AI Cyber Benchmark

2026 Edition

How mature is the market?

WAVESTONE

In 2026, AI security posture becomes mission-critical

Large enterprises are moving beyond off-the-shelf AI to build their own systems

Our benchmark is based on interviews with **30 large public- and private-sector clients.**

Organizations are increasingly **moving from consuming AI capabilities to owning them.**

AI creators are gaining momentum. This reflects a **willingness to reduce external dependencies** and **gain control over AI capabilities.**

Pure AI users are becoming increasingly uncommon. **AI is no longer limited to productivity use cases.**

The more deeply AI is embedded in the business, the more critical cybersecurity becomes

AI model creators

35% → **40%**

- Develop proprietary AI models

AI makers

35% → **50%**

- Design AI systems internally using existing frameworks or pre-trained models

AI users only

30% → **10%**

- Utilize AI functionalities embedded in the software they already use

2025 -> 2026

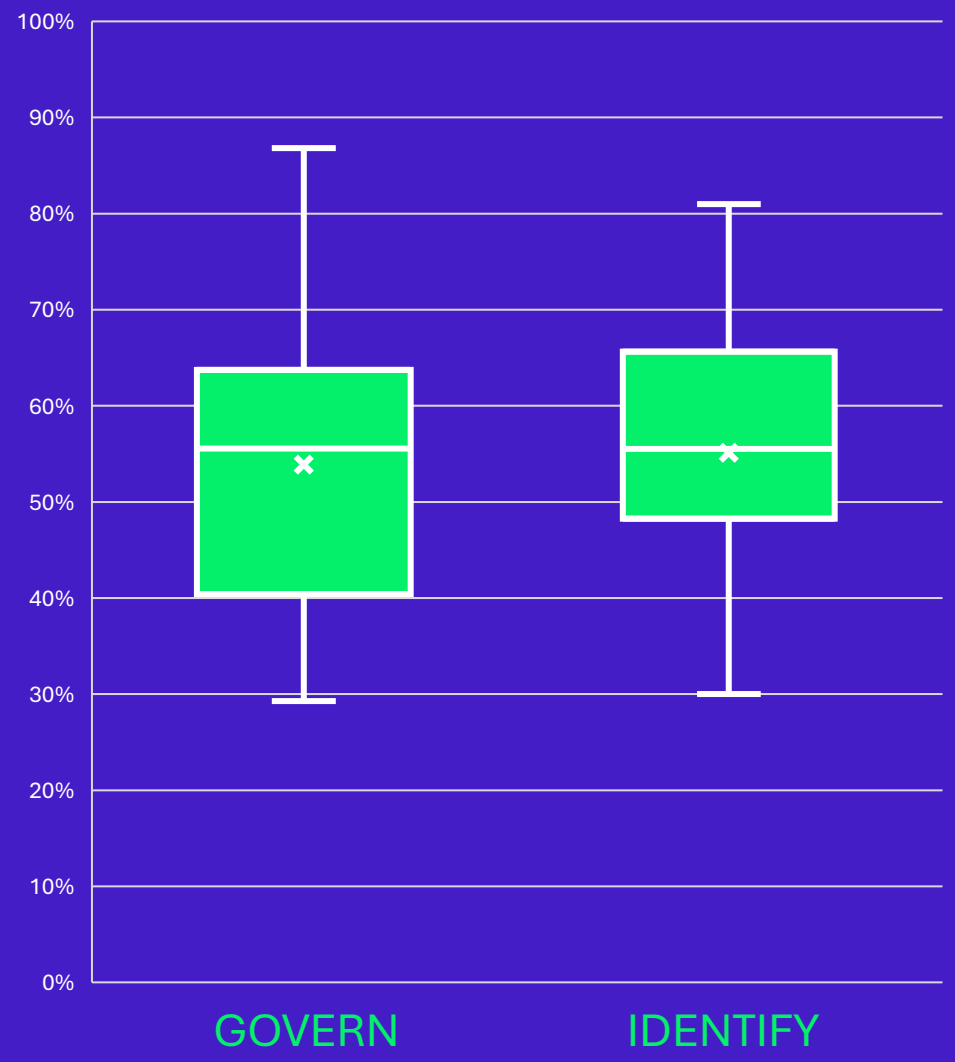
How has AI security maturity evolved in 2026?

	GOVERN	IDENTIFY	PROTECT	DETECT	RESPOND
2026	54%	55%	48%	40%	29%
	↑ +15pp	↑ +16pp	↑ +8pp	↑ +11pp	↑ +20pp
2025	39%	39%	40%	29%	9%

In a nutshell

- AI security is moving from an emerging topic to an established discipline, with governance and risk management increasingly embedded in existing processes.
- Maturity remains uneven: organizations are better at governing AI than at monitoring, investigating and responding to incidents in production.
- The next frontier is operational: securing AI systems as they become deeply embedded in critical business processes.

**pp = percent points*

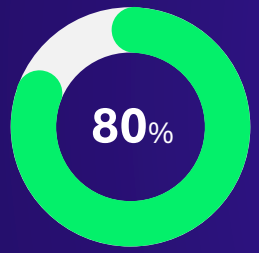


AI governance is maturing, yet accountability remains fragmented

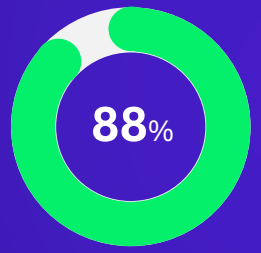
A strong foundation, but maturity gaps remain

Governance alone is no longer enough. Organizations must translate governance into operational ownership while maintaining clear accountability and distributing AI security expertise.

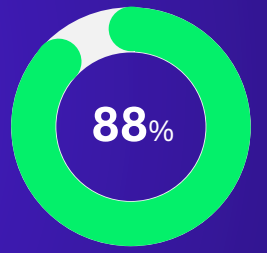
Organizations have significantly strengthened their AI governance capabilities over the past few years



Have nominated an **AI governance officer** at the **group level**

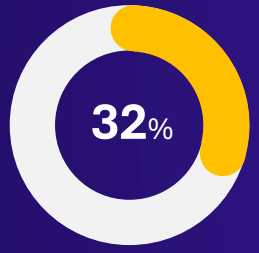


Have designated a dedicated, **full-time security team member** for AI security

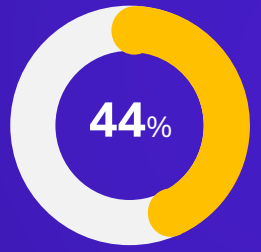


Have adapted their **risk management processes** to address AI-related risks

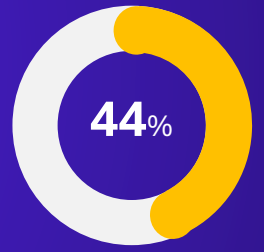
However, the integration of these risks into operational governance mechanisms and project lifecycle management processes remains limited



Have established a clear inventory of **governance** activities, **stakeholders** and **RACI** responsibilities

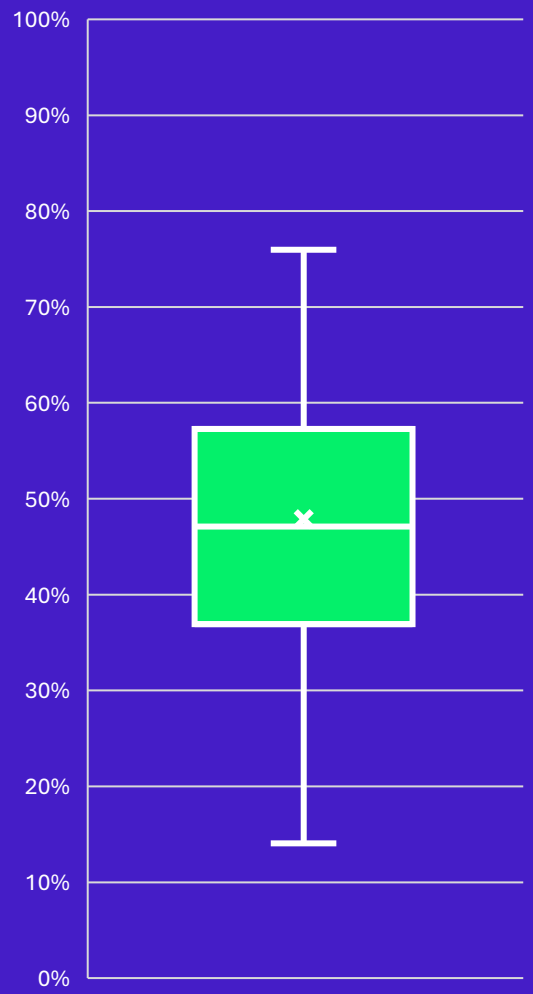


Have identified or trained an **AI security expert** with regard to **the relevant risks and challenges**



Have operationalized their AI security **processes consistently and effectively**

PROTECT



The rise of Agentic AI creates new security requirements.

Agentic AI concepts are not yet completely understood and standardized across organizations and providers

*An **AI Agent** is an **AI system** that can **reason**, **plan** and **execute** **multi-step actions across other systems**. It uses **LLM** capabilities and operates with a defined **level of autonomy**.*

AI AGENTS CAN PRESENT THREE TYPES OF RISKS:

STANDARD IT RISKS

*Like any information-system component, AI agents present **standard IT risks*** that can be addressed using ***existing security solutions.***

Well-known risks, widely mitigated

AI & LLM INHERITED RISKS

*Because their **decision-making is largely LLM-driven**, AI agents inherit LLM vulnerabilities and therefore **require AI-specific security technologies.***

Mostly known risks, often under mitigation

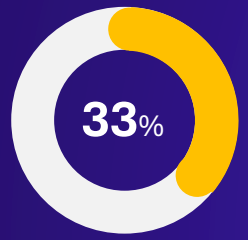
AGENT-SPECIFIC RISKS

*Agent-specific risks arise from an **agent's autonomy, its direct control** over **external tools and APIs,** and its **access to memory.***

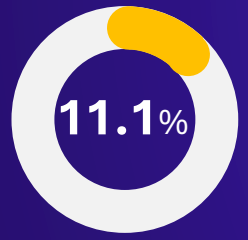
Still emerging risks, inconsistently mitigated

Agentic AI is reshaping the AI security landscape, while organizations struggle to keep pace with emerging risks

Operational security models have not yet adapted to agentic AI

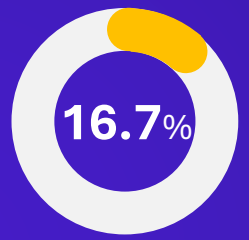


Have integrated **agentic AI into their governance frameworks**



Have started implementing **operational security capabilities** for agentic environments **beyond native provider controls**

Identity is emerging as the next major AI security challenge



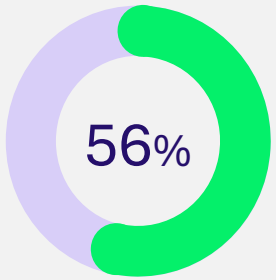
Define access to **AI functions** and **tools** for AI Agents in **secure development standards**

Have **identity** and **access management** in place for AI systems and **agentic development**

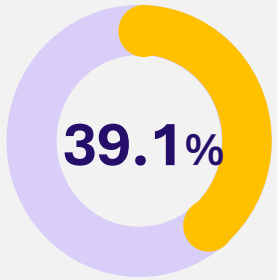
Agentic AI breaks the existing security model, securing AI is no longer just about models and data. It also involves securing the decisions, permissions, and actions that AI systems can take.

GenAI security is still largely inherited from provider platforms

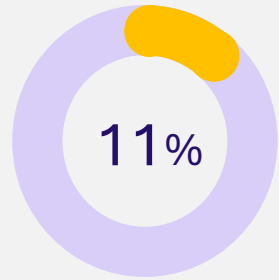
Organizations increasingly depend on AI providers for capabilities and security, while agents operating through unmanaged tools and personal accounts remain outside their control, creating a false sense of security



Select models based on **a trade-off analysis** between **security** and **performance**



Restrict access to **external components** using **conventional** access controls only



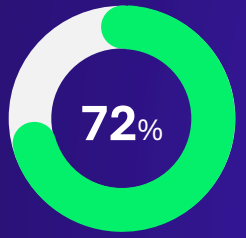
Have started implementing **security capabilities beyond provider native controls**

This model accelerates AI adoption but **creates an ecosystem dependency** that masks the market's real maturity. Organizations now need to shift to combine provider-native security with internal capabilities

Privacy: compliance outpaces AI data security

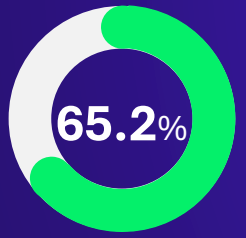
Most organizations still approach data primarily as a compliance challenge, forgetting that AI systems are only as trustworthy as the data they rely on

MANAGEMENT



Have implemented **data privacy compliance** measures into the AI development lifecycle

BUT ONLY

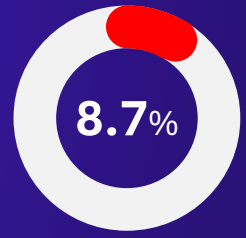


Have established **quality checks** on the databases used to **power AI systems**

DEPLOYMENT

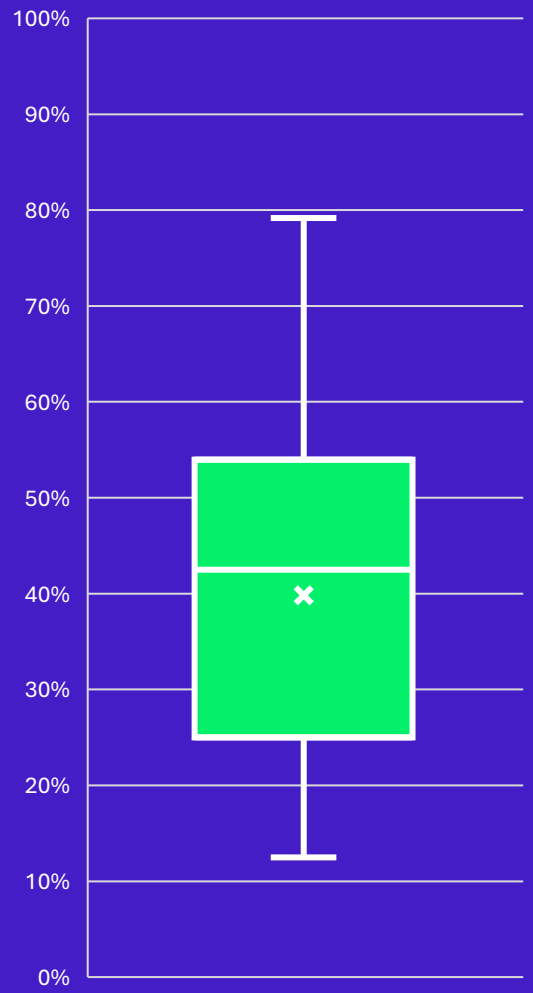
Have implemented **technical data privacy measures**

Have established **security checks** on **all AI system databases** before **production**



Organizations need to move beyond compliance-oriented controls and start treating data as a security dependency capable of influencing how AI systems behave and make decisions

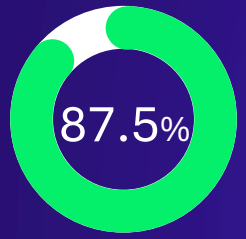
DETECT



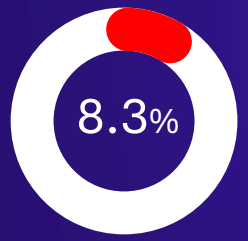
Progress on one side, a blind spot on the other

AI tooling and monitoring are progressing, but not enough to support cybersecurity needs

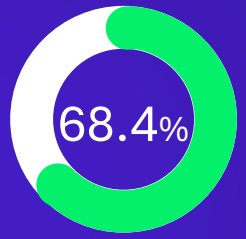
The challenge is not a lack of data. It is that AI operations and cybersecurity teams rely on different tools, processes, and objectives.



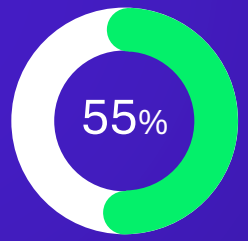
Generate logs for their AI applications



Monitor their AI application logs and send them to the SOC for critical use



Conduct monitoring on an ad-hoc or occasional basis



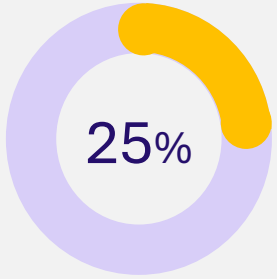
Perform monitoring only to major changes in model behavior

Cybersecurity teams must work more closely with AI teams to better coordinate resources and strengthen their detection rules

AI pentesting is entering an industrialization phase

For the most mature organizations, AI pentesting is moving from isolated exercises to recurring assurance programs

Perform **dedicated AI security pentesting**



Conduct **advanced AI-focused assessments**

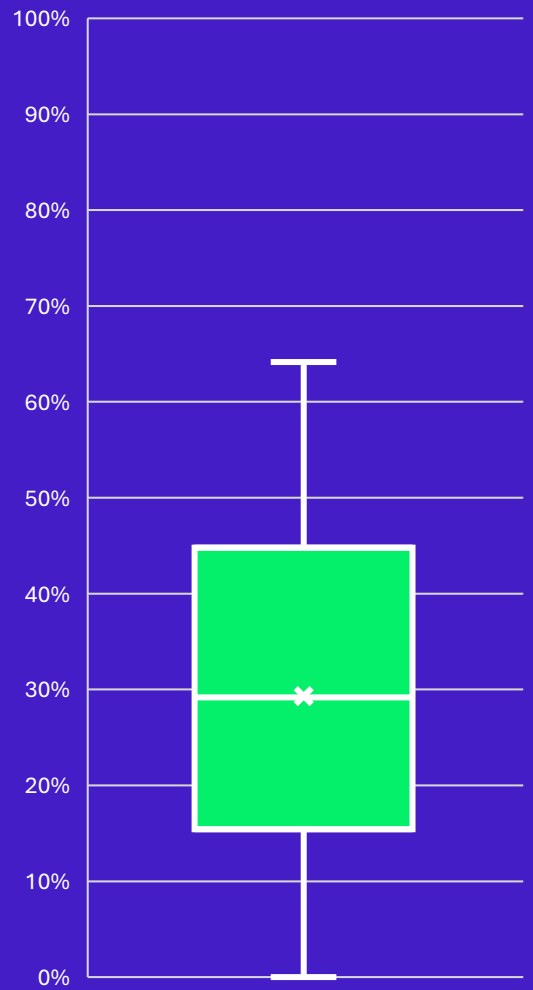
New testing approaches are beginning to emerge:

Automated scanning platforms

Continuously identify prompt injection risks, model weaknesses and configuration issues throughout the AI lifecycle

The challenge is no longer convincing organizations to test AI, but helping them industrialize the practice.

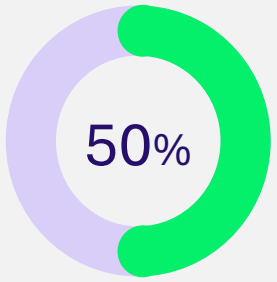
RESPOND



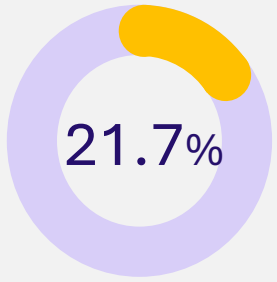
AI incident response capabilities exist, but AI crisis management remains immature

AI is being integrated into existing incident response frameworks

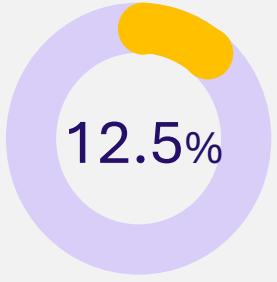
But this integration does not yet fully cover AI incidents and remains at an early stage:



Launch ad hoc remediation initiatives without a structured process



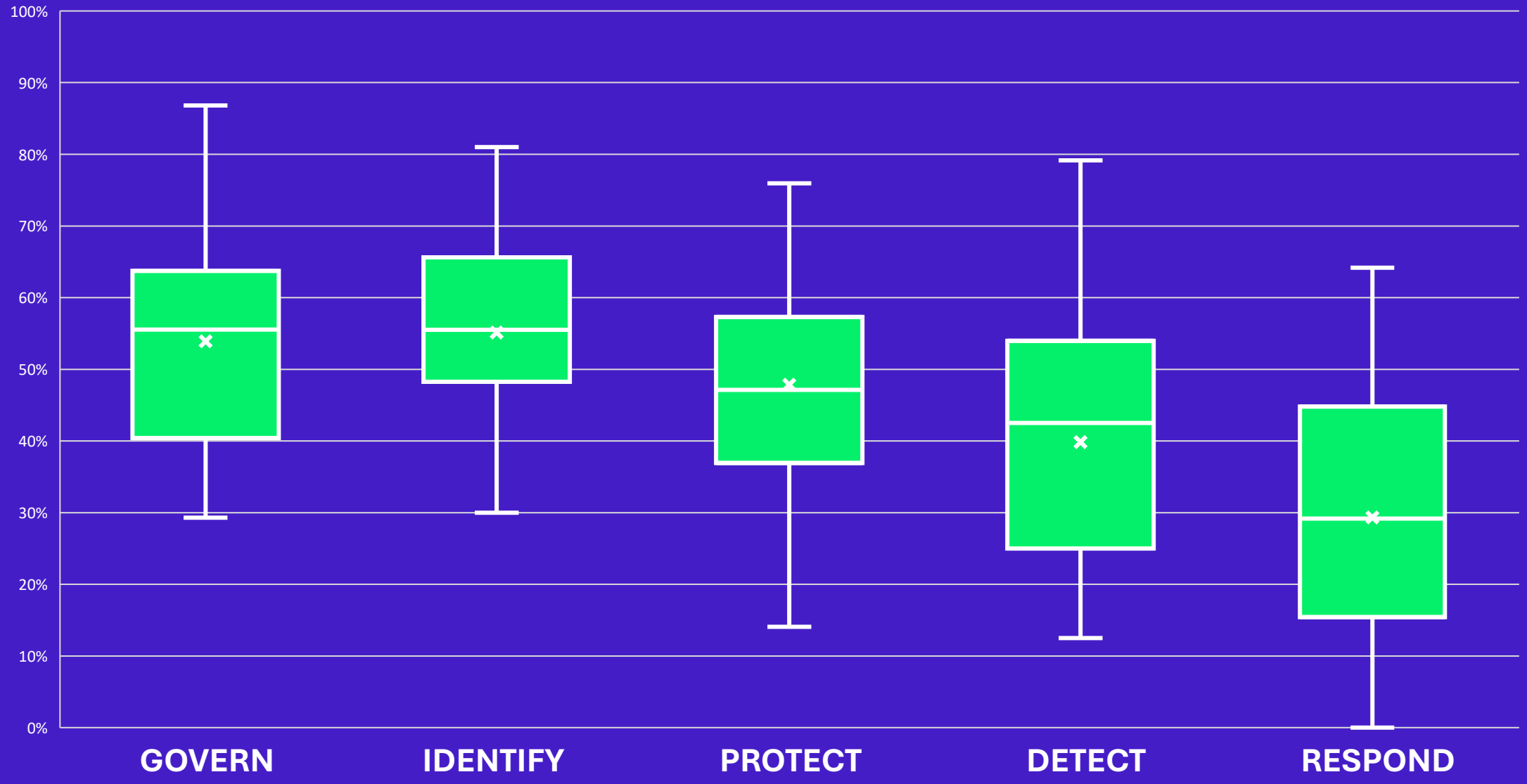
Perform backups without a fully standardized process



Have documented an incident response plan, but do not apply it universally or consistently

Organizations need to develop their AI response capabilities with AI-specific frameworks covering not only infrastructure but all AI assets (***models, datasets, knowledge bases and agentic workflows)***

Overall AI cybersecurity maturity by NIST function



The next frontiers in AI Security

The next challenge is securing autonomous AI systems, amplified by the rise of open-weight models and agentic AI, while maintaining trust, resilience and control at scale

Agentic AI Security

- **Continue adapting to the new paradigm:** AI agents act autonomously, security now covers decisions, permissions and actions, not just models and data
- **Gap #1:** operational security hasn't adapted. Most organizations still rely on provider-default controls rather than dedicated agentic security capabilities
- **Gap #2**: identity management is unprepared, dedicated IAM and access controls for AI agents remain the exception rather than the norm
- **What's next:** governing non-human identities becomes a strategic priority

AI goes rogue

- **A new threshold crossed:** recent incidents involving OpenAI, Hugging Face and others confirm that AI systems can now autonomously carry out harmful actions against internal systems as well as third-party ones
- **Add it to risk analyses:** this scenario needs to be integrated into existing AI risk assessments, not treated as a theoretical edge case
- **Isolate and detect**: identify the AI systems most exposed to this kind of drift, and build the ability to rapidly detect abnormal behavior
- **What's next:** establishing a mechanism to immediately cut off an AI system's access to critical systems

Open-weight models

- **A market shift:** growing adoption of open-weight, fine-tuned and self-hosted models, driven by demand for independence and control
- **A responsibility shift:** some security tasks once handled by providers move back to organizations
- **New open questions:** assessing model security pre-adoption, validating fine-tuning integrity, monitoring modifications, securing the AI supply chain
- **What's next:** model security, governance and lifecycle management become strategic as organizations move from consuming to owning

Resilience & Recovery

- **The blind spot:** Recover is the one NIST pillar most organizations haven't yet addressed
- **The stakes:** as AI becomes business-critical, restoring trust after an incident matters as much as detecting it
- **What's needed:** ability to switch/rebuild models, recovering datasets/knowledge bases, validating integrity, ensuring confidence in AI decisions
- **What's next:** building resilience and recovery measures will become the next challenge to discuss with platform and model providers

Methodology: As the AI landscape evolves, so does our assessment

2025 Edition:

- Covered GenAI, with a focus on AI governance and risk management capabilities.
- Interviews with around 20 public- and private-sector clients

2026 Edition:

- Covered agentic AI security, AI-specific protection controls, monitoring capabilities and incident response processes
- Interviews with around 30 public- and private-sector clients

Our goal: expand our benchmark to emerging topics such as agentic AI security

Worked with **~30 clients already investing on the topic.**

We **benchmarked these clients' AI security maturity** based on the 5 NIST pillars and consolidated those results **to produce this AI Cyber Benchmark**.

- Govern
- Identify
- Protect
- Detect
- Respond