FAIRDINKUM

LEVERAGING INNOVATION WITH TECHNOLOGY

AI Gone Rogue?

When It Comes to Generative and Agentic AI, We Don’t Always Know Who’s Controlling the Machines

A year ago, every conversation in cybersecurity included some discussion about the concerns around generative AI. Today, Agentic AI has been added to the conversation.

Both AI tools are becoming more common in the workplace, each bringing good (like automating preventive security procedures) and bad (like Shadow AI) for cybersecurity teams.

AI is a disruptor, and organizations that don’t adopt the technology will lose a competitive edge and fall behind. But like with any type of technological transformation, the approach to onboarding generative and agentic AI must be done thoughtfully, understanding both the strengths and weaknesses and, perhaps most importantly, recognizing its wide-ranging cybersecurity risks.

A growing number of security experts see AI as the next big attack vector, but before integrating the technology into their systems, organizations need to understand how generative and agentic are similar yet different, the security risks, and how to best use them as a security tool.

Think of AI agents as a digital bodyguard for your network.

Generative AI Brings AI to the Mainstream

AI isn't new. Artificial intelligence was first introduced in the 1950s, and it was first used as a cybersecurity tool in the 1980s. Tools like Alexa and Siri and chatbots brought the power of AI into the hands of everyday users, but it was the introduction of generative AI in 2022 that changed the way most of us see the technology.

Both generative and agentic AI are built on large language models (LLM) that are trained on large data sets which drive intelligence. They both depend on prompts to determine their output and can track conversation and prompt history, as well as simulate decision making processes.

What makes generative AI different from other AI tools is its ability to create. It can produce something—text, images, music, code—based on prompts that are trained into the system.

Smart companies are jumping on the AI boom to improve their defenses and automation.

Agentic AI Poised to Become a Major Player

Agentic AI builds on generative AI capabilities, but where gen AI is about creation, agentic AI is about goals. *"A generative AI model like OpenAI's ChatGPT might produce text, images or code, but an agentic AI system can use that generated content to complete complex tasks autonomously by calling external tools,"* according to IBM.

Think of AI agents as a digital bodyguard for your network. They are trained to monitor for any anomalies that are found in the system and across devices, including unusual user behaviors, and can then address those anomalies without human interaction. Agentic AI can quarantine threats, alert security teams and roll back the system to pre-threat configuration.

Both types of AI are useful for security teams, and for the best defense, they work in tandem. For example, in a phishing attack, generative AI can read the email, create a summary of the attack and draft a report about the incident for the security team. Agentic AI takes action to quarantine the message, blocks the sender and adjusts to add this information to improve its detection capabilities.

Jen Easterly, former CISA director, told an audience at SailPoint Navigate conference in the fall of 2025 that the new generations of AI offer the best opportunity to find and fix flaws in software. Smart companies are jumping on the AI boom to improve their defenses and automation.

However, cybercriminals also see advantages of generative and agentic AIs. Gen AI allows more professional-looking phishing attacks and makes it easier to write malicious code. Agentic AI is used to launch and execute complex cyberattacks autonomously.

Threat actors are increasingly weaponizing both generative AI and agentic AI, not just to scale existing tactics but to invent new, highly targeted attacks.

The Differing Security Threats

The two AI formats complement each other, but they do have very different functions. In addition to its primary role of creating content, generative AI is *reactive* intelligence. It can't do anything until it has a prompt and because it is focused on immediate output, its memory is short term, and tools are limited in capabilities.

AI agents, on the other hand, are all about *action*—executing tasks and making decisions. It is proactive intelligence as it works towards programmed goals and maintains long-term context so it can adapt and revise as tasks require.

The Security Risks around Each Type of AI Differ as Well

Generative AI security risks include:

- Misinformation, disinformation and deepfake
- Shadow AI use and unknown sharing of sensitive data
- Malicious code generation
- Prompt injection attacks and insecure code generation
- Authentication exploits

Security risks for agentic AI include:

- Supply chain vulnerabilities
- API overreach
- Privilege escalation
- Compromised identity of AI agents
- Indirect prompt injection

Threat Actors Use AI Too

Threat actors are increasingly weaponizing both generative AI and agentic AI, not just to scale existing tactics around the technology, but to invent new and highly targeted attacks.

Hyper-personalized and more realistic phishing attacks may be the most familiar AI-generated attacks. And they aren't just coming via email. Thanks to deepfakes, AI has been used to mimic celebrities and politicians—or your boss and co-workers—in voice calls and in social media videos.

Generative AI tools have made it easy for threat actors to build malicious code, but now security researchers have found that the bad guys are building entire infrastructures to launch attacks and then destroy the infrastructure to wipe their tracks and avoid detection.

An increasing number of cyberattacks involve credential theft, so it isn't surprising that threat actors are using chatbots as a way to harvest credentials of both human and non-human identities—and that includes AI agent identities. Yes, generative AI is being used nefariously to attack agentic AI.

Cybercriminals have been developing their own as-a-service options, like ransomware-as-a-service. They now use agentic bots for attacks-as-a-service, with predefined autonomized attack playbooks.

Protecting Your AI from the Bad Guys

In addition to the general cybersecurity tools every company should have in place (endpoint detection and response, security information and event management, identity and access management, and data loss prevention), there are some **AI-specific security tools,** as well as vital governance tools that are designed to meet the unique challenges that generative and agentic systems present.

These tools include:

- **Model integrity monitoring,** which detects any unauthorized changes to AI models.
- **Adversarial input detections**, which identifies inputs designed to exploit AI models for adversarial purposes.
- **Prompt injection defense for LLMs,** which filters inputs to prevent manipulation of data.
- **Model watermarking and/or fingerprinting,** which inserts identifiers of authorized users into the AI models.
- **AI red teaming platforms,** which simulate attacks to find vulnerabilities in the AI models.
- **Policy enforcement engines,** which apply governance and compliance rules across the AI model.

Adding generative and agentic AI will accelerate the capabilities of [existing cybersecurity tools] that can go at the scale and pace of the current threat landscape.

Platforms like Azure offer multi-layered AI cybersecurity that combines many of the tools listed above. For example, Azure's Defender for Cloud includes real-time threat detection and alerts for generative AI models. Governance is covered with role-based access control, audit logging and monitoring and token scanning. Overall platform security for AI workloads includes creating security baselines for AI resources, prompt injection defense and model integrity protection.

Using AI to Build Better Cyber Security Defenses

Whether you are an existing Fairdinkum customer or have just recently come to know us, your existing cybersecurity program likely uses some forms of AI/ML for continuous monitoring and detection. Adding generative and agentic AI will accelerate the capabilities of these tools that can go at scale and pace of the current threat landscape. Specific action plans you and your IT provider should or may already be considering adopting to address the generative and agentic AI risks your organization will face are:

- **Define and establish governance and ethical boundaries.** Governance frameworks detail all of the stakeholders who oversee AI deployment, security and compliance.

- **Know the AI use in your organization and develop risk scores** for use cases. That means decreasing the use of Shadow AI and auditing existing AI deployments to know where and how AI models are used across the company.

- **Build resilience across controls**. In addition to the recommended tools designed to detect, prevent and mitigate threats to AI, you also need a workforce that understands the risks. AI-centric security training for employees and third-parties is essential.

AI is already a gamechanger, and companies who aren't adopting generative and agentic AI are going to fall behind. That's why you'll see and hear more about AI as a security tool from Fairdinkum in 2026. But like any digital transformation, AI adoption requires understanding the risks as well as the benefits—don't forget, that anything you can do with AI, threat actors are doing too to get inside your network. ***It's our job to help you use AI tools to stay one step ahead.***

FAIRDINKUM

LEVERAGING INNOVATION WITH TECHNOLOGY