Human Error, Reliability, Resilience, and Performance, Vol. 213, 2026, 57–66
https://doi.org/10.54941/ahfe1007554

Scenario-Based Human Factors Modelling of Safety-Security Escalation in Critical Socio-Technical Systems

Eylem Thron[1], Duncan Ki-Aries[2], Martin Freer[1], Huseyin Dogan[2], and Shamal Faily[3]

[1]Mima, London, UK
[2]Bournemouth University, Bournemouth, UK
[3]Defence Science Technology Laboratory, Wiltshire, UK

ABSTRACT

Cyber incidents in safety-critical systems rarely produce accidents through technical failure alone. Instead, they alter operational conditions by increasing cognitive demands, degrading information quality, and reshaping human–system interactions. These changes influence operator performance, creating escalation pathways through which cybersecurity disruptions propagate into safety-critical outcomes. Conventional safety and cybersecurity risk assessments often treat human performance as stable or secondary, limiting their ability to anticipate socio-technical failure mechanisms. This paper presents a human factors-driven, scenario-based modelling approach for analysing safety-security interactions in complex socio-technical systems. The method represents operator tasks, cognitive workload, interface behaviour, automation responses, and organisational influences within scenarios contrasting stable and cyber-degraded operations. By modelling how Performance Shaping Factors (PSFs) shift during disruption, the approach makes human performance visible as the mediating layer between cyber events and safety consequences. Implementation within the CAIRIS socio-technical modelling environment demonstrates how degraded operational states, human performance pressures, hazards, and escalation pathways can be captured in structured engineering artefacts. Illustrative rail signalling and defence supervisory control examples show how the approach reveals latent vulnerabilities in interfaces, procedures, and automation reliance often overlooked in traditional assessments. The work contributes a practical socio-technical modelling method that complements system-theoretic analyses by explicitly representing performance variability during cyber-disrupted operations and linking these dynamics to safety risk.

Keywords: Human factors, Cybersecurity, Safety-security interaction, Scenario-based modelling, Safety-critical systems, Socio-technical systems

INTRODUCTION

Safety-critical systems increasingly depend on interconnected digital technologies, automation and networked services. While these developments enhance efficiency and capability, they expand the cyber-attack surface and create new pathways through which cyber disruptions may propagate into operational and safety-critical failures. In domains such as rail and defence,

Received February 20, 2026; Revised April 4, 2026; Accepted April 23, 2026; Available online July 20, 2026

incidents unfold within socio-technical environments where operators must act under time pressure, uncertainty and degraded system conditions.

Safety consequences following cyber incidents are rarely the result of technical failure alone; they are mediated by human performance. Operators must manage degraded automation, conflicting alarms and incomplete information while maintaining safety and continuity under disrupted operational conditions. Under such conditions, cybersecurity disruptions reshape operational contexts by increasing cognitive demand, degrading information quality, and altering human-system interaction. This is where human actions often become the mechanism through which cybersecurity risks manifest as safety hazards (Thron & Faily, 2022). Performance Shaping Factors (PSFs) such as workload, time pressure, fatigue, interface usability and organisational constraints influence operator behaviour under these conditions, increasing the likelihood of unsafe adaptations during disruption. Many human errors reflect socio-technical design conditions rather than individual failings (Vicente, 1999). Poor interface design, unrealistic task assumptions, and organisational pressures increase the likelihood of unsafe adaptations during disruption.

Traditional safety methods, including FMEA, Bow-Tie, Event Trees, and HAZOP, are largely component-centric and assume normal human performance (Lees, 2012). System-theoretic approaches such as STPA conceptualise accidents as control failures within hierarchical structures (Leveson, 2011), yet human controllers are typically represented abstractly. Consequently, degraded operational contexts and performance variability are not systematically integrated into safety-security risk artefacts.

Despite mature HF theory, structured mechanisms for embedding performance variability within cybersecurity and safety engineering remain limited. Existing approaches frequently treat "operator error" as an outcome rather than a mechanism and provide little support for analysing how degraded operational states alter the reliability of human-centred safety barriers.

This paper addresses this gap by introducing a scenario-based modelling approach that operationalises PSF-driven performance variability within system engineering representations of tasks, control actions, hazards, and risk. Demonstrated through rail signalling and defence supervisory control scenarios and implemented within the CAIRIS socio-technical modelling environment, the approach enables traceable linkage between cyber-induced operational degradation, human performance, and safety outcomes, supporting Secure-by-Design development of safety-critical socio-technical systems.

METHODOLOGY: SCENARIO-BASED HF MODELLING

Escalation Mechanism

The modelling approach is grounded in a human-mediated view of escalation. Cyber events are not treated as direct precursors to hazards; instead, they reshape operational conditions, altering human performance and enabling

escalation to unsafe states. Shifts in PSFs influence cognition, behaviour, and control actions within socio-technical systems (Figure 1).

Figure 1: Human-mediated safety-security escalation mechanism in socio-technical systems.

Escalation is therefore modelled as a layered socio-technical process. Cyber events alter operational conditions, which reshape PSFs such as workload, time pressure, and information quality. These shifts influence cognitive processes including attention, situational awareness, and decision strategies. Changes in performance then affect control actions and the effectiveness of human-centred safety barriers. Hazards emerge through this human-mediated propagation, making performance variability the central risk escalation mechanism.

This model complements STPA (Leveson, 2011) by detailing the performance layers preceding unsafe control actions. Rather than stopping at "unsafe control action," the approach explains why operators may drift toward such actions under degraded conditions.

Making Human Performance Visible

Traditional risk models often represent human involvement as "operator error." In contrast, the proposed approach decomposes human involvement into task demands, PSFs, cognitive effects, and behavioural outcomes.

For example, rather than modelling "operator misroutes train," the model represents:

- **Task context:** Route confirmation under degraded signalling
- **PSF shift:** Workload increase, ambiguous feedback
- **Cognitive effect:** Reduced situational awareness, expectation bias
- **Behavioural tendency:** Acceptance of route state without full verification

This decomposition makes human performance an analysable system element rather than an abstract failure source.

CASE APPLICATIONS

Rail Signalling Scenario

A rail signalling example illustrates how a cybersecurity-induced interface degradation can propagate into safety risk through its effects on human performance. Signallers operate in time-critical environments that depend on reliable system feedback, stable automation behaviour, and clear alarm information to maintain safe train separation. In this scenario, a cyber-induced latency and feedback inconsistency in the signalling interface do not directly create a hazardous state; rather, it alters the operational context in which the signaller must make decisions. The resulting shifts in workload, information quality, and confidence in system state reshape cognitive processes such as situation awareness and expectation formation. This example demonstrates how a safety barrier that is normally treated as stable - operator route verification - becomes performance-sensitive under degraded conditions, enabling escalation from cyber disruption to collision risk.

Table 1 illustrates how the model captures the escalation pathway for the scenario.

Table 1: Example escalation pathway (rail scenario).

Stage	Representation in Model	Effect
Cyber event	Signalling interface latency manipulation	Interface state unreliable
Operational state	Degraded feedback, alarm ambiguity	Operator uncertainty increases
PSF shift	Workload (increase), information quality (decrease)	Cognitive demand rises
Human performance effect	Reduced SA; reliance on expectation	Verification quality reduced
Control action	Route confirmation accepted	Safety barrier weakened
Hazard	Conflicting route set	Separation margin compromised
Safety consequence	Collision risk	Collision

The model demonstrates how a safety barrier typically assumed reliable becomes performance-sensitive under degraded PSFs, revealing latent vulnerabilities often overlooked in conventional assessments (Thron & Faily, 2022; Thron et al., 2024).

Defence Supervisory Control Scenario

The defence supervisory control scenario demonstrates how degraded automation behaviour creates safety and mission risk through its effects on operator performance. Supervisory roles involve monitoring multiple information sources and intervening when anomalies occur. In this case, cyber-induced degradation increases monitoring demand and reduces confidence in automation outputs.

The resulting shifts in PSFs - particularly workload, divided attention and perceived automation reliability- affect anomaly detection and response timing. This example shows how automation support, normally a resilience feature, becomes a vulnerability when performance capacity is exceeded.

The defence supervisory control escalation pathway is summarised in Table 2.

Table 2: Example escalation pathway (defence supervisory control scenario).

Stage	Representation in Model	Effect
Cyber Event	Degraded automation behaviour and display reliability	Automation outputs inconsistent or delayed
Operational State	Increased monitoring demands; reduced confidence in automation	Operator must manually cross-check system state
PSF Shift	Workload (increased); divided attention (increased); information reliability (reduced)	Cognitive resources stretched
Human Performance Effect	Delayed anomaly detection and prioritisation	Emerging issues not recognised promptly
Control Action	Late or absent supervisory intervention	Corrective action delayed
Hazard	System operates outside safe / mission limits	Safety or mission risk
Safety Consequence	Safety or mission risk escalates	Safety violation

TOOL-SUPPORTED IMPLEMENTATION USING CAIRIS

The modelling approach is implemented in Computer Aided Integration of Requirements and Information Security (CAIRIS), a socio-technical modelling environment that supports structured representation of tasks, contexts, PSFs, threats, vulnerabilities, hazards, events, and risks (Faily, 2018).

Figure 2 defines how socio-technical elements such as tasks, contexts, PSFs, events and hazards are structurally connected and represented within CAIRIS, and how socio-technical elements are structurally connected. Tasks are associated with contexts and PSFs. Threats and vulnerabilities influence operational states. Hazards are linked to events and risk scores. This structure enables traceability from degraded operational conditions to safety outcomes.

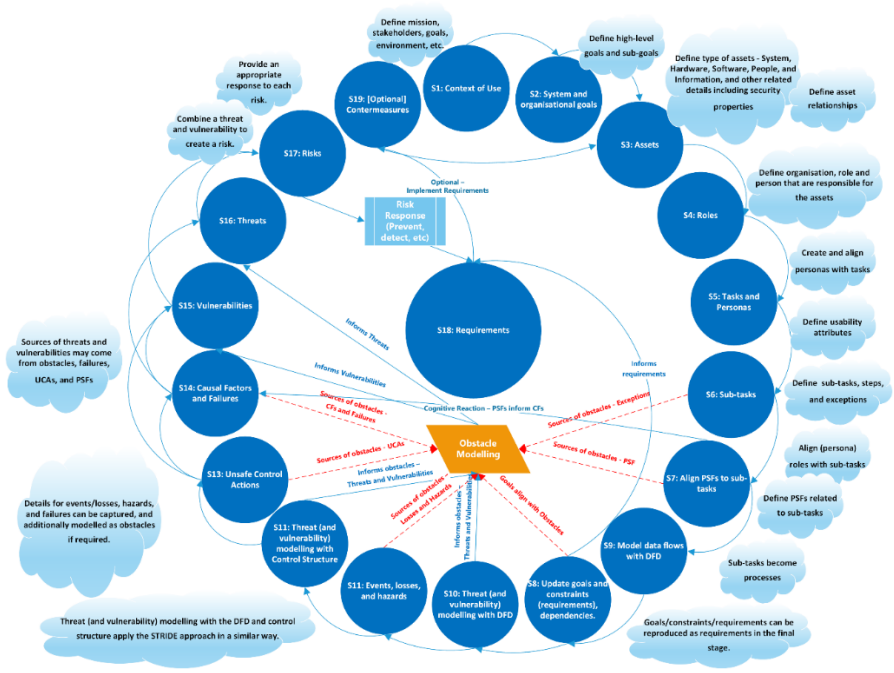

Figure 2: CAIRIS conceptual model.

Defence Scenario Representation

Figure 3 illustrates the interplay between system and human factors. Threats and vulnerabilities influence automation and overall system reliability, while tasks are linked to PSFs capturing workload, loss of situational awareness, and divided attention. Hazards emerge through delayed intervention, and events are connected to safety or mission-level risk, providing an integrated view of how human performance and system conditions combine to shape operational outcomes. This mapping is critical for HF-informed cyber risk analysis, as it allows potential escalation pathways to be anticipated and mitigated before they compromise safety or mission objectives. Importantly, it demonstrates how performance variability can be encoded directly within engineering artefacts, rather than treated separately from risk models.

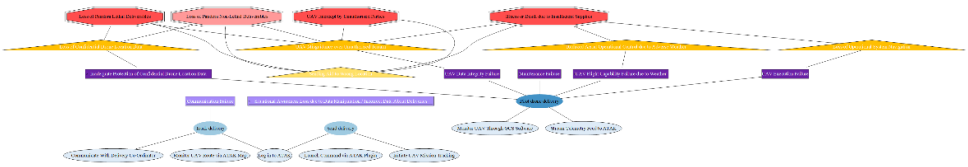

Figure 3: Integrated CAIRIS representation of degraded defence supervisory control scenario.

Rail Scenario Representation

Figure 4 shows the CAIRIS representation of rail tasks under degraded conditions. Elevated PSF levels indicate increased cognitive demand and reduced information reliability.

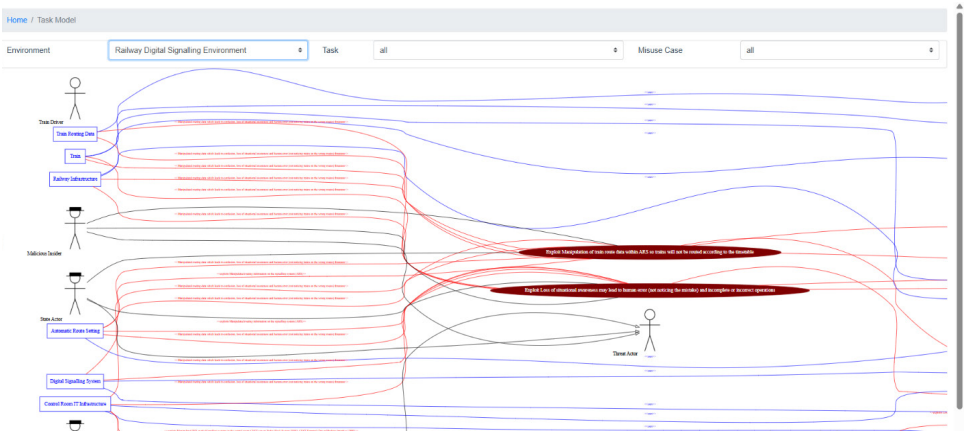

Figure 4: Rail operational tasks under cyber-degraded conditions.

Figure 5 presents the event model linking PSF-driven performance degradation to hazards and safety consequences. This traceability demonstrates how route verification reliability decreases under specific PSF configurations, affecting hazard likelihood.

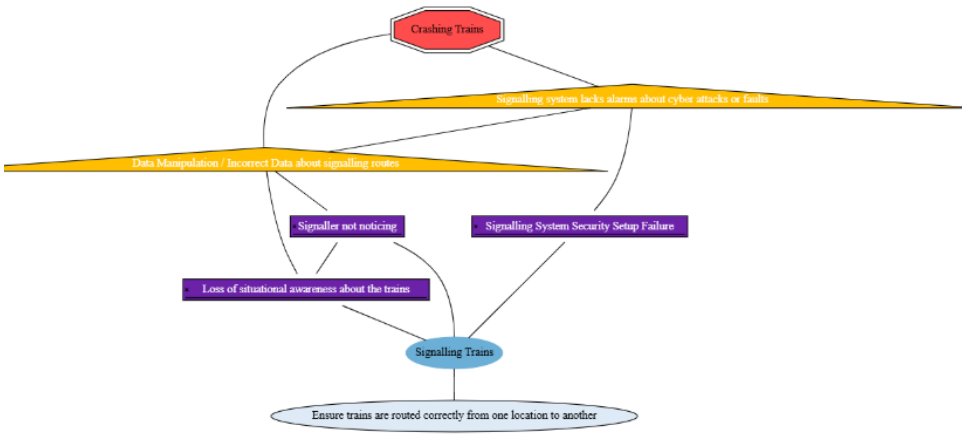

Figure 5: CAIRIS event model linking degraded human performance to rail safety outcomes.

Together, Figures 4 and 5 operationalise the escalation mechanism presented in Figure 1.

HOW THE APPROACH DIFFERS FROM TRADITIONAL METHODS

The proposed approach differs from traditional safety and cybersecurity methods in several key respects. Conventional analyses often treat human involvement as "operator error" and assume stable performance. In contrast, the HF scenario method models human performance as a cognitive mechanism shaped by context-dependent Performance Shaping Factors (PSFs).

Traditional methods typically represent barriers as fixed and event propagation as a technical sequence of failures. The proposed approach instead treats barriers as performance-sensitive and escalation as socio-technical, showing how cyber events reshape operational conditions, alter PSFs, and influence behaviour before hazards emerge. By explicitly modelling performance variability, the method reveals latent vulnerabilities that remain hidden in static risk representations.

Table 3: Key conceptual differences between traditional risk methods and the HF scenario-based approach.

Aspect	Traditional	HF Scenario Method
Human role	"Operator error"	Cognitive mechanism
Performance assumption	Static	PSF-dependent
Barriers	Stable	Performance-sensitive
Event propagation	Technical	Socio-technical
Latent vulnerabilities	Hidden	Revealed

TRANSLATING HF INSIGHTS INTO ENGINEERING DECISIONS

Figure 6 presents a high-level view of how human factors expertise is systematically translated into traceable engineering inputs.

Figure 6: Scenario-based human factors (HF) to engineering solutions.

The following table summarises two illustrative scenarios and the corresponding HF-informed design actions.

Table 4: Translating human factors insights into engineering decisions via PSF-driven performance analysis.

Scenario	HF Insight / PSF Shift	Effect on Barrier / Performance	Engineering Implication	Design Action
Rail signalling	Verification degrades under high workload and ambiguous feedback (workload ↑, info quality ↓)	Safety barrier "route confirmation" becomes unreliable	Route confirmation cannot be treated as fully reliable during degraded operations	- Improve interface state transparency - Prioritise alarm design - Introduce degraded-mode procedural support
Defence supervisory control	Divided attention and automation reliance delay anomaly detection (divided attention ↑, confidence ↓, workload ↑)	Monitoring tasks exceed cognitive capacity, delaying detection and corrective action	Monitoring tasks exceed cognitive capacity under degraded automation	- Redistribute monitoring tasks - Improve alert salience - Define degraded-mode staffing requirements

By explicitly linking PSF-driven performance effects to barrier reliability and hazard likelihood, the approach transforms HF expertise into traceable engineering inputs, including interface requirements, procedural redesign, training priorities, and degraded-mode operational strategies.

DISCUSSION

Scenario-based HF modelling reveals escalation pathways that technology-centric assessments often overlook. Rail and defence examples show how cyber degradation reshapes PSFs, influencing cognitive processes and barrier reliability. The method complements system-theoretic approaches (Leveson, 2011) by detailing the human performance mechanisms preceding unsafe control actions. In contrast to descriptive HF analyses, the approach encodes performance variability directly within system representations of tasks, hazards and risk, allowing degraded operational conditions to be analysed within the same artefacts used for safety and security assessment.

The analysis produces operationally actionable guidance: workload and task complexity can inform staffing and task allocation, procedural redesign reduces reliance on error-prone workarounds, and automation support and alerting strategies enhance situational awareness and decision-making under degraded conditions. Scenario-based modelling thus bridges the gap between HF expertise and engineering practice, providing a structured framework to anticipate escalation pathways and strengthen safety and cybersecurity resilience before deployment.

Case studies illustrate how PSF-driven performance variability exposes vulnerabilities in assumptions about human reliability, particularly under

degraded automation and interface conditions. Importantly, the method complements, rather than replaces, system-theoretic approaches.

CONCLUSION

This paper presented a scenario-based HF modelling approach that explicitly represents performance variability as the mediating layer between cyber disruptions and safety consequences. Implemented within CAIRIS and demonstrated through rail and defence scenarios, the method enables traceable linkage between PSF shifts, control actions, hazards, and risk.

By translating HF insights into structured engineering artefacts, the approach supports Secure-by-Design development of safety-critical socio-technical systems. Future work will explore enhanced visualisation, AI-assisted scenario generation, and validation in operational contexts. The approach is applicable across multiple critical national infrastructure domains and supports human-centred safety and security assurance activities of safety-critical systems.

ACKNOWLEDGEMENT

We thank Egemen Öner for their support in creating the CAIRIS visualizations.

REFERENCES

Faily, S., 2018. *Designing usable and secure software with IRIS and CAIRIS*. New York, NY, USA: Springer International Publishing.

Lees, F., 2012. Lees' Loss prevention in the process industries: Hazard identification, assessment and control. Butterworth-Heinemann.

Leveson, N. (2011) Engineering a Safer World: Systems Thinking Applied to Safety. Cambridge, MA: MIT Press.

Thron, E., Faily, S., Dogan, H., and Freer, M., 2024. Human factors and cyber-security risks on the railway–the critical role played by signalling operations. Information & Computer Security, 32(2), pp. 236–263.

Thron, E. and Faily, S., 2022. Automation and Cyber Security Risks on the Railways-the Human Factors implications. Contemporary Ergonomics & Human Factors 2022, p. 355.

Vicente, K.J. (1999) Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Boca Raton, FL: CRC Press.