Coming to the Journal in April

Artificial Intelligence for Risk Analysis and the Risks of AI – Part 1

Vicki Bier, bier@engr.wisc.edu, University of Wisconsin-Madison
Emanuele Borgonovo, PhD emanuele.borgonovo@unibocconi.it, Bocconi University
Tony Cox, tony.coxjr@ucdenver.edu, UC Denver
Cynthia Rudin, cynthia.rudin@duke.edu, Duke University

The revolution created by artificial intelligence and machine learning (AI-ML) is affecting our society in a profound way. The use of machine intelligence is entering an increasing number of human activities, with AI helping to automate tasks and augmenting human abilities. Although AI-ML has the potential to benefit society, helping us recognize risks more promptly and manage them better, the disruption caused by these new technologies also presents new challenges to society. In particular, unthinking or uninformed application of these technologies in decision-making can result in biased and unreliable decisions. In addition, concerns about data privacy and security pervade the use of AI-ML tools. The associated risks and the potential consequences are difficult to assess and predict. This special issue collects a series of papers that showcase recent advances in research on AI-ML technologies and their relationship to the practice of risk analysis and risk-informed decision-making.

The papers clearly show that the transfer of knowledge and methods between risk analysis and AI-ML occurs in both directions. On the one hand, risk-analysis methods and theories contribute to framing and controlling the threats to society and human activities posed by AI-ML. On the other hand, AI-ML methods and tools are starting to be used in risk analysis.

The use of AI-ML in risk analysis is the subject of a thorough investigation in Stødle et al. (2025). The authors frame this integration as an input-output process that helps to accomplish the three main tasks of consequence identification, uncertainty characterization, and knowledge management. Through this framework, they analyze current applications and discuss potential future uses of AI-ML in risk assessment. They conclude with several recommendations for risk researchers and practitioners, highlighting the opportunities and limitations of the use of AI-ML in risk analysis.

Paté-Cornell (2025) identifies and investigates the concern that AI recommendations may not be consistent with the preferences and risk attitudes of the individuals involved in a given decision. Some of the questions that emerge are how to identify the risk attitude implied by an AI-ML tool, and how the results should be communicated and possibly modified before being integrated into the risk-analysis process.

Baum (2025) proposes a risk assessment of AI takeover (i.e., a potential takeover of key decisions by generative AI), with potentially catastrophic consequences for humanity. The author develops a stylized model that compares the type of capabilities that generative AI would need to have for such a takeover to happen against the capabilities of current large language models (LLMs). Based on that assessment, the author provides recommendations on whether more aggressive governance of LLMs is needed. We leave the answer to the paper.

Collier et al. (2025) investigate whether LLMs can play a role in product risk assessment. The authors began by using the popular ChatGPT to provide suggestions on tasks such as performing failure modes and effects analysis and recommending risk mitigations. The authors then presented the results to safety experts to evaluate ChatGPT's output. The same analysis was performed for additional LLMs. The expert examination identified significant limitations in product risk assessment, producing inconsistent, generic, and sometimes unrealistic guidance. However, researchers suggest these models may still be useful for initial ideation, with experts focusing on critically reviewing and refining AI-generated content to improve the overall risk assessment process.

Faddi et al. (2025) fill in a gap in the literature by developing quantitative methods to assess the reliability and resilience of AI-ML models to be applied in safety-critical domains. The work highlights the dynamic nature of the problem, since reliability and performance change over time. The authors demonstrate the approach by applying it to an image-recognition system subject to adversarial attacks.

Issa et al. (2025) study how the integration of AI technologies affects the work and experiences of finance professionals. The authors use a transactional-stress framework to examine how six technology-related stressors influence both positive and negative stresses experienced by finance professionals when using AI. They identify a new element, techno-accountability, as a significant risk driver. Techno-accountability is a new type of technological stressor that addresses the unique responsibility and liability challenges arising from AI-driven decision-making across various stakeholders. It focuses on the ethical and legal implications of unexpected AI outcomes, recognizing that users, developers, and organizations can be held accountable for decisions made by autonomous systems.

Madsen et al. (2025) study the use of AI tools and information coming from near misses to improve the prediction of maritime accidents. They use data from the U.S. Coast Guard’s Marine Information for Safety and Law Enforcement database, analyzing incidents from 2007 to 2022. After a systematic analysis, they test alternative machine-learning models and select the best-performing ones, whose high level of accuracy shows the effectiveness of using data from near misses to predict future incidents. These results show the potential for

implementation of the tool for the US Coast Guard, as well as for other industries whose risk-analysis practices use incident reporting.

Osman and El-Gendy (2025) study the potential economic impacts of AI-driven cyberattacks on global trade and supply chains. They apply computable general equilibrium modeling to carefully selected simulation scenarios to demonstrate the cascading effects of such cyberattacks across highly interconnected regions and industrial sectors.

Thekdi et al. (2025) address the problem of integrating advanced analytics and AI-ML tools for disaster risk management. They focus on the compatibility of these new tools with traditional methods and established practices in risk assessment and management. They develop a research framework, and evaluate the methods' compatibility using a survey of the community of risk researchers and practitioners.

Winter et al. (2025) develop an approach for analyzing non-technical dimensions of risk for an AI system. The system is a swarm of ten robots developed by engineers at the University of Bristol to work in a public cloakroom designed to receive, store, and return belongings. The method nests technological risks with other sources of risk encompassing human and contextual factors. The risks range from physical/safety hazards (collision between robots or between the robots and humans), to operational risks (how the robots deal with human belongings) or technical and organizational hazards.

In AI, causal modeling plays a central role in different domains, ranging from healthcare to financial applications of AI (the so-called causal AI). Yu and Smith (2025) propose a new graphical representation that nests causal elements in probabilistic graphs. These new chain event graphs are applied to understand and model system failures and repair processes. The new methodology is based on event trees and uses Bayesian inference techniques to accommodate complex and asymmetric system failure processes. The authors introduce a class of remedial interventions to model the causal effects of maintenance.

In the area of healthcare, Macrae (2025) performs an in-depth analysis of sociotechnical risks and resilience factors related to the use of AI-ML in healthcare, based on 40 in-depth interviews with professionals involved in the development, management, and regulation of AI. Finally, Li and Wang (2025) and Guo and Zhang (2025) successfully use AI-ML methods to enhance risk assessment in rescue operations and tunnel projects, respectively.

These works will illustrate the rich conceptual overlap and flow of ideas that link AI to risk analysis. These works represent the initial phases of these research intersections and provide an exploration of an exploding research field, with many research opportunities coming up as both AI and risk analysis rapidly evolve. We anticipate having a second volume soon with additional works continuing this important exploration.

References

Baum, S. D. (2025). Assessing the risk of takeover catastrophe from large language models. *Risk Analysis*. https://doi.org/10.1111/risa.14353

Collier, Z. A., Gruss, R. J., & Abrahams, A. S. (2025). How good are large language models at product risk assessment? *Risk Analysis*. https://doi.org/10.1111/risa.14351

Faddi, Z., da Mata, K., Silva, P., Nagaraju, V., Ghosh, S., Kul, G., & Fiondella, L. (2025). Quantitative assessment of machine learning reliability and resilience. *Risk Analysis*. https://doi.org/10.1111/risa.14666

Guo, K., & Zhang, L. (2025). Multisource information fusion for safety risk assessment in complex projects considering dependence and uncertainty. *Risk Analysis*. https://doi.org/10.1111/risa.17651

Issa, H., Dakroub, R., Lakkis, H., & Jaber, J. (2025). Navigating the decision-making landscape of AI in risk finance: Techno-accountability unveiled. *Risk Analysis*. https://doi.org/10.1111/risa.14336

Li, X., & Wang, X. (2025). Rescue path planning for urban flood: A deep reinforcement learning–based approach. *Risk Analysis*. https://doi.org/10.1111/risa.17599

Macrae, C. (2025). Managing risk and resilience in autonomous and intelligent systems: Exploring safety in the development, deployment, and use of artificial intelligence in healthcare. *Risk Analysis*. https://doi.org/10.1111/risa.14273

Madsen, P. M., Dillon, R. L., & Morris, E. T. (2025). Using near misses, artificial intelligence, and machine learning to predict maritime incidents: A U.S. Coast Guard case study. *Risk Analysis*. https://doi.org/10.1111/risa.15075

Osman, R., & El-Gendy, S. (2025). Interconnected and resilient: A CGE analysis of AI-driven cyberattacks in global trade. *Risk Analysis*. https://doi.org/10.1111/risa.14321

Paté-Cornell, E. (2025). Preferences in AI algorithms: The need for relevant risk attitudes in automated decisions under uncertainties. *Risk Analysis*, *44*(10), 2317–2323. https://doi.org/10.1111/risa.14268

Stødle, K., Flage, R., Guikema, S., & Aven, T. (2025). Artificial intelligence for risk analysis—A risk characterization perspective on advances, opportunities, and limitations. *Risk Analysis*. https://doi.org/10.1111/risa.14307

Thekdi, S., Tatar, U., Santos, J., & Chatterjee, S. (2025). On the compatibility of established methods with emerging artificial intelligence and machine learning methods for disaster risk analysis. *Risk Analysis*. https://doi.org/10.1111/risa.17640

Winter, P., Downer, J., Wilson, J., Abeywickrama, D. B., Lee, S., Hauert, S., & Windsor, S. (2025). Applying the “SOTEC” framework of sociotechnical risk analysis to the development of an autonomous robot swarm for a public cloakroom. *Risk Analysis*. https://doi.org/10.1111/risa.17632

Yu, X., & Smith, J. Q. (2025). Causal chain event graphs for remedial maintenance. *Risk Analysis*. https://doi.org/10.1111/risa.14308