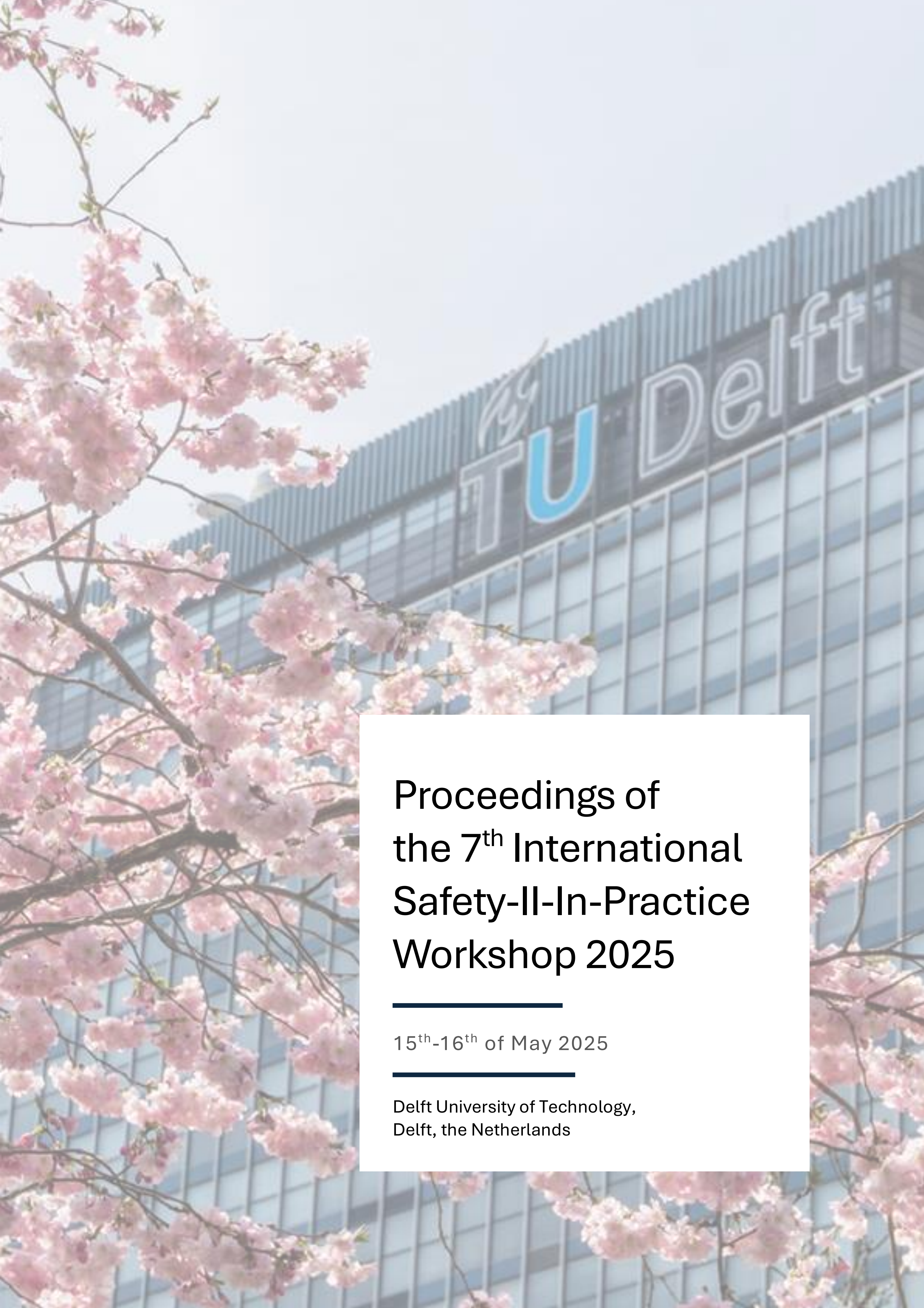
AI/ML for exploring future variability scenarios in complex sociotechnical systems under Safety-II principles

Trierveiler, Heron; Woltjer, Rogier; Sell, Denilson

2025

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Proceedings of the 7th International Safety-II-In-Practice Workshop 2025

15th-16th of May 2025

Delft University of Technology,
Delft, the Netherlands

Organisation Committee

Arie Adriaensen | *TU Delft*
Martina Ivaldi | *University of Genova*
Monique van der Toorn-Fennema **|** *TU Delft*
Natalie van der Wal | *TU Delft*
Paulina Zurawska | *TU Delft*
Wendela Nooteboom | *TU Delft*

Scientific Committee

Arie Adriaensen | *TU Delft*
Anthony Smoker | *Lund University*
David Slater | *Cardiff University*
Gesa Praetorius | *VTI*
Jean-Christophe Le Coze | *INERIS*
Maria Carlo Gonzalez | *University of Nottingham*
Riana Steen | *BI Norwegian Business School*
Riccardo Patriarca | *Sapienza University*
Rogier Woltjer | *Lund University*

Emeritus Member

Erik Hollnagel

Abstract Titles and Authors

Thursday, 15th of May

Safety-II-in-Practice 2025

Day 1

Safety-II-in-Practice 2025
Session 1

Session chair: James Norman | Independent Researcher and Airline Pilot

AI/ML FOR EXPLORING FUTURE VARIABILITY SCENARIOS IN COMPLEX SOCIOTECHNICAL SYSTEMS UNDER SAFETY-II PRINCIPLES

Heron Jader Trierveiler[1,2], Rogier Woltjer[1], Denilson Sell[2,3]

[1] Lund University, Lund, Sweden

[2] Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

[3] Santa Catarina State University, Florianópolis Brazil

Keywords: work-as-done (WAD); adaptative monitoring; performance variability; predictive modeling.

Introduction: In complex sociotechnical systems, the persistence of a gap between Work-as-Imagined (WAI) and Work-as-Done (WAD) is a recognized source of vulnerability [1], [2], [3]. While Safety-I approaches focus on analyzing failures after they occur, the Safety-II perspective emphasizes the need to understand and manage performance variability before it compromises safety [4]. A central challenge remains how to anticipate and act upon potential future deviations driven by both internal (e.g., organizational changes) and external (e.g., regulatory, environmental) dynamics. Traditionally, this anticipation relies heavily on expert judgment and experience – valuable but often limited in scalability and even on responsiveness [4], [5]. This research explores how Artificial Intelligence (AI) and Machine Learning (ML) methods, especially forecasting techniques, anomaly detection, and probabilistic modeling, can be leveraged to identify early signals of variability in operational contexts. These techniques allow organizations to process extensive datasets and detect emerging patterns that may indicate future WAD scenarios, thus reinforcing resilience and decision-making. While expert judgment remains critical, we hypothesize that AI/ML can serve as a complementary analytical layer to improve foresight in line with Safety-II principles.

Methods: The methodological foundation of this study is a structured review of the literature on AI/ML applications for detecting and forecasting variability in complex systems. The focus is threefold: (a) time-series forecasting models capable of detecting performance shifts over time [6], (b) anomaly detection algorithms that highlight rare or unexpected operational behaviors [7], and (c) probabilistic models, such as Bayesian networks, that support reasoning under uncertainty and scenario-based inference [8], [9], [10]. These methods are being analyzed both from a technical perspective and in terms of their suitability for integration into socio-technical systems that demand resilience and adaptability. In parallel, we are designing illustrative case-based examples to demonstrate how selected models can be applied in practice.

Results and Discussion: As the study is ongoing, the results presented are preliminary and intended to guide future stages of development and validation. Initial findings from the literature review suggest that AI/ML methods have strong potential to enhance the early detection of variability patterns that may not be evident through conventional monitoring approaches. For instance, time-series models [6], [11] show promise in identifying subtle trends and deviations in system behavior, while anomaly detection [7], [12] algorithms can uncover atypical signals at scale and in near real-time. Bayesian networks and other probabilistic frameworks allow analysts to model causal relations and explore potential future scenarios under different conditions [8], [10]. These capabilities are particularly relevant for addressing both endogenous variability (e.g., process drift, workload changes) and exogenous factors (e.g., market shocks, regulatory shifts, climate change) [13], [14], which may otherwise lead to normalization of deviance if left unchecked [15], [16]. Upcoming phases of the research will focus on applying the selected models to domain-relevant datasets and establishing protocols for collaborative validation with experts. Key concerns include data quality, the interpretability of ML outcomes, and the need to bridge the epistemic gap between algorithmic reasoning and practical knowledge [17], [18].

Conclusion: This work contributes to the field of safety and resilience engineering by investigating how AI/ML methods can be employed to predict operational variability in ways that are aligned with Safety-II principles. While still under development, the proposed approach emphasizes the proactive use of data to anticipate emerging risks and enhance adaptability in high-reliability organizations. The study also points to future opportunities for advancing the methodological integration between predictive analytics, expert knowledge, and safety-oriented practices, particularly in the design of more structured, adaptive, and context-aware safety systems.

Acknowledgments: This work was supported by CNPq - Brazil under project number 200953/2024-9.

REFERENCES:

[1] A. Adriaensen, Patriarca ,Riccardo, Smoker ,Anthony, and J. and Bergström, 'A socio-technical analysis of functional properties in a joint cognitive system: a case study in an aircraft cockpit', *Ergonomics*, vol. 62, no. 12, pp. 1598–1616, Dec. 2019, doi: 10.1080/00140139.2019.1661527.

[2] P. E. Hollnagel, 'Resilience engineering and the systemic view of safety at work: Why work-as-done is not the same as work-as- imagined', in *Bericht zum 58.*, Dortmund, 2012, pp. 19–24.

[3] R. Patriarca *et al.*, 'Framing the FRAM: A literature review on the functional resonance analysis method', *Safety Science*, vol. 129, p. 104827, Sep. 2020, doi: 10.1016/j.ssci.2020.104827.

[4] E. Hollnagel, *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*. Farnham, Surrey, UK England ; Burlington, VT: Ashgate, 2012.

[5] J. Rasmussen, 'Risk management in a dynamic society: a modelling problem', 1997, doi: 10.1016/S0925-7535(97)00052-0.

[6] G. Shmueli, 'To Explain or to Predict?', *Statist. Sci.*, vol. 25, no. 3, Aug. 2010, doi: 10.1214/10-STS330.

[7] V. Chandola, A. Banerjee, and V. Kumar, 'Anomaly detection: A survey', *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: 10.1145/1541880.1541882.

[8] R. Daly, Q. Shen, and S. Aitken, 'Learning Bayesian networks: approaches and issues', *The Knowledge Engineering Review*, vol. 26, no. 2, pp. 99–157, May 2011, doi: 10.1017/S0269888910000251.

[9] P. Krause and D. Clark, *Representing Uncertain Knowledge*. Dordrecht: Springer Netherlands, 1993. doi: 10.1007/978-94-011-2084-5.

[10] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 1. Aufl. s.l.: Elsevier Reference Monographs, 2014.

[11] R. J. Hyndman and G. Athanasopoulos, 'Forecasting: Principles and Practice'.

[12] M. Goldstein and S. Uchida, 'A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data', *PLoS ONE*, vol. 11, no. 4, p. e0152173, Apr. 2016, doi: 10.1371/journal.pone.0152173.

[13] E. Hollnagel, *Safety-I and Safety-II*. Boca Raton, FL: CRC Press, 2014.

[14] D. D. Woods and M. Branlat, 'Basic Patterns in How Adaptive Systems Fail', in *Resilience Engineering in Practice*, 1st ed., E. Hollnagel, J. Pariès, D. Woods, and J. Wreathall, Eds., CRC Press, 2017, pp. 127–143. doi: 10.1201/9781317065265-10.

[15] S. Dekker, *The field guide to understanding 'human error'*, Third edition. Farnham, Surrey, England ; Burlington, VT, USA: Ashgate, 2014.

[16] D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press, 1996.

[17] D. D. Woods, E. S. Patterson, and E. M. Roth, 'Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis', *Cognition, Technology & Work*, vol. 4, no. 1, pp. 22–36, Apr. 2002, doi: 10.1007/s101110200002.

[18] D. Woods and E. Hollnagel, *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Boca Raton, FL: CRC Press, 2006.

SAFETY AND RESILIENCE IN RESPONSIBLE, SUSTAINABLE ENGINEERING DESIGN, BELT AND BRACES

Ben J. M. Ale[1], **David H. Slater**[2]

[1] *Technical University of Delft, Mekelweg 5, 2628 CD Delft, The Netherland*

[2] *Cardiff University, Cardiff, UK*

Keywords: **Safety, Safety II, Resilience, Responsibility, Engineering Design**

Introduction: This paper explores the evolution of safety and resilience in responsible and sustainable engineering design, tracing the transition from traditional prevention-based approaches (Safety I) to resilience-focused strategies (Safety II). It reviews historical safety models, such as the