Business AI Alliance

Briefing: Accessible AI Security for UK SMEs

Jeff Watkins

07/09/2025

UK SMEs are starting to embrace AI, especially generative AI like LLMs, for growth and innovation, not only in applications, but at the organisational level, with smaller organisations moving towards being "AI-first". However, this adoption introduces new cybersecurity risks such as prompt injection, data leakage, model poisoning, and model Denial of Service (DoS). Current cybersecurity standards and certifications (such as Cyber Essentials) cover general IT cybersecurity hygiene, but there is a need for AI-specific, lightweight frameworks to safeguard the extra attack surface these innovations expose.

Background

Generative AI is expanding rapidly worldwide, including in the UK. While it creates significant opportunities, it also widens the attack surface for organisations. Meanwhile, cybercrime continues to grow, and attacks on AI systems handling sensitive data can cause reputational damage, service outages, and regulatory fines. Because Generative AI is often exposed to end users, it introduces risks such as prompt injection, model theft, and data poisoning, leading to data loss, ethical failures, harmful content, and breaches. Its emergent nature means threats like model inversion and hallucinations are novel, poorly structured, and beyond the reach of traditional cyber controls. Threat-modelling methods remain underdeveloped.

At the same time, GDPR and incoming AI-specific rules demand risk assessments and data-processing safeguards. Public trust is also essential, especially in sensitive sectors like finance and life sciences. A pragmatic approach to securing and assuring UK AI adoption is therefore critical, one that is robust enough to offer real protection but lightweight enough for SMEs to adopt without friction.

What SMEs Need for Sufficient AI Security

Adopting Generative AI in SME organisations introduces unique threats not covered by existing security schemes and tooling. These new AI threats contribute to an already extensive and continually growing cybersecurity risk landscape, potentially overwhelming SMEs with excessive risks and administrative compliance tasks. However, there is a lack of in-depth AI and Cybersecurity knowledge in smaller organisations, meaning that most SMEs are unable to model, measure and mitigate these risks without outside assistance. The resource constraints that stem from organisational size and budget make full external technical audits impractical and would deter AI adoption. There is therefore a need for a self-paced self-assessment mechanism, guided by accessible frameworks for low-risk organisations. There is also an additional need for SME-appropriate standardised assurance mechanisms for cases where a more robust review is required.

It is worth noting that generative AI adoption doesn't necessarily mean bespoke implementation, but can include the use of existing 3rd party and SaaS AI-based software within an organisation's estate. It is envisaged that, like CyberEssentials, many organisations would benefit from organisation-level assurance of AI security hygiene.

Existing Frameworks & Toolkits

There are multiple existing frameworks, certifications, and toolkits on the topics of cybersecurity threat modelling & assurance, as well as AI security & management. A non-exhaustive list of the existing landscape is shown in Appendix A. There are more frameworks and toolkits, but there is currently a gap between the AIME toolkit, Cyber Essentials and OWASP/MITRE ATLAS, etc. that needs to be addressed to enable SMEs to safely implement AI in their organisations, without the heavyweight requirements of implementing and being assessed to ISO 27001 and 42001 standards.

Proposed Solution: Cyber Essentials Equivalent Certification

It is recommended that a group be formed to collaborate with the NCSC, DSIT, and IASME to define self-assessment questions, test the efficacy and then integrate into the existing Cyber Essentials online infrastructure, with the ability to issue Cyber-AI Essentials certification for SMEs. This group would also create the frameworks and infrastructure to implement the Cyber-AI Essentials Plus audits, alongside the ability to certify organisations, with yearly renewals, etc. This could be treated as a separate certification or rolled into an existing scheme.

AI Cyber Security Basic (Self-Assessment)

This certification is a companion to Cyber Essentials, focusing on AI security, with areas including input and output validation and sanitisation, access control over AI keys/services, logging & anomaly detection, fail-safe modes & human oversight, Data lifecycle governance and GDPR alignment. Based on the “traffic light” system of risk levels, for lower risk implementations, this could be a simple online questionnaire that can be completed in under an hour. Minimal cost, integrated with existing portals. SMEs receive a “Cyber-AI Essentials” badge for completion, with a yearly requirement to recomplete the questionnaire.

AI Cyber Security Plus (Light Audit)

For implementations that require higher assurance, the “Plus” version of the proposed AI Cyber Security certification would include external review, combining a documentation review on the AI provider and patching, guided OWASP LLM Top 10 scenario testing (e.g., prompt injection), basic resilience testing (Model DoS, failover to prevent over-reliance), etc. This would also include a process that reviews evidence of adequate logging, monitoring and alerting, alongside sufficient AI incident response processes and controls. These would be performed by an external auditor using some automation to reduce costs.

Potential Integration into Cyber Essentials Plus

Eventually embed AI controls into CE Plus audit (e.g., sampling API security checks). This could become part of full CE Plus verification as an optional module for those implementing AI.

Funding for AI Cyber Security Plus Certification

- This could be funded via Innovate UK/NCSC, extending the funded voucher scheme to SMEs implementing AI
- This should also subsidise spot audit costs (~£500–£1k)
- Expand assessor training to include AI-specific audit protocols

It is also recommended to close the loop and extend CE Plus audit requirements to signpost to AI elements (secure API, logging, sandboxing) and to encourage those implementing AI to engage with the AI Cyber Security programme.

Ongoing Education & Awareness

- Fund microlearning (<1 hr) modules on prompt injection, model security
- Publicise via NCSC webinars and industry associations
- Support community-led red-teaming events using OWASP LLM tools

Provide Supporting Toolkits

- Access to structured learning tools on AI threats in plain English
- Provide scenarios/pathways tailored to sectors, e.g. iGaming, finance, health, where risks are greater
- Signpost to industry frameworks, including the OWASP LLM Top 10 Checklist, MITRE ATLAS, and STRIDE threat modelling
- Self-Assessment Worksheets, aligned with Cyber-AI Essentials

Provide these elements via a central AI security portal, with interactive guidance ("e.g. if you use LLM in finance, start here") and a pathway to certification.

Recommendations

Implement the measures discussed, which will ensure that small businesses can adopt and innovate with AI responsibly and securely, without being hindered by excessive processes. This would accelerate the safe adoption across sectors, enhance AI trustworthiness throughout the UK, and equip SMEs to meet emerging AI regulatory and procurement demands.

Key proposals:

1. Build upon the good work with the AIME (AI Management Essentials)
2. Extend Cyber Essentials with Cyber-AI Essentials (self-certification layer)
3. Offer Cyber-AI Essentials Plus (light technical audit) for implementations that require it
4. Provide a portal of supporting educational material and toolkits for threat modelling (STRIDE), adversarial AI threat mapping (MITRE ATLAS), and LLM-specific vulnerabilities (OWASP LLM Top 10).
5. Launch government-led guidance, vouchers, and the centralised portal mentioned above to support uptake.

In conclusion, the proposed certifications offer a pragmatic and scalable approach to AI security tailored to SME capacity by leveraging existing knowledge and schemes (Cyber Essentials, OWASP, MITRE), embedding lightweight AI-specific considerations, providing curated toolkits and support, and crucially offering low-cost pathways to allow SMEs to implement AI at the organisational level with confidence.

Appendix A - Security and AI Standards, Frameworks, Certifications & Toolkits

Framework / Tool	**Type**	**Level**	**AI Specific?**	**Description**
Cyber Essentials	Assurance standard	Basic	No	Low cost (~£300), but no AI coverage
Cyber Essentials Plus	Technical audit	Intermediate	No	Medium cost (~£1.4k), no AI coverage
ISO 27001	Assurance Standard	Advanced	No	An international standard that provides a framework for an Information Security Management System (ISMS), focusing on the confidentiality, integrity, and availability of information.
ISO 41001	Assurance Standard	Advanced	Yes	
STRIDE	Threat Framework	Basic	No	A simple framework mapping potential threats across an application. This is Ideal for SMEs to create quick threat maps using STRIDE worksheets, but it is not AI-specific.
OWASP LLM Top 10	Threat framework	Basic	Yes	AI-specific extension/addition to the other OWASP Top 10 lists. adapted from OWASP's generative AI project, the LLM Top 10 highlights common vulnerabilities. Using this, SMEs can adopt mitigation practices, including input/output sanitisation, sandboxing, and adversarial testing. However, it is a “top 10” and does not cover holistically. Open-source, available via OWASP site, requires some knowledge to leverage, not comprehensive
MITRE ATLAS	Adversarial library	Intermediate	Yes	AI- specific, open-source, available via MITRE site. A live, evolving matrix of adversarial AI-related TTPs, from reconnaissance to exfiltration. It complements OWASP by mapping how attacks unfold in real-world scenarios. It is useful for SMEs doing threat modelling or red-teaming, but requires a reasonable understanding of cybersecurity.
AIME	Self assessment	Basic to Intermediate	Yes	AI-specific, but no in-depth security advice. The AI Management Essentials Toolkit - A self-assessment tool that aims to help organisations assess responsible AI management systems and processes.