IBM NCEE News Room

Trustworthy AI: Navigating Governance and Security Risks in Enterprise AI Projects

Why AI Success Depends on More Than Just Algorithms — It Requires Oversight, Security, and Strategy.

AI is not just technology – it's a Strategic Business Risk

Artificial Intelligence (AI) is rapidly transforming business operations, decision-making, and customer interactions. But with this transformation comes a new class of risks – ones that are not purely technical, but deeply strategic. AI can generate misleading content, operate without oversight, and expose sensitive data through misconfigurations. These risks are not confined to IT departments; they directly impact regulatory compliance, customer trust, and business continuity.

> *"AI doesn't fail because of bad algorithms — it fails because of unclear ownership, poor governance, and unmanaged risk. Trustworthy AI starts with visibility, accountability, and action."*

Leadership cannot manage what it cannot see. That's why trustworthy AI must be built on a foundation of governance and security – not just algorithms and dashboards.

Governance vs. Security: Two Sides of AI Risk

To understand the full spectrum of AI-related risks, it's essential to distinguish between**AI Governance** and **AI Security**.

AI Governance	AI Security
Focuses on oversight, accountability, and process integration	Focuses on defending against cyber threats and malicious actors
Deals with passive anomalies (e.g., QA misses, undocumented usage)	Deals with active anomalies (e.g., prompt injection, system abuse)
Requires clear ownership, documentation, and auditability	Requires threat detection, escalation, and technical remediation
Strategic risk: flawed decisions, compliance gaps	Operational risk: data breaches, system compromise

"How AI Can Misbehave" clearly illustrates this distinction, categorizing risks into passive (Governance) and active (Security) anomalies.

Why AI Projects Fail – And What That Tells Us

According to Gartner, **over 40% of agentic AI projects will be cancelled by the end of 2027**, primarily due to escalating costs, unclear business value, and inadequate risk controls

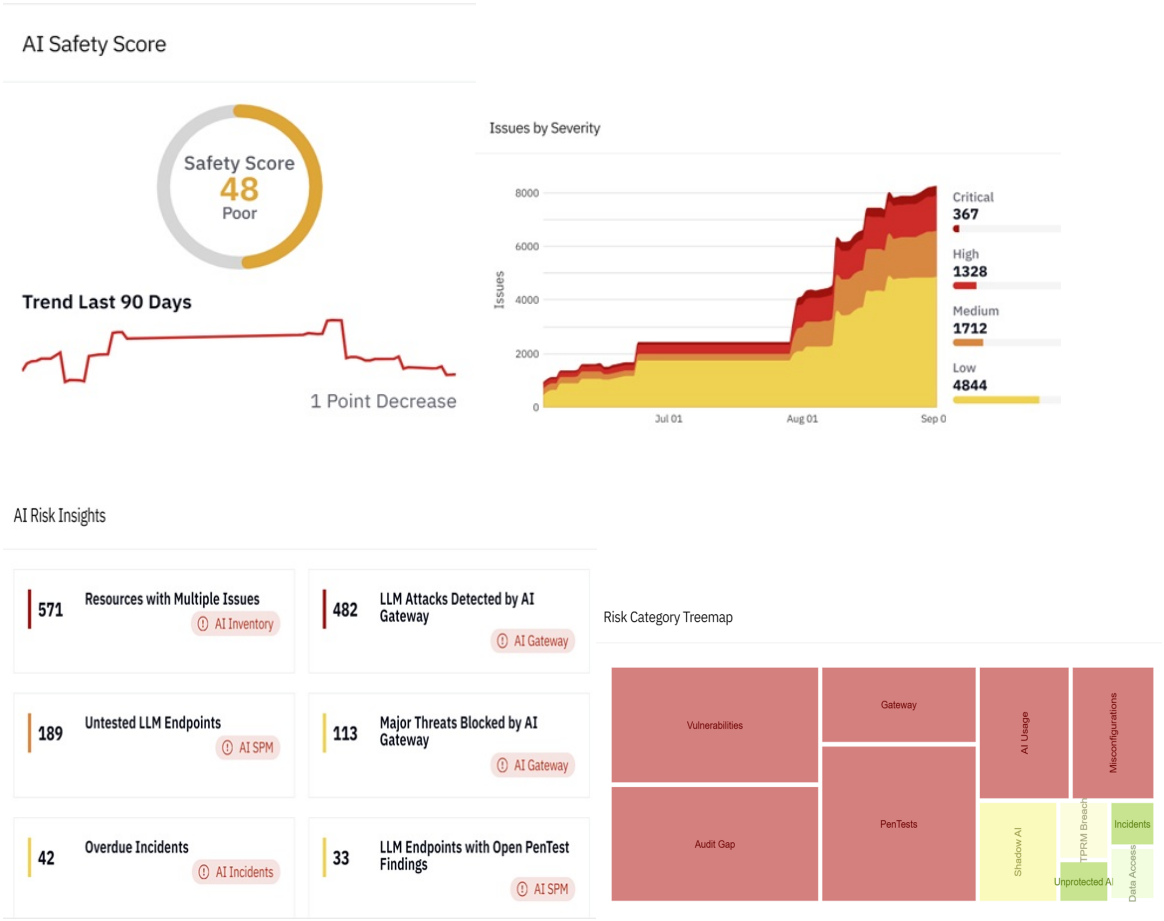
These are not technical failures – they are governance failures. When AI is deployed without clear ownership, defined

outcomes, or risk management structures, it is destined to stall or collapse.

From Dashboards to Decisions: What AI Security Metrics Really Reveal

AI dashboards offer a wealth of technical insights – but without context, they risk becoming noise. Let's explore what these dashboards reveal, and why leadership must interpret them through both a **security** and **governance** lens.

1. AI Safety Score: A Warning Signal, Not Just a Metric

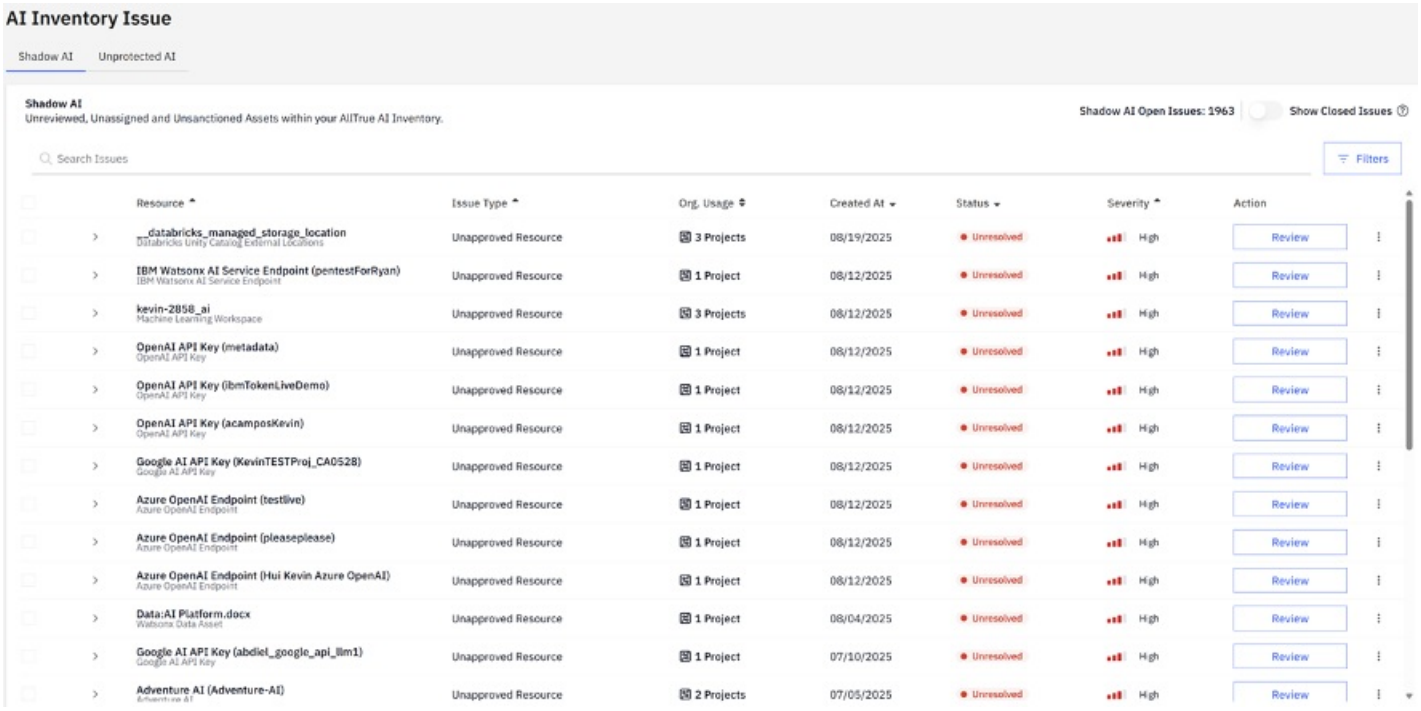


AI Safety Score Dashboard

- **Safety Score: 48 (Poor)**
- **Critical issues: 367**, **High: 1328**, **Medium: 1712**
- **LLM Attacks Detected: 482**, **Major Threats Blocked: 1139**

These numbers reflect systemic issues in how AI is deployed and governed. A poor safety score is not an IT problem – it's a business risk.

2. Shadow AI: The Governance Gap in Plain Sight



AI Inventory Issue Dashboard

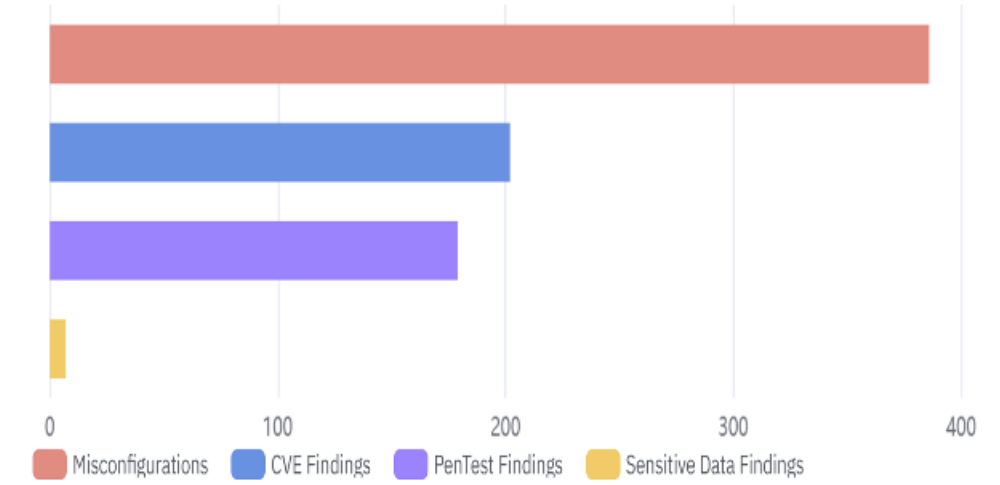
The dashboard lists **unapproved resources** such as OpenAI API keys and Azure endpoints, all marked as **“Unresolved”** and **“High” severity**. Tabs like **“Shadow AI”** and **“Unprotected AI”** highlight the lack of visibility and control.

Example: A marketing team uses an external AI tool with customer data. No one approved it. No one anonymized the data. The result? A GDPR violation and reputational fallout.

Shadow AI is not a rogue behavior – it’s a governance failure.

3. Security Posture Management: Misconfigurations with Business Consequences

Open Issues by Type



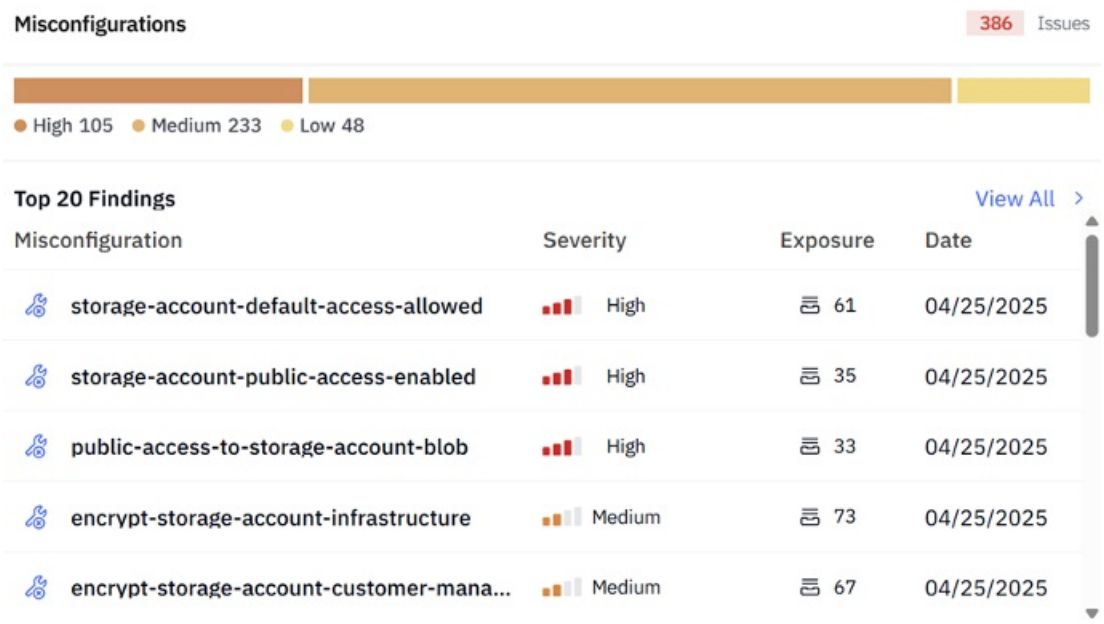
Sensitive Data Findings

7 Issues

Top 20 Findings

View All >

Finding	Severity	Exposure	Date
Sensitive Information Found in Dataset	Medium	2	08/09/2025
Sensitive Information Found in Notebook Ce	Medium	1	06/11/2025
Sensitive Information Found in Notebook Ce	Medium	1	06/11/2025
Sensitive Information Found in Notebook Ce	Medium	1	06/11/2025
Sensitive Information Found in Notebook Ce	Medium	1	06/11/2025



AI Security Posture Dashboard

- **Misconfigurations: 386**
- **Sensitive Data Findings: 40**
- **High-severity issues** include:
  - storage-account-default-access-allowed
  - public-access-to-storage-account-blob

Example: A storage account used by an AI model is publicly accessible and contains customer identifiers. The dashboard flagged it – but no one was responsible for acting on it.

4. Why Dashboards Alone Are Not Enough

Across all these dashboards, one theme emerges: **visibility without governance is not protection**. Dashboards can show you where the risks are – but they can't tell you:

- Who owns the AI model?
- What business process it supports?
- Whether it has been audited?
- If its use is even approved?

This is the governance gap. And it's where many AI projects fail – not because the technology is flawed, but because the organization lacks the structures to manage it.

Conclusion: From Technical Metrics to Strategic Risk Management

AI dashboards are powerful tools – but they are only as valuable as the decisions they inform. Misconfigurations, Shadow AI, and sensitive data exposures are not just technical anomalies. They are signals of deeper organizational vulnerabilities.

To manage AI as a **strategic business risk**, organizations must move beyond technical monitoring and embrace **business-driven governance and security**.

Three Imperatives for Leadership

- **Continuously evaluate AI usage through processes and accountability**
- **Identify and address Shadow AI – and understand its root causes**
- **Expand Security Posture Management into a strategic tool**

Trustworthy AI is not just secure – it is governed.

And governance is not just policy – it is visibility, accountability, and action

---

https://ncee.newsroom.ibm.com/news?item=122454