Abstract Title: Human factors principles for the design and operation of artificial intelligence (AI) systems in rail

Author 1: Paul Michael Christopher LEACH
Department: Human Factors - Head of Human Factors
Institution / Company: Rail Safety and Standards Board
City: London
Country: UK

Author 2: Owen MCCULLOCH
Department: Human Factors – Human Factors Specialist
Institution / Company: Rail Safety and Standards Board
City: London
Country: UK

Author 3: Barbro ARNES
Department: Human Factors – Human Factors Specialist
Institution / Company: Rail Safety and Standards Board
City: London
Country: UK

Topics: *Innovation and Technologies*

Keywords: *Artificial Intelligence, Technical System Integration and Interaction, Human Factors*

Background / Context

Artificial intelligence (AI) has the potential to transform rail operations, with potential applications in customer service, optimisation of complex railway systems, and improving safety and security of rail networks. To do this, AI systems will need to maximise human strengths, mitigate human limitations, and optimise human performance and decision making. Human factors provides the methods, theoretical frameworks, and empirical data to achieve this.

While other sectors (healthcare and aviation) have taken steps to formalise human factors principles for AI, the UK rail industry currently lacks sector-specific human factors principles to inform AI development and deployment. As a result, there is a risk that AI systems may be introduced without fully accounting for how they affect human roles and operational contexts. This, in turn, could undermine AI's potential to deliver operational improvements and introduce new safety vulnerabilities.

Objectives

The Rail Safety and Standards Board in the UK completed research to identify and describe a set of rail human factors principles that can be applied to the design and operation of rail-specific AI systems. These can help:

- increase the likelihood that AI systems in rail optimise human performance and decision making
- reduce the risk that they deteriorate and impair human performance.

The paper and presentation will describe the principles, how they were developed and how they can be applied to optimise the design, implementation and operation of rail-specific AI systems.

Methods / Approach

A literature review was conducted to explore existing best practices, potential risks, and human factors enablers, considerations, and principles associated with AI implementation. The review focused on the automotive, healthcare, and aviation industries, where AI applications are more established compared with the UK rail sector. A total of 187 papers were reviewed.

A draft set of principles were developed and then tested and refined through a series of semi-structured interviews. The interviewees were experts in AI, rail, and/or human factors from safety-critical industries across the UK and Europe.

The final set of human factors principles was then developed, shaped by the literature, expert input, and internal validation.

Results / Findings

Seven human factors principles were created that can help facilitate the safe and effective integration of AI into railway operations. These are:

1. **User empowerment** - how AI systems should complement human strengths and mitigate human limitations through task allocation, decision-making support, and workload management.
2. **Ongoing capability development** – how humans remain competent, resilient, and prepared as AI capabilities evolve, rather than becoming passive or overly dependent users of AI.
3. **Collaborative human-AI teamwork** - the creation and implementation of human-AI teams and crafting an integrated partnership, with AI as an active teammate.
4. **Transparency and explainability** - the creation of clarity between AI systems and their users so that users can engage, trust and where required intervene with the AI.
5. **Calibrated trust** - designing AI so that it provides users with confidence in and reliance upon the decisions, actions, and outputs of the system.
6. **Ethical awareness** - how AI should align with, respect, and uphold social and cultural norms.
7. **Proactive organisational readiness** - the importance of preparing both the organisation and its people for the introduction of AI.

Conclusions

The adoption of AI in the rail industry introduces a new set of challenges and opportunities. Unlike highly automated systems, which operate on fixed rules and behaviours, AI systems are capable of learning, adapting, and making autonomous decisions based on complex data. These characteristics can make AI more powerful, but also less predictable, increasing the need for careful human factors consideration.

By focusing on user empowerment, organisations can ensure that AI systems complement human strengths and mitigate limitations, fostering a resilient human-AI partnership. Ongoing capability development keeps human skills aligned with evolving AI functionalities, preventing skill degradation and maintaining operational effectiveness.

Collaborative human-AI teamwork requires clear communication, shared mental models, and social acceptance. Transparency and explainability of AI systems build trust and enable users to understand and critically evaluate AI decisions. Calibrated trust means users neither overrely on nor underutilise AI and instead maintain a balanced and effective interaction.

Ethical awareness helps highlight and then address potential biases and moral implications of AI training data and decisions. Proactive organisational and system readiness help prepare both the organisation and its people, fostering a supportive environment for AI adoption.

The seven principles identify represent an initial contribution towards identifying and describing human factors challenges associated with rail AI and providing up-to-date guidance. Their future application to practical industry use cases will provide opportunities to evaluate and refine these principles in real-world settings.