

AI (artificial intelligence)

This article is more than 1 year old

## Elon Musk says all human data for AI training 'exhausted'

Tech boss suggests move to self-learning synthetic data though some warn this could cause 'model collapse'

Business live - latest updates



Elon Musk: 'The cumulative sum of human knowledge has been exhausted in AI training. That happened basically last year.' Photograph: Allison Robbert/Reuters

Artificial intelligence companies have run out of data for training their models and have "exhausted" the sum of human knowledge, [Elon Musk](#) has said.

The world's richest person suggested technology firms would have to turn to "synthetic" data - or material created by AI models - to build and fine-tune new systems, a process already taking place with the fast-developing technology.

"The cumulative sum of human knowledge has been exhausted in AI training. That happened basically last year," said Musk, who [launched his own AI business](#), xAI, in 2023.

AI models such as the GPT-4o model powering the [ChatGPT chatbot](#) are "trained" on a vast array of data taken from the internet, where they in effect learn to spot patterns in that information - allowing them to predict, for instance, the next word in a sentence.

Speaking in an interview livestreamed on his social media platform, X, Musk said the "only way" to counter the lack of source material for training new models was to move to synthetic data created by AI.

Referring to the exhaustion of data troves, he said: "The only way to then supplement that is with synthetic data where ... it will sort of write an essay or come up with a thesis and then will grade itself and ... go through this process of self-learning."

Meta, the owner of Facebook and Instagram, has used synthetic data to fine-tune its biggest Llama AI model, while Microsoft has also used AI-made content for its Phi-4 model. Google and [OpenAI](#), the company behind ChatGPT, have also used synthetic data in their AI work.

However, Musk also warned that AI models' habit of generating "hallucinations" - a term for inaccurate or nonsensical output - was a danger for the synthetic data process.

He said in the [livestreamed interview](#) with Mark Penn, the chair of the advertising group Stagwell, that hallucinations had made the process of using artificial material "challenging" because "how do you know if it ... hallucinated the answer or it's a real answer".

Andrew Duncan, the director of foundational AI at the UK's Alan Turing Institute, said Musk's comment tallied with a recent academic paper estimating that publicly available data for AI models [could run out as soon as 2026](#). He added that over-reliance on synthetic data risked "model collapse", a term referring to the outputs of models deteriorating in quality.

"When you start to feed a model synthetic stuff you start to get diminishing returns," he said, with the risk that output is biased and lacking in creativity.

Duncan added that the growth in AI-generated content online could also result in that material being absorbed into AI data training sets.

High-quality data, and control over it, is one of the legal battlegrounds in the AI boom. OpenAI admitted last year it would be impossible to create tools such as ChatGPT [without access to copyrighted material](#), while the [creative industries](#) and [publishers](#) are demanding compensation for use of their output in the model training process.

Explore more on these topics

AI (artificial intelligence) Elon Musk Computing Consciousness ChatGPT news

Dan Milmo *Global technology editor*

Thu 9 Jan 2025 17:14 GMT

Share

Prefer the Guardian on Google

AI-generated 'slop' is slowly killing the internet, so why is nobody trying to stop it?

Arwa Mahdawi



Read more

### Most viewed



A rock star welcome and human rights protests to greet India's prime minister Narendra Modi in Australia



Search under way after Boeing 737 cargo plane goes missing off Pakistan coast



Live Middle East crisis live: Trump says memorandum of understanding with Iran is 'over'



US charges Indian criminal gang leader with organising murder of Canadian Sikh activist



Egypt manager Hassan to switch off World Cup after Argentina 'injustice'

## More from News



### Middle East crisis live

Trump says memorandum of understanding with Iran is 'over' but talks can continue

8m ago



### World Cup 2026

Tearful Messi inspires Argentina's great escape in World Cup thriller with Egypt

14h ago | 446



### France

Marine Le Pen to run for presidency and appeal against conviction in France's highest court

13h ago



### Pakistan

Search under way after Boeing 737 cargo plane goes missing off Pakistan coast

3h ago



### Europe

Greece imposes not in Denial at Na

18m ago



## Most viewed

### Across The Guardian

### In Technology

1

Egypt manager Hassan to switch off World Cup after Argentina 'injustice'

2

**Live** Middle East crisis live: Trump says memorandum of understanding with Iran is 'over' but talks can continue

3

Search under way after Boeing 737 cargo plane goes missing off Pakistan coast

4

From legal threats to 'the worst haircut you can think of': 25 years of The Office

5

ICE agent fatally shoots motorist during traffic stop in Houston

6

Italian village to impose fines of up to €200 on tourists with bare chests or in swimwear

7

Tearful Messi inspires Argentina's great escape in World Cup thriller with Egypt

8

Trump renews call for US to take over Greenland as he arrives for Nato summit

9

Switzerland knock out Colombia on penalties to reach World Cup quarter-finals

10

**Live** World Cup 2026: Egypt fury after Argentina win thriller, Switzerland triumph in shootout - live



## It's your choice

When we make the Guardian available to you online, we and our partners may use cookies and similar technologies to help us to do this. Some are necessary to help our website work properly and can't be switched off, and some are optional but support the Guardian and your experience in other ways. To do this we work with a cross section of 139 [partners](#).

Cookies and other similar technologies may be used to access personal data, including page visits and your IP address. We use this information about you, your devices and your online interactions with us to provide, analyse and improve our services. Depending on your choice, we may also use your data to personalise content or advertising.

### We use cookies and similar technologies for the following purposes:

- ✓ Store and/or access information on a device
- ✓ Personalised advertising, advertising measurement, audience research and services development
- ✓ Personalised content and content measurement

Learn more in our [privacy policy](#) and [cookie policy](#), and manage the choices available to you at any time by going to 'Privacy settings' at the bottom of any page.

### Are you happy to accept cookies?

To manage your cookie choices now, or to opt out where our partners rely on legitimate interests to use your information, [click on manage cookies](#).

Yes, I accept

No, thank you

Manage cookies