24/25 MIAA Audit Committee Insight

Technology and Data Analytics Risk Update

EU Artificial Intelligence Act and its wider implications

October 2024

1 Background

The rapid evolution of Artificial Intelligence (AI) presents both significant opportunities and challenges. According to the UK Cyber Security Council[1], AI is still in its early stages, yet it shows great potential. Regulation around AI is advancing rapidly, with the UK seeking to balance innovation and safety through its Pro-Innovation AI Framework, which is part of the broader National AI Strategy. This framework is built on five guiding principles for responsible AI including:

- Safety, security and robustness
- Appropriate transparency and explainability
- Fairness
- Accountability and governance
- Contestability and redress

These principles are echoed by several industry standards and frameworks, including those from Microsoft and Rolls Royce, and the work of the UK AI Standards Hub, established to maintain and encourage responsible AI innovation. There is now a global trend towards convergence in AI regulation, notable across the EU, US and UK.

2 What is the EU AI Act?

March 2024: the European Parliament passed the AI Act, which focuses on managing the risks associated with AI, particularly around biometric categorisation, manipulation of human behaviour, and stricter regulations on generative AI.

May 2024: The Council gave final approval to the AI Act setting a "risk-based approach" to AI regulation, whereby higher-risk AI applications face stricter requirements to operate in EU.

August 2024: the AI Act became law across all 27 EU member states, and the enforcement of the majority of its provisions will commence on 2nd August 2026.[2]

Enforcement: is being clarified / under development.

3 What are the potential cyber security risks of AI?

The National Cyber Security Centre (NCSC) has flagged potential cyber security risks of AI[3], including:

- AI hallucination: where AI systems generate incorrect information.
- Bias and gullibility: AI systems are susceptible to influence through leading questions.
- Prompt injection attacks: where attackers manipulate AI inputs to generate harmful or unintended outputs.
- Data poisoning: attackers tamper with training data to produce biased or malicious outcomes.
- These vulnerabilities make AI systems targets for cyberattacks, particularly prompt injection and data poisoning, which can severely compromise both security and trust in AI applications.

[1] Ethics in Artificial Intelligence (ukcybersecuritycouncil.org.uk)

[2] https://www.theregister.com/2024/07/31/eu_ai_act/

[3] https://www.ncsc.gov.uk/guidance/ai-and-cyber-security-what-you-need-to-know

Prompt injection attacks are one of the most widely reported weaknesses. It is when an attacker creates an input designed to make the model behave in an unintended way. By accepting unchecked input, it may cause AI to generate offensive content, reveal confidential information, or trigger unintended consequences in a system.

Data poisoning attacks occur when an attacker tampers with the data that an AI model is trained on to produce undesirable outcomes (both in terms of security and bias). As models are increasingly used to pass data to third-party applications and services, the risks from these attacks will grow, as described by NCSC in their blog: Thinking about security AI Systems

4 What are the potential implications for healthcare?

There is a need to balance innovation against ethical and safety considerations. Digital Health reported[4]:

- Under the Act, any AI system that is a Class IIa (classification for medical devices) or higher medical device, or uses an AI system as a safety component, is designated as “high risk”.
- The Act also specifies certain types of healthcare AI systems as high risk, whether or not they are medical devices, such as AI systems used by public authorities to evaluate the eligibility of people for essential public services, and AI systems that are emergency healthcare patient triage system.
- Unlike the Medical Devices Regulation (MDR) and In Vitro Diagnostics Medical Devices Regulation (IVDR), which place responsibilities on economic operators in the supply chain, the AI Act also puts responsibilities onto the **deployers** of AI systems, **being any person using an AI system** in the course of a business or professional activity, **such as hospitals or clinicians.**

Key obligations for AI system deployers (e.g. hospitals, clinicians) include:

- Taking appropriate technical and organisational measures to ensure that AI systems are used in accordance with their instructions for use
- Assigning human oversight to competent, trained people
- Continuous monitoring and surveillance, and maintaining system logs

[4] Digital Health - EU regulation: AI Act will mean a raft of new requirements for ‘high-risk’ systems 24th April 2024, https://www.digitalhealth.net/2024/04/eu-regulation-ai-act-will-mean-a-raft-of-new-requirements-for-high-risk-systems/

- Conducting, where applicable, data protection impact assessments.

Additional requirements for AI systems (not already in EU MDR / IVDR), include:

- Governance and data management requirements for training and testing data sets
- New record-keeping requirements, including the automatic recording of events (logs) over the system's lifetime
- Transparent design requirements so deployers can interpret the output and use it appropriately
- Human oversight design requirements
- Accuracy and cybersecurity requirements.

Even AI systems that aren't medical devices, such as those used by public authorities for essential public service assessments, are subject to stringent regulations.

5 UK's AI Risk Treaty

Balancing Innovation, Regulation, and Ethics

The UK's AI Risk Treaty aims to regulate AI development while fostering innovation. Striking this balance is critical, as AI can transform industries from healthcare diagnostics to financial services. However, the ethical deployment of AI, ensuring fairness, transparency, and accountability, remains essential to prevent biased or harmful outcomes, as seen in cases involving discriminatory hiring practices or biased law enforcement algorithms.

Maintaining a "human in the loop" (HITL) is vital to mitigating these risks. Human oversight ensures that AI decisions, especially in high-stakes sectors like healthcare and cybersecurity, are reviewed and validated. For instance, while AI can flag potential health concerns, human doctors must interpret the findings within the broader context of a patient's medical history.

The Need for Flexible Regulation

Overly rigid regulatory frameworks could hinder AI innovation. For instance, delayed regulatory approvals for AI-driven healthcare tools could reduce the potential for early disease detection, while stringent cybersecurity regulations might slow down the deployment of real-time fraud detection algorithms.

An adaptive regulatory approach is essential to ensure regulations evolve alongside technological advancements, as emphasized by the NCSC in its Intelligent Security Tools[5] report. AI-driven tools must remain flexible and forward-thinking to respond to emerging threats and opportunities.

Growing Cybersecurity Threats

AI-powered cyberattacks are on the rise, with a 300% increase in AI-driven phishing attacks in 2023 alone. A notable incident involved a financial institution losing £28 million to an AI-generated voice scam, underscoring the growing sophistication of AI-enhanced cybercrime.

To counter these threats, the development of AI-enhanced cybersecurity tools is crucial. The AI Risk Treaty should encourage the creation of these defensive tools while ensuring they are deployed responsibly.

[5] Intelligent security tools - NCSC.GOV.UK

International Cooperation and Economic Impact

AI's potential to contribute to the global economy is immense, with projections suggesting it could add £320 billion to the UK economy by 2030, boosting productivity by 30%. However, the lack of international cooperation on AI regulation, particularly concerning high-risk AI technologies like autonomous weapons, could destabilize global security.

Aligning with international standards, such as the EU AI Act, is critical to ensuring responsible AI development and mitigating cross-border cyber threats. At the same time, over-regulation could stifle innovation, slowing down job growth and inhibiting the economic benefits that AI promises.

Summary

The UK's AI Risk Treaty marks a crucial step toward developing AI ethically and safely. A dynamic and adaptive regulatory approach is essential to protect public interests while encouraging innovation. As AI continues to shape the future, the key to success lies in balanced regulation, international cooperation, and ensuring human oversight in critical decision-making processes.

6 Questions to Ask Your Organisation

NCSC suggest managers, board members and senior executives use the following questions to help understand how an organization is dealing with the AI / Machine Learning (ML) threat:

1. Do you understand where accountability and responsibility for ML/AI security sit in your organisation?
2. Does everyone involved in ML/AI deployment, including board members and/or senior executives, know enough about ML/AI systems to consider the risks and benefits of using them?
3. Does security factor into decisions about whether to use ML/AI products?
4. How do the risks of using ML/AI products integrate into your existing governance processes?
5. What are your organisation's critical assets in terms of ML/AI and how are they protected?
6. What is the worst case (operationally or reputationally) if an ML/AI tool your organisation uses fails?
7. How would you respond to a serious security incident involving an ML/AI tool?
8. Do you understand your data, model and ML/AI software supply chains and can you ask suppliers the right questions on their own security?
9. Do you understand where your organisation may have skills or knowledge gaps related to ML/AI security? Is a plan in place to address this?

Find out more:

If you have any queries or feedback on this briefing, please contact:

Paula Fagan, Head of Technology Risk

Email: paula.fagan@miaa.nhs.uk

Catherine Watts, Principal Digital Risk Consultant

Email: catherine.watts@miaa.nhs.uk

Andrew Bowdler, Principal Data Analyst

Email: andrew.bowdler@miaa.nhs.uk