FLETCHLABS
SOFTWARE & SECURITY

AI Security Assessment & Automation

Secure your AI systems, then put AI to work securing everything else.

SERVICE PACKAGE
August 2026

fletchlabs.ca • info@fletchlabs.ca • Cambridge, Ontario, Canada

Executive summary

This service has two connected components. First, we assess the security of the AI you have built or adopted: the integrations, the data flows, and the governance around them. Second, where it makes sense, we design and build AI-assisted automation for your security operations, with guardrails baked in so the tools we deliver never reintroduce the risks we just assessed. Adoption without governance is how organizations end up in the failure statistics below; we exist to keep you out of them.

When AI implementations backfire

These are not hypotheticals. They are documented outcomes from well-known organizations:

Secrets pasted into public chatbots

In 2023, multiple Samsung employees reportedly pasted confidential source code, internal tooling, and meeting material into ChatGPT to debug or summarize work, sending proprietary data to a third-party service with retention and training implications. Samsung later restricted generative AI use.[3][4]

Massive data exposure from misconfigured AI research assets

Microsoft AI researchers accidentally exposed tens of terabytes of internal data, including credentials, backups, and internal communications, via a misconfigured SAS URL while publishing open-source material. AI and ML workflows amplify traditional configuration and secrets-management mistakes.[5]

Internal breach of AI org communications

A threat actor reportedly accessed OpenAI's internal messaging systems and stole information about AI technology from employee chats: a reminder that AI vendors and adopters alike face classic security and insider-threat challenges alongside new AI-specific risks.[6]

Copilot-style tools and prompt-driven data theft

Researchers have demonstrated prompt-injection and related attacks against AI assistants integrated into productivity suites, showing how user prompts and document context can be abused to exfiltrate data when controls are weak. Deployment and policy matter as much as the model.[7]

The ROI problem is real, and solvable

~95%	MIT's Project NANDA studied 300 public AI deployments, surveyed 350 employees, and interviewed 150 leaders: roughly 95% of enterprise generative-AI pilots deliver little to no measurable P&L; impact. The root cause is not model quality; it is flawed enterprise integration and a "learning gap" in tools and organizations.[1]
~67% vs 1/3	The same research found externally purchased and partnered AI solutions succeed about 67% of the time, while purely internal builds succeed only about one-third as often. Experienced outside help materially changes the odds.[1]
30%+	Gartner predicted at least 30% of generative-AI projects would be abandoned after proof of concept by the end of 2025, citing poor data quality, inadequate risk controls, escalating costs, and unclear business value, with deployment costs ranging from $5 million to $20 million.[2]

Doing AI right means picking a concrete pain point, integrating deeply with the workflows that exist, and putting risk controls in from the start: precisely the combination the failed pilots lacked. That is how we structure every engagement.

Component one: AI Security Assessment

- Review of AI/LLM integrations for prompt injection, data leakage, insecure tool and agent permissions, model supply-chain risk, and unsafe output handling.
- AI usage governance: what data reaches AI tools, under what controls, with clear boundaries between external APIs, corporate AI, and on-prem models.
- Recommendations aligned to the OWASP LLM Top 10 and the NIST AI Risk Management Framework.
- Model hosting, privacy, and compliance alignment reviewed in context.
- Prioritized, actionable remediation guidance you can sustain.

Component two: Automated Workflow Development

- AI-assisted automation for security operations: triage, enrichment, reporting, and repetitive investigative work.
- Integration with your existing security tooling to reduce analyst workload and response time.
- Guardrails built into the automation itself, so the tools we create do not reintroduce the risks being assessed.
- Secure delivery: authentication, secrets, dependencies, and production readiness.
- Pragmatic operational documentation so the automations survive team turnover.

How we work with you

You are matched with a **named Fletch Labs representative** who leads the assessment, designs the automation with your team, and remains your point of contact as models and tooling evolve. Recommendations are vendor-neutral (we do not resell or mandate specific products), and governance is sized to your organization, not paperwork for its own sake.

White-label arrangements for MSPs

Clients are adopting AI faster than most managed service providers and consultancies can staff for it. We support those partners with **white-label delivery**: the same assessment and automation work described above, performed as an extension of your practice rather than as a competing vendor.

- **Assessments delivered under your brand**: findings, remediation guidance, and executive summaries issued in your templates and your voice.
- **Automation built for your stack**: workflows we develop can ship as part of your managed offering, documented for your engineers to operate.
- **Your client relationship stays yours**: we work through you, and we do not approach or solicit your clients.
- **Flexible visibility**: fully behind the scenes, or introduced as a named specialist on your team when a client wants to meet the practitioner.
- **Confidentiality in writing**: partner NDAs and non-solicitation terms are standard, not an upsell.

It is a practical way to add an AI security line to your catalogue without building the specialization in-house, and to keep the client, the contract, and the margin.

Engagement & pricing

A scoped assessment paired with an automation build, with a retainer available for ongoing automation maintenance as your environment changes. Pricing is subject to the scale and scope of the arrangement, which we identify together in a discovery call, and provide a tailored quote for your requirements.

Start a conversation

Start with a scoped AI security assessment, an automation build, or both. Tell us what you have adopted or built and we will map the risks and the opportunity. If you are an MSP or consultancy, ask about white-label delivery under your brand.

Reach us at info@fletchlabs.ca or through fletchlabs.ca/contact.

Sources

1. MIT Project NANDA, The GenAI Divide: State of AI in Business 2025 (reported by Fortune) — https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/

2. Gartner press release, July 29, 2024 — https://www.gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025

3. The Register, Samsung ChatGPT leak coverage — https://www.theregister.com/2023/04/06/samsung_reportedly_leaked_its_own/

4. Mashable, Samsung ChatGPT leak details — https://mashable.com/article/samsung-chatgpt-leak-details

5. TechCrunch, Microsoft AI researchers' data exposure — https://techcrunch.com/2023/09/18/microsoft-ai-researchers-accidentally-exposed-terabytes-of-internal-sensitive-data/

6. Reuters, OpenAI internal breach report — https://www.reuters.com/technology/cybersecurity/openais-internal-ai-details-stolen-2023-breach-nyt-reports-2024-07-05/

7. OECD AI Incidents Monitor, Microsoft Copilot vulnerability record — https://oecd.ai/en/incidents/2024-08-27-975d