Building Trust in the Age of AI:

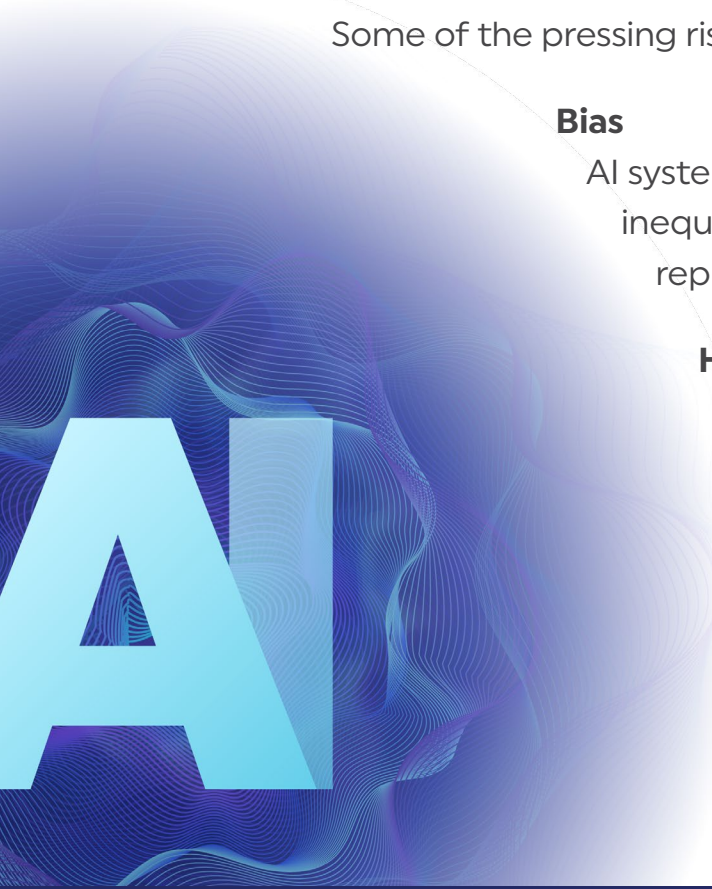
A Practical Approach to Secure and Responsible AI

HITRUST®

The AI Opportunity and Emerging Risk Landscape

AI is reshaping how businesses operate. Generative AI is gaining traction for its ability to automate tasks, create content, and drive smarter decisions across industries. From finance and healthcare to retail and logistics, organizations are tapping into AI to unlock efficiencies, reduce costs, and boost innovation.

But alongside this potential lies a complex and fast-evolving set of risks. Generative AI models and machine learning systems introduce new challenges that traditional security frameworks weren't built to address.



Some of the pressing risks include the following:

Bias

AI systems learn from the data they're trained on. If that data contains historical inequalities, underrepresented populations, or skewed outcomes, the AI is likely to replicate or even amplify those patterns, resulting in discriminatory decisions.

Hallucinations

Generative AI models are designed to produce humanlike responses. But they don't understand the truth in the way people do. They generate content by predicting likely word sequences, which means they can confidently produce fabricated information that can be dangerously misleading.

Data Poisoning

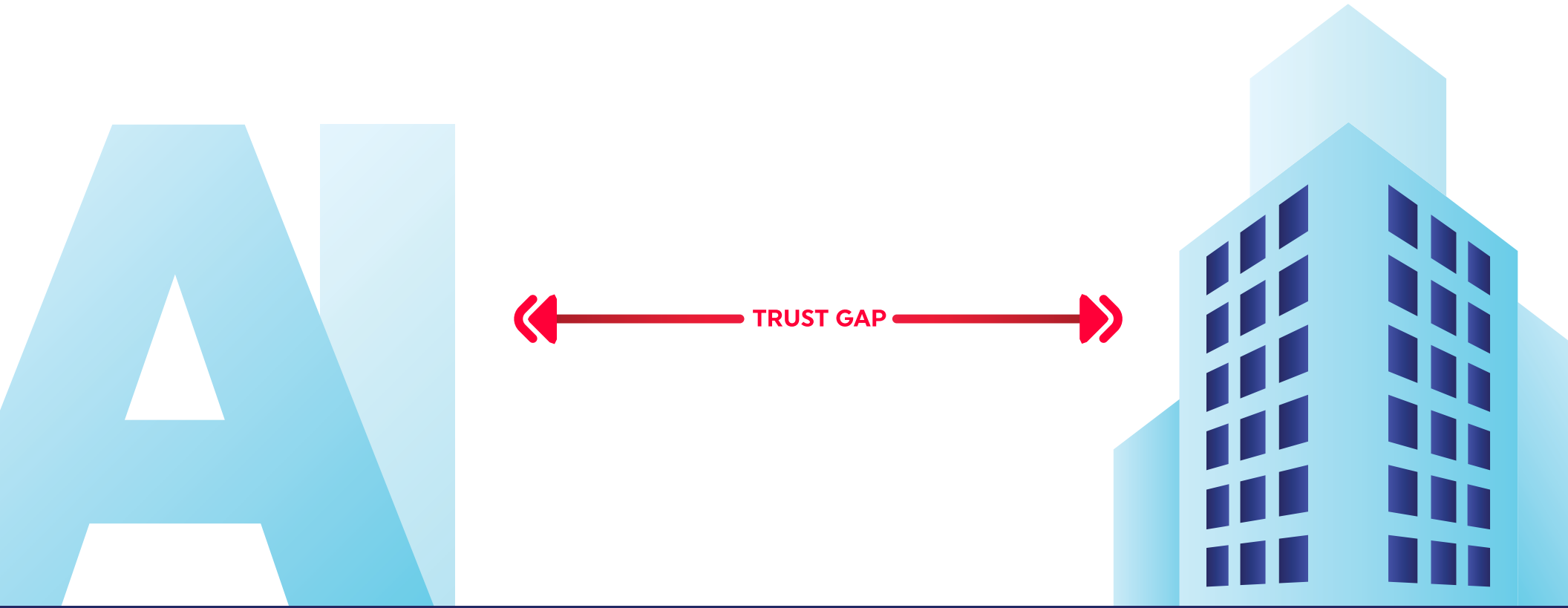
AI systems are only as reliable as the data they're trained on. Malicious actors can intentionally inject flawed, misleading, or manipulated data into a model's training pipeline, which can cause the model to behave incorrectly.

Regulatory Gaps

Only a few jurisdictions have clear, enforceable standards. This creates uncertainty for organizations about what compliance looks like, how liability should be handled, and what constitutes responsible use of AI. Without consistent rules, businesses navigate a patchwork of emerging policies, often at their own risk.

These issues contribute to a growing trust gap. Many organizations hesitate to scale the use of AI. The reason is simple: they're unsure whether they can trust what the AI is doing, how it's making decisions, and whether it's secure.

This hesitation isn't just about cybersecurity. It's about a new category of risk — one that requires deeper governance, broader vigilance, and smarter assurance.



Understanding Trustworthiness vs. Security in AI

Understanding trustworthiness and security is essential for organizations looking to deploy AI responsibly.

Trustworthiness in AI

Trustworthiness in AI refers to whether the system behaves in a way that is reliable, fair, and aligned with expectations. It's about the ability to confidently depend on the AI to do what it's supposed to in real-world situations.

Some of the pillars of a trustworthy AI system include:

- **Repeatability** – It produces consistent results when used in the same context.
- **Transparency** – It offers visibility into how decisions are made.
- **Fairness** – It avoids bias or unintended discrimination.
- **Accountability** – It can be governed, monitored, and improved over time.

Security in AI

Security is focused on protecting the AI system and the data, infrastructure, and processes that support it. It ensures that the system is guarded against threats and performs safely under a range of conditions.

Some of the key elements of AI security include:

- **Confidentiality** – Prevent unauthorized access to data, models, and systems.
- **Integrity** – Ensure data and model outputs are accurate and unaltered.
- **Availability** – Keep AI systems readily accessible and functional when needed.

Why Both Matter

Trustworthiness and security are different, but they are deeply connected. A system can be secure but still untrustworthy. For example, a system may be free from cyber threats but produce biased results. Conversely, a trustworthy model isn't helpful if it can be compromised, misused, or brought down by attackers.

In practice

Security ensures the **safety** of the AI system.

Trustworthiness ensures the **reliability** of its behavior.

Both must work together to give organizations, customers, and regulators confidence in AI.

Enter HITRUST — A Proven Approach to AI Assurance

Trust in AI doesn't happen automatically. It requires structure, management, and a commitment to transparency. That's where HITRUST comes in.

For more than 17 years, HITRUST has helped organizations manage security, compliance, and risk with confidence. Its widely adopted framework, the HITRUST CSF, has become a gold standard in sectors like healthcare, finance, and technology.

HITRUST is known for its ability to deliver

- Prescriptive, harmonized controls based on 60+ global standards.
- A frequently updated framework that evolves with emerging threats.
- Independent, third-party validated and centralized quality-assured assessments.
- Proven methodologies trusted by over 100 assessment firms.
- Scalable assessments enabled through its SaaS platform, MyCSF.

In 2024, 99.41% of HITRUST-certified environments reported no breaches as per the HITRUST 2025 Trust Report. That's not just a statistic — it's evidence that HITRUST reduces real-world risk.

Built for Today's Challenges

Now, HITRUST is applying its legacy of leadership and proven approach to AI.

AI introduces a new category of risk. HITRUST addresses it through a purpose-built assurance program that aligns with emerging standards, regulatory expectations, and practical business needs.

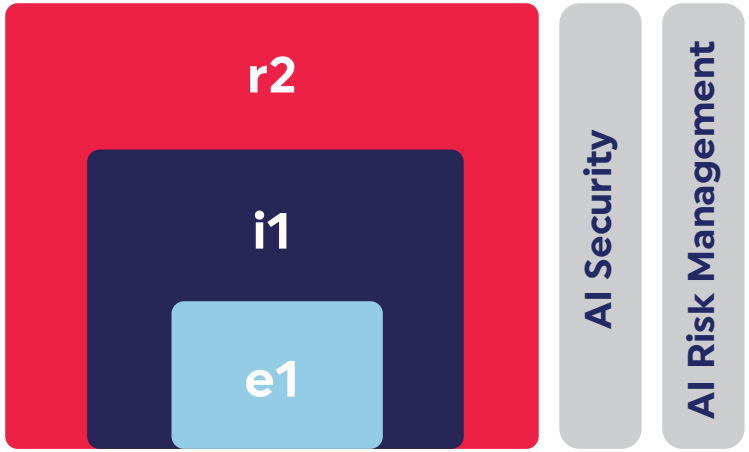
Organizations can now assess, manage, and certify AI systems using a framework they already trust.

The HITRUST AI Security Assessment and Certification

As AI systems move from experimentation to full-scale deployment, organizations need more than internal confidence — they need externally validated proof that their AI implementations are secure. This is where the HITRUST AI Security Assessment and Certification steps in. It offers formal, certified validation that an organization's deployed AI systems meet robust cybersecurity standards.

Built on HITRUST's widely trusted assessment model, the AI Security Certification is not a standalone offering. Instead, it is designed to be a part of a broader HITRUST e1, i1, or r2 assessment, ensuring that the AI system is evaluated in the context of its full technical environment.

Systems undergo independent testing and centralized quality review before getting a certification to share with customers, regulators, and business partners.



AI Security

HITRUST®
CERTIFIED

Security Focus

The AI Security Assessment is built around AI security topics such as:

- AI security threat management and governance
- Development practices for AI software
- Legal and compliance obligations specific to AI
- Resilience of the AI system

The control set features 44 prescriptive requirements focused on the following:

- Securing model integrity (e.g., adversarial attacks and poisoning defenses)
- Data security (e.g., access controls and encryption for training data)
- Monitoring and responding to security incidents in AI environments

These areas align with global AI security guidance from over 20 authoritative sources, including ISO/IEC 23894:2023, OWASP AI Exchange, and Guidelines from the Cybersecurity and Infrastructure Security Agency (CISA).

Ideal For?

This certification is ideal for security and risk leaders, legal and compliance teams, regulators, and AI deployers who need a clear and validated picture of how AI systems are secured. It's not just a technical requirement — it's a strategic advantage, showing the world that your organization takes AI security seriously and has the confidence to prove it.

The HITRUST AI Risk Management Assessment

The AI Security Certification provides assurance for deployed systems. However, not every organization is at that stage. Many are still in the early phases of building, experimenting, or evaluating how AI fits into their operations. The challenge for these organizations is understanding and managing AI-specific risks.

That's where the HITRUST AI Risk Management Assessment helps. This solution provides a structured way for organizations to evaluate their AI risk posture. It doesn't require certification, making it a low-barrier option for teams seeking to map their practices against leading standards.

The output is the HITRUST AI Risk Management Insights Report. It's a powerful and professional-grade deliverable that offers detailed scorecards, gap analyses, and mapped observations.

The HITRUST AI Risk Management Assessment

Risk Focus

The AI Risk Management Assessment includes 51 curated controls, harmonized with two sources:

- ISO/IEC 23894:2023
- NIST AI Risk Management Framework (RMF) v1.0

These controls address a wide spectrum of AI-related concerns, from ethical governance and explainability to accountability mechanisms and continuous risk monitoring.

Ideal For?

This assessment is relevant for product leaders, governance teams, and internal risk groups shaping the organization's AI strategy. Whether you're planning to deploy AI or already experimenting with it, the AI Risk Management Assessment helps you establish a baseline and chart a path toward responsible, informed adoption.

Quick Comparison: AI Security vs. AI Risk Management

Feature	HITRUST AI Security Certification	HITRUST AI Risk Management Assessment
Type	Certified, validated solution	Self-directed or assessor-led evaluation
Purpose	Prove that AI systems are secure	Understand and improve AI risk posture
Controls	Up to 44 controls	51 harmonized controls
Output	HITRUST AI Security Certification	HITRUST AI Risk Management Insights Report
Required assessment type	e1, i1, or r2	Can be standalone or added to any type
Ideal for	Security and compliance leaders, AI deployers	Governance, product, and internal risk teams

The Business Value of AI Assurance

AI is a business enabler. But with that opportunity comes a growing demand from stakeholders for proof that AI systems are secure, ethical, and well-managed. HITRUST's AI assurance program delivers meaningful business value.

Build Confidence with Stakeholders

- Organizations can demonstrate that their AI systems are secure to build trust.
- Regulators get proof of governance and compliance.
- Boards and executives get assurance of reduced risk and reputational protection.

Accelerate AI Sales and Adoption

- Customers are more likely to adopt AI-powered services when safeguards are in place.
- Sales teams gain an edge with shorter sales cycles when they can offer certified proof of security.
- HITRUST assurance acts as a competitive differentiator in a noisy market.

Streamline TPRM and Audit

- HITRUST AI assessments can be shared across the supply chain.
- Vendors and partners benefit from standardized, credible assessments.
- Audit and compliance teams save time and reduce complexity.

Enable Innovation with Fewer Barriers

- Strong assurance helps teams adopt new AI tools faster.
- HITRUST aligns security and risk teams with innovation efforts.
- Organizations are empowered to scale AI responsibly and confidently.

The Path to Trustworthy AI Starts Today

AI is here. It's powerful, promising, and risky.

Organizations that move forward without a plan for trust and security expose themselves to legal, ethical, and operational consequences. But those who invest in assurance will lead with confidence.

HITRUST offers a mature, practical, and proven path forward, whether you're building, buying, or integrating AI systems. Now is the time to take action.

Ready to build trust and secure your AI? Visit the HITRUST AI Hub.