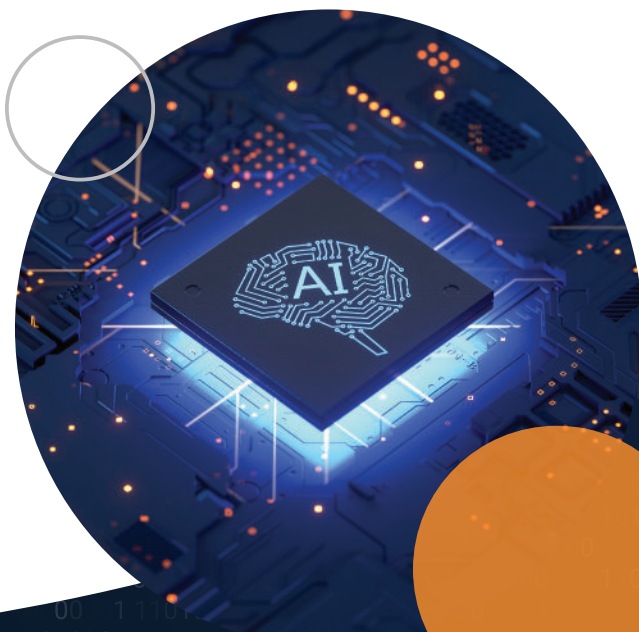


# Securing the AI-enabled future

## Navigating AI risks and rewards in healthcare

As artificial intelligence (AI) is increasingly integrated into daily work across industries, the healthcare ecosystem is no exception. AI stands to reshape healthcare by elevating patient outcomes, reducing costs, and enhancing operational efficiency. Yet, AI also brings on unique cybersecurity challenges that healthcare organizations must fully understand and address before they can reap the rewards.

**In this whitepaper, we will provide a foundational view of AI and its current applications in the healthcare ecosystem, along with the primary risks. We also introduce some of the exemplary regulatory procedures put in place to combat these issues, provide specific risk mitigation strategies, and outline considerations for the future.**



# An Overview: Defining AI and its use cases

According to NIST, "An AI system is an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy."<sup>1</sup>

According to Wharton, AI is "A branch of computer science dealing with the simulation of intelligent behavior in computers, or the capability of a machine to imitate intelligent human behavior."<sup>2</sup>

## Common subfields used interchangeably with AI and how they differ



### Computer vision

Automated interpretation of visual data, including object detection, facial recognition, action recognition, and pose estimation



### Expert systems

Interpretable programs using expert-provided domain knowledge to generate conclusions from a knowledge base for decision-making



### Machine learning

Enables systems to learn predictive patterns from data without explicit programming to make predictions or decisions autonomously



### Multi-agent systems

Distributed system of autonomous entities (agents) that collaborate for beneficial outcomes in a complex, dynamic environment



### Natural language processing

Use of computational techniques to learn, understand and produce content in human language



### Robotics

Development of adaptable physical machines that autonomously interact with their environment by use of perception, planning and execution loops

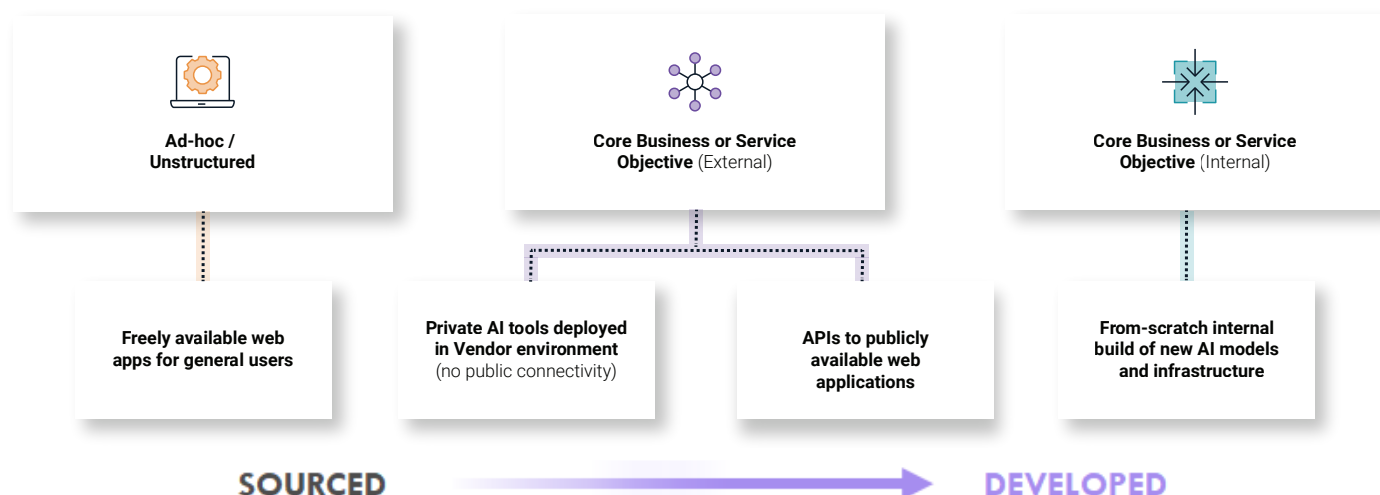


### Speech recognition

Methods for translating spoken language into text, enabling communication and interaction between humans and machines

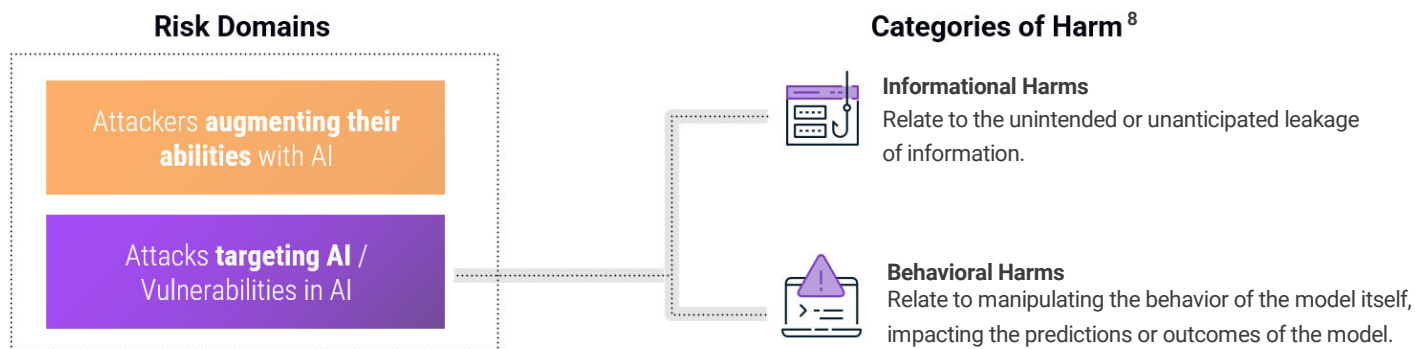
## General AI use case categories

According to the 2023 AI Index Report from Stanford University,<sup>3</sup> the industry with the largest AI investments is healthcare (\$6.1 billion), followed by data management, processing, and cloud (\$5.9 billion), and FinTech (\$5.5 billion). The manner in which organizations leverage AI can vary significantly, and for healthcare organizations, it's important to fully understand their AI scope to evaluate the level of risk it may present. **We can break AI use cases down into three main categories:**



# Breaking down the risks

**Before we explore actionable ways to mitigate AI risks, we must understand the risks that exist and what they really mean for healthcare.** At a fundamental level, AI risk can be categorized in the following ways:



For an example in practice, if the training data used to develop an AI algorithm is not actually representative of reality or contains biases or falsehoods, the AI systems may produce inaccurate or discriminatory results, leading to disparities in healthcare outcomes. Addressing the potential for these biases in advance and ensuring fairness in AI algorithms is crucial to prevent unintended consequences and ensure equitable healthcare. There is also the fear that healthcare professionals may become overly dependent on AI-generated insights, and as such, it is essential to maintain a human-led approach where AI serves as a valuable tool to support decision-making rather than replacing the expertise and judgment of healthcare professionals.

However, privacy and security concerns remain the most pressing risk related to AI usage in the healthcare landscape. Using sensitive patient data to train AI models raises questions about data protection and the potential for unauthorized access.

## Data breaches and patient privacy concerns

Protecting patient data from breaches and ensuring compliance with privacy regulations is a significant concern, particularly as healthcare organizations store vast amounts of sensitive patient data that third-party vendors can access. This presents a concern specifically in relation to AI systems as they require access to this data for analysis, making them prime targets for cyberattacks.

From a compliance standpoint, use cases for AI are subject to the same guidelines as any other tool or technology that deals with processing personal information, including fairness, use limitation and purpose specification, data minimization and storage limitations, and more.

And the access to this information that AI would gain once integrated into the ecosystem poses significant regulatory and compliance risks. Striking a balance between leveraging data for improved healthcare outcomes and safeguarding patient privacy is a critical challenge that healthcare organizations must navigate.

The risks of AI within the healthcare ecosystem can be broken down into five key categories. Let's dive deeper into the details of these specific AI risks and what they mean for healthcare organizations and vendors.

**The International Association of Privacy Professionals (IAPP) estimates that more than half of AI governance approaches are being built on top of existing privacy programs, and only 20% of self-identified 'mature' organizations have already begun rolling out some formalized AI practices and guidelines.<sup>4</sup>**

## Data leakage considerations

Use of free tools (e.g., ChatGPT) warrants minimum necessary information gathering that could assist with reconnaissance as those tools are targeted for breach. Failure to properly address the acceptable use associated with information scraping or personnel involvement with generative AI tools represents a significant leakage risk, particularly for intellectual property and other secrets.

## Malicious attacks on AI models

AI models themselves can be targeted. Adversarial attacks can manipulate AI algorithms to provide incorrect diagnoses or treatment recommendations, potentially endangering patients' lives.

**Some more specific examples of malicious attacks are as follows:**



### Membership reference attack

Determine if a data record was in the model's training dataset. This is a non-repudiation problem where the individual cannot deny being a member of a sensitive group (e.g., cancer patient, an organization related to a specific sexual orientation, etc.).<sup>5</sup>



### Input manipulation attack

Fool AI models with deceptive input data, which can be done by experimenting with the model input (black box), by introducing maliciously designed input based on analysis of the model parameters (white box), and by basing the input on data poisoning that has taken place.<sup>5</sup>



### Data poisoning attack

Can change training data and manipulate the behavior of the AI model, either sabotaging the model or causing the model to make decisions in favor of the attacker.<sup>5</sup>



### Model inversion attack

Estimates the training data with varying degrees of accuracy, which is concerning if the training data contains sensitive or copyrighted information.

## Ransomware and malware attacks

Ransomware attacks on healthcare institutions have become increasingly common, and documented instances of AI-augmented malware or social engineering campaigns only serve to illustrate an impending boom in an already rampant environment of malicious activity.

In a recent survey conducted by CyberArk, AI-supported malware was listed as a top concern by security professionals because malicious software augmented with machine learning (ML) provides a more capable means of traversing domains versus legacy malware.<sup>6</sup>

Hwas, a research firm, developed and tested an AI-generated malware called 'BlackMamba,' which successfully bypassed industry-leading endpoint detection and response tools in test environments. While the BlackMamba malware was only tested as a proof-of-concept and does not live in the wild, its existence signals the urgency of the risks associated with AI augmentation.

### Generative AI Augmentation Risks

The development of inauthentic materials, recordings, or even virtual interactions that heavily mimic what a user would expect to see enables 'counterfeit reality' as an emerging threat that fundamentally changes the enterprise attack surface. Highly sophisticated AI facilitates convincing spoofing of a company brand, personal image, video, or voice at scale.<sup>7</sup>

## Supply chain compromise

Many healthcare organizations rely on third-party vendors for AI solutions. These vendors may introduce vulnerabilities into the healthcare system, and any compromise in the supply chain can have far-reaching consequences.

**Here are some questions that healthcare payors and providers should ask their vendors today to ensure suitability for contracting.**

- ? What is the name of the AI service/tool the vendor is incorporating into their solution?
- ? Does the vendor provide an opt-out option to its customers?
- ? Where is the AI model storing my company's data?
- ? Does the vendor monitor for behavioral harms, such as bias, hallucination, and incorrect conclusions, offensive content, etc?
- ? What is the use case category of the vendor's AI (Public AI, Private AI, From-scratch AI Model Build)?
- ? Does the vendor perform traditional cyber red teaming on its AI?
- ? Does my company's data become part of the public domain?
- ? Does the vendor perform full-scope Responsible AI (RAI) red teaming on its AI?
- ? Is my company's data used to train the model?

## Lack of AI cybersecurity expertise

Given AI's rapid growth and evolution, healthcare IT professionals may lack the expertise to effectively secure AI systems. This knowledge gap can result in misconfigured systems that do not adequately address and protect against AI cyber threats. Inability to properly acknowledge or understand and implement key AI governance mechanisms may lead to significant organizational turmoil and liability in the event of model abuse. Similarly, failure to prepare technical personnel for the unique risks associated with defending AI systems only serves to increase likelihood of a successful attack against sensitive AI systems and relevant data.



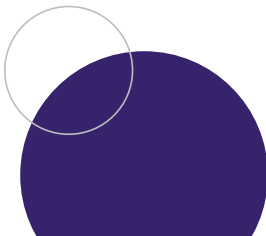
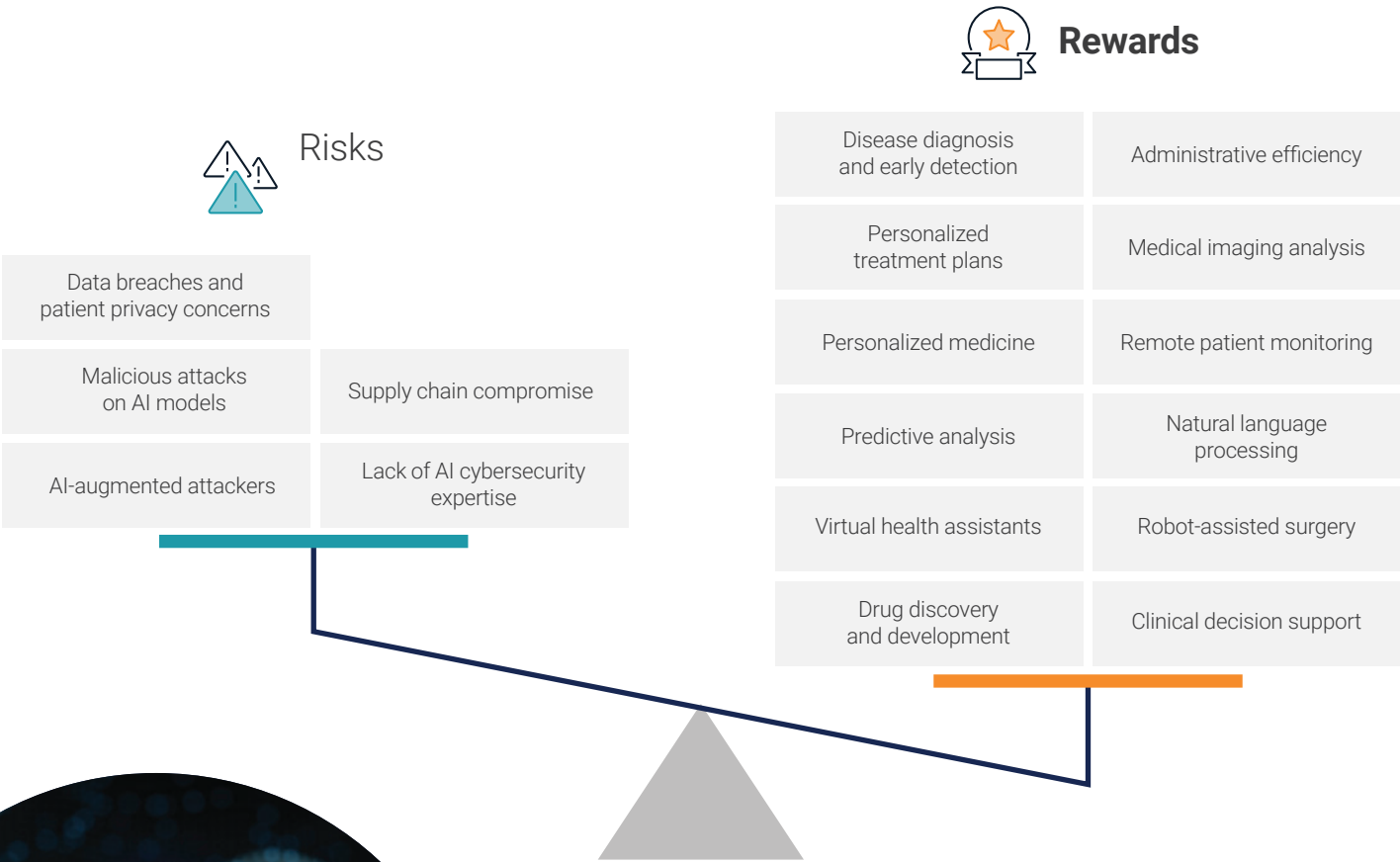
# Weighing risks and rewards in healthcare

As the healthcare industry progressively integrates AI into care models and workflows it's important to understand the power of this technology, along with the potential risks.

In general terms, one of AI integration's significant rewards within the healthcare ecosystem is the potential for improved diagnostics and treatment outcomes. This is due to AI-driven algorithms' ability to analyze vast amounts of medical data, leading to more accurate diagnoses, more comprehensive monitoring practices, and more timely life-saving interventions. By enabling healthcare professionals to make informed, data-driven decisions in a fraction of the time, AI can become integral to improving patient care quality.

AI can also bring increased efficiency and cost-effectiveness to healthcare operations. Automation of routine tasks, such as administrative processes and data entry can free up valuable time for healthcare professionals and others in the ecosystem, allowing for better resource utilization and more time for patient care.

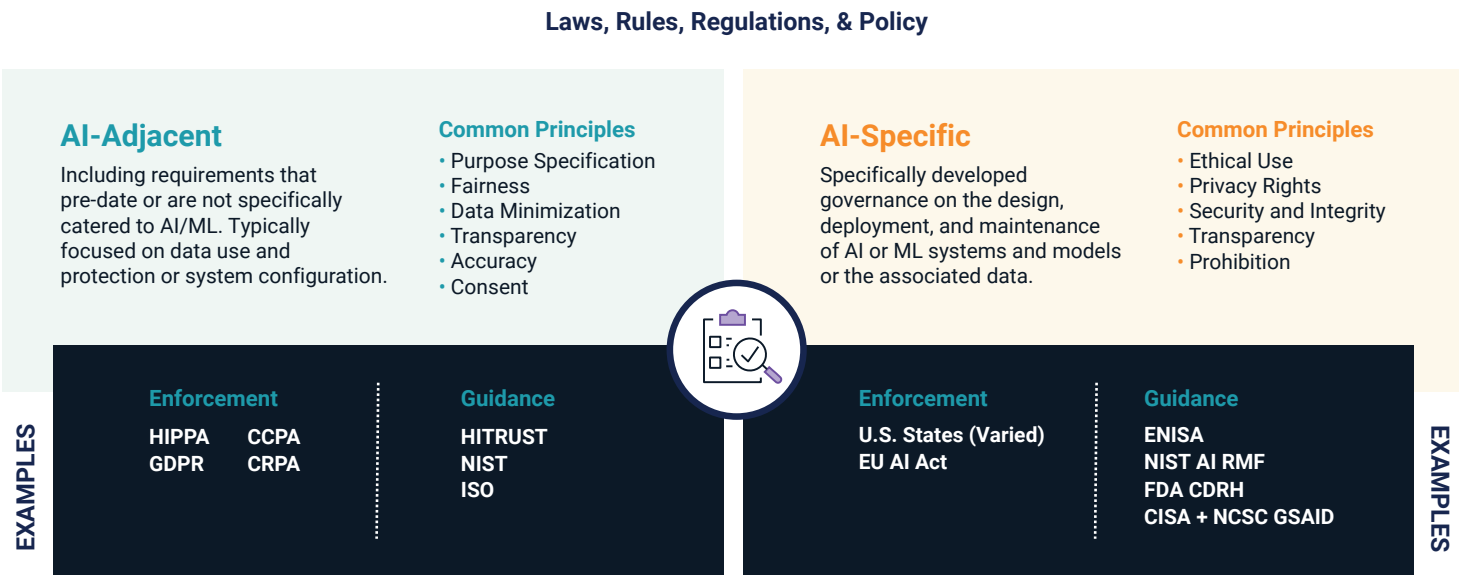
At a glance, we can see that the potential rewards of AI far outweigh the risks. However, we must fully understand the risks, their implications, and how to mitigate them before we can truly reap the rewards.



# The evolving AI regulatory landscape

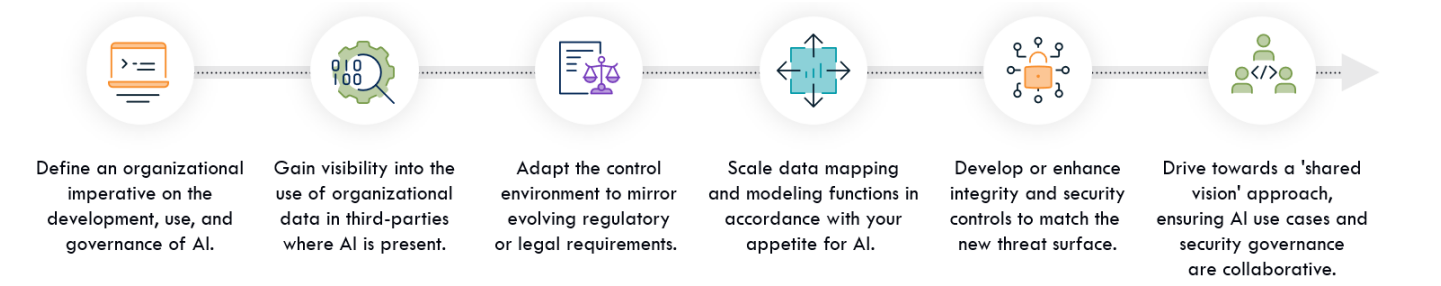
Key regulatory and standards bodies are incorporating new guidance around AI to support organizations in mitigating risk as they adopt AI on a larger scale and to work towards a more secure ecosystem for all. In addition, many organizations have existing requirements in place that were not specifically designed for AI, but place emphasis on secure data use and system configuration in general.

By understanding the regulatory landscape as it relates to AI, organizations can better position themselves to tap into AI's full potential while remaining secure and compliant.



# Strategies for mitigating AI cybersecurity risks in healthcare

AI's potential for healthcare is significant, but organizations must develop an effective strategy that allows them to mitigate the risks of AI while reaping the rewards. Below is an approach organizations can take that maximizes transparency between healthcare organizations and their key partners while enabling more effective and secure AI utilization for all.



There are other steps healthcare organizations and vendors can take, depending on the way they are using AI. Let's take a closer look at some specific controls organizations should consider, whether they are actively developing or configuring AI models or tools or taking advantage of tools that are already freely available:

#### **Training and user awareness**

Expand end-user awareness of deepfake technology and data leakage by augmenting specialized training and security awareness programs.

#### **Credentialing and personnel controls**

Ensure that the personnel working with these systems are qualified and properly vetted. This includes robust access controls.

**For non-developing organizations** that are not actively developing or configuring AI models or tools

**For any organization**

**For AI-developing organizations** responsible for data modeling and data science for AI use

#### **Enhanced data integrity controls**

Leverage available technology (such as cryptographically generated digital signatures) to ensure the authenticity and legitimacy of content.

#### **Elevated procurement security**

Ensure third-party controls are updated and data use agreements are current.

#### **Evolved governance**

Ensure policies, committees, and other key activities include AI considerations and input from all levels.

#### **Strategic planning**

Organizational direction and expectations for the use and adoption of AI should be formally defined and evaluated on an ongoing basis.

#### **Data quality and assurance**

Implement controls around the data utilized, ensuring integrity and security of the pipeline is maintained.

#### **Data model attack prevention**

Implement hyper-focused data security and attack prevention controls centered around the data pipeline and novel attack techniques.

#### **Model behavior curation**

Conduct an intentional review process with automated and manual controls to prevent unintended harm to AI model outputs.

#### **AI red teaming**

Partner with experts to conduct exhaustive penetration testing and scenario execution against AI models and underlying systems.

# A secure and AI-enabled future for healthcare

As AI continues to secure its place within the healthcare ecosystem, the cybersecurity landscape is undergoing a rapid evolution. While AI holds the power to save costs, improve patient care, enhance efficiency, and strengthen decision-making, healthcare organizations must have the controls in place to tap into its potential without putting sensitive data at risk. Ongoing collaboration between healthcare payors, providers, vendors, and policymakers is critical to ensuring that AI is applied securely and responsibly for the benefit of all. A partner with deep expertise across the healthcare risk landscape can empower organizations to minimize the risks of AI and maximize the rewards.

At CORL and Meditology, we are 100% focused on the unique cybersecurity requirements of the healthcare ecosystem. Together, we empower healthcare organizations and vendors to elevate their overall security posture and efficiently evaluate and mitigate third-party risks. As technology evolves and new risks arise, we're committed to helping our clients reap the benefits while avoiding the risks.



**Looking to securely  
apply AI in your  
organization?**

Visit **meditologyservices.com** and  
**corltech.com** to learn more.



1. <https://www.nist.gov/itl/ai-risk-management-framework>
2. <https://aiab.wharton.upenn.edu/research/artificial-intelligence-risk-governance/>
3. <https://aiindex.stanford.edu/report/>
4. <https://iapp.org/resources/article/professionalizing-organizational-ai-governance-report-summary/>
5. <https://owasp.org/www-project-machine-learning-security-top-10/>
6. <https://www.cyberark.com/press/cyberark-establishes-artificial-intelligence-center-of-excellence-to-combat-attacker-innovation-with-identity-security-innovation/>
7. <https://www.gartner.com/en/topics/generative-ai>
8. [https://fpf.org/wp-content/uploads/2019/09/FPF\\_WarningSigns\\_Report.pdf](https://fpf.org/wp-content/uploads/2019/09/FPF_WarningSigns_Report.pdf)