Secure-by-design through Integrated Security, Safety and Human Factors

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

[1]Mima Group, [2]Bournemouth University, [3]Defence Science and Technology Laboratory

SUMMARY

Cyber-attacks increasingly threaten critical infrastructure, where interactions between security, safety, and human-system behaviour create complex socio-technical risks. If not managed early, these interactions can produce latent vulnerabilities and unsafe operational states.

This paper presents a Minimum Viable Product (MVP), developed by Bournemouth University and Mima and funded by the Defence Science and Technology Laboratory (Dstl), to operationalise Secure-by-Design through integrated Human Factors (HF), safety, and cybersecurity analysis. The MVP combines System-Theoretic Process Analysis (STPA) with Hierarchical Task Analysis (HTA), Cognitive Task Analysis (CTA), Performance Shaping Factors (PSFs), and Human Attributes analysis to generate a structured and traceable User Requirements Document (URD) from a Defence specification exemplar.

Results demonstrate that integrating HF, safety, and cybersecurity during early capability definition enables identification of cross-domain risks and supports derivation of coherent, traceable Secure-by-Design requirements for cyber-physical systems.

KEYWORDS

Secure-by-Design, Human Factors, Cybersecurity, Safety, Minimum Viable Product, Critical National Infrastructure

Introduction

Cyber-physical systems within Critical National Infrastructure (CNI) and defence environments operate at the intersection of digital systems, physical processes, and human operators. Cyber-attacks increasingly target CNI, intensifying tensions between security, safety, functionality and, in military contexts, survivability. Decisions affecting authentication, automation, safety interlocks, and operator workload are often made in parallel during early capability definition, limiting visibility of cross-domain interactions and introducing latent vulnerabilities.

Secure-by-Design promotes embedding cybersecurity from the earliest stages of system development. However, guidance on integrating safety engineering and Human Factors (HF) analyses alongside cybersecurity remains limited (Altaf et al., 2021, Ki-Aries et al., 2018; 2022; Thron and Faily, 2022). HF, safety, and cybersecurity are typically conducted as separate activities using different artefacts and modelling approaches. As a result, interactions - such as how degraded human performance may undermine safety barriers or expose cyber vulnerabilities - are rarely captured within a unified framework.

This paper presents a project delivered by Bournemouth University and Mima, funded by Dstl, to develop an MVP demonstrating how HF, safety, and cybersecurity analyses can be aligned within a Secure-by-Design methodology. The MVP was applied to a Defence specification exemplar and used to generate a structured URD embedding capability, safety, security, and HF requirements.

Background

Secure-by-Design encourages early integration of cybersecurity into system architecture rather than treating it as a late lifecycle add-on (Ministry of Defence, 2025). In defence acquisition and other safety-critical domains, this aligns with expectations for early hazard analysis, traceable requirements, and structured assurance. Despite the availability of mature HF, safety and cybersecurity methods, each discipline often implements these methods independently. Outputs do not naturally converge into a coherent set of design constraints, limiting early detection of conflicts between safety, usability and security requirements.

Interactions between safety, security and HF generate inherent design tensions. Security controls, such as strong authentication or monitoring may increase cognitive demand, potentially delaying time-critical safety actions. Safety mechanisms, such as fail-safe states or remote access for maintenance, may inadvertently expand attack surfaces (Kriaa et al., 2015). Automation reduces routine workload, but may erode situational awareness (SA), increasing reliance on operators during abnormal or cyber-degraded conditions (Thron et al., 2024). Reduced SA also means that operators may be unprepared to handle abnormal or cyber degraded conditions. Procedural assumptions often fail to capture real-world human performance, where fatigue, workload and organisational pressures can cause deviations. Without early integration, these conflicts typically surface during detailed design or operation, where redesign is costly.

Research Objective and Approach

The project aimed to operationalise Secure-by-Design by integrating HF, safety and cybersecurity analyses within a unified workflow. The objective was to make socio-technical risk pathways explicit and generate structured, traceable requirements during early capability definition.

Rather than evaluating the MVP against an existing URD, the process was applied to a Defence specification exemplar (case study). The exemplar defined operational intent and context. Through integrated modelling, the MVP transformed this into a structured, multi-domain URD incorporating safety, security, and HF considerations.

Role of Analytical Methods Within the MVP

The MVP integrates established analytical techniques within a structured socio-technical model.

HF analysis provides the behavioural foundation. HTA decomposing operational activities into structured subtasks to clarify task goals and dependencies (task hierarchy and subtasks) (Kirwan & Ainsworth, 1992; Stanton et al., 2005). CTA examines decision-making processes, mental models and workload demands, revealing steps where errors or delays may arise (Knisely et al., 2021). PSFs such as fatigue, workload and environmental conditions, characterise factors influencing human performance variability and are key inputs to human reliability and performance analysis (UK Government Cyber Security Advisory, 2025). Human Attributes analysis captures operator skills, training and cognitive limitations. Together, these analyses reveal task vulnerabilities and performance sensitivities relevant to safety and cybersecurity.

Safety analysis is conducted using STPA, a systems-theoretic hazard analysis method that models control structures, supervisory relationships and feedback loops to identify Unsafe Control Actions (UCAs) and causal pathways to hazards (Leveson, 2011). Within the MVP, UCAs are explicitly

linked to task steps and performance conditions identified through HF analysis for traceable hazard linkage.

Cybersecurity analysis adopts a socio-technical perspective: system boundaries, roles and goals are defined using personas to contextualise stakeholders and motivations in threat scenarios (Altaf et al.,2021; Dev et al., 2023); assets are identified and mapped to tasks; data flow modelling represents information movement across components and trust boundaries, supporting systematic threat enumeration and risk assessment and threat modelling identifies vulnerabilities and attack vectors linked to control structures and human interaction points within the socio-technical system (Fortinet, 2025; UpGuard, 2025).

Integration across these methods enables identification of how degraded performance, unsafe control actions and cyber vulnerabilities may interact. This structured linkage forms the basis for requirement derivation.

Artefacts in the MVP

Table 1 summarises the key analytical artefacts across Human Factors, safety, and cybersecurity, and their respective roles in the integration framework.

Table 1: Artefact Types and Roles (Inputs to the MVP)

Domain	Artefact	Role in Integration
Human Factors	Tasks (HTA), Decisions (CTA), PSFs, Human Attributes	Identify performance variability and cognitive demands
Safety	Control Structures, UCAs and Failures, Hazards, Events	Identify unsafe system states
Cybersecurity	Assets, Tasks/Sub-tasks, Goals, Roles, Personas, Dataflow Diagram (DFD), Threat modelling (and Obstacle modelling), Threats, Vulnerabilities, Risks, Requirements	Identify attack surfaces and exposures

Together, these domain-specific artefacts provide the structured basis for identifying cross-domain risks and deriving traceable Secure-by-Design requirements.

MVP Architecture and Workflow

The MVP operates as a structured integration process transforming a specification exemplar into a traceable URD. Initial cybersecurity scoping informs architectural understanding and asset identification. HF and safety analyses then model operational tasks and control relationships. These artefacts are integrated into a socio-technical model from which Secure-by-Design requirements are derived.

Figure 1 presents the high-level workflow, illustrating that Secure-by-Design is not a linear security activity but an iterative socio-technical process. Cybersecurity scoping informs task and control modelling, while HF and safety analyses refine risk understanding. Requirement derivation emerges from iterative cross-domain integration rather than serial disciplinary processing.

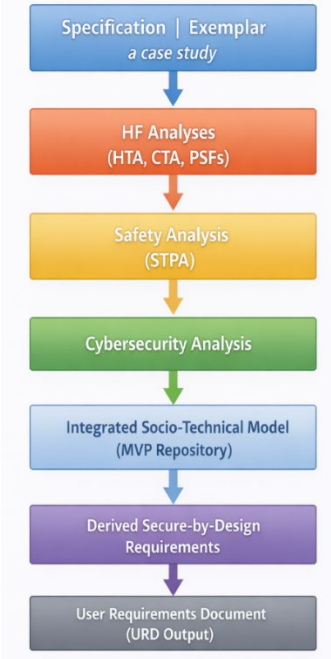

Figure 1: High-level MVP Workflow

Figure 2 presents the detailed MVP socio-technical workflow. It shows how cybersecurity scoping, Human Factors analysis (HTA, CTA, PSFs and Human Attributes) and STPA-based safety modelling interact iteratively rather than sequentially. Task and decision modelling inform identification of UCAs, which are then examined for associated vulnerabilities and threat pathways. The figure highlights how cross-domain linkages are established within the socio-technical model, enabling traceable derivation of Secure-by-Design requirements.

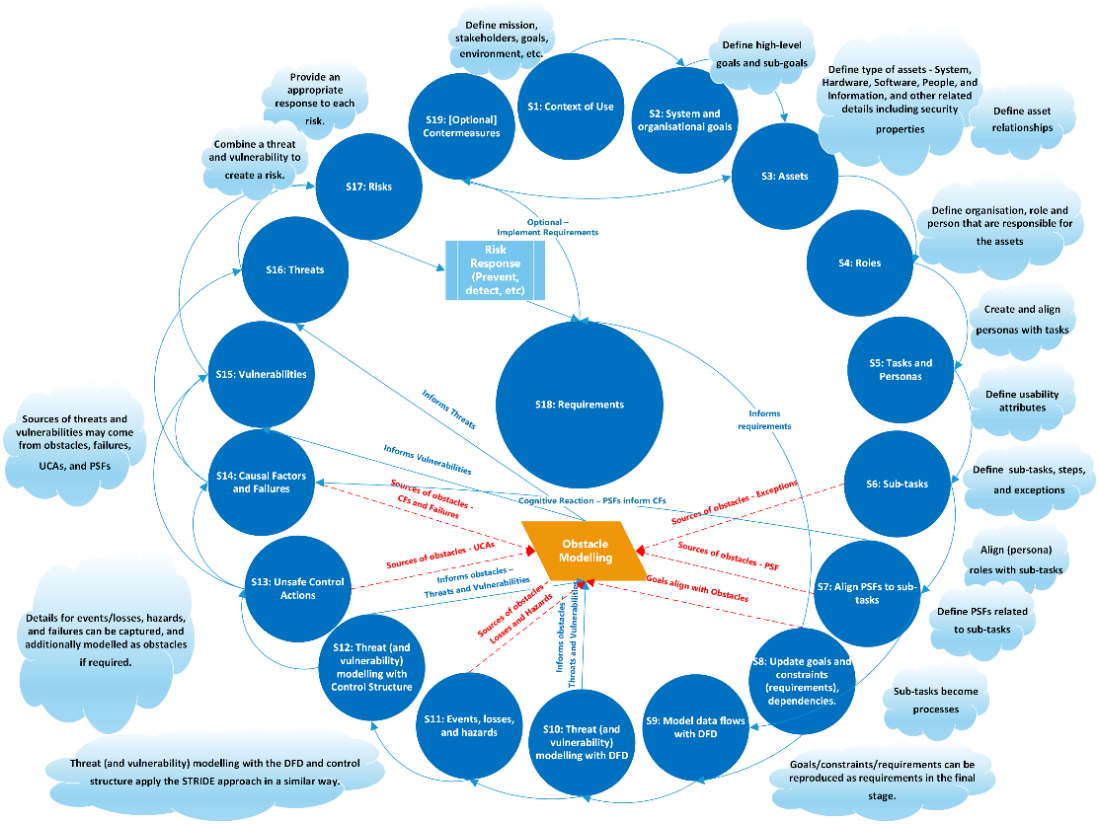

Figure 2: MVP Workflow

CAIRIS implementation

The MVP was implemented using CAIRIS (Computer Aided Integration of Requirements and Information Security), an open-source socio-technical modelling platform designed to maintain explicit traceability between goals, tasks, assets, threats, risks and requirements (Faily, 2018). The aligning process was applied assuming roles taken by a requirements manager, central to the data specification and analysis, whilst including a security analyst and a HF/safety analyst to perform related activities and analysis with the requirements manager.

Within the MVP, CAIRIS was extended to incorporate HF and safety artefacts alongside existing cybersecurity elements. The extended task interface and models represented operational behaviour and capturing related PSFs. Whereas, STPA-derived control structures captured supervisory relationships, feedback mechanisms and UCAs. Based on tasks performed, the Data Flow Diagram (DFD) (Figure 3) represents system components, information exchanges and trust boundaries, supporting identification of assets and potential attack surfaces.

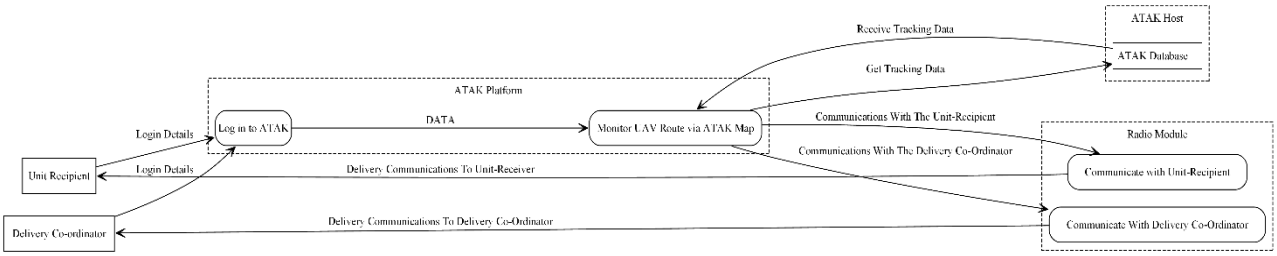

Figure 3: Data Flow Diagram

The DFD aligns with the Control Structure Model (Figure 4), which represents supervisory relationships between controllers and controlled processes. UCAs identified through STPA are embedded within this structure, linking behavioural and system-level failures to hazards. These artefacts were linked directly to assets, vulnerabilities and risk scenarios, forming a coherent socio-technical model rather than parallel analytical outputs.

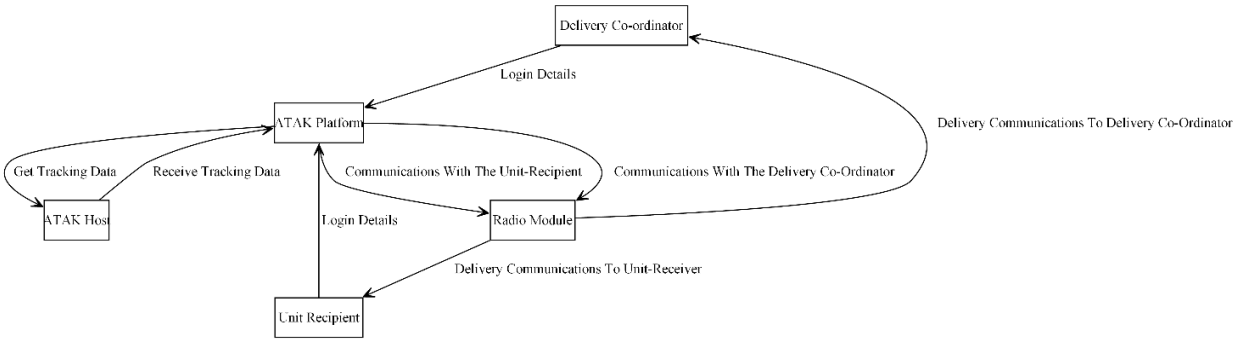

Figure 4: Control Structure Model

To demonstrate certain findings, the Filtered Event Model shown in Figure 5 demonstrates the alignment of one task and its subtasks, where conditions specified in the interface lead to the realisation and specification of a potential failure and hazards, that combined could lead to a loss event (accident or incident).

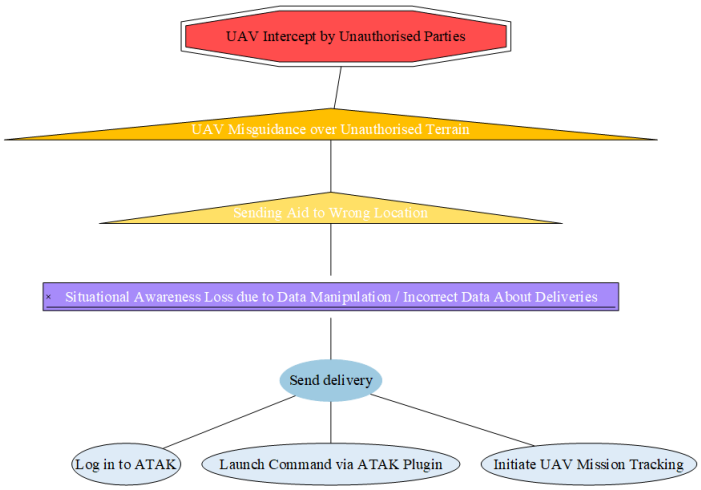

Figure 5: Filtered Event Model – A task with sub-tasks to Failures to Hazard to Event

The Filtered Risk Model shown in Figure 6 demonstrates integration by presenting a traceable pathway connecting task-related failures, hazards and a related event, and the associated security risk. *Note, the full risk model showing all elements (assets, tasks, attackers, threats, vulnerabilities, risks, failures, hazards, events, requirements) is too complex to illustrate in a readable small image.*

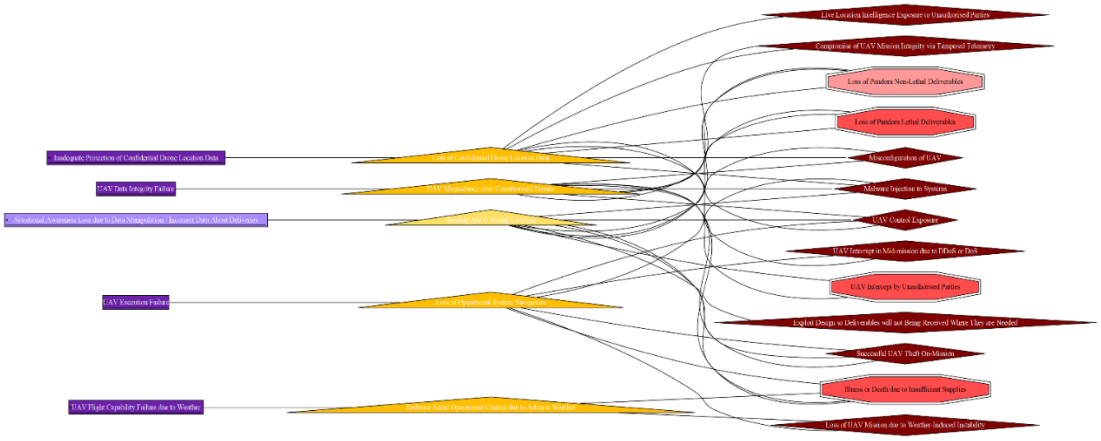

Figure 6: Filtered Risk Model - Failures to Hazard to Event and Risk

Together, these figures show some aspects about how the MVP extends standard CAIRIS usage by embedding HF and safety constructs directly within the cybersecurity modelling framework. Rather than treating usability, safety and security as separate analytical layers, the implementation demonstrates a unified socio-technical structure supporting Secure-by-Design requirement derivation.

Requirement Derivation Framework

The central contribution of the MVP is its integrated socio-technical risk-to-requirement pathway. Figure 7 illustrates how task performance, UCAs, vulnerabilities, and hazards propagate to derived Secure-by-Design requirements.

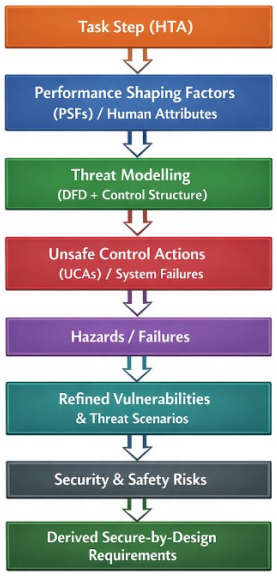

Figure 7: Integrated Socio-Technical Risk-to-Requirement Pathway

This structured pathway makes explicit how degraded human performance, combined with cyber exposure, can propagate to safety consequences across the full spectrum of capability development and delivery and vice versa. To ensure that all relevant socio-technical dimensions are considered, the pathway explicitly incorporates Ministry of Defence Lines of Development (DLODs) - such as Personnel, Training, Logistics, Communications/Information, Organisation, Doctrine & Concepts, Equipment, and Infrastructure - since these collectively define the ecosystem that enables a capability to operate effectively and safely through its lifecycle. In the UK defence context, DLODs provide a coherent, through-life taxonomy for capability delivery and support which must be addressed in planning, design and assurance activities (Ministry of Defence, 2025).

Requirements are categorised into system and control constraints, interface/HMI requirements, procedural and organisational measures, cybersecurity controls and trade-off requirements addressing cross-domain tensions (Table 2).

Table 2: Secure-by-Design Requirement Classes

Requirement Type	**Examples / Description**
System / Control	Interlocks, automated validation of critical actions, confirmation logic (e.g., equipment robustness)
Interface / HMI	Differentiation of security-critical alerts from operational or routine alarms (supporting personnel and information flow)
Procedural / Organisational	Staffing assumptions, escalation procedures, training provisions (personnel, training, organisation)
Security	Hardening of controls linked to safety dependencies, alignment with threat models (information, doctrine & concepts)
Trade-off	Usability constraints to balance security measures and operator performance (affecting personnel, equipment, training)

Bi-directional traceability ensures that each requirement can be justified from its originating artefacts (see Table 1) and supports review, verification, and iterative refinement across these socio-technical and support dimensions.

Generation and Assessment of the User Requirements Document

The MVP was applied to the specification exemplar to generate a structured URD. Analysts modelled operational tasks and control structures, identified human performance sensitivities, mapped cybersecurity exposures and analysed cross-domain tensions.

The URD served as the primary assessment artefact for the MVP. Completeness was examined by assessing whether derived requirements addressed capability, safety, security and HF dimensions. Traceability was evaluated by verifying bidirectional linkage between requirements and originating analytical artefacts. Coherence was assessed qualitatively through cross-domain review, ensuring that requirements emerging at disciplinary intersections were captured explicitly. The resulting URD integrates capability requirements with safety constraints derived from STPA, HF requirements linked to task reliability, cybersecurity controls aligned with safety dependencies and trade-off requirements resolving design tensions.

Results and Evaluation

Application of the MVP demonstrated that significant requirements emerged at intersections between HF, safety and cybersecurity rather than within isolated analyses. The integrated workflow enabled early identification of security controls that increased workload during time-critical tasks and highlighted potential reductions in situational awareness under degraded conditions. Some of these would not have been evident without the additional HF and safety analysis.

The process embedded safety, security and human performance considerations at capability definition stage rather than during detailed design, reduced the likelihood of late-stage redesign and improved transparency in cross-domain decision-making.

Discussion

The study demonstrates that integrating HF, safety and cybersecurity within a unified socio-technical model enables structured identification of cyber-physical risks - hazards arising from interactions between digital systems, physical processes and human operators that may impact safety, security, or mission outcomes.

By explicitly linking human performance variability to unsafe control actions and cyber vulnerabilities, the MVP exposes risk pathways that would otherwise remain implicit. This supports earlier mitigation through interface refinement, control logic adjustments, training provisions and architectural safeguards.

The structured visualisation of cross-domain interactions facilitates shared understanding among analysts and engineers and strengthens traceability for assurance activities.

Conclusion

This study demonstrates a practical method for operationalising Secure-by-Design in complex socio-technical systems through integrated HF, safety, and cybersecurity analysis. Using a specification exemplar and an MVP-supported workflow, the approach systematically generates structured URDs

that embed safety, security and HF considerations from inception. By making cross-domain risk pathways explicit, it produces traceable and justifiable requirements, supports early-stage design assurance, and reduces the likelihood of costly redesign.

The MVP elevates HF from an advisory input to a structured driver of requirement generation, enabling designers to anticipate conflicts between safety, security and human performance. Analysts can visualise cyber-physical risks - hazards arising from interactions among digital systems, physical processes and human operators- which facilitates proactive mitigation via interface enhancements, control logic adjustments and organisational provisions. Addressing these risks early reduces lifecycle rework, supports cost containment and strengthens the resilience of critical cyber-physical systems.

As automation and AI increasingly shape critical infrastructure and military systems, this approach provides a practical pathway for embedding Secure-by-Design practices at early stages. Future work will enhance visualisation, automate data ingestion and requirements generation, and evaluate the MVP in operational environments to achieve higher Technology Readiness Levels (TRLs), further reinforcing structured, traceable, and actionable Secure-by-Design requirements in highly cybersecurity-sensitive domains.

References:

Altaf, A., Faily, S., Dogan, H., Thron, E. and Mylonas, A. (2021) 'Integrated design framework for facilitating systems-theoretic process analysis', in *European Symposium on Research in Computer Security*, Cham: Springer International Publishing, pp. 58–73.

Dev, J., Rashidi, B. and Garg, V., 2023, April. Models of applied privacy (map): A persona based approach to threat modeling. In Proceedings of the 2023 CHI Conference on human factors in computing systems (pp. 1-15).

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

Fortinet, *Threat modeling*. Available at: https://www.fortinet.com/uk/resources/cyberglossary/threat-modeling (Accessed: 10 February 2026).

Ki-Aries, D., Faily, S., Dogan, H. and Williams, C. (2018) 'Assessing system of systems security risk and requirements with OASoSIS', *Proceedings - 2018 5th International Workshop on Evolving Security and Privacy Requirements Engineering, ESPRE 2018*, pp. 14–20.

Ki-Aries, D., Faily, S., Dogan, H. and Williams, C. (2022) 'Assessing system of systems information security risk with OASoSIS', *Computers and Security*, 117.

Kirwan, B. and Ainsworth, L.K. (1992) *A guide to task analysis*. London: Taylor & Francis.

Knisely, B.M., Joyner, J.S. and Vaughn-Cooke, M., 2021. Cognitive task analysis and workload classification. MethodsX, 8, p.101235.

Kriaa, S., Pietre-Cambacedes, L., Bouissou, M. and Halgand, Y., 2015. A survey of approaches combining safety and security for industrial control systems. *Reliability engineering & system safety*, *139*, pp.156-178.

Leveson, N. (2011) *Engineering a safer world: Systems thinking applied to safety*. Cambridge, MA: MIT Press.

Ministry of Defence (2025) *Secure by Design: Design for security from the start*. Available at: https://www.digital.mod.uk/policy-rules-standards-and-guidance/secure-by-design (Accessed: 10 February 2026).

Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C. and Jenkins, D.P. (2005) *Human factors methods: A practical guide for engineering and design*. Aldershot: Ashgate.

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.

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.

UK Government Cyber Security Advisory (2025) *Managing tensions between security, safety, and human factors: Requirements analyses*.
Available at: https://www.gov.uk/government/publications/managing-tensions-between-security-safety-and-human-factors/cyber-security-advisory-managing-tensions-between-security-safety-and-human-factors-requirements-analyses (Accessed: 09 February 2026).

UpGuard (2025) *Threat modelling in practice*.
Available at: https://www.upguard.com/resources/threat-modeling (Accessed: 12 February 2026).