



IRSC 2025

***Advancing Railway Safety through
Innovations and Collaborations***

Hong Kong, China 25th - 30th November 2025

**Optimising human/AI performance by
integrating human factors into the
design and operation of rail AI enabled
technology**

Paul Leach

Head of Human Factors

Rail Safety and Standards Board (RSSB)

Why Human Factors and AI?

- Artificial Intelligence (AI) adoption is growing across safety-critical domains, including rail
- Effective AI requires understanding and designing for human–AI interaction in real operational contexts
- RSSB set out to identify a rail-tailored set of HF principles for AI



Hosted by:



機電工程署
EMSD



AI vs Highly Automated Systems

AI Systems

- Learn & adapt; improve with data
- Handle unstructured tasks; probabilistic outputs
- Behaviour can change post-deployment

Highly Automated Systems

- Preprogrammed & static
- Narrow/repetitive tasks; predictable outputs
- Change only via human re-tuning

- AI's adaptivity/probabilistic reasoning reduces predictability and transparency
- Mental models are harder to form and maintain
- Raises need for trust, explainability and careful human-AI teaming
- Human–autonomous teams: “dynamic, interdependent coupling” needing collaboration and coordination to complete tasks (Cuevas et al., 2007)

Developing the Principles

Our focus

- Human-centred design principles relevant to AI systems
 - Operational implications for human–AI interaction
 - HF principles for introducing AI systems
 - Change-management approaches specific to AI integration
- Draft principles based on literature insights (187 academic papers)
 - Validation with international experts
 - Result = 7 Human Factors Principles



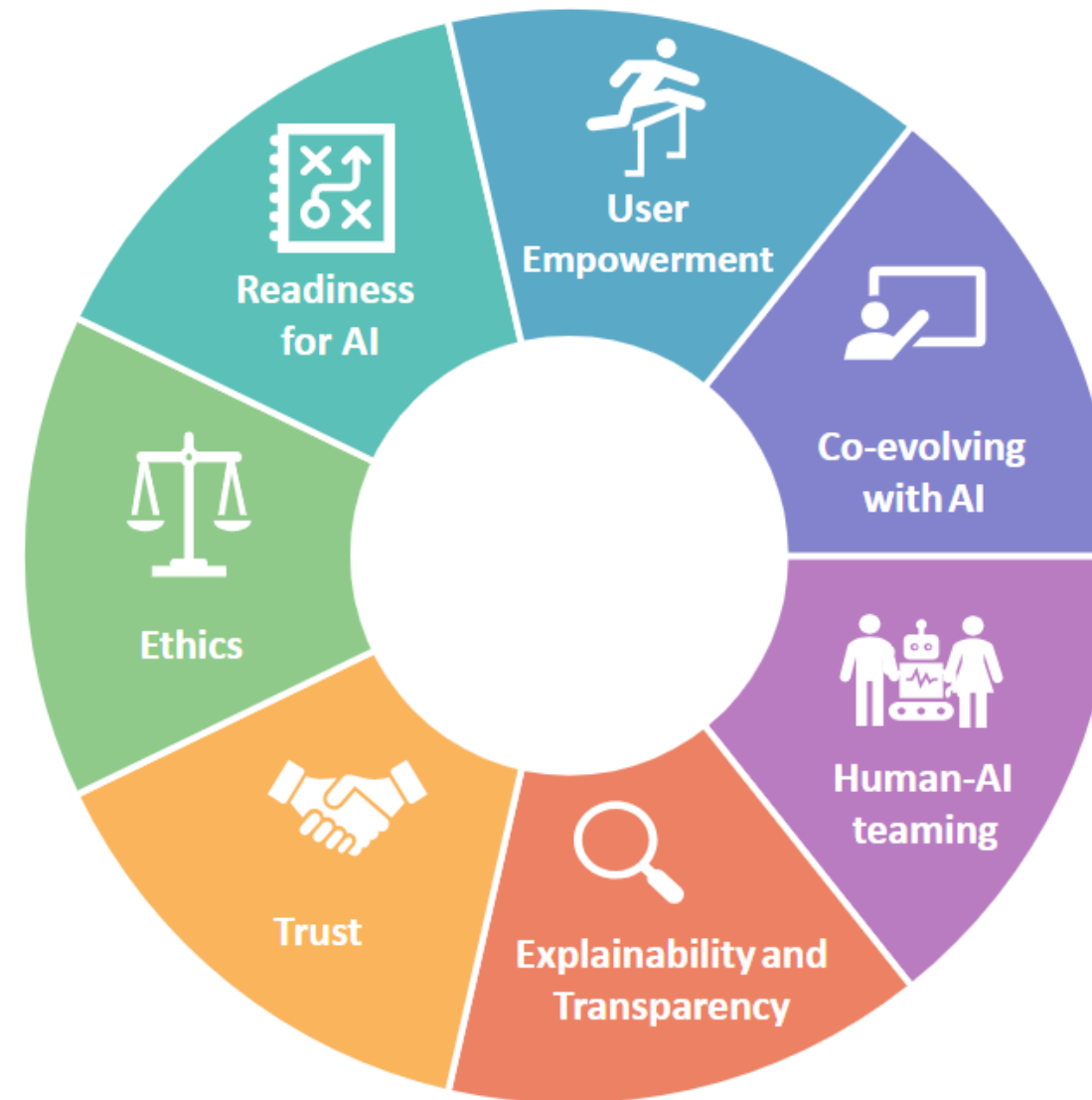
Hosted by:



機電工程署
EMSD



The Human Factors Principles



AI in Rail: A scenario



An AI-powered CCTV system monitors railway platforms for safety risks.

It was pre-audited to prevent bias from protected characteristics (e.g race, gender).

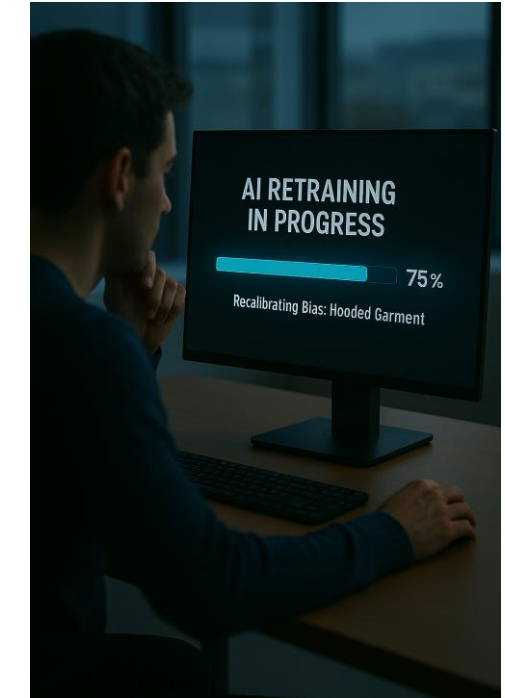


However, Safety Officer Anna has noticed that it repeatedly flags people wearing hoodies as high-risk. Anna escalates this for investigation.



The investigation reveals why:

Though the AI never used protected characteristics, it had learned to associate hoodies with trespass risk.



The AI is recalibrated to correct the bias.

From that point onward, the bias is tracked to monitor for recurrence.



Hosted by:



AI in Rail: A scenario



Principle 4: Set appropriate levels of explainability and transparency

This shows how explainability can help to detect unknown, emerging biases. By revealing the factors that the AI used to detect trespass risk, the system made its flawed reasoning easy to identify and correct.



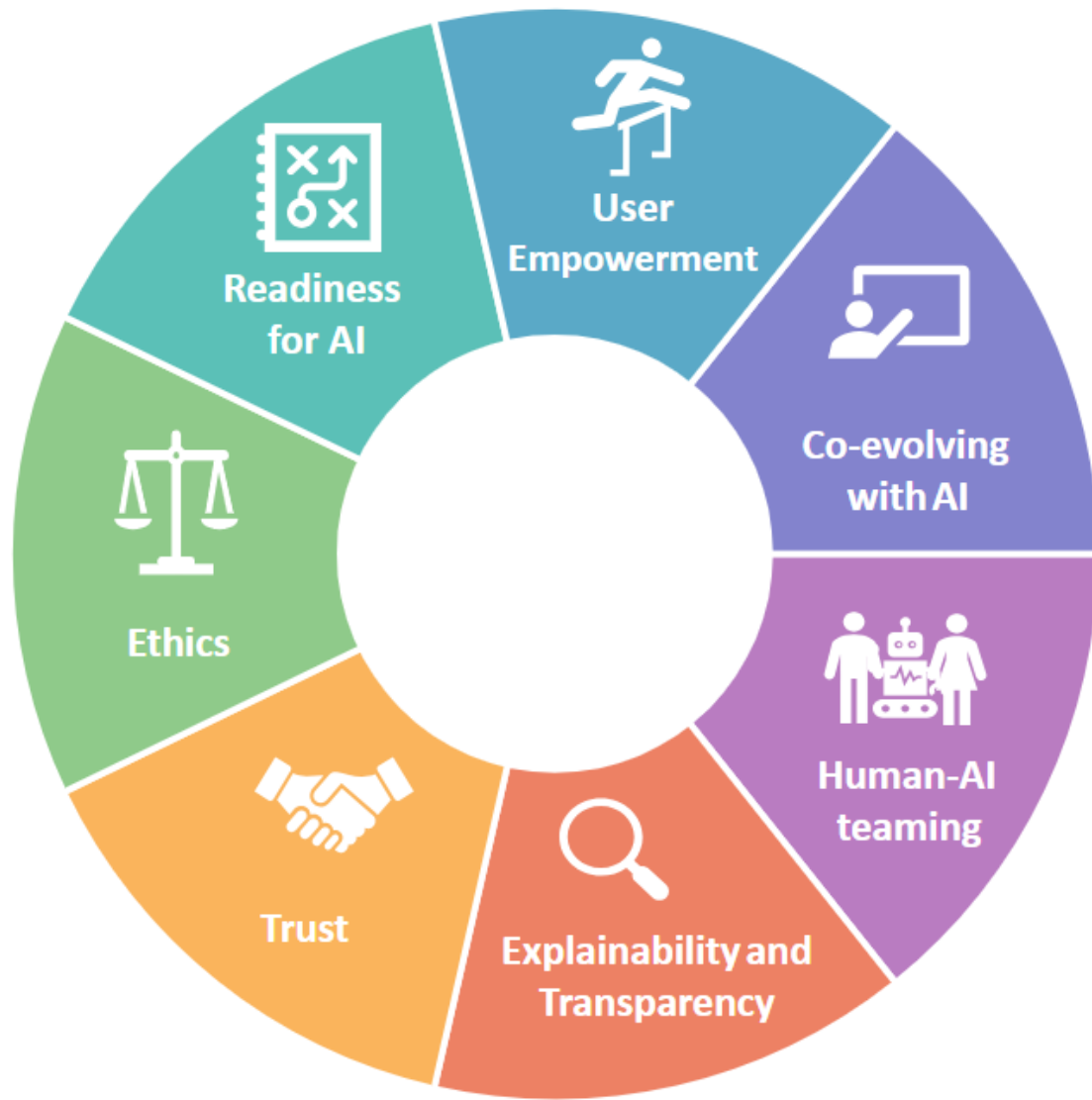
Principle 6: Define and overcome ethical challenges

The AI had learned to associate neutral features with risk (proxy-based bias). This risk was prepared for by empowering staff like the Safety Officer to report suspicious patterns freely.

The resolution combined short-term technical fixes (model recalibration) with structural safeguards (bias tracking) to uphold ethical integrity.

Future direction

- These 7 principles underline how AI and humans can work together and optimise one another
- The rail industry has an opportunity to embed robust HF practice from the outset with AI
- In parallel, RSSB is developing an AI evaluation toolkit to support effective decisions regarding AI
- Next steps is to apply these principles to real rail use cases



Exploration of human factors principles for the design, build, and operation of AI systems in rail (T1362)

<https://www.rssb.co.uk/research-catalogue/CatalogueItem/T1362>



www.irsc2025.com

Hosted by:

