Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27875-cd

Adapting Human Reliability Analysis to Autonomous Control for Small Modular Reactors

Roberto Mascherona

Mott MacDonald, France. E-mail: roberto.mascherona@mottmac.com

Martina Artioli

Mott MacDonald, France. E-mail: martina.artioli@mottmac.com

Alexis Jourdain

Mott MacDonald, France. E-mail: alexis.jourdain@mottmac.com

The increasing interest in autonomous and AI-supported control systems for Small Modular Reactors (SMRs) introduces profound changes to the future role, responsibilities, and actions of human operators. While higher levels of automation promise enhanced operational efficiency and reduced staffing, they also reshape traditional human–system interaction patterns, creating new pathways for human error and challenging the applicability of existing Human Reliability Analysis (HRA) approaches. Current HRA methods—developed for operator-centric control architectures—struggle to capture the dynamic, supervisory, and intervention-based roles that humans assume in highly automated, AI-assisted environments. This paper examines the emerging human reliability challenges associated with AI-enabled autonomous control in SMRs, focusing on key aspects as shifts in situational awareness, trust calibration, and transitions of control between automated and human agents. We review the suitability of established HRA methodologies, identify gaps in modelling cognitive load and human–automation interaction, and propose updates and methodological adaptations needed to evaluate human performance within this new operational paradigm. The aim is to provide a structured foundation for future HRA approaches capable of supporting licensing, safety analysis, and design of AI-driven control systems in next generation nuclear reactors.

Keywords: Human reliability, Risk analysis, Small Modular Reactors, Autonomous Control Systems, Artificial Intelligence.

1. Introduction

Small Modular Reactors (SMRs) are being developed with operational concepts that differ from conventional nuclear power plants, including reduced staffing, multi-unit supervision, and increased reliance on automation and Artificial Intelligence (AI)–supported control. Advanced SMR concepts envisage higher levels of autonomy, shifting operators from continuous manual control toward supervisory monitoring and intervention-based decision-making. This shift challenges assumptions embedded in established Human Reliability Analysis (HRA) methods, which were largely developed for operator-centric control rooms characterised by frequent manual actions and relatively stable procedural contexts. Autonomous control environments, by contrast, are dominated by prolonged nominal operation with limited human interaction, punctuated by infrequent but cognitively demanding interventions during degraded conditions—an "irony of automation" pattern long recognised as a source of human reliability challenges (Bainbridge, 1983). These issues should be understood on a continuum with modern digitalised and highly automated control rooms, including plants already in operation, where computer-based procedures and digital I&C modernisation have introduced supervisory monitoring demands, mode-awareness issues, and context-dependent changes in operator actions (Xing and Chang, 2023; Fernandes et al., 2024). Autonomous SMR concepts amplify and reconfigure these phenomena through reduced staffing, extended autonomous operation, multi-unit oversight, and the potential use of adaptive or less transparent AI-supported functions (Bainbridge, 1983). The objective of this paper is to examine how AI-enabled autonomous control in SMRs affects the applicability of existing HRA approaches and to identify how HRA can be adapted in a disciplined and regulator-acceptable manner. Rather than proposing a new methodology, the paper adopts an evolutionary strategy: it positions cognitive, context-dependent HRA as a defensible baseline and identifies targeted extensions required to address autonomy-specific challenges.

2. Evolution of Human Roles in Autonomous SMR Control

Advanced SMR concepts increasingly rely on automation to maintain plant stability during normal operation, shifting the operator role from continuous control toward supervisory oversight and intervention when automation performance degrades. This transition does not reduce human responsibility for safety; instead, it concentrates safety-significant human contribution into monitoring effectiveness, interpretation, decision timing, and management of control authority across automation modes.

Extended periods of nominal automated operation raise challenges related to vigilance, engagement, and maintenance of accurate mental models of plant state. When abnormal conditions occur, operators may be required to reconstruct system context rapidly and decide whether intervention is necessary, increasing susceptibility to delayed detection or delayed intervention. Official work on HRA in digital I&C modernisation explicitly notes that digital environments introduce new operator actions and change the context of actions, motivating careful assessment of method applicability under changed work processes (Xing and Chang, 2023).

AI-supported automation further complicates supervisory control by introducing adaptive or opaque system behaviour, affecting trust calibration—particularly where explainability and transparency of AI-based functions are limited (Hall et al., 2024). Over-trust can delay necessary intervention, while under-trust can lead to unnecessary overrides and increased workload. Experimental work on automation levels in nuclear simulated procedure execution reports that automation can influence workload and trust in complex and context-sensitive ways (Schreck et al., 2025). A particular safety-significant aspect of autonomous operation is the transition of control between automation and humans: takeover and handover episodes combine elevated cognitive demand, potential ambiguity of authority, and compressed decision windows, and have been identified as dominant risk contributors in SMR automation concepts (Mascherona et al., 2025). These supervisory demands are not unique to SMRs and have been observed in modern digitalised and highly automated control rooms currently in operation (Xing and Chang, 2023; Schreck et al., 2025). Autonomous SMR concepts intensify these tendencies through longer low-interaction periods, reduced staffing with potential multi-unit supervision, and increased reliance on automation-initiated mode changes or fallback behaviours. Consequently, human reliability becomes dominated by supervisory cognition and transition management, motivating HRA approaches that explicitly represent omission/delay vulnerabilities and mode-dependent context variability.

3. Implications for Human Reliability Analysis

Traditional HRA methods tend to focus on discrete human actions, procedural compliance, and relatively stable task contexts. However, the role described above imposes specific conditions on HRA for autonomous SMRs:

- dominant failure modes may involve omission, delayed intervention, or mis-interpretation, rather than incorrect action execution.
- task context is dynamic, changing with automation mode and system state.
- human–automation interaction, including trust and transparency,

becomes a central determinant of performance.
- uncertainty is higher due to first-of-a-kind designs and limited operating experience.

These issues have been explicitly identified as gaps in current HRA practice for advanced reactors (Blackett, 2024; Blackett et al., 2022; Bye, 2023). Evidence from digitalised environments reinforces that automation and digital systems can alter the work process and hence the conditions under which human error occurs, requiring explicit attention to applicability and evidence support (Xing and Chang, 2023; Park et al., 2024). Action-centric and static Performance Shaping Factor (PSF) based HRA approaches remain valuable for bounded manual tasks but can be insufficient as primary frameworks for autonomous control environments where human reliability is dominated by cognitive and contextual factors. Work on adaptive automation in nuclear control rooms similarly highlights that increased automation can change operator roles and performance and must be analysed with attention to safety challenges arising from computerized and adaptive procedures (Fernandes et al., 2024). Within this landscape, a cognitive, context-dependent baseline is advantageous for structuring supervisory control and decision-making under uncertainty. CREAM is adopted in this paper as a defensible baseline "spine" because of its explicit focus on context-dependent cognitive control and performance variability, consistent with the need to evaluate human reliability under changing operational conditions. However, this baseline must be extended to represent autonomy-specific challenges.

3.1. Continuities with modern highly automated reactor operation

The supervisory-control challenges highlighted for autonomous SMRs—such as loss of engagement during extended monitoring, trust calibration, mode awareness, and automation surprise—have precedents in highly automated and digitalised nuclear control rooms. Modernisation and adaptive automation work emphasise that digital systems introduce new operator actions, change action context, and require careful evaluation of method applicability and evidence support (Xing and Chang, 2023; Fernandes et al., 2024). Experimental findings also indicate that increased automation can shift workload and trust in non-linear ways depending on task structure and operator strategy (Schreck et al., 2025). Autonomous SMR concepts nonetheless represent an important shift in operating regime by intensifying known supervisory-control issues and reshaping dominant error pathways: the human role becomes increasingly defined by omission/delay, multi-unit situation awareness, and short-notice transitions from low-engagement monitoring to high-consequence intervention. The HRA challenge is therefore not to treat these phenomena as entirely novel, but to represent how autonomy amplifies known issues and changes how they manifest in accident sequences.

3.2. Demand-reducing SMR design features and reshaped HRA demands

A balanced HRA discussion should also consider that several SMR design features may reduce certain traditional human reliability stressors. For example, passive safety features, extended coping times, and designs that reduce the need for rapid, high-workload manual actions during some emergency scenarios can decrease time pressure and compress fewer safety-critical operator actions into short windows. This may reduce the prevalence of classic action execution errors in specific sequences, while shifting emphasis toward detection, diagnosis, and supervisory decision-making.

Importantly, these features may reshape rather than eliminate human reliability challenges. Longer coping times and fewer urgent actions can reduce acute workload but may increase the relative importance of vigilance, maintenance of accurate mental models, and avoidance of delayed recognition of degradation, particularly under prolonged autonomous operation (Bainbridge, 1983). Consequently, HRA for autonomous SMRs should explicitly represent both (i) potential reductions in time-critical manual actions and (ii) potential increases in omission/delay vulnerabilities under low engagement conditions.

4. Mapping Established HRA Methods to Adaptation Needs for Autonomous SMRs

Established nuclear HRA methods were developed for operating contexts in which safety-significant human contribution is primarily associated with procedural task execution, manual control actions, and clearly bounded interventions. Across these approaches, analysis typically involves: (i) identifying human actions relevant to accident sequences, (ii) characterising contextual influences on performance, and (iii) linking results to probabilistic safety assessment models. Emerging autonomous and AI-supported SMR operating concepts challenge these foundations by changing not only human–machine function allocation but also the temporal structure, cognitive demands, and dominant failure modes of human involvement. Operators are increasingly positioned as supervisors of automated processes, with safety-significant contribution concentrated in monitoring, interpretation, prioritisation, and control transitions rather than continuous action execution. This shift exposes recurring tensions between established HRA assumptions and supervision-dominated operation:

- From action to non-action: Safety-relevant human failures increasingly involve delayed intervention, failure to intervene, or inappropriate reliance on automation.
- From stable to dynamic context: Human performance context changes abruptly with automation mode and during control handover and takeover.
- From local to global cognition: Operators must maintain situation awareness across multiple systems or units while engaging selectively and intermittently.
- From procedural certainty to interpretative judgement: AI-mediated information requires operators to assess confidence, credibility, and intent rather than follow deterministic cues.

No single class of established HRA methods fully addresses all these challenges. Instead, each contributes partial capabilities that become strained when applied to highly automated, supervision-dominated concepts. Adapting HRA for autonomous SMRs is therefore less a matter of replacing existing methods than of re-balancing analytical focus toward the human functions and failure modes that become dominant under autonomy. Action-centric HRA methods were developed for contexts in which safety-relevant human contribution is expressed through discrete, observable actions embedded in procedures and performed within relatively stable operating modes. They provide strong traceability to PSA models and remain regulator-familiar, and they continue to be applicable in autonomous SMR concepts for bounded manual activities that persist despite increased automation (e.g., maintenance/testing, manual alignments, local field actions, defined recovery steps). However, as automation absorbs routine control functions and human involvement becomes more intermittent, safety-significant contribution increasingly shifts toward supervisory monitoring, interpretation, and decisions about whether and when to intervene. In this regime, omission/delay, mode misunderstanding, and inappropriate reliance on automation can dominate over execution errors. Action-centric methods can represent these effects only indirectly and are therefore necessary but insufficient as a primary analytical lens for supervision-dominated operation. Performance Shaping Factor (PSF)-based approaches extend action-centric reasoning by accounting explicitly for contextual influences such as workload, time pressure, information quality, and training. They remain valuable—particularly for screening and for bounded manual tasks—but autonomous SMR operation challenges their typical assumptions of relatively stable context. Long periods of low-engagement monitoring may be followed by short, high-consequence decision windows during automation degradation or takeover, meaning that performance drivers can shift rapidly with mode changes. Traditional PSF formulations often lack mechanisms to represent abrupt context shifts, mode awareness, and automation-interaction factors (e.g., trust, transparency, predictability) that shape supervisory intervention behaviour. PSF-based approaches therefore remain informative, but benefit from being embedded within a higher-level structure that explicitly represents mode-dependent context variability. Cognitive, context-dependent HRA approaches were developed to address situations in which human performance cannot be represented adequately as a sequence of independent actions. They focus on how operators adapt control strategies in response to context, uncertainty, and time pressure, treating performance variability as intrinsic rather than residual. This perspective aligns naturally with supervisory roles in autonomous SMRs, where operators must interpret ambiguous information, maintain situation awareness across modes and units, and decide whether intervention is necessary under incomplete knowledge. Rather than replacing action-centric or PSF-based reasoning, cognitive approaches provide an integrative layer in which actions, omissions, and delays can be interpreted as manifestations of underlying cognitive control shaped by changing context. Overall, mapping autonomous SMR characteristics to established HRA capabilities supports a reframing in which action-centric and PSF-based analyses are retained for residual manual tasks but are embedded within a broader context- and control-mode-oriented structure

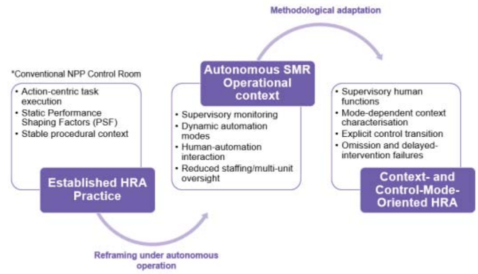

Figure 1. Conceptual reframing of established Human Reliability Analysis for AI-enabled autonomous Small Modular Reactor operation.

Figure 1 illustrates this conceptual reframing, which forms the basis for the targeted CREAM-consistent extensions proposed in Section 5.

5. Proposed CREAM Extensions for AI-Enabled Autonomous SMR Control

Sections 3 and 4 established that CREAM provides a defensible cognitive baseline for analysing human reliability in autonomous SMR control environments, while also identifying clear gaps that must be addressed to represent supervisory monitoring, human–automation interaction, and control transitions adequately. This section specifies a set of targeted, CREAM-consistent extensions intended to address these gaps without altering the fundamental structure of the method. The objective is not to redefine CREAM or introduce a new HRA methodology, but to adapt its existing constructs—in particular Common Performance Conditions (CPCs), control modes, and task characterisation—so that they can represent the operational realities of AI-enabled autonomous SMRs.

5.1. *Explicit treatment of supervisory monitoring and non-action failures*

5.1.1. *Monitoring as a safety-critical task*

In autonomous SMR concepts, extended periods of nominal operation are dominated by supervisory monitoring rather than active control. Under such conditions, safety-significant human failures are often expressed as omissions—for example, failure to detect a degrading trend, failure to recognise a fallback trigger, or failure to intervene within an available time window. These failure modes differ qualitatively from incorrect procedural execution and must therefore be treated explicitly in HRA.

Within a CREAM framework, this requires explicit identification of monitoring tasks as primary cognitive activities, rather than implicit background functions. Monitoring tasks should be defined with clear performance objectives (e.g., detection thresholds, timeliness requirements) and associated with CPC profiles that reflect vigilance demand, information presentation quality, and automation reliability.

5.1.2. *Omission and delay as primary phenotypes*

CREAM's existing error taxonomy can represent omission-type failures, but traditional applications often focus on action errors. For autonomous SMRs, non-action and delayed action should be elevated to first-class phenotypes, particularly for takeover initiation and abnormal condition recognition. This extension is consistent with findings from automation and digital environment studies showing that automation changes not only what operators do, but also how and when failures occur.

5.2. *Mode-dependent and dynamic Common Performance Conditions*

5.2.1. *Automation mode as a contextual driver*

Autonomous SMR operation involves distinct operational modes—such as normal autonomous supervision, degraded automation, and manual takeover—each associated with different cognitive demands, information availability, and time pressure. Treating CPCs as static across these modes' risks obscuring the dominant drivers of human performance variability.

A key proposed extension is therefore to make CPC assessment explicitly mode-dependent. For example, the same CPC (e.g., adequacy of information, available time, or system predictability) may reasonably take different qualitative values during normal supervision versus during a takeover phase. This approach is consistent with nuclear research highlighting that automation effects on human performance are context-sensitive and cannot be captured by a single, static characterisation.

A mode-dependent CPC assessment also provides a structured way to reflect SMR design features that may reduce traditional HRA stressors. For example, in scenarios where passive safety and longer coping times extend decision windows, the CPC related to available time/time adequacy may reasonably improve compared with conventional time-critical response contexts. At the same time, prolonged low-engagement supervision can increase vulnerability to delayed recognition and delayed intervention. Treating CPCs as mode-specific therefore allows the analysis to represent both risk-reducing and risk-reshaping effects coherently.

5.2.2. *Dynamic re-evaluation during transitions*

In addition to mode-dependent CPC profiles, explicit re-evaluation of CPCs at control transitions is proposed. Transitions of control are characterised by abrupt changes in workload, situational awareness demands, and time pressure, and they frequently dominate risk in autonomous concepts. Treating takeover as a distinct analysis phase with its own CPC configuration improves traceability and avoids conflating steady-state supervision with transient, high-stress intervention conditions.

5.3. *Representation of human–automation interaction factors*

5.3.1. *Trust calibration and automation transparency*

AI-supported automation introduces challenges related to trust calibration, automation surprise, and interpretability of system behaviour. Experimental studies in nuclear automation demonstrate that higher levels of automation can both reduce and increase workload depending on task and context, and that trust miscalibration can significantly affect intervention behaviour.

Within a CREAM-based analysis, these effects can be addressed by extending the CPC set to include explicit descriptors related to:

- automation transparency and predictability,
- clarity of automation state and authority,
- feedback quality during autonomous operation.

Such extensions do not alter CREAM's conceptual basis but enrich the context description so that cognitive control modes reflect human–automation interaction realities rather than purely human-centric task environments.

5.3.2. Interface-mediated cognition

Advanced human–system interfaces, decision support tools, and automated procedures shape how operators perceive and reason about plant state. Evidence from digital control room research indicates that interface design and information integration can significantly influence error pathways and situational awareness. Accordingly, CPC assessment should explicitly consider interface-mediated cognition rather than treating information quality as an abstract or static factor.

5.4. Control transitions as a distinct analytical construct

A central insight from prior SMR automation work is that control transitions represent a primary automation hazard, rather than a secondary operational detail. In HRA terms, this implies that takeover and handover should be modelled as distinct cognitive episodes, characterised by:

- rapid situation reconstruction,
- ambiguity in authority and responsibility,
- elevated time pressure,
- potential mismatch between automation logic and human expectations.

Within a CREAM-anchored approach, this can be achieved by defining transition-specific task contexts, with dedicated CPC profiles and explicit omission/delay phenotypes. This treatment improves alignment between HRA outputs and broader safety arguments around takeover risk acceptability, without requiring a separate hazard analysis methodology.

5.5. Use of simulator-based and experimental evidence

Given the first-of-a-kind nature of autonomous SMR control concepts, empirical operating data will be limited. Official NRC and INL work highlight the role of simulator-based studies in providing evidence on how digitalisation and automation affect human reliability. In the proposed framework, such evidence is not used to validate a new method, but to inform CPC characterisation, support sensitivity analysis, and justify assumptions regarding monitoring performance and transition timing.

This approach aligns with regulatory expectations that early safety cases rely on structured argumentation supported by experimental and simulator evidence, rather than on spurious numerical precision, consistent with international guidance on human reliability analysis for nuclear installations (IAEA, 2025).

6. Conclusions

The introduction of AI-enabled autonomous control in Small Modular Reactors fundamentally reshapes, rather than eliminating, the human contribution to nuclear safety. As automation increasingly assumes routine control functions, human involvement becomes concentrated in supervisory monitoring, interpretation of system behaviour, and management of control authority during degraded conditions and transitions. These roles are safety-significant, yet they challenge key assumptions embedded in traditional Human Reliability Analysis (HRA) methods that were developed for operator-centric, action-dominated control environments.

At the same time, many of the supervisory-control phenomena discussed here have precedents in modern digitalised and highly automated control rooms, including plants currently in operation, where automation has shifted operator work toward monitoring, mode awareness, and managing automation behaviour. Autonomous SMR concepts extend these challenges in degree and coupling through reduced staffing, longer low-interaction supervision, and potentially more adaptive or less transparent AI-supported functions. In parallel, SMR design features such as passive safety and longer coping times may reduce some time-critical manual execution demands in specific scenarios, while shifting HRA emphasis toward timely detection, interpretation, and escalation decisions and increasing the relative importance of omission/delay pathways.

This paper has shown that the emerging human reliability challenges associated with autonomous

SMR operation cannot be adequately represented by action-centric or static context-based approaches alone. At the same time, these challenges do not necessitate the introduction of an entirely new HRA paradigm. Instead, they motivate a disciplined evolution of established methods, grounded in cognitive, context-dependent reasoning and aligned with existing nuclear safety practice.

Building on this perspective, the paper proposed targeted extensions to a cognitive HRA baseline that refine how context is characterised, how supervisory functions are defined, and how autonomy-specific failure modes are identified, in this way the approach enables a more faithful representation of human reliability in supervision-dominated and first-of-a-kind SMR operational concepts.

From a safety assessment and licensing perspective, the findings emphasise the importance of explicitly addressing control handover and takeover as primary safety concerns, rather than as secondary operational details. The proposed adaptations support transparent argumentation, conservative treatment of uncertainty, and traceability between automation design claims and human reliability assumptions, which are essential in the absence of extensive operational experience.

Overall, this work demonstrates that established HRA methods can remain relevant and defensible for autonomous SMRs when applied with a clear understanding of changing human roles and adapted in a structured, regulator-credible manner. As SMR designs and autonomous control concepts continue to mature, such evolutionary approaches provide a practical foundation for integrating human reliability considerations into safety analysis, design decisions, and regulatory dialogue.

References

Bainbridge, L. "Ironies of Automation." *Automatica* 19, no. 6 (1983): 775–779.

Blackett, C. "Adapting Human Reliability Analysis for Small Modular Reactors." Conference paper, June 2024.

Blackett, C., M. H. R. Eitrheim, and A. Bye. "The Challenge of Assessing Human Performance and Human Reliability for First-of-a-Kind Technologies." In *Proceedings of the 16th Probabilistic Safety Assessment and Management Conference (PSAM 16)*, Honolulu, HI, 2022.

Bye, A. "Future Needs of Human Reliability Analysis: The Interaction between New Technology, Crew Roles and Performance." Safety Science 158 (February 2023): 105962. https://doi.org/10.1016/j.ssci.2022.105962.

Fernandes, A., R. Bisio, and A. Bye. "Adaptive Automation in Computerized Procedures: Discussing Safety Challenges in Nuclear Control Rooms." In *Proceedings of PSAM17 / ASRAM 2024*, 2024.

Hall, A., P. Murray, R. L. Boring, and V. Agarwal. "Human-Centered and Explainable Artificial Intelligence in Nuclear Operations." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2024.

International Atomic Energy Agency. *Human Reliability Analysis for Nuclear Installations*. Vienna: International Atomic Energy Agency, 2025.

Mascherona, R., M. Artioli, R. Puisa, and A. Nelson. "Automation Criterion for Small Modular Reactors." In *Proceedings of ESREL & SRA-E 2025*, 2025.

Park, J., R. R. Boring, and T. A. Ulrich. *Simulator Data Analysis to Inform Digitalized Environment Impacts on Human Reliability*. Idaho National Laboratory report prepared for the U.S. Nuclear Regulatory Commission, 2024.

Schreck, J., G. Matthews, J. Lin, S. Mondesire, D. Metcalf, K. Dickerson, and J. Grasso. "Levels of Automation for a Computer-Based Procedure for Simulated Nuclear Power Plant Operation: Impacts on Workload and Trust." *Safety* 11, no. 1 (2025).

Xing, J., and Y.-J. Chang. "Application of Human Reliability Analysis to Digital I&C Control Room Modernization." In *Proceedings of NPIC & HMIT 2023*. U.S. Nuclear Regulatory Commission, 2023.