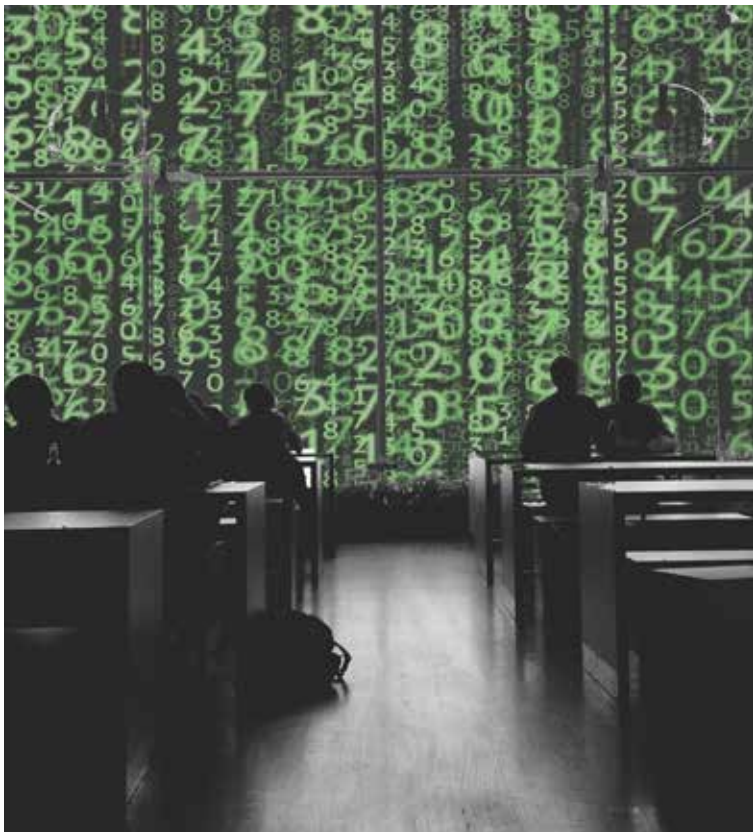
FAKING IT

Chris Newton-Smith *reports on worrying trends in weaponised artificial intelligence*

Businesses are under increased threat from weaponised artificial intelligence. More than one in four surveyed organisations in the UK and US have fallen victim to AI data poisoning in the past year, wherein hackers corrupt the data that trains AI systems, planting hidden backdoors, sabotaging performance or manipulating outcomes to their advantage. The consequences are far-reaching, and poisoned models can quietly undermine fraud detection, weaken cyber defences and open the door to large-scale attacks, putting both businesses and the public at risk.



The IO State of Information Security Report, conducted among 3,001 cyber security and information security managers in the UK and USA, worryingly found that 20 percent of organisations also reported experiencing deepfake or cloning incidents in the last 12 months. In line with this, 28 percent of respondents highlight deepfake impersonation in virtual meetings as a growing threat for the next 12 months, showing how AI is increasingly being weaponised to target people directly and undermine trust in everyday business interactions.

Beyond deepfakes, AI-generated misinformation and disinformation tops the list of emerging threats for the next 12 months, cited by 42 percent of security professionals concerned about scams and reputational harm. Generative AI-driven phishing (38 percent) and shadow AI misuse are also on the rise, with more than a third (37 percent) of respondents reporting that employees use generative AI tools without permission or guidance, creating risks of data leaks, compliance breaches, and reputational damage.

Shadow IT in general – downloading or accessing unapproved software or services – is already an issue for 40 percent of organisations, and generative AI is exacerbating the problem – especially when it is used without human oversight. 40 percent of those who are currently facing challenges in information security cited tasks being completed by AI without human compliance checks as a key challenge. If businesses are not fast enough to address this problem, employees may well continue to find insecure workarounds and shortcuts, putting sensitive data at risk.

AI has always been a double-edged sword. While it offers enormous promise, the risks are evolving just as fast as the technology itself. Too many organisations rushed in and are now paying the price. Data poisoning attacks, for example, don't just undermine technical systems, but they threaten the integrity of the services we rely on. Add shadow AI to the mix, and it's clear we need stronger governance to protect both businesses and the public.

AI adoption has surged, and more than half of organisations (54 percent) admit they deployed the technology too quickly and are now struggling to scale it back or implement it more responsibly. In line with this, 39 percent of all respondents cited securing AI and machine learning technologies as a top challenge they are currently facing, up sharply from 9 percent last year. Meanwhile, 52 percent state that AI and machine learning are hindering their security efforts. Although the statistics show that AI may not yet be on the side of the defender, encouragingly, 79 percent of UK and US organisations are using AI, machine learning or blockchain for security, up from just 27 percent in 2024. A further 96 percent have plans to invest in GenAI-powered threat detection and defence, 94 percent will roll out deepfake detection and validation tools, and 95 percent are committing to AI governance and policy enforcement in the year ahead.

The UK's National Cyber Security Centre has already warned that AI will almost certainly make cyber attacks more effective over the next two years, and our research shows businesses need to act now. Many are already strengthening resilience, and by adopting frameworks like ISO 42001, organisations can innovate responsibly, protect customers, recover faster and clearly communicate their defences if an attack occurs ●

AI-generated misinformation and disinformation tops the list of emerging threats for the next 12 months

Chris Newton-Smith is CEO of IO.