Risk Mitigation

FDA Digital Health Advisory Board meeting
November 20th, 2024

Michael Schlosser, MD, MBA
Senior VP Care Transformation & Innovation

HCA Healthcare℠

Risk mitigation discussion

- Governance
  - Framework to understand unique risks created by Generative AI solutions.
  - Evaluation of which models to approve or to deploy based on the unique risks they present
- Feedback mechanisms
  - Designing model systems to allow for real time feedback, in workflow – **Human-in-the-loop**
  - Nurse Handoff tool example
- Final Comment on Data use rights

HCA’s Responsible AI Framework:

Responsible AI Program (Policy)

- A robust governance framework and set of policies to ensure we use AI technology with the utmost safety, effectiveness, and accountability.

Our Purpose

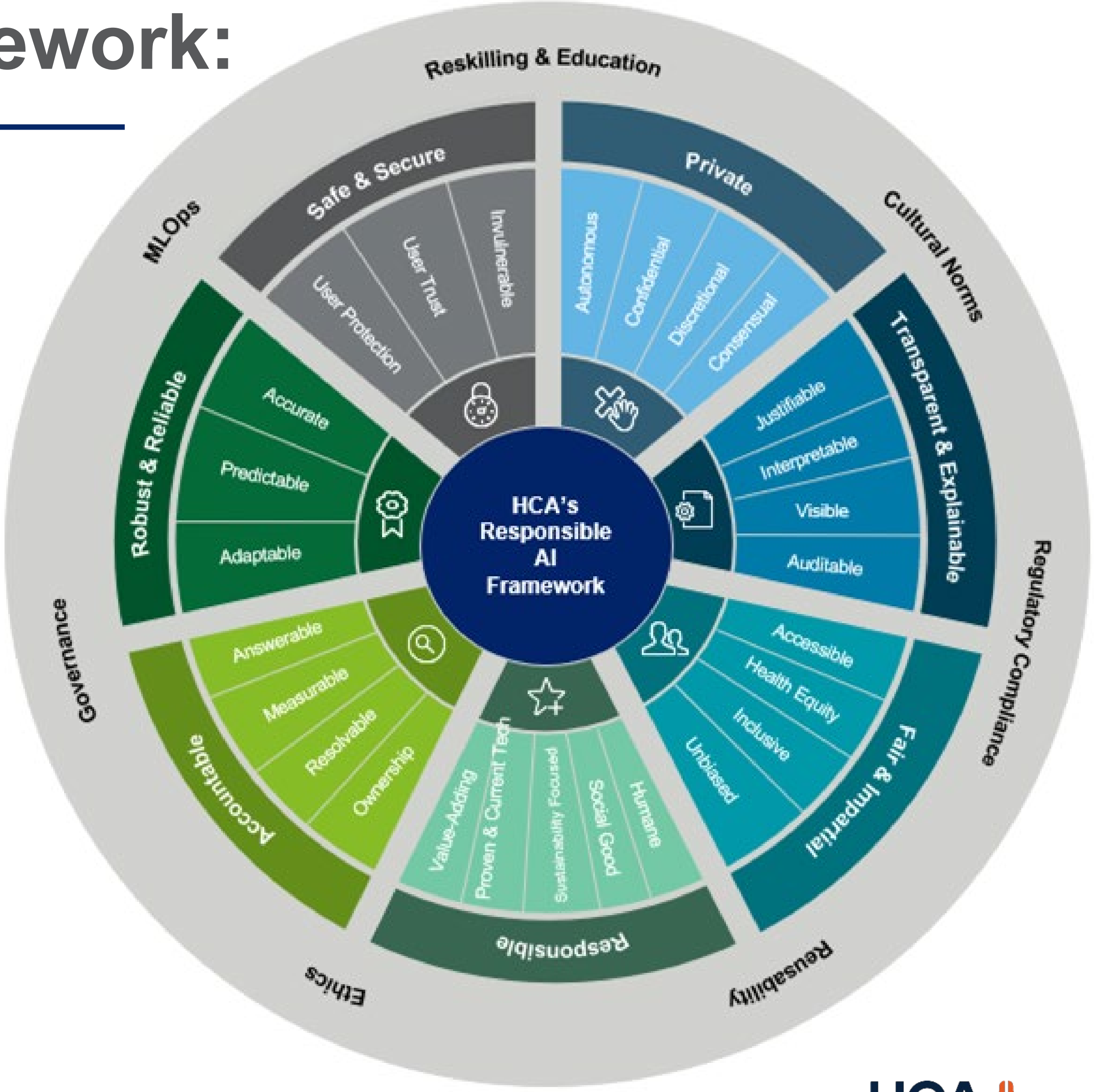
- To lay the groundwork to bring AI technology to our entire healthcare system.

Our Vision

- Unlock benefits of AI while upholding the highest standards of safety, security, trust

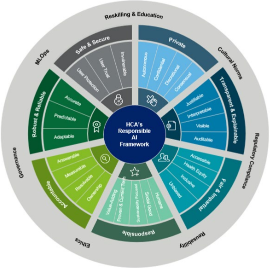
Our Approach

- Integrate AI technology in a way that’s consistent with who we are across HCA Healthcare. To do so, all will be responsible for upholding the key pillars.



CONFIDENTIAL – Contains proprietary information. Not intended for external distribution.

RAI Framework



Safe & Secure	We will establish careful and intentional guidelines to protect patient data against risk and ensure data is trusted and interpreted accurately.
Private	We are required by law to maintain the privacy of protected health information and advise patients at the time of admission/registration regarding how we will permissibly use their information for treatment, payment, and healthcare operations purposes.
Transparent & Explainable	We will be transparent and visible when we use AI to augment our work and will partner with stakeholders across the system to have real and honest conversations about how we're using the tools.
Fair & Impartial	We will use AI to be fairer and more equitable as we take steps to ensure applications are inclusive and accessible.
Responsible	We will use the technology to do good, ensuring any technology is proven research, reduces risk, and adds value to patients or the system at-large.
Accountable	We will empower users, patients and colleagues, to get the full benefit from any AI technology, and we will each take implicit ownership of the results while holding our partners and vendors to the same high standards we impose upon ourselves.
Robust & Reliable	We will double- and triple-check the validity of any results, with ongoing checks to ensure the technology is delivering the level of accuracy we expect.

DESCRIPTION

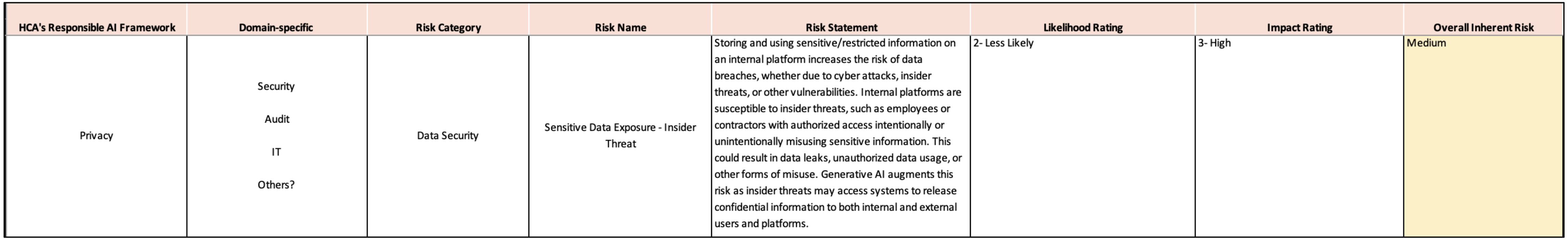
AI Risk Management Plan | Risk Register

The AI Risk Register identifies and maps key risks related to AI development / usage to HCA's Responsible AI framework. This enables the Risk function within the AI Governance Council to elicit informed decisions around risk and mitigation priorities.

INTENDED USERS

HCA Leadership and the **AI Governance Council** can understand and analyze the inherent risk related to AI developments

Enterprise RAI risk management can leverage the risk register to assign risk owners / risk leaders*



HCA's Responsible AI Framework	Domain-specific	Risk Category	Risk Name	Risk Statement	Likelihood Rating	Impact Rating	Overall Inherent Risk
Privacy	Security Audit IT Others?	Data Security	Sensitive Data Exposure - Insider Threat	Storing and using sensitive/restricted information on an internal platform increases the risk of data breaches, whether due to cyber attacks, insider threats, or other vulnerabilities. Internal platforms are susceptible to insider threats, such as employees or contractors with authorized access intentionally or unintentionally misusing sensitive information. This could result in data leaks, unauthorized data usage, or other forms of misuse. Generative AI augments this risk as insider threats may access systems to release confidential information to both internal and external users and platforms.	2- Less Likely	3- High	Medium

- ✓ Identifies **40+ AI / Generative AI Risks** relevant to HCA's operations
- ✓ Utilizes a **risk category** to map each risk across four domains – technology, business, data security, and malicious use
- ✓ Categorizes each risk to the corresponding pillar within **HCA's responsible AI framework**
- ✓ Provides an initial estimate of a) the **likelihood of occurrence** and b) potential **impact** to build an overall **inherent risk score**

**A risk owner / risk leader is a role within enterprise risk to assign ultimate accountability for managing key risks*

AI Risk Management Plan | Risk Register

Risk Name	Risk Statement	HCA's Responsible AI Framework	Risk Category	Risk Score
IP Infringement	Generative AI models are built on massive amounts of data, which may include input data that is protected by copyright law (such as books, art, code, etc.). The use or development of a generative AI model may lead violations of legal protection including copyright and intellectual property infringement.	Transparent and Explainable	Data Security	**High**
Inaccurate Summarization or Question Answering	The underlying content summarization or question answering capability misrepresents and/or fails to capture the source content, intention, and purpose. This is specific to generative AI as the model is dependent on properly intaking large datasets to return outputs aligned to the prompt.	Robust and Reliable	Technology	**Medium**
Hallucination	A generative AI model produces confident, plausible results that are factually incorrect or misrepresentative. Hallucination occurs when algorithms and deep learning neural networks produce outputs that are not real, do not match any data the algorithm has been trained on, or any other identifiable pattern, resulting in generated false information.	Robust and Reliable	Technology	**Medium**
Sensitive Data Exposure	Use of sensitive / restricted data (Confidential, e.g., employee & patient personal information, electronic health information) on a public platform leads to increased security or privacy exposure (e.g., data leakage, inappropriate access, notice, transparency, right to be forgotten) or regulatory non-compliance. The risk of this is magnified by generative AI as users will have access to share information on public tools, risking exposure of confidential information if proper controls are not in place.	Privacy	Data Security	**High**
Overreliance on AI Systems	User overreliance on AI-generated outputs through acceptance of incorrect recommendations, or inappropriate modification of human action to match the AI system recommendations. There is a potential strain on originality and innovation if users depend on AI-generated content as a source of truth.	Responsible	Business	**Low**

Feedback mechanisms – Human-in-the-loop

To design effective human-in-the-loop systems that keep users engaged and prevent overconfidence:

- **Promote Transparency:** Use explainable AI to help users understand AI decisions.
- **Display Uncertainty:** Show confidence levels to inform users about the reliability of outputs.
- **Encourage Active Participation:** Design interfaces that require user input and validation.
- **Provide Training:** Educate users about AI limitations and potential errors.
- **Design for Trust Calibration:** Ensure systems foster appropriate levels of trust.
- **Implement Feedback Mechanisms:** Use user feedback to continually improve AI systems.

Nurse Handoff: Product Description

Develop a product that:

Enhances the nurse handoff process by **generating an automated shift report** that provides a consolidated clinical picture of a patient's care, based on both discreate nursing documentation and non-discrete patient notes from within MediTech.

Key Product Features

- ❑ Generation of a **consolidated summary** of a patient's encounter from admission to current date
- ❑ Generation of a **timeline to show progression of care** and significant events throughout a patient's encounter
- ❑ Generation of a **prioritized list of tasks and key considerations** for a given nursing shift
- ❑ Presentation of **pertinent handoff data** such as assessments, lab results, imaging, etc.
- ❑ **Customizable** (template, widget-based); 80/20 or 60/40 approach to standardization versus customizability
- ❑ Solution is **interactive** and supports iterative workflow and cognitive processing.

I-PASS Framework

I	Illness Severity
P	Patient Summary
A	Action List
S	Situation Awareness & Contingency Planning
S	Synthesis by Receiver

Assessment Framework Overview

Does the summary and timeline contain any issues regarding:

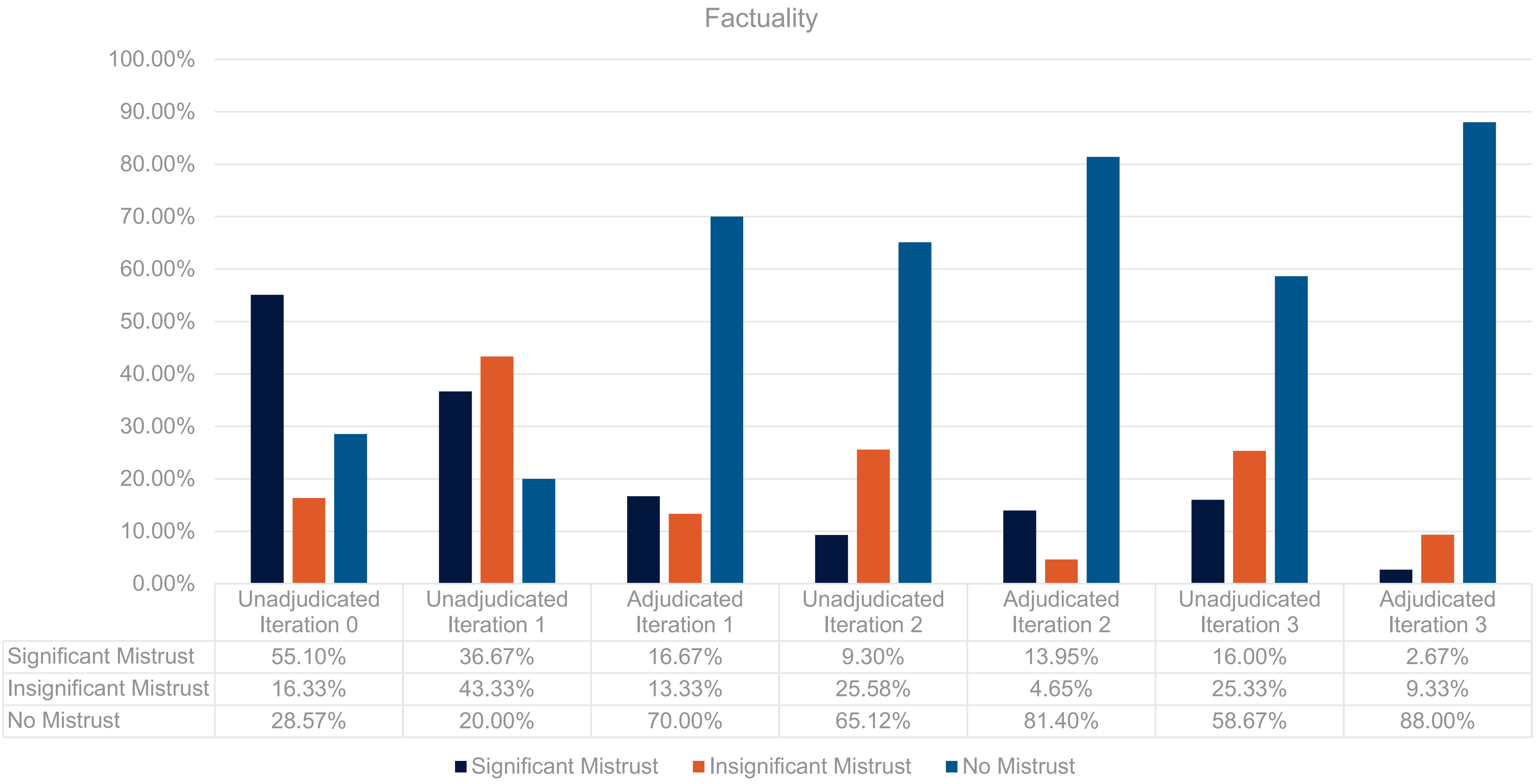
- Factuality (Hallucinations)
- Coverage (Omission)
- Organization (Misstructured)
- Conciseness
- Helpfulness

🔴 **Yes – Significant or critical issues exist**

🟡 **Yes – Minor issues exist**

🟢 **No**

Summary Factuality



	Unadjudicated Iteration 0	Unadjudicated Iteration 1	Adjudicated Iteration 1	Unadjudicated Iteration 2	Adjudicated Iteration 2	Unadjudicated Iteration 3	Adjudicated Iteration 3
Significant Mistrust	55.10%	36.67%	16.67%	9.30%	13.95%	16.00%	2.67%
Insignificant Mistrust	16.33%	43.33%	13.33%	25.58%	4.65%	25.33%	9.33%
No Mistrust	28.57%	20.00%	70.00%	65.12%	81.40%	58.67%	88.00%

References

- Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- Mosier, K. L., et al. (1998). Automation bias in intelligent time-critical decision support systems. *NASA Technical Report*.
- Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5-27.
- Simpkin, A. L., & Schwartzstein, R. M. (2016). Tolerating uncertainty—the next medical revolution? *New England Journal of Medicine*, 375(18), 1713-1715.
- Holzinger, A., et al. (2019). Interactive machine learning: Experimental evidence for the human in the algorithmic loop. *Applied Intelligence*, 49(7), 2401-2414.