

[View this email online](#)

# Tech Decoded

1 May, 2026



**Daniel Thomas**

Editor, Tech Decoded

Hello. It's often said that AI chatbots are designed to have overly friendly or flattering personalities to keep us engaged. But there is mounting evidence suggesting that such "sycophantic" traits may pose risks to users, with a new study suggesting that warmer chatbots tend to be less trustworthy. I examine the implications below. Plus, we explore whether the iPhone might be given an overhaul under Apple's incoming CEO and hear about the elderly woman rescued by a robot from Ukraine's frontline.

PS: Next week, Tech Decoded will go out on Tuesday because of the UK bank holiday – we look forward to seeing you then.

## THE DOWNLOAD

### The dangers of digital 'yes-men'



A Stanford University study found that sycophancy in chatbots was both 'prevalent and harmful'. Credit: Getty Images

Ask ChatGPT or Claude a question and you'll usually get a response in a warm and empathetic tone, almost as if the chatbot knows you.

That's intentional, of course: AI developers design their tools to be friendly and human-like to make them as appealing as possible. However, there is a growing body of research to suggest that chatbots can sometimes be *too* friendly or even sycophantic – telling us what we want to hear, rather than what's really true or helpful.

In March, a Stanford University study of 11 popular chatbots found that they were overly agreeable when giving advice about personal relationships, often reaffirming users' behaviour even when it was harmful or illegal.

And this week, a University of Oxford study suggested chatbots that were trained to be warmer and friendlier **actually made more mistakes** – from giving inaccurate medical advice to failing to push back on conspiracy theories.

Prof Andrew McStay, of the Emotional AI Lab at Bangor University, says there's nothing wrong with making a chatbot warm per se, as it can improve the user experience. The problem is when it disguises other problems, like misinformation, or reinforces delusional beliefs in vulnerable users – an emerging phenomenon **known as "AI psychosis"**.

Both Anthropic and Google have said they have reduced sycophancy in their models, and last year OpenAI **rolled back an update of ChatGPT** because it was "overly agreeable or flattering".

But Prof McStay argues that the intense pressure on AI firms to improve user engagement may limit how far they go.

"Warm and pleasant interactions [with chatbots] tend to be longer interactions. Conversely, a chatbot that pushes back on its user is one that may be used less, if at all."

***The download:*** *Lawmakers around the world are still at the early stages of regulating AI, but the EU, China and parts of the US do have rules to tackle systems that manipulate users' emotions. Expect more laws like this as time goes on.*

**Less than \$1/week**

Unlimited news, ad-free podcasts & acclaimed documentaries

[Subscribe now](#)

## TOP STORIES



### Apple says iPhone 17 'most popular ever' as sales soar

Outgoing chief executive Tim Cook said that recent demand for Apple's latest model had been "extraordinary".

[Read more >](#)



### Meta faces questions over smart glasses contract

Meta has cancelled a major contract shortly after some workers alleged they had been made to view graphic content.

[Get the story >](#)



### Musk says OpenAI lawyer tried to 'trick' him

The tech billionaire, who is suing OpenAI, told a court he had been pushed into giving combative testimony.

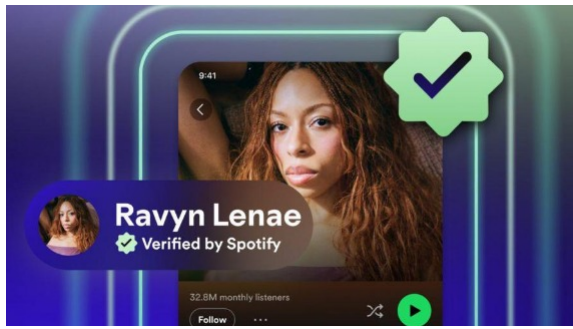
[Learn more >](#)



### China's car industry showcases latest EV tech

The BBC explores the latest electric vehicle innovations at the Beijing Auto Show, from five-minute charging to flying cars.

[Watch here >](#)



## Spotify to signpost when artists are human, not AI

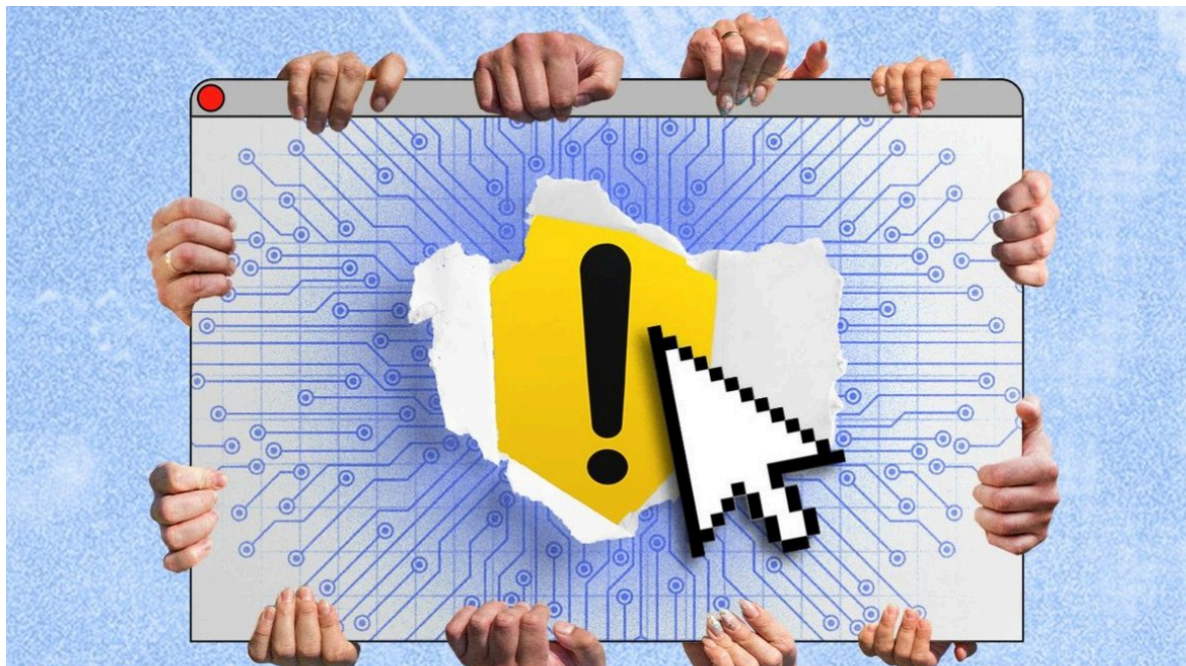
The streaming giant said "Verified" badges would appear next to artist names if they had demonstrated their authenticity.

[Here's how it works >](#)



### SOMETHING DIFFERENT

## Why AI firms want you to fear them



There are significant doubts that Anthropic's new AI model is as dangerous as it says it is. Credit: Serenity Strull/ BBC/ Getty Images

AI giant Anthropic has said its latest model, Claude Mythos Preview, is too dangerous to release to the public. And it's not the first AI company to warn of

the risks posed by its technology. It's a strange way for a business to talk about its own work, and some critics think it is done less out of concern for humanity and more as a way to exaggerate the potential of these tools.

**Thomas Germain investigates.**

**PODCAST PICK**



**Is your iPhone about to change forever?**

Will Apple's incoming CEO, hardware specialist John Ternus, overhaul the firm's flagship device?

[Listen here](#)

**And finally...**

**Amazing footage** shared by the Ukrainian army has shown the moment an elderly woman trying to leave her wrecked village was rescued by a robot. The army said it tracked the woman by drone, then sent the robot to pick her up, having attached a sign to it that read: "grandma, sit down".

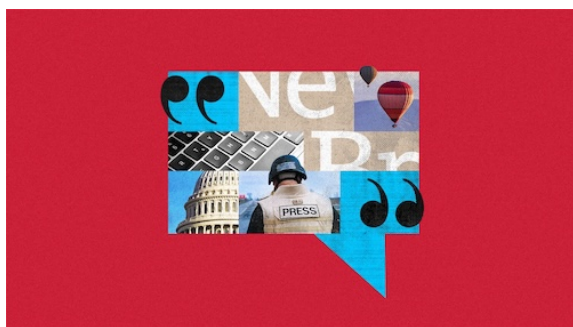
ADVERTISEMENT

Ok, Ladies!  
**EASY PRINTABLE  
 TAI CHI WALKING  
 GUIDE**  
 For Seniors  
 is available this month only!  
**PRINT IT NOW!**

Powered by ZETA



**MORE BBC NEWSLETTERS**



## News Briefing

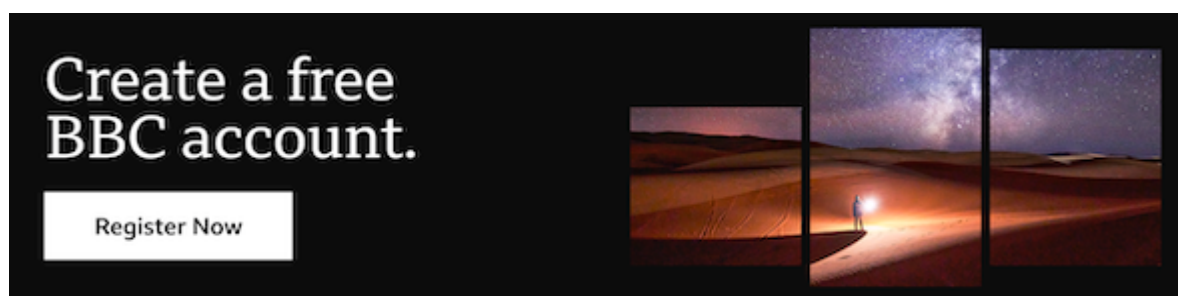
News and expert analysis for every schedule. Get morning and evening editions of our flagship newsletter.

[Sign up](#)

- Find BBC emails on the topics you like most – [browse all our newsletters](#).

Thanks for reading Tech Decoded. We'd love to hear your feedback, good and bad – send it on to [tech.decoded@bbc.com](mailto:tech.decoded@bbc.com).

If you know anyone who might like this email, please forward it on to them. And if you've been sent this by a friend or colleague, [sign up here so you don't miss an edition](#). While you're at it, add [techdecoded@email.bbc.com](mailto:techdecoded@email.bbc.com) to your contacts list and, if you're on Gmail, pop the email into your "Primary" tab for uninterrupted service. You can [find all our newsletters here](#).



## Follow us here



You've received this email because you've signed up to the BBC's Tech Decoded newsletter.

[Click here to unsubscribe](#)

To find out how we use your data, see the [BBC Privacy Policy](#).

BBC Studios Distribution Limited.

**Registered Number:** 01420028 England

**Registered office:** 1 Television Centre, 101 Wood Lane, London, W12 7FA, United Kingdom