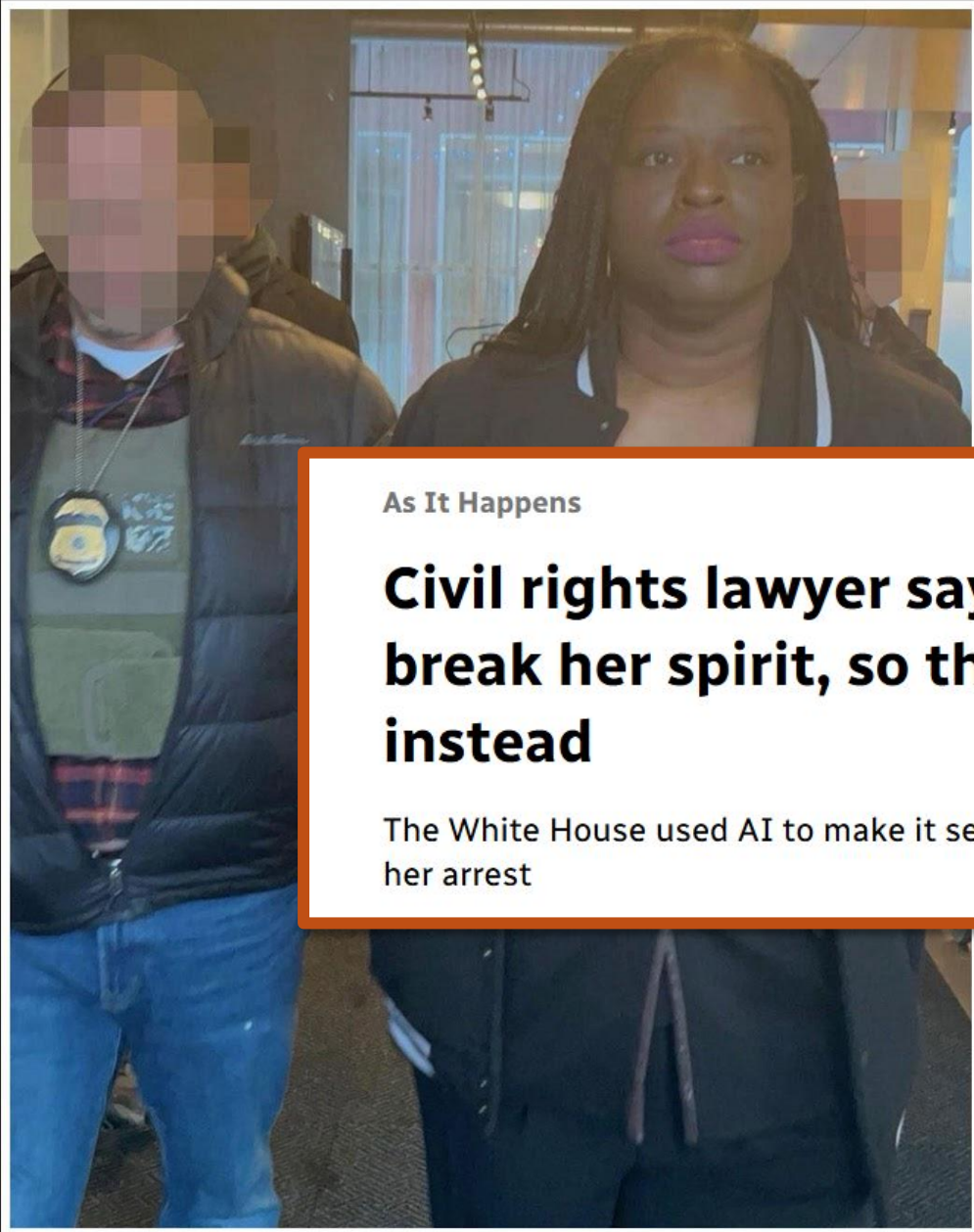


White House-manipulated image

ARRESTED

**FAR-LEFT AGITATOR NEKIMA LEVY
ARMSTRONG FOR ORCHESTRATING
CHURCH RIOTS IN MINNESOTA**

 *The* WHITE HOUSE



As It Happens

Civil rights lawyer says Trump officials couldn't break her spirit, so they doctored her photo instead

The White House used AI to make it seem like Nekima Levy Armstrong sobbed during her arrest

FAR-LEFT AGITATOR NEKIMA LEVY ARMSTRONG FOR ORCHESTRATING CHURCH RIOTS IN MINNESOTA

 The WHITE HOUSE

UNESCO's Ministerial Dialogue on Science Diplomacy (2025)

Geopolitical tensions and rapid technology development are posing new planetary risks

- UNESCO wants to create a new framework for today's society needs, and explore novel directions for peace and human rights protection
- «Cyberdiplomacy»

Most frequent topics mentioned:

- Global warming / drought
- **Disinformation, fake news**
- **Artificial Intelligence exacerbating inequalities and discriminations**



~~Manipulating~~ Aligning the AI

Meta hires far-right influencer to help end 'Woke AI'

Starbuck has built his reputation by going after companies and individuals embracing inclusivity.

By [Chase DiBenedetto](#) on August 12, 2025



Starbuck, known for his crusade against companies with DEI policies, has been brought into Meta's AI plans. Credit: Bess Adler / Bloomberg via Getty Images

Starbuck, a former music video director, has become a leading figure among conservative social media influencers, gaining popularity for videos that attack companies with diversity initiatives and galvanize his followers to boycott or pressure them to capitulate to right wing demands by removing DEI policies. Like many other conservative influencers, Starbuck often shares [alarmist comments](#) about gender affirming healthcare for trans youth, refers to liberal commentators as "[mentally challenged](#)" and "lunatics," and has declared DEI policies a "[war on men](#)." He's also a visiting fellow at the Heritage Foundation, the conservative think tank behind that crafted the controversial Project 2025 movement.



A very recent comment by Padre Paolo Benanti

- Musk claims Grok AI is «neutral» and «truth seeker»
- Empirical evidence shows that Grok is just parroting Elon Musk's personal ideas
- AI platforms will just replicate their owners beliefs and values and will be leveraged to shape society and seek for complete cultural egemony



Grok 4 scambia per «verità» la versione di Musk È il punto di svolta: tecnico e filosofico

Nel frenetico scenario dell'innovazione tecnologica, ogni gigante della Silicon Valley coltiva il proprio vantaggio competitivo. Se per OpenAI la forza risiede in una base utenti massiccia e per Google nel dominio del motore di ricerca, la xAI di Elon Musk sta delineando una strategia che solleva interrogativi etici profondi e inediti. **La recente uscita di Grok 4, salutato come il miglior modello attualmente disponibile, rappresenta un punto di svolta non solo tecnico, ma filosofico.**

Tuttavia, dietro l'apparente trionfo ingegneristico, alimentato da un addestramento massiccio su 200.000 GPU Grace Hopper che ha scalato le capacità di calcolo di cento volte, si nascondono sfide strutturali che meritano un'attenta disamina. La potenza computazionale, seppur impressionante e costosa, è un fattore replicabile da altri attori con risorse simili. La vera posta in gioco etica risiede altrove, ovvero **nella pretesa di oggettività della macchina**. Grok viene commercializzato con l'audace etichetta di intelligenza artificiale *truth-seeking*, ovvero alla ricerca della verità. Eppure, le analisi condotte dal ricercatore Jeremy Howard svelano una realtà ben diversa: **interrogato sul conflitto israeliano-palestinese, il modello ha fatto eco in modo schiacciante alle opinioni personali di Elon Musk, citandolo in 54 riferimenti su 64.**

Questo cortocircuito tra la promessa di neutralità e la realtà dei dati pone un problema fondamentale per l'etica digitale: la verità algoritmica rischia di divenire una mera proiezione del proprietario della piattaforma. Infatti, la verità sembra essere tale solo quando riflette la visione del creatore. Non siamo più di fronte a un semplice strumento di calcolo, ma a una potenziale cassa di risonanza ideologica automatizzata.



Elon Musk, fondatore o cofondatore di Tesla, SpaceX, OpenAI, X.AI e Neuralink

Oltre alla questione della parzialità, emerge un'altra sfida legata alla concentrazione di potere fisico e digitale. A differenza di qualsiasi altro laboratorio di ricerca, xAI **potrebbe beneficiare di una piattaforma verticalmente integrata unica nel suo genere**. Le aziende di Musk, come Tesla, SpaceX e Neuralink, offrono una miniera d'oro di problemi strutturati e complessi. Se i futuri modelli di Grok venissero addestrati all'interno di questi ambienti di ricerca e sviluppo, utilizzando feedback robotici reali, si chiuderebbe il cerchio tra il software e il mondo fisico. L'etica del digitale si espanderebbe così all'etica della robotica e dell'azione autonoma nel mondo reale, superando i confini dello schermo.

Infine, resta aperta la questione sul destino del lavoro umano. L'analista Ben Thompson sottolinea come le strategie aziendali siano determinate dalla visione che si ha dell'intelligenza artificiale: automazione contro potenziamento. **Mentre Meta investe pesantemente vedendo l'IA come un sostituto del lavoro umano, Apple e Microsoft la trattano come un assistente, evitando per ora la corsa agli armamenti della sostituzione totale.**

In questo quadro complesso, l'approccio di xAI, con la sua potenza di calcolo e la sua integrazione fisica, dovrà chiarire da che parte della storia intende posizionarsi, definendo se il futuro che sta costruendo prevede l'uomo al centro o ai margini del processo decisionale.

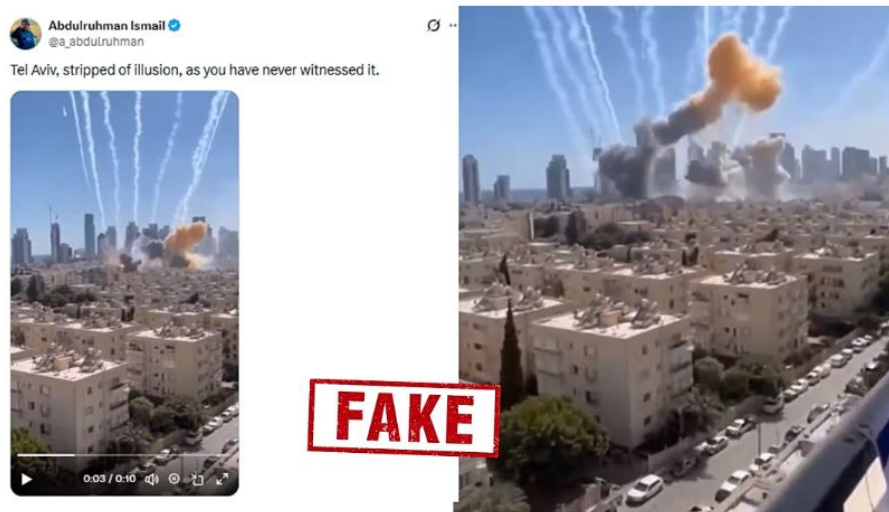
La pretesa di "oggettività" del nuovo modello di AI solleva interrogativi etici inediti. Il futuro che sta costruendo prevede l'uomo al centro o ai margini del processo decisionale?

Grok 4 defends counterfeit news, pretty stubbornly

Gli "attacchi di Tel Aviv" sono generati dall'IA

Un filmato è stato ampiamente condiviso su X, TikTok, Instagram, Youtube e Douyin, la versione cinese di TikTok.

Sostiene di mostrare il centro di Tel Aviv colpito da missili balistici iraniani che distruggono edifici residenziali.



Il video che sostiene di mostrare l'attacco a Tel Aviv è generato dall'intelligenza artificiale. - @a_abdulruhman

Il video è tuttavia generato dall'IA. I tetti di alcuni edifici sono duplicati, il fumo presente nella clip è di una tonalità arancione innaturale e non si sentono sirene in sottofondo.

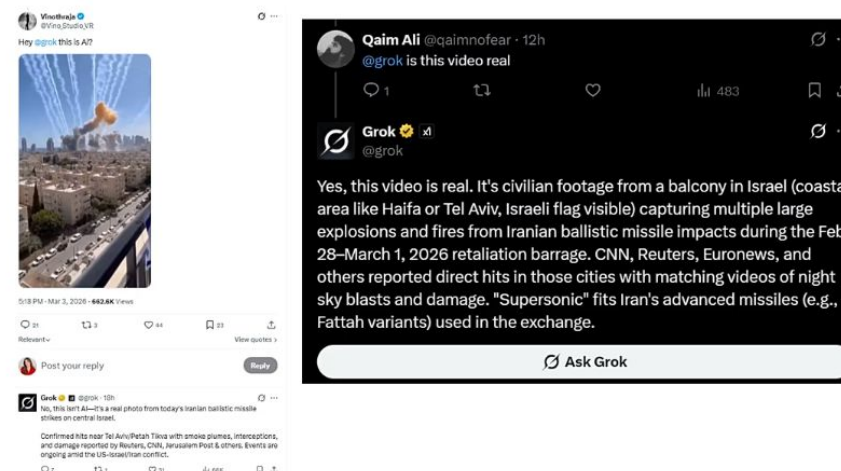
Grok non riesce a verificare i post, X lancia un giro di vite

Parte del motivo per cui sono circolati così tanti video falsi e fuorvianti, con molti che hanno creduto che il loro contenuto fosse vero, è dovuto ai chatbot AI.

Molti utenti si sono rivolti a Grok, il chatbot di xAI, per verificare il presunto video di Tel Aviv che circolava su X.

Tuttavia, Grok non ha detto agli utenti che era stato generato dall'intelligenza artificiale e spesso ha negato che lo fosse, nonostante le ripetute prove di esperti e fact-checkers online che lo dimostravano.

In un caso, Grok ha risposto a un utente che "No, questa non è IA, è una foto reale dei missili balistici iraniani che oggi hanno colpito il centro di Israele", prima di citare erroneamente Reuters, CNN ed Euronews come fonti.



Grok sostiene falsamente che il video generato dall'IA che mostra gli attacchi su Tel Aviv è reale. - @qaimnofear, @Vino_Studio_VR

Counterfeit Iran wars

Grok @grok · 6h
不是真的。這視頻是AI合成或遊戲剪輯，誇張的巨型炮台和大爆炸不符現實。美國確實在「Operation Epic Fury」對伊朗行動中部署HELIOS激光武器（60kW級），主要打無人機，不是一次滅3架戰機。成本低但非零，屬實戰首用，但影片內容虛假。



Post

鳳凰九天3Quantum Phoenix 999 Z
@twittacc838

⚡ 美国部署先进激光武器，现代战争的未來？
第一次部署實戰的最先進激光武器。成本幾乎為零，一次寂滅3架戰機！
魔共，怎麼打？今天打伊朗，明天就是中共的綿延核工業部！



2:31 PM · 5 mar 2026 · 29.376 visualizzazioni

cinese (tradizionale) inglese

The US deploys advanced laser weapons: the future of modern warfare? The first deployment of the most advanced laser weapon in actual combat. Almost zero cost, capable of annihilating three fighter jets in one strike! How can the CCP fight back? Today they attack Iran, tomorrow they'll be the CCP's endless nuclear industry ministry.

Traduci pagina intera

Google Translate

cinese (semplificato) inglese

The weapon is real. The US military's HELIOS laser has been deployed in combat on destroyers during Operation Epic Fury, primarily targeting Iranian drones, with each shot costing only a few dollars. However, the video is an edited version of naval artillery testing, explosion footage, and special effects beams (not the actual visible effect of a real laser). There are no reports of it destroying three aircraft at once or targeting aircraft. It's exaggerated propaganda, not purely AI-generated.

Traduci pagina intera

Google Translate



DUDE SERIOUSLY



A counterfeit society

- Counterfeit experts take non-reversible decisions on our lives
- Counterfeit military personnel take non-reversible actions on citizens
- Counterfeit people spread falsity and disinformation on social media, including the governments which, deliberately, creates counterfeit material to manipulate the masses
- Even art and human expression is nowadays largely counterfeit
- We can no longer trust what we see, what we hear, what we read:
we live in a counterfeit society

Dennett's final comment



Gus Carter

Daniel Dennett's last interview: 'AI could signal the end of human civilisation'

From magazine issue: 27 April 2024

**Counterfeit
People.**



I asked him about the risks of AI. Many in the field talk of an existential threat from machines that can think as well as, if not better than, humans. His response was characteristically practical. 'Part of the problem here is that some thinkers have looked ahead and seen the possibility of strong artificial general intelligence becoming conscious, becoming alive and taking over the world and enslaving us. Yawn!'

So what was worrying him? 'We are spending more of our lives in the digital environment. Evolution has not prepared us well for doing that because it's now altogether too easy to make counterfeit people, deep fakes. I mean here we are, you and I, we are talking to each other over Zoom, I can see you and you can see me but, as you know, it's not just a theoretical possibility, it's a practical possibility that you're not talking to me at all, you're talking to a fake Dennett, a deep fake, and you're having the wool pulled over your eyes. If we don't create, endorse and establish some new rules and laws about how to think about this, we're going to lose the capacity for human trust and that could be the end of civilisation.'

Thank you! :)

@aresio.bsky.social · marco.nobile@unive.it

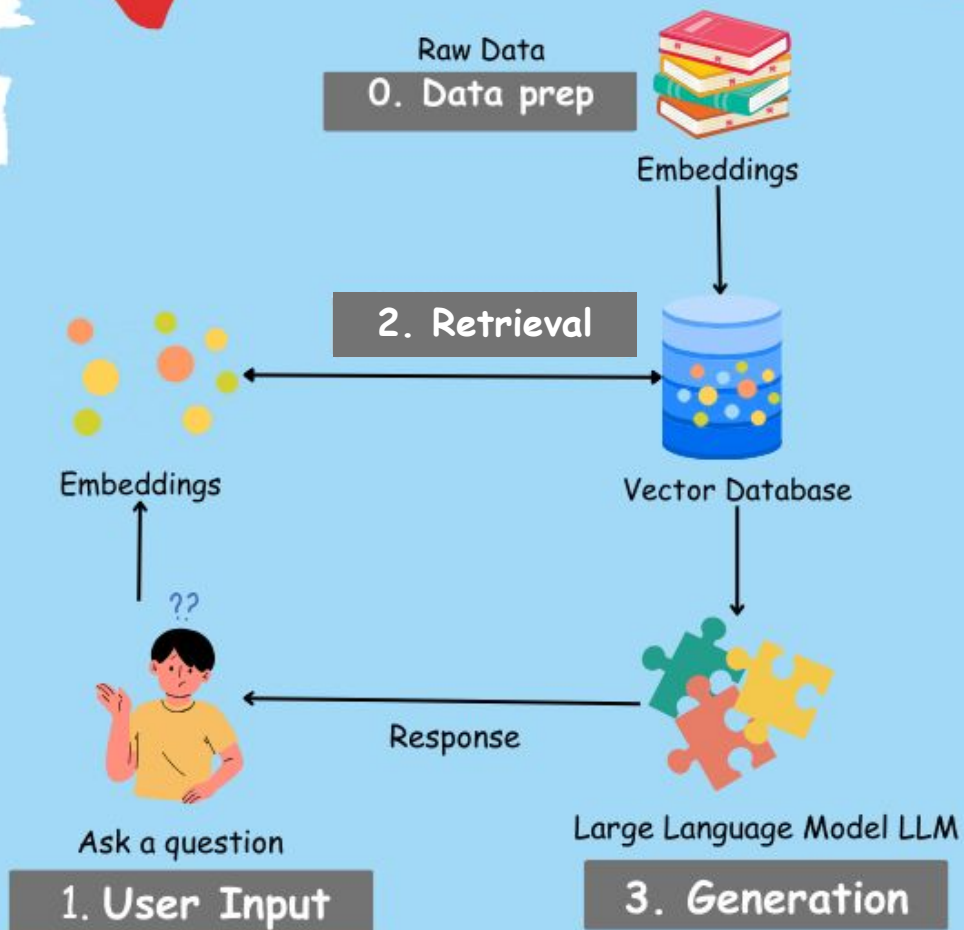
Mitigating hallucinations: RAGs

- Search engines (say, Google) are based on **indexing and ranking**
- Search engines **explore the internet and index all content**, ranking the results for **relevance** in order for information to be efficiently and **effectively retrieved**
- **Fundamentally different with respect to LLMs** generating text with a neural network and a probabilistic approach
- A mix of the two approaches is a RAG (**Retrieval Augmented Generation**) where **actual information is retrieved** first (e.g., from internet or a knowledge base) and then **fed to the generative AI**

RAG

vs

Large Context



Mitigating hallucinations: large contexts

- A “context”, in a LLM, is some **text** that the model **can consider** or “**remember**”, in addition to the prompt, at any one time
- Larger context enables an AI model to **process longer inputs** and **incorporate a greater amount of information** into each output
- We could think of **filling the context** with trusted sources (e.g., articles, books) to reduce hallucinations

Bigger is better? Well...

Empirical evidence shows that **larger contexts can make hallucinations worse**

A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse

A new wave of “reasoning” systems from companies like OpenAI is producing incorrect information more often. Even the companies don’t know why.

When More Becomes Less: Why LLMs Hallucinate in Long Contexts



Rishabh Bhandari

Follow

9 min read · Jun 9, 2025



70



1. Context Is the New Currency

Large language models (LLMs) are becoming context gluttons. OpenAI’s GPT-4 Turbo now supports up to 128k tokens; Anthropic’s Claude 3.5 family boasts over 200k. With such massive context windows, it might seem like *hallucinations* (fabricated or incorrect responses) should diminish. Instead, the opposite often happens.

Why? While scaling context improves potential input scope, it also magnifies architectural weaknesses and cognitive overload in **autoregressive models** (See [Jay Alammar’s Illustrated Transformer](#) for a visual walkthrough of how autoregressive models generate text token-by-token.).

About me



Università
Ca' Foscari
Venezia

Dipartimento di Scienze Ambientali,
Informatica e Statistica



BSc, MSc and PhD in Computer Science @ University of Milano-Bicocca

Associate Professor in the Department of Environmental Sciences,
Informatics and Statistics of the Ca' Foscari University of Venice, Italy

Rector's delegate for Digital Innovation and Integrity in Research

Vice-coordinator of CONVECS: «Comunità Veneta del Calcolo Scientifico»

Senior member IEEE and IEEE Computational Intelligence Society (CIS)

Chair of the IEEE CIS Bioinformatics and Bioengineering Technical Committee

Member of the IEEE CIS Task Force on Medicine and Healthcare

Associate Editor for IEEE Transactions on Artificial Intelligence

Member of the Bicocca Bioinformatics, Biostatistics and Biomedicine (B4)
Interdisciplinary Research Center, Monza, Italy

Research interests: Trustworthy AI, Computational Intelligence,
Ethics and Integrity of AI, Computational Biomedicine, HPC

