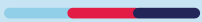# A Path to Trustworthy AI

## HITRUST's Strategy for Secure Use of AI

*HITRUST Office of Strategy, Research, and Innovation*

# The Promise of Artificial Intelligence

The implementation and use of AI algorithms and systems, such as natural language processing, the resulting large language models, and other machine learning applications, has significant potential. These systems are being imagined and applied in novel and creative ways and are evolving at a breakneck pace. However, like many new, disruptive, and exciting technologies, the use of AI introduces new terminology, concepts, and enterprise risks that must be understood and addressed. These considerations center around whether the AI system is trustworthy: operating with the expected quality and integrity needed to deliver the expected information with sufficient quality.

The use of AI also raises important governance, ethical, and legal considerations which must be considered when deploying and operating an AI system. There are a number of operational considerations, including the quality of the service and its reliability, that must be addressed for trustworthy and consistent operation. Risk management, security and assurance for AI systems is only possible if the multiple organizations contributing to the system share responsibility for identifying the risks to the system, managing those risks, and measuring the maturity of controls and safeguards.

AI systems are made up of the system that is using, or consuming, AI technologies, but also the provider or providers that are providing the AI service and, in many cases, additional data providers supporting the machine learning system and large language model underpinning the system. The context of the overall system on which AI is delivered and consumed is critical to understand, as is the benefit of partnering with high-quality AI service providers that provide clear, objective, and understandable documentation of their AI risks and how those risks, including security, are managed in their services.

Users of AI services can leverage the capabilities of those high-quality service providers as part of their overarching risk management and security system accompanying their deployment of AI with resulting increase in efficiency and trustworthiness of their systems if the provider is committed to an approach that supports inheritance and shared responsibility. The rapid pace of AI adoption requires industry leadership to deliver assurances that scale and bring stakeholders together to demonstrate that the combined system is trustworthy. HITRUST has years of experience bringing leaders across the private sector together to focus on practical shared responsibility based upon an inheritable control framework proven daily in security compliance and cloud computing. Shared AI Assurances between stakeholders are essential to maintaining trust in AI systems based on proven, practical, and achievable approaches.

AI systems must be designed, implemented and managed in a trustworthy manner. HITRUST is supporting this shared accountability by building upon proven experience in applying risk management and security frameworks to the important requirements of actionable risk management, reliable and objective security assurances, and measurable outcomes.

# The HITRUST AI Assurance Program

HITRUST is launching the AI Assurance Program: the first and only assurance program able to demonstrate and enable sharing of security control assurances for Generative AI and other emerging AI model applications. HITRUST and industry leader partners are identifying and delivering practical and scalable assurance for AI risk and security management through key initiatives providing organizations with the leadership needed to achieve the benefits of AI while managing the risks and security of their AI deployments. These include the following:

## 1. Prioritizing AI Risk Management as a Foundational Consideration using the HITRUST CSF

AI systems require thoughtful analysis in order to assess potential harms from a catalog of potential risks including risks to the information incorporated into AI systems, risks associated with improper or incorrect outcomes from AI models, and operational risks when AI systems fail to meet their expected potential.

Beginning with the release of HITRUST CSF v11.2 in October 2023, HITRUST is incorporating AI risk management and security dimensions in the HITRUST CSF. This provides an important foundation that AI system providers and users can use to consider and identify risks and negative outcomes in their AI systems with regular updates available as new controls and standards are identified and harmonized in the framework and available through HITRUST assurance reports.

AI systems are also built upon existing IT systems with proven security patterns. HITRUST's commitment to timely and regular updates of the CSF will provide a needed foundation to manage risks to AI adoption and achieve the promised benefits based on a proven system to document, test, and measure alignment with and assurance for those principles.

HITRUST CSF version 11.2 includes two risk management sources with plans to add additional sources through 2024:

- NIST AI Risk Management Framework – the NIST AI Risk Management Framework ("RMF") provides for considerations of trustworthiness in the "design, development, use and evaluation of AI products, services and systems."

- ISO AI Risk Management Guidelines – ISO Risk Management Guidelines (ISO 23894) provides "guidance on how organizations that develop, produce, deploy or use products, systems and services that utilize artificial intelligence (AI) can manage risk specifically related to AI."

## 2. Providing Reliable Assurances around AI Risks and Risk Management through HITRUST Reports

Beginning in 2024, HITRUST assurance reports will include AI risk management so that organizations can address AI risks through a common, reliable, and proven approach. This will allow organizations that are implementing AI systems and the AI model and service providers to understand the risks associated and reliably demonstrate their adherence with AI risk management principles with the same transparency, consistency, accuracy, and quality available through all HITRUST reports.

AI risk management certifications will be supported on top of the HITRUST Essentials (e1), HITRUST Leading Practices (i1), and HITRUST Expanded Practices (r2) reports. This allows organizations to provide assurances that they have considered risks from their adoption and use of AI while also demonstrating the maturity of the underlying system that supports the AI platform.

HITRUST Insight Reports will also be available to support organizations that wish to demonstrate the breadth, coverage, and quality of their AI Risk Management efforts to relying parties, including customers, that are seeking to understand efforts that the organization has undertaken to understand and manage AI risks and to govern their AI systems in a trustworthy, responsible, and reliable manner.

The use of existing and proven HITRUST reports and the HITRUST assurance system demonstrates that the security of the underlying technology systems supporting the AI system has also been considered including transparency around the identification and documentation of risks, consistency in assessment results, and independent verification and quality assurance of the testing of the risk management to ensure that it is aligned with the expected outcomes.

---

[1] National Institute of Standards and Technology (NIST), AI Risk Management Framework. https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF.
[2] International Standards Organisation (ISO), ISO 23894 Information Technology, Artificial Intelligence, Guidance on Risk Management. https://www.iso.org/standard/77304.html.

## 3. Embracing Inheritance in Support of Shared Responsibility for AI

HITRUST has served as an industry leader in the development and operationalization of a robust inheritance and shared responsibility model with significant adoption in cloud computing. AI governance is similar, and it benefits from this proven expertise. The providers of AI services, such as large language model providers and the users of AI in their systems and services, must clearly understand how they jointly identify the risks to the AI system and then share in the management accountability for those risks.

As HITRUST adds AI risk management and security controls to the HITRUST CSF, the HITRUST Shared Responsibility Model will allow AI service providers and their customers to agree on the distribution of AI risks and allocation of shared responsibilities. It is important to consider those areas where the parties share risk management considerations, such as when both parties have responsibility for model training, tuning, and testing with different contexts.

Shared responsibilities for AI include the identification of risks and control dimensions for the large language models and other AI services the service provider is delivering, which of those risks and controls are implemented by the service provider, and identification of items that are the responsibility of others including the organizations that are deploying AI in their systems with the provider's AI service. These parties must demonstrate that they have considered and addressed important questions, including but not limited to:

- Identification of the training data used by the AI system

- Consideration that training data is relevant and of the expected quality

- Safeguards are in place to prevent poisoning of data with impacts to the integrity of the system

- Recognition, identification of, and managing to minimize bias

- Clarity from model providers on the responsibilities of model users, including testing to evaluate whether the model is appropriate for the intended business outcome and further tuning of the model

- Identification of required distinctions where companies choose to create their own large language models or use their organization's data to refine or extend the model

AI Shared Responsibility and Inheritance is achievable starting now. The use of the existing and proven shared responsibility and inheritance system available from HITRUST and cloud service providers, who are now also AI service providers, allow companies using AI to validate, today, that their systems are being built with robust, understood, and provable security safeguards.

This combination of HITRUST and industry leading cloud and AI service providers make shared responsibility and inheritance of AI risk management and security a reality, all based on the extensible HITRUST CSF and reliable AI Assurance Reports provided by HITRUST

## 4. Leading Industry Collaboration

HITRUST will use its long-standing experience in control frameworks, assurance, and shared responsibility to drive responsible and industry-led solutions for AI risk management and security.

Leaders from Microsoft and Databricks are already engaged, and others are invited to join us as we establish relevant and achievable safeguards that maximize the benefits of AI and allow consuming organizations to reap those benefits in a safe and predictable manner.

Microsoft Azure OpenAI Service supports HITRUST maintenance of the CSF and enables accelerated mapping of the CSF to new regulations, data protection laws, and standards. This in turn supports the Microsoft Global Healthcare Compliance Scale Program, enabling solution providers to streamline compliance for accelerated solution adoption and time-to-value.

"At Microsoft, we are committed to a practice of responsible AI by design, guided by a core set of principles: fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. We are putting these principles into practice across the company to develop and deploy AI that will have a positive impact on society," said John Doyle, Global Chief Technology Officer, Healthcare and Life Sciences at Microsoft.

"AI has tremendous social potential and the cyber risks that security leaders manage every day extend to AI. Objective security assurance approaches such as the HITRUST CSF and HITRUST certification reports assess the needed security foundation that should underpin AI implementations," says Omar Khawaja, Field CISO of Databricks.

"Databricks is excited to be working with HITRUST to build on this important foundation and to significantly reduce the complexity of risk management and security for AI implementations across all industries," he added.

# Use of AI for Industry Benefits and Quality

HITRUST has been an early adopter of AI for greater efficiency and to sustain quality. This has created a number of benefits for the community already with several areas of research and innovation focus as HITRUST leads through use of this exciting technology:

### Completed

| | |
|---|---|
| **Mapping of Assessed Entity Policies** *(Patent Pending)* | The use of AI to help map written policies to the HITRUST CSF and relevant safeguard and requirement statements supports reduced assurance effort for assessed entities<br><br>Helping assessed entities and assessors to quickly evaluate the comprehensiveness and quality of their policies with a reduction in assessed entity, assessor, and quality assurance analysis time |
| **Quality Assurance Efficiency** | AI accelerating HITRUST's ability to do robust quality assurance of assessments |
| **CSF Quality and Updates** | Use of AI to map authoritative sources into the HITRUST CSF and keep the CSF relevant as security requirements are continually changing and new authoritative sources are identified |

### Innovation in Progress

| | |
|---|---|
| **Support for Assessed Entities and Assessors** | Leverage AI helpers/co-pilots as tools to help assessed entities and assessors identify required controls and evidence for assurance purposes |
| **Support for Relying Parties** | Leverage AI helpers/co-pilots as tools to support relying parties in understanding their ecosystem of suppliers, their security assurance and maturity levels, and areas of concern |
| **Support for Industries** | Leverage AI helpers/co-pilots as tools to support benchmarking for industries and cohorts, helping leaders assess how their programs compare to others |
| **Supporting Threat-Adaptive Principles** | Use AI to progress the HITRUST CSF as a living framework that is continually updating itself, considering new and changing threats and control specifications |

## Enabling Trustworthy AI

A trustworthy approach to AI is aided by existing and proven approaches to risk, security, and compliance management, all supported by a reliable and scalable assurance system. Thousands of organizations already access and use the HITRUST CSF and rely on HITRUST assurance reports to demonstrate their adherence and compliance to security expectations with numerous industry-leading companies relying on HITRUST reports from their suppliers as evidence of security and compliance maturity. All of the major cloud service providers who are leading the adoption of AI also leverage the HITRUST Shared Responsibility and Inheritance Program to provide reliable assurances for their subscribers today.

The leadership of these existing service providers as they create the future of AI is a key accelerator to providing practical and scalable AI risk management and security assurances. The addition of AI risk management considerations to the HITRUST ecosystem and shared industry leadership will enable secure and scalable AI in support of the expected benefits for companies adapting AI, the service providers providing compelling features such as large language models, and relying parties seeking evidence that the companies and services they work with are responsible in their use of AI.