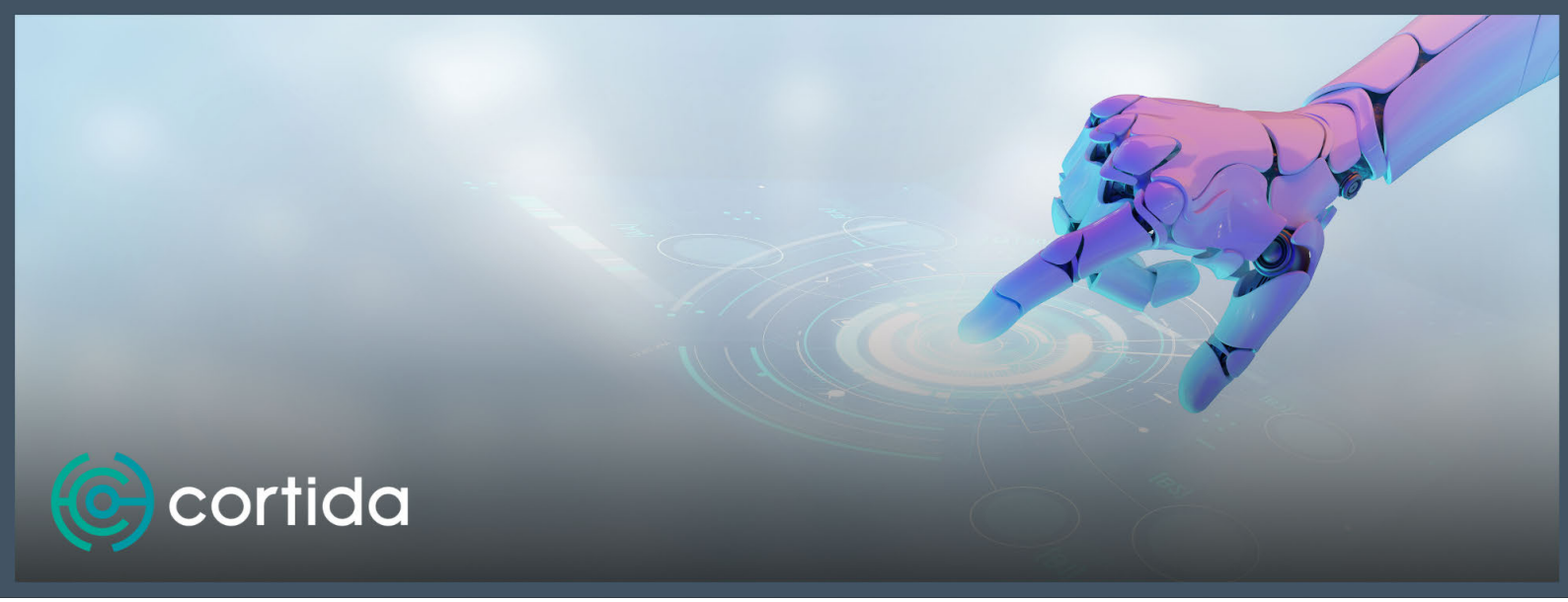


# ChatGPT: a double-edged sword?

*A recent report claims that artificial intelligence models could boost global GDP by 7%.[1] However, Elon Musk has called for a six-month halt in testing amid fears of a threat to humanity.[2] And Italy has banned it.[3] In the first in a series of articles, we sort the hype from the happening around generative AI and ChatGPT in particular, plus examine the security implications.*

Artificial intelligence, or AI, is essentially making computers do what minds do. The sub-set of generative AI uses algorithms to create new content, such as text, code, images, audio and video. ChatGPT – the GPT stands for generative pretrained transformer – is one such AI chatbot developed by OpenAI, a US tech start-up.

When it was launched in November 2022, ChatGPT-3 was considered to be the best chatbot yet. Users could ask questions in plain text or input images and the chatbot generated convincing-sounding content.

That's because such large language models (LLMs) are trained on huge amounts of text-based data. Algorithms analyse the relationships between words and turn this into a probability model. When queried, a chatbot answers based on the relationships of the words in its model.

**Data quality and assurance**

The machine-learning element within LLMs means that models 'learn' from data patterns without human direction. So, while the model can update itself in response to things it has done and outcomes it has observed. It can also hallucinate – invent seemingly plausible responses, be gamed into creating toxic content, and amplify existing biases.

At Cortida, we'd recommend monitoring data quality to avoid skewing the training models. Similarly, monitoring outcomes to help prevent models 'drifting' or providing inaccurate answers or those which reinforce existing inequalities or exclusions.

**Data security and privacy**

Who is asking and when, could also be sensitive, depending on context. For example, disclosures that a chief executive queried how to cut 30% of the workforce could affect the company share price and reputation as well as staff morale and productivity.

How data is stored, processed and transmitted is clearly as important for chatbots as for other business processes. This is irrespective of whether the chatbot is used internally among colleagues or externally with customers.

The tried-and-true principles of assessing and understanding what data is held, where it is, who has access to it and how it is protected apply as much to emerging tech as the established ones and are a good starting place for those considering adoption of these exciting new capabilities.

**Cortida view on AI**

Technology is morally neutral – for now, at least. So, artificial intelligence will increasingly be used with both good and bad intent in the future. The direction of travel is clear, though.

Cybercriminals will use AI for routine attacks initially, such as writing phishing e-mails and guessing passwords. This is before they create

more sophisticated attacks, using embedded malware to circumvent defensive controls and devise exfiltration routes, depending on the countermeasures in place. Over time, AI will become more automated and dangerous in the hands of criminals.

However, the upside benefits of AI to companies will also increase. These include better customer segmentation and targeting with hyper-personalised messages and content, better root-cause analysis of customer problems, better self-service, risk identification and decision-making in real time.

The use of generative AI, and ChatGPT in particular, in a corporate environment should be carefully considered and controlled. Plus be supported by a formal risk assessment plan, privacy and data strategy.

For more information on Cortida embedded consultant services please get in touch: info@cortida.com

[1] 'Goldman Sachs claims AI could increase annual global GDP by 7%', Investing.com, 27 March 2023, https://uk.investing.com/news/stock-market-news/goldman-sachs-claims-ai-could-increase-annual-global-gdp-by-7-432SI-2964837

[2] 'Elon Musk among experts urging a halt to AI training', BBC news, 31 March 2023, https://www.bbc.co.uk/news/technology-65110030

[3] 'ChatGPT banned in Italy over privacy concerns', BBC news, 01 April 2023, https://www.bbc.co.uk/news/technology-65139406

[4] 'ChatGPT outage: here's what happened', OpenAI blog post, 24 March 2023, https://openai.com/blog/march-20-chatgpt-outage

+44 (0) 207 164 669
info@cortida.com
www.cortida.com
The Shard, 32 London Bridge Street, London SE1 9SG