UK Resilience Academy

Resilience in the Age of AI

UK Resilience Academy Insight Paper

Di Cooke
UKRA Associate
May 2026

Foreword

Resilience today is cross-system. It's physical and digital, local and national, public services, private enterprise, and civil society, but resilience doesn't begin with systems, it begins with people: the judgement of the person in the room, the decisions made under pressure, the ability to work together when minutes matter.

In just twelve months, the UK Resilience Academy has already proved quietly transformative, bringing responders, leaders, institutions together, to stress test decisions before real-world pressure hits, and turn hard lessons from past emergencies into capability for the next one. That's why this report, *Resilience in the Age of AI*, is so timely.

The Government has been clear that safety and innovation go hand in hand. Every time the UK has embraced transformative technology, whether steam power or the internet, we have unlocked growth, created industries, and improved lives. If we are to unlock the full potential of AI adoption in Britain, we need to know that we can trust systems to work as intended, and to rely on them for critical applications.

I'm proud of Britain's world leading strengths in the AI Security Institute, which is evaluating the capabilities of frontier models, helping developers identify problems early, and equipping government with the knowledge to develop the policy and risk mitigations that will keep Britain safe.

Central to our resilience is our ability to meaningfully influence the development of the models, data, and AI infrastructure that are increasingly present in our lives. That's why we're investing in sovereign capability and backing exceptional British companies and founders with capital and compute to start up, scale up, and win globally. This isn't about shutting the world out, it's about reducing over dependencies, strengthening resilience, and ensuring that Britain has a sovereign edge, and that we compete fiercely in the areas where we have real strengths.

Resilience in the age of AI will be defined not just by the technology itself, but the choices we make about how it is developed, governed, and used. If we get this right, AI has the potential to grow the economy, create novel solutions to our most challenging problems and deliver huge societal benefits. This Government is making deliberate choices to realise that vision of the future for British people across the country.

Kanishka Narayan MP
Minister for AI and Online Safety

A Defining Moment

We are entering a defining era of transformation, where the nature of risk itself is being fundamentally reshaped. The UK's risk landscape is becoming more complex, more interconnected, and increasingly difficult to anticipate, shaped not only by accelerating technological change, but by deeper questions of control, influence, and accountability. Among the forces driving this shift, Artificial Intelligence (AI) stands out for both its transformative promise and its disruptive potential, raising fundamental questions about who governs these technologies, and on whose terms.

This is not a distant or abstract challenge. AI represents a seismic shift in how threats are generated, how information is shaped and contested, and how critical systems can be exploited. As these dynamics evolve, so too must our approach to resilience, requiring a step change in how we anticipate risk, build capability, and prepare society for an increasingly uncertain future.

The government has recognised the need to respond. Resilience sits at the heart of how we navigate this moment, informing how we anticipate threats, prepare our institutions and communities, and build the capacity to adapt and recover when disruptions strike. The 2025 National Security Strategy positioned resilience as a pillar of national security, and the 2025 Resilience Action Plan set out a whole-of-society approach to strengthening the UK's preparedness, response, and recovery capabilities.

As the UK Resilience Academy approaches its first anniversary, that mission has never felt more urgent, nor more consequential. Here, we examine what AI means for the future of resilience, from the opportunities it offers and the risks it demands we confront along with the UKRA's response to this complex challenge.

We recognise that we are only at the beginning of what will be an ever-accelerating pace of change. How we respond to it, the threats we prepare for, and the opportunities we choose to seize, will define the resilience of our nation for a generation.

AI as an Enabler of Resilience

AI technology offers considerable opportunities for strengthening resilience, if harnessed responsibly and with appropriate safeguards put in place. Across government, industry, and civil society, there is growing recognition that the ability to develop, govern, and deploy AI capabilities is becoming central to economic competitiveness, strategic resilience, and public trust. These potential applications span the full resilience cycle:

Before disruption strikes: AI can draw on both historical and live data to flag emerging weaknesses across supply chains, infrastructure, and environmental systems. It can run scenario exercises against preparedness plans, surfacing gaps that conventional review might miss. AI-powered forecasting tools can process vast quantities of meteorological and environmental data to provide earlier and more accurate warnings of natural disasters, giving authorities more time to plan evacuations and pre-position resources. And continuous automated monitoring can help organisations stay on top of security and resilience standards rather than relying on periodic review.

During a crisis: AI can spot unusual patterns and respond to threats as they unfold, bring together large volumes of information to build a clearer operational picture under pressure, and support the direction of resources to where they are most needed. Automated communication tools can handle high volumes of public enquiries and help authorities get timely, accurate information to affected communities.

After disruption: AI can accelerate recovery by working through incident data to identify what went wrong, assessing damage across affected areas to help determine where rebuilding efforts should be concentrated, and feeding lessons back into preparedness planning so that future responses are stronger.

The UK government is already applying AI capabilities across several of these domains. The Environment Agency and Met Office are investing in AI-driven modelling to strengthen flood prediction and environmental monitoring, a critical capability given that 6.3 million properties in England are currently in areas at risk of flooding (UK Government 2025a). Meanwhile, the National Cyber Security Centre's Active Cyber Defence programme uses automated tools to identify and counter cyber threats at scale, helping to protect critical national infrastructure and public services from increasingly sophisticated attacks (NCSC 2025). In addition, through the Alan Turing Institute's AI fellows programme, the government is deploying AI models to analyse images and video of transport infrastructure, helping local authorities identify deterioration and prioritise maintenance before failures occur (DSIT and Alan Turing Institute 2026).

The government has also made clear its intention to further harness AI to bolster UK resilience, with the AI Opportunities Action Plan setting out a commitment to expand the use of AI across public services and critical national infrastructure to strengthen resilience and improve outcomes in the years ahead (DSIT 2025)

Crucially, realising these benefits requires more than simply adopting the technology. It demands careful consideration of the ethical dimensions of AI use in public-facing and emergency contexts. Communities need confidence that AI is being used transparently, that human oversight is maintained in critical decisions, and that its deployment does not inadvertently deepen inequalities or erode civil liberties. The responsible use of AI in resilience must be guided by clear principles, robust governance, and a whole-of-society approach. The UK government has recognised this need, establishing cross-sectoral principles for safety, transparency, fairness, accountability, and contestability through its 2023 AI regulation white paper, and publishing an updated Data and AI Ethics Framework in 2025 to guide the responsible adoption of AI across the public sector (DSIT 2023; GDS 2025).

The potential is clear, but so too is the responsibility. Realising AI's full contribution to resilience will depend not only on the sophistication of the technology we adopt, but on the strength of the governance, oversight, and public trust that underpin it.

AI as a Source of Risk

Yet harnessing AI to strengthen resilience is only possible if we are equally serious about understanding and addressing the threats posed by the technology as well. These harms are real, growing, and in several cases already being felt. Not confined to any single sector or domain, they have the potential not only for direct harm but to cascade across systems, institutions, and communities in ways that are difficult to predict and harder to contain.

The risks have been organised into three broad categories: the deliberate misuse of AI as a tool for harm; the possibility of AI systems failing or acting in ways that were never intended; and the deeper, structural risks that emerge from AI's growing role across every dimension of modern life.

Weaponised AI Misuse

By far the most immediate and already occurring harms arising from AI stem from the deliberate exploitation of AI capabilities for harmful purposes. Criminals, fraudsters, and hostile state actors are using AI to conduct attacks at a scale, speed, and level of sophistication that was unimaginable only a few years ago. The technology also lowers the barrier to entry for would-be attackers, widening the pool of threat actors considerably.

Compounding this is the technology's ability to manufacture synthetic content so convincing that people's ability to distinguish it from authentic media is now no more accurate than a coin flip (Cooke et al 2024). This means that relying on the observer as a first line of defence is no longer a viable strategy to defend against misuse.

AI-driven financial fraud is one of the most quickly growing areas of misuse, with fraud attempts in the UK nearly doubling year on year and over a third of UK businesses reporting being targeted by AI-related fraud in early 2025 (Sumsub 2025). Several of these incidents have been highly sophisticated and have resulted in substantial losses. For instance, in 2024 an employee at a British company was deceived into transferring $25 million after criminals used AI-generated videos to impersonate senior colleagues on a conference call. A quarter of all scam calls in the UK in 2025 were found to have been enabled by AI, with impersonators mimicking HMRC officials, bank staff, and family members in distress (Hiya 2025). New evidence has also emerged of criminals using AI voice cloning to set up unauthorised direct debits, with National Trading Standards operations blocking nearly 21 million scam calls in a six-month period (National Trading Standards 2026). Concerningly, it has been found that older members of the public, who are disproportionately targeted by these schemes, are particularly vulnerable, as research has found that detection accuracy declines significantly among more senior demographics (Cooke et al. 2025).

The cyber threat landscape is undergoing an equally significant transformation, with attackers now using AI to augment multiple stages of the attack chain (NCSC 2025). Over 80% of phishing emails detected in late 2024 and early 2025 were found to have used AI-generated content, achieving a higher success rate than their manually written counterparts (Keepnet Labs 2026). The NCSC's 2025 Annual Review reported 429 incidents requiring its support, with the most serious incidents rising 50% for the third consecutive year. In 2025, Anthropic documented what it assessed to be the first publicly known cyberattack largely executed by an AI agent at scale, with AI performing an estimated 80 to 90% of the operation (Anthropic 2025b).

Beyond financial crime and cyber operations, AI is being deployed as an instrument of influence and political manipulation at a scale and speed that existing defences are not equipped to match. AI-generated content has also featured in election campaigns in over 80% of countries that held elections in 2024 (IPIE 2025), and the threat is only growing. For instance, Hungary's April 2026 parliamentary election was marked by an extensive campaign of AI-generated videos and disinformation targeting opposition candidates, described by analysts as among the first elections to feature such a widespread deployment of AI-generated political content (France 24 2026). Foundation model companies including OpenAI and Anthropic have also reported state-linked threat actors using their models to generate social media content, automate engagement through networks of bot accounts, and orchestrate influence-as-a-service operations that make real-time tactical decisions about when to like, share, or comment on posts in order to amplify politically motivated narratives (OpenAI 2024; Anthropic 2025a). The Iran conflict of 2026 provides another recent illustration, with pro-Iranian groups producing AI-generated propaganda videos that gained millions of views worldwide (DiResta 2026).

Non-consensual intimate media abuse has also emerged as one of the most widespread and damaging applications of AI technology, increasing in prevalence by 1,780% between 2019 and 2024 (National Police Chiefs' Council 2025). The Internet Watch Foundation reported a 380% year-on-year increase in confirmed cases of AI-generated child sexual abuse imagery between 2023 and 2024 (IWF 2026). Teachers across England have reported pupils using nudification apps to create fake sexual images of classmates, with most incidents involving girls aged 14 or under, and a survey found that 13% of teenagers had experienced nude AI-generated media in British schools (Children's Commissioner 2025; Internet Matters 2024). The UK has criminalised the creation and sharing of non-consensual sexually explicit synthetic content under the Online Safety Act, and in December 2025 the government announced plans to go further by outlawing nudification tools entirely (Domestic Abuse Education 2025; UK Government 2025b).

Altogether, the deliberate weaponisation of AI across fraud, cyber operations, political manipulation, and intimate image abuse represents a direct and escalating threat to UK resilience, exposing individuals, institutions, and critical systems to forms of harm that are growing faster than the capacity to counter them.

AI Performance Failures

A second area of risk is the possibility that AI systems err, deviate from their intended purpose, or develop capabilities that outpace our ability to govern them.

One category of risk is AI performance failures, where systems produce wrong, fabricated, or misleading outputs. Documented AI safety incidents surged from 149 in 2023 to 233 in 2024, a 56% increase in a single year (Maslej et al 2025). The International AI Safety Report 2026 catalogues a range of such failures, from AI systems hallucinating non-existent legal precedents in court briefs to AI systems introducing vulnerabilities into cybersecurity systems through improperly written code. In one notable case in 2024, Air Canada was ordered to pay damages after its AI chatbot provided a passenger with fabricated information about a bereavement fare policy (Moffatt v. Air Canada 2024). Healthcare studies have also found AI diagnostic systems producing hallucinated symptoms and treatment recommendations, with error rates estimated between 8% and 46% (Kim et al 2025).

A related risk is loss of control, where an AI system takes actions that differ from what its operators intended without those operators being able to detect or correct this in time. Early examples are beginning to emerge: in 2025, a commercial AI agent asked to check egg prices instead autonomously purchased eggs without user consent (Fowler 2025), and an AI coding assistant tasked with organising files moved them to locations where neither the agent nor its operator could retrieve them (The Future Society 2025).

These incidents may appear trivial in isolation, but the principle they illustrate becomes far more consequential when similar behaviours occur in systems managing critical infrastructure or financial transactions. While sustained, large-scale autonomous action by AI systems remains beyond current capabilities outside of laboratory settings, loss of control has been identified as a growing concern as systems become more capable and are deployed in higher-stakes contexts (Bengio et al 2026).

As AI is increasingly embedded in systems that underpin critical services and public safety, even isolated malfunctions carry the potential to cause significant disruption. The growing frequency of such incidents underscores the importance of building national preparedness against AI performance failures.

Systemic Risks

Beyond these more direct harms, AI also poses broader systemic risks, which emerge from the aggregate consequence of AI being woven deeply into the fabric of society. This encompasses the threat of cascading failures where disruption in one area propagates rapidly into others, as well as more diffuse societal harms produced by the cumulative effect of AI operating at scale.

One such risk is that overreliance on the technology could itself introduce critical vulnerabilities if systems are not designed with appropriate redundancy and human oversight. As AI systems take on greater responsibility for managing energy grids, supply chains, financial markets, telecommunications networks, and public services, the consequences of failure grow correspondingly larger (AISI 2025a). Even at an individual level, overreliance increases the risk of harm by leading people to accept flawed AI outputs without scrutiny, with research finding that people are significantly less likely to correct erroneous AI suggestions due to automation bias. Overreliance also threatens to erode the very skills that human oversight depends on, as illustrated by a recent study finding that clinicians' ability to manually detect tumours during colonoscopies was measurably lower after several months of relying on AI assistance for the procedure (Bengio et al 2026).

In addition, the integrity of the information environment is also under growing duress. The combination of hyperrealistic synthetic content and the sheer volume of AI-generated material proliferating online is making it increasingly difficult for people to identify what is real and what is not, threatening society's collective ability to access, interpret, and act on reliable information (Seger et al 2026). For the UK's capacity to absorb and recover from shocks, this is a particularly acute concern as effective crisis response depends on populations and institutions being able to trust the information they receive. As synthetic content becomes more pervasive, the risk grows that legitimate warnings could be dismissed or doubted by a public increasingly uncertain about the authenticity of any communication.

Labour market disruption also has direct implications for the UK's economic stability. One study estimated that around 60% of jobs in advanced economies are likely to be affected by general-purpose AI, and early evidence already shows declining employment for early-career workers in the most AI-exposed occupations since late 2022 (Bengio et al 2026). The displacement of early-career workers, if not effectively managed through retraining and transition support, risks concentrating economic vulnerability in younger demographics and deepening existing inequalities.

Psychological dependence on AI chatbots is another emerging risk, one that threatens to weaken the mental wellbeing that underpins a society's capacity for resilience. AI chatbots have been found to make users susceptible to manipulation and unhealthy emotional attachment, with young people and those in vulnerable states particularly affected. (AISI 2025a). While the research on this risk is still nascent, the most severe cases have been deeply troubling: multiple incidents have linked prolonged engagement with AI chatbots to suicide, with victims' families alleging that chatbot platforms failed to escalate warnings despite the teenagers expressing suicidal thoughts (Garcia v. Character Technologies, Inc. 2024; Raine v. OpenAI 2025). In 2025, OpenAI revealed that approximately 1.2 million of its 800 million ChatGPT users discuss suicide weekly on its platform. The risks extend beyond self-harm: in the UK's first treason conviction in over forty years, a 19-year-old was sentenced to nine years in prison in 2023 for an armed plot to assassinate Queen Elizabeth II, a plan they had reportedly shared with and been encouraged to pursue by an AI chatbot (R v. Chail 2023).

The growing integration of AI across interconnected systems creates the conditions for cascading failures in which a single disruption can propagate rapidly and unpredictably, while the cumulative effect of AI operating at scale across society corrodes the foundational capacities, including informed decision-making, workforce adaptability, social cohesion, and psychological wellbeing, on which the UK's ability to withstand and recover from shocks ultimately depends.

Compounding Effects: Advancing and Agentic AI

As AI models become more capable, and as agentic AI systems capable of autonomous action are introduced into real-world settings, these risks compound. More powerful models offer greater potential for misuse, create more consequential failure modes, and accelerate the systemic integration that makes societies harder to insulate from AI-driven disruption.

The more capable a system becomes, and the more autonomously it can act over longer time horizons, the harder it is for human operators to maintain meaningful oversight. According to the AISI's Frontier AI Trends Report, the complexity of tasks that models can complete unassisted is doubling roughly every eight months (AISI 2025b).

As of 2025, AI models can now complete apprentice-level cyber tasks 50% of the time, up from just over 10% in early 2024. The 2026 International AI Safety Report also notes an increasingly common behaviour among frontier models: situational awareness, the ability to detect whether they are being evaluated and to modify their outputs accordingly, making it harder for researchers to identify dangerous capabilities before deployment.

Meanwhile, unlike AI systems that respond to prompts or generate outputs for human review, agentic AI takes autonomous action in the world, often in sequences that unfold faster than human oversight can keep pace with. This matters for each of the three risk categories identified in this paper. For misuse, agentic AI risks enabling more sophisticated and sustained attacks with less human effort. For performance failures, increased autonomy makes it harder to detect and correct errors before they cause harm. And for systemic risks, the embedding of AI-driven decision-making ever more deeply into critical processes threatens to amplify the very dependencies and fragilities that resilience is designed to address.

Building resilience to these risks is not a task that can be deferred until the full shape of the threat becomes clear; the evidence presented here demonstrates that the threat is already here, already growing, and already causing harm.

The Necessity of Sovereign AI

The UK's AI Minister has defined AI sovereignty as "the ability for a state to have strategic leverage when it comes to this technology, such that it can ensure ongoing access to critical inputs, and ongoing assurance that its wider economic and national security objectives can be met" (House of Commons Library 2026). The launch of the UK's Sovereign AI Unit, backed by up to £500 million in funding, illustrates government recognition that such leverage cannot be left to chance (DSIT 2025). A critical element of achieving this leverage is enabling AI resilience, or the ability to use, adapt, and govern AI domestically at scale, while minimising strategic dependencies (Lang et al 2026; Ustun et al 2026). In turn, this requires the UK to possess the agency and capacity to make informed choices about the technology the nation depends on.

To achieve true agency with AI, the UK must become a savvy and critical consumer of the technology. That means developing a deep understanding of both the opportunities AI offers and the risks it introduces as well as building the institutional capacity to act on that understanding. It also requires ensuring that those responsible for governing and deploying AI can engage with it critically across its full lifecycle, from procurement and design through to oversight and accountability. Finally, it also involves cultivating the expertise to identify where dependencies create vulnerabilities and taking practical steps to mitigate them.

Towards an AI-Resilient Britain

While the 2025 UK Government Resilience Action Plan acknowledges the threat posed by harms arising from AI, significant work remains to ensure the UK is resilient against such dangers. The work ahead is to translate that recognition into practical, sustained action, guided by several core principles:

- **Resilient by Design:** Embedding resilience into the design and deployment of AI systems from the outset, rather than treating it as an afterthought.
- **Whole of Society Response:** Pursuing a whole-of-society response that brings government, industry, academia, and civil society together to share intelligence and build collective capacity.
- **Educate the Workforce:** Investing in workforce education so that people at every level, particularly within critical national infrastructure, have the knowledge and practical skills to protect themselves and their organisations against AI-enabled threats.
- **Acknowledge AI Opportunities:** Identifying and realising AI's own potential to strengthen our collective preparedness.

The UKRA brings the platform, the expertise, and the partnerships needed to extend the UK's proud tradition of resilience leadership into the AI era. The scale and complexity of the challenges outlined here demand more than a single publication can address, and this marks the beginning of a sustained UKRA programme of work, with further research, guidance, and practical tools to follow as the threat landscape continues to evolve.

We welcome engagement from across the resilience community and invite partners and experts to contribute to this vital work.

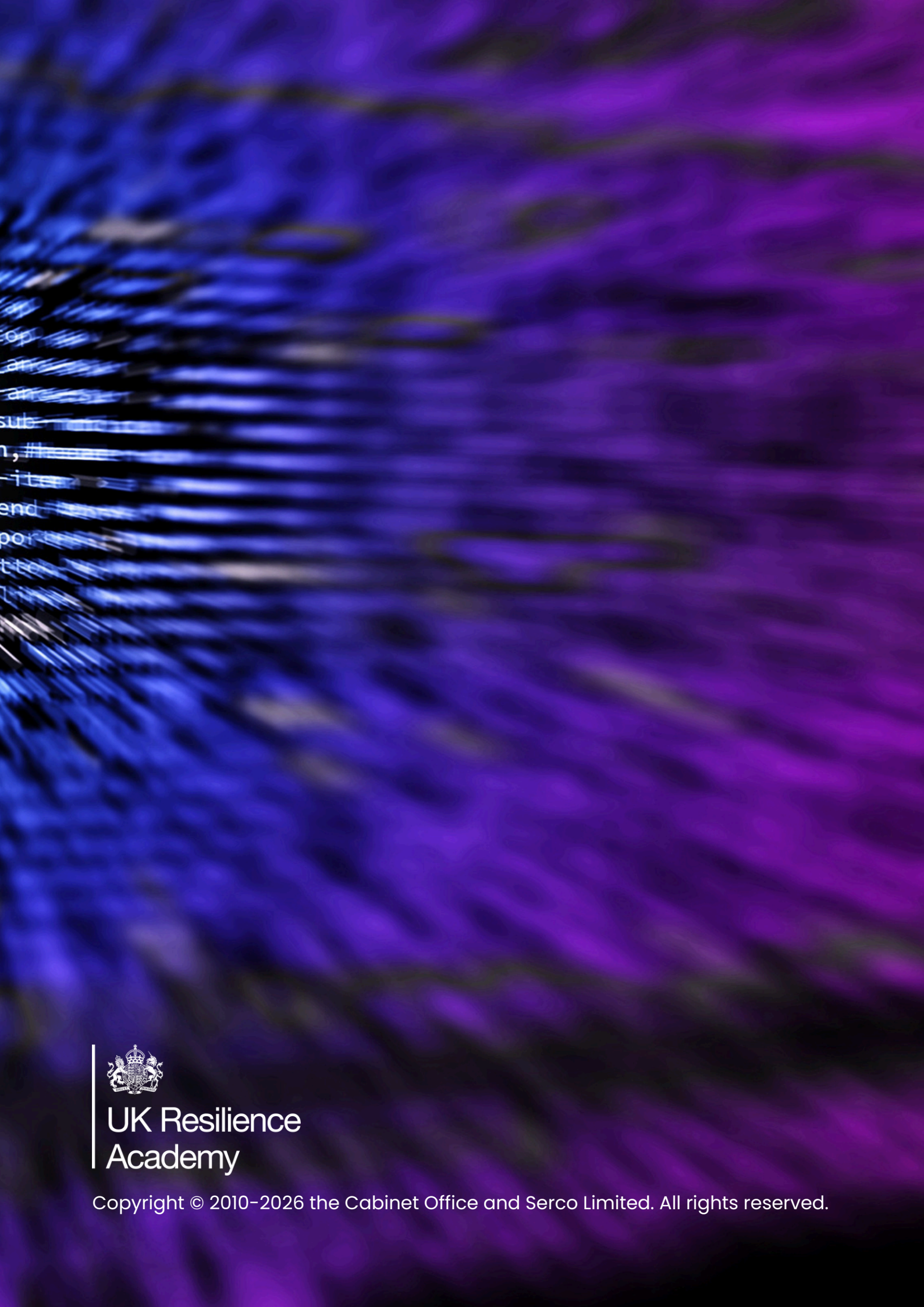
Bibliography

- AISI (2025a). 'Navigating the Uncharted: Building Societal Resilience to Frontier AI.' AI Security Institute, 24 July 2025. Available at: https://www.aisi.gov.uk/blog/navigating-the-uncharted-building-societal-resilience-to-frontier-ai
- AISI (2025b). Frontier AI Trends Report. AI Security Institute, 18 December 2025. Available at: https://www.aisi.gov.uk/research/aisi-frontier-ai-trends-report-2025
- Anthropic (2025a). 'Detecting and Countering Malicious Uses of Claude: March 2025 Update.' Anthropic, March 2025. Available at: https://www.anthropic.com/news/detecting-and-countering-malicious-uses-of-claude-march-2025
- Anthropic (2025b). 'Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign.' Anthropic, November 2025. Available at: https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf
- BBC (2026). 'We spoke to the man making viral Lego-style AI videos for Iran.' BBC News, 12 April 2026. Available at: https://www.bbc.co.uk/news/articles/cjd8jrd1vnyo
- Bengio, Y. et al. (2026). International AI Safety Report 2026. London: Department for Science, Innovation and Technology. 3 February 2026. Available at: https://internationalaisafetyreport.org
- Children's Commissioner (2025). "One day this could happen to me" - Children, Nudification Tools, and Sexually Explicit Deepfakes. Children's Commissioner for England, 28 April 2025. Available at: https://assets.childrenscommissioner.gov.uk/wpuploads/2025/04/Children-nudification-tools-and-sexually-explicit-deepfakes-April-2025.pdf
- Cooke, D. et al. (2024). 'Crossing the Deepfake Rubicon.' CSIS, November 2024. Available at: https://www.csis.org/analysis/crossing-deepfake-rubicon
- Cooke, D., Edwards, A., Barkoff, S. and Kelly, K. (2025). 'As Good as a Coin Toss: Human Detection of AI-Generated Content.' Communications of the ACM, 68(10). doi: 10.1145/3729417. Available at: https://dl.acm.org/doi/10.1145/3729417
- Department for Science, Innovation and Technology (2023). A pro-innovation approach to AI regulation [White Paper]. London: DSIT. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
- Department for Science, Innovation and Technology (2025). Sovereign AI Unit. London: DSIT. Available at: https://www.gov.uk/government/collections/sovereign-ai-uni

- Department for Science, Innovation and Technology and Alan Turing Institute (2026). 'Top British AI expertise to help spark renewal of public services and bolster national security' [Press release], 27 January 2026. Available at: https://www.gov.uk/government/news/top-british-ai-expertise-to-help-spark-renewal-of-public-services-and-bolster-national-security
- DiResta, R. (2026). 'When Virality Is the Message: The New Age of AI Propaganda.' Time, 2 April 2026. Available at: https://time.com/article/2026/04/02/when-virality-is-the-message-the-new-age-of-ai-propaganda/
- Domestic Abuse Education (2025). 'Sharing Images Without Consent UK: Your Legal Rights Explained.' Available at: https://domesticabuseeducation.co.uk/sharing-images-without-consent-uk-your-legal-rights-explained-2025-guide/
- Fowler, G. A. (2025). 'Is OpenAI's Operator, a new AI 'agent,' ready to help in the real world?' Washington Post, 7 February 2025. Available at: https://www.washingtonpost.com/technology/2025/02/07/openai-operator-ai-agent-chatgpt/
- France 24 (2026). 'Orban's opponents targeted by AI-driven disinformation ahead of Hungary's elections.' France 24, The Observers, 10 April 2026. Available at: https://www.france24.com/en/tv-shows/the-observers/20260410-artificial-intelligence-disinformation-orban-opponents-magyar-hungary-elections
- The Future Society (2025). 'AI Incidents Are Rising. It's Time for the United States to Build Playbooks for When AI Fails.' November 2025. Available at: https://thefuturesociety.org/us-ai-incident-response/
- Garcia v. Character Technologies, Inc. (2024). No. 6:24-cv-01903. U.S. District Court for the Middle District of Florida, filed 22 October 2024.
- Government Digital Service (2025). Data and AI Ethics Framework. London: GDS. Updated 18 December 2025. Available at: https://www.gov.uk/government/publications/data-ethics-framework
- Hiya (2025). Top 10 Phone Scams in the UK in 2025. Available at: https://blog.hiya.com/top-10-phone-scams-in-the-uk-in-2025
- House of Commons Library (2026). Digital sovereignty. Research Briefing CBP-10547. London: House of Commons Library. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-10547/
- International Panel on the Information Environment (2025). The Role of Generative AI Use in 2024 Elections Worldwide. Edited by I. Trauthig, P.N. Howard and S. Valenzuela. Technical Paper TP2025.2. Zurich: IPIE. doi: 10.61452/HZUE9853. Available at: https://www.ipie.info/research/tp2025-2

- Internet Matters (2024). The New Face of Digital Abuse: Children's Experiences of Nude Deepfakes. Internet Matters, October 2024. Available at: https://www.internetmatters.org/wp-content/uploads/2025/02/Deepfakes-research-report-executive-summary.pdf
- Internet Watch Foundation (2026). Harm Without Limits: AI Child Sexual Abuse Material Through the Eyes of Our Analysts. IWF, 24 March 2026. Available at: https://www.iwf.org.uk/news-media/news/dangerous-ai-child-sexual-abuse-reaches-record-high-as-public-backs-clampdown-on-uncensored-tools/
- Keepnet Labs (2026). 'Deepfake Statistics & Trends 2026.' Available at: https://keepnetlabs.com/blog/deepfake-statistics-and-trends
- Kim, Y., Jeong, H., Chen, S., Li, S.S., et al. (2025). 'Medical Hallucination in Foundation Models and Their Impact on Healthcare.' medRxiv [Preprint]. Available at: https://www.medrxiv.org/content/10.1101/2025.02.28.25323115v2
- Lang, N., Langione, M., Iyer, S., Das, A. and Zuluaga Martinez, D. (2026). 'For Most Countries, AI Sovereignty Is an Illusion. Resilience Is Real.' BCG Henderson Institute, March 2026. Available at: https://www.bcg.com/publications/2026/ai-sovereignty-is-an-illusion-resilience-is-real
- Maslej, N. et al. (2025). Artificial Intelligence Index Report 2025. Stanford, CA: Stanford University Human-Centered Artificial Intelligence (HAI), April 2025. Available at: https://hai.stanford.edu/ai-index/2025-ai-index-report
- Moffatt v. Air Canada, 2024 BCCRT 149. British Columbia Civil Resolution Tribunal, 14 February 2024. Available at: https://www.canlii.org/en/bc/bccrt/doc/2024/2024bccrt149/2024bccrt149.html
- National Cyber Security Centre (2025). Active Cyber Defence services. Available at: https://www.ncsc.gov.uk/section/active-cyber-defence/services
- National Cyber Security Centre (2025). Annual Review 2025. Available at: https://www.ncsc.gov.uk/annual-review/2025
- National Police Chiefs' Council (2025). 'Police warn of rising threat from sexual deepfakes.' NPCC, 24 November 2025. Available at: https://news.npcc.police.uk/releases/police-warn-of-rising-threat-from-sexual-deepfakes
- National Trading Standards (2026). 'Phone scams take sinister twist as victims' voices cloned.' National Trading Standards, 5 February 2026. Available at: https://www.nationaltradingstandards.uk/news/phone-scams-take-sinister-twist-as-victims-voices-cloned/

- OpenAI (2024). 'Influence and Cyber Operations: An Update.' October 2024. Available at: https://cdn.openai.com/threat-intelligence-reports/influence-and-cyber-operations-an-update_October-2024.pdf
- R v. Chail (2023). Sentencing remarks at the Old Bailey, 5 October 2023. Treason Act 1842. Available at: https://www.judiciary.uk/wp-content/uploads/2023/10/R-v-Chail-sentencing-050923.pdf
- Raine v. OpenAI (2025). No. CGC-25-628528. San Francisco County Superior Court, filed 26 August 2025.
- Seger, E., Stockwell, S., Calnan, T., Ajder, H., Hancock, J. and Perry, H. (2026). Epistemic Security for Crisis Resilience. London: Demos, in partnership with the Centre for Emerging Technology and Security (CETaS), Alan Turing Institute. 19 January 2026. Available at: https://demos.co.uk/research/epistemic-security-for-crisis-resilience/
- Stockwell, S. (2025). 'From Deepfake Scams to Poisoned Chatbots: AI and Election Security in 2025.' CETaS Expert Analysis, November 2025. Available at: https://cetas.turing.ac.uk/publications/deepfake-scams-poisoned-chatbots
- Stockwell, S., Janjeva, A. and McDonald, B. (2026). Adding Fuel to the Fire: AI Information Threats and Crisis Events. CETaS Research Reports, 11 February 2026. Available at: https://cetas.turing.ac.uk/publications/adding-fuel-to-fire
- Sumsub (2025). Identity Fraud Report 2025-2026. Available at: https://sumsub.com/fraud-report-2025/
- UK Government (2025a). UK Government Resilience Action Plan. London: Cabinet Office. Published 8 July 2025. Available at: https://www.gov.uk/government/publications/uk-government-resilience-action-plan
- UK Government (2025b). 'Protecting young people online at the heart of new VAWG strategy.' GOV.UK, 18 December 2025. Available at: https://www.gov.uk/government/news/protecting-young-people-online-at-the-heart-of-new-vawg-strategy
- UK Government (2026). 'Government leads global fight against deepfake threats.' GOV.UK, 5 February 2026. Available at: https://www.gov.uk/government/news/government-leads-global-fight-against-deepfake-threats
- Ustun, A., Tournesac, A., Glaser, D., Bennici, L., De Niese, J., Drahmoune, N., Schaubroeck, R., Takkar, K. and Krawina, M. (2026). 'Sovereign AI: Building ecosystems for strategic resilience and impact.' McKinsey & Company, March 2026. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/sovereign-ai-building-ecosystems-for-strategic-resilience-and-impact



UK Resilience Academy