Human-AI interaction in safety-critical network infrastructures – extended abstract

Samira Hamouche[1], Toni Waefler[2]

[1, 2] University of Applied Sciences and Arts Northwestern Switzerland, School of Applied Psychology
samira.hamouche@fhnw.ch
toni.waefler@fhnw.ch

1 Introduction

Artificial Intelligence (AI) has substantial potential to support the operation of safety-critical network infrastructures, including power grids, railway networks, and air traffic management. These infrastructures are currently undergoing major transformations driven by climate change, energy transition, increasing mobility and energy demand, and the need for more resilient and efficient use of existing assets[1-3]. AI-based decision systems may help operators anticipate disruptions, explore alternative actions, optimize resources, and respond more flexibly to complex and dynamic situations. However, introducing AI into safety-critical infrastructures is not merely a technical challenge[4-11]. These systems are socio-technical in nature: they involve technical artefacts, human operators, institutions, rules, safety procedures, and multiple stakeholders. In such contexts, AI must not only be accurate and efficient, but also robust, reliable, transparent, trustworthy, and designed in ways that preserve meaningful human involvement. This is particularly important when decisions are time-sensitive, uncertain, and may have cascading consequences across an entire network. Therefore, the central question is not simply how AI can automate critical infrastructure operations, but how AI can collaborate with human operators in ways that improve decision-making while maintaining human agency, trust, learning, and safety[12-22]. Based on the original article, which conceptually brings together different perspectives, this extended abstract summarizes shared decision-making challenges across three safety-critical domains and outlines a psychologically grounded perspective on complementary human-AI collaboration. We will particularly highlight the latter part in our presentation at the Upper Rhine AI Symposium.

2 Shared decision-making challenges across domains

To understand decision dynamics in safety-critical network infrastructures, our article compares decision-making processes across power grid operations, railway network management, and air traffic management. Based on scenarios developed and analyzed with industrial stakeholders[23], we identify several shared challenges across these domains. Although the operational contexts differ, decision-making is shaped by similar characteristics: external events and uncertainty, network capacity constraints, coordination among multiple operators and stakeholders, large action spaces, tight time constraints, and trade-offs between conflicting objectives.

Decisions are rarely isolated; rather, they emerge from an iterative process in which human expertise and AI-driven insights are combined to explore, validate, and adapt possible courses of action. For example, in railway network operations, densely planned timetables are frequently disrupted by delays, infrastructure defects, or short-term maintenance. Dispatchers must

therefore make rapid rescheduling decisions while balancing punctuality, safety, capacity, and downstream effects across the network.

3 Designing enhanced human-AI systems

Our article argues that the design of AI-based decision systems requires the integration of technical robustness and human-centered interaction design. On the AI side, approaches such as distributed, hierarchical, and knowledge-assisted decision-making can help reduce the limitations of black-box AI by structuring decisions in ways that are more transparent and understandable for human operators[35-55]. Across the three domains, AI systems may therefore take different forms: recommending congestion management strategies in power grids, supporting or automating rescheduling in railway operations, or assisting airspace sectorization in air traffic management.

However, transparency cannot be reduced to explaining AI outputs. From a psychological perspective, AI also needs to support the cognitive processes through which operators monitor dynamic systems, build situation awareness, make decisions, and update mental models of the environment, other humans, themselves, and the AI[56-62]. Figure 1 illustrates how AI can provide transparency through explanation, exploration, animation, mirroring, and intuitive interface design. In this sense, AI design should not only make the system more understandable, but also support human learning, calibrated trust, motivation, and meaningful involvement in decision-making.

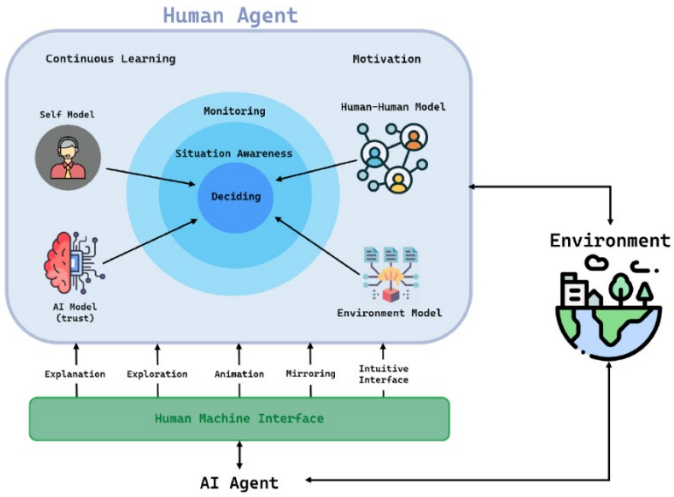


Figure 1. Model of human decision-making illustrating how AI can provide transparency to support human-AI collaborative decision-making through different forms of interaction.

The aim of our presentation at the Upper Rhine AI Symposium is to provide a more detailed overview of this framework and to emphasize the psychological aspects of human-AI collaboration.

4 Conclusion

Our work argues that AI-based decision systems in critical infrastructures should not focus on automation alone, but on optimizing the quality of decision support and human-AI interaction. This requires aligning AI systems with human cognitive processes and limitations, operational requirements, and rigorous safety constraints. In this view, AI is not merely a tool for automation, but a collaborative partner in safeguarding critical systems.

At the same time, effective human-AI collaboration requires more than explainable AI outputs. AI systems should provide transparency that supports operators in building accurate mental models, maintaining situation awareness, reflecting on decisions, and learning over time. By supporting these cognitive processes, AI and humans can act as a joint cognitive system that leverages the complementary strengths of both.

5 References

All references according to the original source: Mussi, M., Metelli, A. M., Restelli, M., Losapio, G., Bessa, R. J., Boos, D., Borst, C., Leto, G., Castagna, A., Chavarriaga, R., Dias, D., Egli, A., Eisenegger, A., El Manyari, Y., Fuxjaeger, A., Geraldes, J., Hamouche, S., Hassouna, M., Lemetayer, B., Leyli-Abadi, M., Liessner, R., Lundberg, J., Marot, A., Meddeb, M., Schiaffonati, V., Schneider, M., Stadelmann, T., Usher, J., Van Hoof, H., Viebahn, J., Waefler, T., & Zanotti, G. (2025). Human-AI interaction in safety-critical network infrastructures. *iScience, 28(9), 113400.* https://doi.org/10.1016/j.isci.2025.113400