NIST Internal Report
NIST IR 8445

Panel Summary Report: Risk Management & Industrial Artificial Intelligence, INFORMS Annual Meeting, 2021 (Virtual)

Mehdi Dadfarnia
Michael Sharp

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8445

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Internal Report
NIST IR 8445

Panel Summary Report: Risk Management & Industrial Artificial Intelligence, INFORMS Annual Meeting, 2021 (Virtual)

Mehdi Dadfarnia
Michael Sharp
Smart Connected Systems Division
Communications Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8445

September 2022

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST Technical Series Policies
Copyright, Fair Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2022-09-28

How to cite this NIST Technical Series Publication:
Dadfarnia M, Sharp M (2022) Panel Summary Report: Risk Management & Industrial Artificial Intelligence, INFORMS Annual Meeting, 2021 (Virtual). (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (IR) NIST IR 8445.
https://doi.org/10.6028/NIST.IR.8445

NIST Author ORCID iDs
Mehdi Dadfarnia: 0000-0003-2401-7184
Michael Sharp: 0000-0002-8014-3612

Abstract

Risk management is an important topic in any domain. With the growing adoption of Industrial Artificial Intelligence (IAI), academic and industrial communities are taking a more serious look at the risks and rewards of using IAI. This paper summarizes insights generated at a panel at the 2021 Institute for Operations Research and the Management Sciences (INFORMS) Annual Meeting, addressing the risk-based evaluation of AI use from a performance and business impact perspective. The panel was held virtually on Sunday, October 24, 2021. The summarized insights highlight identified gaps and barriers in the evaluation and adoption of IAI technologies. The authors provide additional context and suggestions regarding paths forward.

Keywords

Artificial intelligence, economic analysis, reliability engineering, risk assessment, risk management.

Table of Contents

1. Introduction

Industry understands the overall importance of risk vs return studies to their technical engineering and economic bottom line. Frameworks such as Probabilistic Risk Assessment (PRA)[1] or Reliability Centered Maintenance (RCM)[2] use risk analysis and defined system-level requirements to evaluate and qualify potential threats for tools, procedures, and events that can occur within a facility. Business logistics and finance decision-makers similarly use risk analysis when determining the investment potential of new efforts or resources. With these long-standing accepted uses for risk-based evaluation, it is natural to examine new technologies under similar criteria when determining if they can meet functional requirements with the necessary reliability and socio-economic return.

Artificial Intelligence (AI)-enabled tools and technologies are a prime candidate for risk-based evaluation. This class of tools and technologies attracts scrutiny as it becomes prevalent across many industrial domains. The recognized need for better understanding and management of AI-based technologies highlights a lack of effective standardized evaluation regarding the technical and economic impact of these digital tool. Creating risk-centered management and evaluation methodology fulfills the need to provide informative feedback about the impact of an Industrial AI (IAI) tool with domain-relevant language and context.

Discussions from a targeted panel at the 2021 Institute for Operations Research and the Management Sciences (INFORMS) Annual Meeting explored IAI technologies' risk-based evaluation and management. The intent of this panel was to discuss the prevalence of risk-based evaluation techniques for different technologies, the existing evaluation practices for IAI-based technologies, and the opportunities, gaps, and barriers for extending risk-based evaluation to apply to IAI tools and technologies. The discussions highlighted the following:

- Technical accuracy of a tool's performance is only a single aspect that an industrial practitioner considers when deciding to adopt or continue to invest in an IAI tool.
- Trustworthiness, user buy-in, economic return, safety, and effective impact are some factors that influence industrial decision-making.
- Risk-based evaluation assesses the impact of a tool's performance on factors that decision-makers highly value, such as economic returns and safety.
- Understanding and managing the technical, financial, and other risks of concern can allow stakeholders to make the most informed and confident decisions regarding the use and development of IAI-based technologies and tools.

The panel was held virtually on Sunday, October 24, 2021, as part of the virtual INFORMS Annual Meeting.

1.1. Schedule & Format

Table 1 presents the agenda of the talks and the panel. The session began with a series of technical presentations from the panelists. Each presenter, considered an expert in their respective area, used their unique technical background to share views on the status quo of risk-based evaluation in industries and future implications for risk assessment and management with AI-based technologies. Following these short presentations, Dr. Michael Sharp led a panel prompting the panelists to discuss the implications of risk-based evaluations with IAI-based tools and technologies. They were guided through a series of preconstructed questions to understand the similarities and differences between panelists' industries regarding the topics. One presenter that could not be in attendance provided a recorded presentation.

Table 1. Schedule for the 2021 INFORMS Annual Meeting Panel on Risk Management & Industrial Artificial Intelligence

Time (ET)	Topic	Presenter
11:00	Welcome & Introduction	Dr. Michael Sharp (NIST)
11:05	Invited Presentation	Dr. Enrico Zio (Polytechnic University of Milan)
11:20	Invited Presentation	Dr. Katrina Groth (University of Maryland-College Park)
11:35	Invited Presentation	Dr. Askin Guler Yigitoglu (Oak Ridge National Laboratory)
11:50	Invited Presentation	Dr. Sola Talabi (Pittsburgh Technical)
12:05	Panel Discussion (Moderator: Dr. Sharp)	Drs. Zio, Groth, Guler Yigitoglu, Talabi, & Dr. Fan Zhang (Georgia Tech)
12:35	Recorded Presentation	Dr. Jamie Coble (University of Tennessee)

2. Abstracts of Invited Panelist Presentations

The submitted abstracts from each panelist presentation are quoted below.

2.1. Enrico Zio: Intelligent Risk Management by Artificial Intelligence

Affiliation: Polytechnic University of Milan, Milan, Italy

In this talk, I will dare to state how artificial intelligence can help risk assessment and management, and underline the main characteristics that artificial intelligence solutions must have to make risk management intelligent. I will, then, address some research and development directions that are emerging in the area of risk assessment and management supported by artificial intelligence.

2.2. Katrina Groth: Exploring The Connection Between IAI And PRA

Affiliation: University of Maryland-College Park, College Park, Maryland

Ensuring the safety of complex engineering systems ranging from power plants and pipelines is a challenging problem rife with uncertainties. Probabilistic risk assessment (PRA) plays an important role in because of its ability to handle uncertainty and complexity. Advances in data analytics have positioned IAI as an important tool for component failure monitoring. However, single point hardware failures rarely lead to catastrophic failure; instead, complex systems fail due to the interplay between hardware, software, humans, the environment, and the physical constraints. Is there a solution at the intersection of these two approaches?

2.3. Askin Guler Yigitoglu: Artificial Intelligence for Risk Assessment of Complex Engineering Systems: Nuclear Industry Applications

Affiliation: Oak Ridge National Laboratory, Oak Ridge, Tennessee

This talk will focus on state-of-art the AI applications for risk assessment in nuclear industry. Integrating AI methods and tools to risk assessment for advanced reactors designs at different stages of the analysis from pre-assessment (e.g., component health assessment, data analysis) to post-processing (e.g., generating surrogate models of high-fidelity simulations, uncertainty assessment) will be discussed.

2.4. Sola Talabi: Fail-safes And Risk Mitigation Strategies For Mission Critical IAI Systems - Lessons Learned From The Nuclear Power Industry

Affiliation: Pittsburgh Technical, Pittsburgh, Pennsylvania

Nuclear risk and safety management uses the Probabilistic Risk Analysis (PRA) framework to assess and mitigate risk of core damage (Level 1), radioactivity release (Level 2) and consequences (Level 3). PRA may be applied in other industries that require a robust framework to characterize nascent issues and uncertainties. Mission Critical Industrial Artificial Intelligence Systems may benefit from a PRA framework. Fail-safe design is required for safety-related nuclear components, which based on a PRA approach, implies a probability of failure at an acceptably low frequency. Nuclear passive safety systems that reduce the probability of core damage and radioactivity release will be explained.

2.5. Jamie Coble: Enhancing Risk Assessment With Greater Situational Awareness

Affiliation: University of Tennessee, Knoxville, Tennessee

Formal logic-based approaches to risk analysis, so-called probabilistic risk assessment, are ubiquitous in the nuclear power industry. Risk assessment results used to support licensing and license amendments rightfully rely on the expected gross behavior over long periods of time. Risk monitors used for short-term decision making at plants, however, should consider the current and near-term conditions at the specific facility. Industrial AI provides the opportunity to integrate greater situational awareness into these risk monitors, giving a more complete view of the risk of performing (or not performing) actions.

3. Six Key Presentation Takeaways

The takeaways that follow are summarized ideas and observations taken from the panelist presentations and subsequent discussions. These takeaways represent recurrent themes and viewpoints articulated by the panelists. They serve both as a record of the dialogue and as a springboard for future conversations centered on risk management and IAI. Each is presented as a good faith distillation of important ideas presented by the panelists. The takeaways should not be considered as developed consensus from the broader communities, nor as formal recommendations or guidance from any of the participants.

> Takeaway 1: Industrial use of AI comes with unique sets of challenges and requirements that are different from other uses of AI tools and technologies.

Industrial AI is a combined set of actionable intelligence for industrial decision-makers provided by static discriminators, adaptable algorithms, and human-in-the-loop investigators. Integrating industrial AI tools and technologies faces the following issues:

- Use and development of Industrial AI tools must adapt to the specific needs, resources, and limitations that arise from the industrial domain that utilizes the tool.
- Industrial users want IAI tools that give them competitive advantage by providing their assets with system awareness, improved quality performance, automated scheduling, and enhanced risk management. IAI tools promise to deliver these advantages with low-cost deployment, ease of use, and the utilization of data collected throughout the industrial ecosystem.
- Some industry stakeholders are hesitant to adopt IAI technologies. They expect a promise of measurable value returns (often economic in nature) to justify the investment of integrating new technologies. This is often compounded by the 'closed-box' nature of IAI tools, which makes it difficult to predict the full scope of their implications for industrial assets. This lack of transparency and understanding creates

skepticism and hesitancy in adopting IAI-based technologies.

- A suite of common problems arise from bad or insufficient training data feeding into an IAI tool. Generally the knowledge extracted from the tools in these cases may be unreliable, incomplete, or trivialized from coincidental patterns in the data. Although the concept of poor data generating poor results is not unique to IAI, its presence can be more easily obscured or overlooked when associated with IAI, and increased diligence in verification of data throughout the life of an IAI tool is required.

> Takeaway 2: A critical challenge that often impedes the implementation of AI-based solutions in industrial contexts is the lack of mechanisms to certify these solutions for their intended use.

Executives and managers will only embed tools and technologies into their assets if they are convinced of benefits and value from using the tool now and in the future. If there is potential for a regulatory body to restrict or add requirements to the use of a tool, industry decision makers become reluctant to invest in technologies that are not guaranteed to meet these criteria when and if they are put forth by the regulatory body. This reluctance in adoption appears most often seen when adding new tools and technologies to safety-critical applications.

IAI-based solutions are at a disadvantage in this regard. The obfuscated internal logic of many of these solutions makes testing and understanding a full set of outputs from possible inputs difficult or impossible in some cases. Lacking the ability to guarantee reliable or safe solutions in every situation is a barrier to gaining certification. For IAI-based solutions, acquiring an a priori proof of benefits and certifying that a solution behaves as intended or required is very challenging [3]. The following considerations can improve the certification of IAI-based solutions:

- Procedures for certifying an IAI-based solution needs to demonstrate the relative impact to an application with and without the IAI solution. Establishing the solution's impacts needs to be both effective and achievable with a comparatively low barrier to make this evaluation and qualifying process worthwhile to managers and executives. In many cases the established benefits will need to be continually or periodically verified to both maintain trust and justify continued investment in the IAI.
- IAI-based solutions should be evaluated and certified from a systems impact level, encompassing all relevant assets. Executives and managers care most about a high level goals and outcomes provided by a solution. Therefore, demonstrating and communicating the value that an IAI-based solution serves at a system or higher level can be a means to justify investment into the solution. Here we refer to a system as a set of connected assets directed towards the same function or a singular high value asset likely comprised of many smaller sub-components whose function is critical but self contained.

  - For example, a facility manager seeks to decrease costs associated with maintaining their assets. They would want to evaluate an IAI-based condition monitoring solution concerning how it can improve the assessment and scheduling of needed upkeep or repairs for assets. Ideally the IAI would help lower unexpected down time and increase the overall availability of a monitored asset, in turn cutting maintenance costs. Then, the manager can develop a business case that justifies the costs associated with the solution, not from a low level accuracy of output from the IAI, but from a higher level evaluation of impacts to the availability of the system.

Takeaway 3: IAI solutions can be used to improve an asset's monitoring and assessment processes.

Assessment tools are collections of software and hardware devices that alert users to the changing conditions of an asset. Terms used in the literature to describe these sets of tools include prognostics & health management (PHM) technologies and condition monitoring systems (CMS). IAI-based solutions can be integrated into these tools to realize the following capabilities:

- IAI solutions can be used to obtain fault or failure detection, diagnosis, and prognostic information. This information can build the foundation of condition-based and predictive maintenance paradigms (see Refs. [4–6] for a more extended discussion).
- IAI solutions can be used to perform a condition-based risk assessment. Condition-based risk assessment can improve risk management by providing vital information about asset operation needs and safety barriers in or near real-time.
- Challenges of integrating IAI-based solutions in an asset's monitoring processes include dealing with real-data anomalies, changing environments, model explainability, and model security.
- Steps to realizing IAI-enabled, condition-based risk assessment may take the following form:
  1. Incorporating knowledge of the asset's component configuration to create a risk analysis model (e.g., an event tree) with failure probabilities for the components.
  2. Obtaining condition monitoring data from the asset, its sub-components, and its sensors. This data is used by IAI solutions to derive asset/component condition state indicators.
  3. Collecting or inferring operational indicators such as environmental conditions. These indicators also can be derived from IAI solutions.
  4. Using an AI solution (e.g., a Bayesian network) to translate the information col-

lected in Steps 2 & 3 into updated failure probability states in the risk analysis model, representing the condition-based risk assessment.

Takeaway 4: Integrating conventional probabilistic risk analysis (PRA) with IAI tools and technologies has potential for system-level risk monitoring.

Achieving continuously-updated system-level risk monitoring allows for data-informed analysis and decision support at the asset or plant level. Complex industrial assets composed of interdependent components significantly benefit from system-level risk monitoring. Component-level monitoring and analysis will not give the complete picture and may even be misleading towards the decision support required for the entire asset. In conjunction with advances in computing and more significant amounts of available industrial component data, IAI tools and technologies can be utilized to realize real-time, system-level risk monitoring. However, setting up these IAI tools for this systems-level decision support requires careful study and needs to consider the following:

- Obtaining system-level risk monitoring needs to leverage the capabilities of IAI tools with the strengths of PRA techniques (for further discussion and a framework, see Ref. [7]). Although IAI-related technologies consist of software- and hardware-based techniques to process real-time component-level data for analysis (e.g., diagnosis and prognosis) and decision-making (e.g., risk mitigation recommendations), most commercially available products lack a systems-level view. PRA can introduce (chiefly online) techniques to understand risk at a systems level.
- Integrating IAI tools in probabilistic risk analysis should avoid turning into a closed-box process. A closed-box form of risk analysis would decrease operator and regulator confidence in these analyses, rendering any resulting system-level risk monitor a less-effective decision support tool.
- Validation and verification techniques are required to determine whether the integration of data processing and analysis done with IAI tools at the components level with risk analysis done at the system-level yields accurate systems-level monitoring of large and complex industrial assets. A significant challenge for complex assets is quantifying and tracking the uncertainty in risk monitoring and prognostications.

Takeaway 5: IAI solutions can enhance simulations that aid in the identification and evaluation of risk in an asset's operations.

Risk associated with an asset is a function of the possible scenarios for the asset, consequences of those scenarios, and any uncertainty associated with their respective likelihood of occurrence. Assessment of these risks relies on available knowledge from both past experience and future expectations regarding these scenarios and an asset's operations. Simulations using this knowledge are typically a cost-effective method to gain knowledge and

intuition about the asset and its associated risks. IAI solutions can enable and enhance these simulations with comparatively low entry barriers, particularly when there is available data and information to train an IAI predictor.

- IAI-enabled simulations can explore possible hazardous conditions and consequences, especially rare-event, high-risk accident scenarios not previously considered part of the asset's risk equation [8]. Identifying these rare-event, high-risk scenarios is significant for assets with strict safety design requirements, such as the risk analysis involved with nuclear reactor design.
- IAI-based solutions can be used as part of discrete-event simulation and Monte Carlo simulation frameworks to estimate the probability of hazard occurrence that can lead to critical scenarios and consequences [8]. Furthermore, this process is replicable for different asset configuration parameters.
- Another benefit is in using AI-based techniques to boost simulation speed. For instance, simulations can be time-consuming when used to find rare-but-critical accident scenarios in the asset's operations. Techniques that address simulation speed include advanced Monte Carlo methods that direct the search for these rare scenarios or methods that replace time-consuming simulations with quicker surrogate models trained by AI tools (e.g., a neural network that gives sufficiently accurate risk assessment of an asset under specified conditions).
- IAI driven simulations can incorporate and evaluate specific what-if scenarios as directed by decision makers to help better understand options, make better decisions managing risks, or to highlight needs for appropriate safety barriers against particular scenarios.

Takeaway 6: AI-integrated digital twins can expand situational awareness of asset maintenance and operations, guiding better inspection, maintenance, and risk mitigation.

Digital twins are digital representations of an asset, built by constructing structural and behavioral models and collecting a stream of data from physical sensors on the asset. The intent is for the digital twin to simulate behavior of assets in its deployed environment [9]. This is more specific and generally higher fidelity simulation than the system and risk evaluation described in the previous section. However a digital twin could also provide risk assessment capabilities. AI-integrated capabilities in digital twins process data and provide models to gain unique insights about the performance of the asset's operations, maintenance, safety guards, and the prevalence of any hazards or risks. These insights can, in turn, be used for decision-making concerning asset operations, maintenance, and risk management. The following is a list of digital twin use benefits and development challenges:

- The increased situational awareness that an AI-enabled digital twin provides about the asset can be especially helpful for risk management in situations where traditional probabilistic risk analysis approaches have shortcomings. Conventional risk assessment methods analyze a large amount of operating history from the asset and its components (or from similar assets and components). Without an extensive history of asset operations, these risk assessments may not be complete or accurate. To overcome this, simulated asset behavior from a digital twin can provide the data and processing necessary to obtain a complete picture of asset and component risk levels.
- Analyzing the models and data incorporated into a digital twin can detect and identify unobservable information about the asset, such as unknown fault modes.
- Knowledge about asset behavior gained from a digital twin can help create more accurate risk assessments and analyses. Whereas traditional risk evaluation only provides a snapshot of the risks to a system, digital twins can extend the risk analyses to evaluate point-in-time risks. With digital twins, real-time data from actual asset operations - the sensory data, maintenance data, and environmental conditions - culminates in an enhanced, real-time or near-real-time risk monitor. The result is a continuously updated risk analysis that incorporates the following:
  - Up-to-date risk initiating event data,
  - Operational and maintenance statuses,
  - Checks for component inter-dependency effects on reliability (especially important for complex industrial assets),
  - Component reliability and availability information, and
  - Safety barrier performance data.
- Digital twins may efficiently provide knowledge about an asset's behavior. Coupled with AI-based techniques that boost simulation speed, digital twins can facilitate quicker assessment of operations, conditions, and behaviors from the modeled asset. Increased simulation speeds at higher accuracies benefit complex industrial assets with interdependent components, such as advanced nuclear reactors, that typically require a substantial effort to make any predictions about their complex behaviors.
- Digital twins can also be used as a testbed to understand potential cybersecurity vulnerabilities that pose risks to asset design, operations, and maintenance.
- AI-enabled digital twins used towards risk analysis require more validation and verification methods to ensure that the incorporated data and models are correct. Methods are especially needed to test the degree of reliability of predictions from the AI-based models to prevent or keep track of uncertainty propagated to risk assessments.
  - These validation and verification measures also help with regulator and asset operator acceptance of conducting risk analysis with digital twins. Operators

ideally should know how much they should trust the resulting analysis and the evidence that sits behind any recommended decisions.

4. INFORMS Panel Discussion

The INFORMS panelists addressed several high-level questions . This section breaks down the three high-level topics used as clarifying discussions for issues raised in the presentations. The summarized responses of the participants below are good-faith efforts to represent the thoughts and assertions made during the event without verifying or qualifying any statements made. We intentionally omitted the speakers' direct quotes and assignments for conciseness to improve readability.

4.1. Risks of AI

4.1.1. Informally, the definition of risk can vary between people and applications. What do you view as notable risks associated with Industrial AI at the various levels of observable impact (i.e. algorithm-level, equipment-level, facility-level, enterprise-level, or society-level)?

Notably, AI presents a considerable risk if it can make decisions outside its verified training region. Similarly, there is a risk of misinterpreting the trustworthiness of an output from an IAI tool if the algorithmic-level metrics are somehow incomplete or biased to the training data. An example of this could occur if the training data is not properly qualified and the reported metrics are oblivious to the data's bias or noise. This dichotomy often leads to false representations of the performance of the AI.

As asset complexity increases, the lack of system observability becomes a greater risk to developing, using, and trusting IAI tools. Connected systems can have a cascading impact on observed effects within a target system. Without special considerations, an IAI could produce unexpected output or actions. In many cases, it is impossible to predict all such interactions and accommodate them a priori. Environmental conditions and reconfigurable connected systems exacerbate this effect.

Lastly, some risks come from antagonistic agents. The cybersecurity threat and the threat of data poisoning, intentionally injecting bad or maliciously altered training data, should also be considered. Most of these threats come towards larger organizations and software tools, but no enterprise is entirely immune and should consider installing preventative procedures in their software and culture.

4.1.2. What are barriers or viewpoints that might make someone resistant to adoption of IAI? Did you find any reasons you have encountered surprising?

Nuclear industry regulations are rigid and slow to change. People fear regulators may disagree on the requirements for using IAI tools in critical or non-critical systems. Future policy changes may require significant investments to comply with digital or AI systems and tools, especially for any early adopters who chose incompatible solutions with the new requirements. Given the rapid pace of the change in IAI technologies, a regulatory body's traditionally long and rigorous approval process presents the fear that any approved technology may already be out of date.

Beyond the nuclear industry, there are similar issues with the ability to certify an AI system in safety-critical applications. The expectation is that any technology would be sufficiently 'proven' before release, but it is unclear what it would mean to 'prove' an AI. Furthermore, it is often infeasible to fail-safe or review every possible state and outcome of an AI. Without ensuring reasonable expectations of mitigating failure or high-risk outcomes, decision-makers are hesitant to invest in AI solutions. This uncertainty also opens the question of liability if such a solution fails. There is no clear answer on who would be responsible in a legal sense at the moment.

Even before an AI technology goes through the complicated process of being regulatory-compliant or otherwise able to be licensed or certified, there is the problem of determining if it is worth it. Evaluation of an IAI solution sometimes comes down to proving usefulness and marketability. There must be a clear value proposition, both for the AI developer and the industry user, that shows that this solution to a problem is better than any currently out there and worth any potential additional risks or problems the solution might introduce. Unfortunately, building the business case for an AI-driven solution is not always straightforward, particularly in domains and applications where users tend to distrust new technologies.

The desire to have a human in the loop also factors into adopting or investing in IAI technologies. The lack of understanding of how an AI tool arrives at a decision or output inclines stakeholders to want to install human oversight. This precedent may complicate the construction of an AI tool by requiring the model's internal decision-making process to have some degree of explainability.

4.1.3. What are methods for mitigating the risks around IAI tools?

Effectively identifying and mitigating threats to and from IAI requires a formal connection to risk management processes such as PRA as an oversight process. Making this connection includes better monitoring of the behavior of an IAI tool and its impact on the asset. More importantly, it also incorporates routinely questioning the algorithm to understand its decision-making process and its implications. Technological solutions for risk mitigation

alone are only effective in the short term. Long-scale solutions include cultural changes in oversight views and the inclusion of humans in that process.

4.2. Evaluating IAI

4.2.1. What is a good measure or indicator of 'worth' when it come to the viability or usability of an IAI solution?

The goal of showing an AI's merits to a broad spectrum of stakeholders is to relay the returns and impacts of an AI solution. For many stakeholders, the bottom line is the monetary value gained or saved from investing in the AI solution. Monetary value can derive from many factors, such as products made, service up-time, or other facility success measures. For example, a success measure could be the ability to free up resources from menial or avoidable tasks or improve safety for the facility's equipment and workers. The AI product's return on investment is its most potent selling point.

4.2.2. Beyond algorithmic performance metrics, what are some informative and intuitive metrics that could be presented to a stakeholder less familiar with AI?

No single metric can completely encompass an IAI tool's performance or expected quality. Instead, there is a need for a revolving suite of metrics that may change a user's individual needs. Some metrics have no direct quantitative accessibility, such as usability, completeness, usefulness, or difficulty of training. Qualitative attempts to assess these issues will be case-dependent and may not translate across specific applications.

Metrics typically do not perform well when the focus of an effort lies in rare occurrences in AI model performance. Where the primary objective of an AI is to capture rare or previously unseen events, it is not easy to develop metrics that can accurately reflect future performance expectations. Furthermore, metrics that address the qualitative aspects of the data, both training and in-situ, do not commonly exist but need to be considered when evaluating an AI that relies on this data for training and operation.

System-based performance metrics are the most informative. However, these metrics can be challenging to construct for complex systems, where systems-level performance may also depend on the performance of many underlying factors and components.

4.2.3. Is there a level that is more appropriate to evaluate IAI from?

The system-level operators need to see value from an applied IAI tool. They are also most likely to recognize the value when presented and envision improvements in operations with the IAI tool. Their intuition can be the starting point for understanding the asset or facility level impact. Their insights can inform AI tool developers and vendors about business case development strategies that users may use to justify investments.

Although business-level assessment can drive many decisions, some areas cannot be solely driven by a business case and financial justification. The cost of investing in some IAI technologies for more long-reaching impacts can be societal. Incorporating these higher-level implications is imperative to IAI-based solutions for applications with significant environmental damage or public safety issues. For example, generational benefits such as climate change are outside most companies' profit planning purview. However, they are necessary for the global sustainment of resources and industry viability.

Given that the AI community needs to consider the societal side, it is difficult to quantitatively assess if users implement algorithms appropriately to achieve societal goals. Attempts could focus on more procedural qualitative assessment, which may require a shifting culture to use methods that give better qualitative interrogations of IAI impacts on social concerns.

5. Next Steps

This panel was part of a series of conversations and interactions among academics and industrial practitioners with expertise and experience in the risk assessment and management of complex systems and the evaluation of IAI-based tools. The goal is to establish a community and form a common understanding of the challenges, gaps, and opportunities relevant to evaluating IAI-based tools and technologies across industrial domains and academic disciplines.

This report describes the takeaways and productive exchanges from this panel and makes them available to the research community. The next steps include:

- Survey evaluation methods for domain-centric IAI technologies that enable tool selection via intuitively informative metric generation and business value justification.
- Create a roadmap to identify, examine, and alleviate barriers to risk-based evaluation and adoption of IAI-based tools and technologies. This effort includes advancing reproducible and repeatable evaluation guidelines and assessing the applicability of developed open-source tools used for this purpose.
- Facilitate additional panels and workshops to serve as a springboard for establishing a community of discourse regarding risk management and IAI. These will have the goal of providing spaces for community and consensus-building toward cross-disciplinary collaborations, general guidelines for risk assessment with IAI-based tools, and domain-centric best practices for evaluating such tools and technologies. These efforts should target a broad sections of stakeholders, including:
  - IAI experts from both industry and academia with experience or developing or evaluating IAI tools.
  - Experts from a wide array of domains and disciplines that have experience with risk assessment and management of complex systems.

- Researchers studying the impacts of AI technologies from a technology performance level to a social impact level.
- Standards communities interested in evaluation methods for IAI-based systems.
- Enterprises that consume or integrate IAI-based tools in their assets and are interested in evaluating their performance and use.

References

[1] Winkler RL Uncertainty in probabilistic risk assessment 54(2):127–132. https://doi.org/10.1016/S0951-8320(96)00070-1. Available at https://linkinghub.elsevier.com/retrieve/pii/S0951832096000701

[2] Nowlan FS, Heap HF Reliability-centered maintenance. Section: Technical Reports Available at https://apps.dtic.mil/sti/citations/ADA066579.

[3] Zio E (2016) Some challenges and opportunities in reliability engineering. *IEEE Transactions on Reliability* 65(4):1769–1782.

[4] Zio E (2022) Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety* 218:108119.

[5] Hu Y, Miao X, Si Y, Pan E, Zio E (2022) Prognostics and health management: A review from the perspectives of design, development and decision. *Reliability Engineering & System Safety* 217:108063.

[6] Compare M, Baraldi P, Zio E (2019) Challenges to iot-enabled predictive maintenance for industry 4.0. *IEEE Internet of Things Journal* 7(5):4585–4597.

[7] Moradi R, Groth KM (2020) Modernizing risk assessment: A systematic integration of pra and phm techniques. *Reliability Engineering & System Safety* 204:107194.

[8] Zio E (2018) The future of risk assessment. *Reliability Engineering & System Safety* 177:176–190.

[9] Haag S, Anderl R (2018) Digital twin–proof of concept. *Manufacturing letters* 15:64–66.