

Original Scientific Paper/Original naučni rad
Paper Submitted/Rad primljen: 31.12.2025.
Paper Accepted/Rad prihvaćen: 20.01.2026.
DOI: 10.5937/SJEM2600049B

UDC/UDK: 004.8:004.3]:005.962.13

Pouzdanost i bezbednost AI-baziranih hardverskih sistema: Značaj ljudske ekspertize i upravljanja tehničkim talentima

Ivana Bojić, MScEE1

¹ Belgrade School of Engineering Management, University “Union-Nikola Tesla”, Belgrade, Serbia,
ivanabojić89@gmail.com

Summary in Serbian: Primena veštačke inteligencije u savremenim hardverskim sistemima donosi nove izazove u pogledu pouzdanosti i bezbednosti. Za razliku od tradicionalnih determinističkih rešenja, AI-bazirani hardverski sistemi pokazuju adaptivno i delimično nedeterminističko ponašanje, što otežava procese verifikacije i procene rizika. Iako AI-asistirani alati značajno unapređuju efikasnost verifikacije, njihova primena ostaje ograničena u složenim i graničnim uslovima rada.

U ovom radu se analizira uloga ljudske ekspertize i upravljanja tehničkim talentima u obezbeđivanju pouzdanosti i bezbednosti AI-baziranih hardverskih sistema. Istraživanje se zasniva na analizi relevantne literature, važećih standarda i industrijskog primera iz prakse digitalne verifikacije hardvera. Poseban akcenat stavljen je na ograničenja automatizovanih alata i značaj stručne procene u interpretaciji rezultata i donošenju tehničkih odluka. Rad pokazuje da upravljanje tehničkim talentima kroz kontinuirano usavršavanje, mentorstvo i zadržavanje iskusnih inženjera predstavlja ključni faktor stabilnosti i pouzdanosti sistema. Zaključuje se da kvalitet i pouzdanost AI-baziranih hardverskih sistema ne zavise isključivo od algoritama i alata, već i od načina na koji su tehnički timovi organizovani, vođeni i podržani.

Ključne reči: veštačka inteligencija, hardverski sistemi, pouzdanost, bezbednost, upravljanje tehničkim talentima

Reliability and Security of AI Models in Hardware Systems: The Role of Expert Knowledge and Technical Talent Management

Abstract: The integration of artificial intelligence into modern hardware systems introduces new challenges related to reliability and security. Unlike traditional deterministic designs, AI-based hardware systems exhibit adaptive and partially non-deterministic behaviour, which complicates verification and risk assessment processes. While AI-assisted tools significantly improve efficiency in hardware verification, their effectiveness remains limited in complex or boundary operating conditions.

This paper examines the role of human expertise and technical talent management in ensuring the reliability and security of AI-based hardware systems. The study combines insights from existing literature with an industry-based case example from digital hardware verification practice. Particular attention is given to the limitations of AI-assisted verification tools and the importance of expert judgment in interpreting results and preventing incorrect design decisions.

The paper further argues that technical talent management, including continuous learning, mentorship and knowledge retention should be treated as a reliability mechanism rather than a purely organizational or human resource's function. The findings suggest that the quality and trustworthiness of AI-based hardware systems depend not only on algorithms and tools but also on how technical teams are developed, guided, and supported within organizations.

Keywords: artificial intelligence, hardware systems, reliability, security, human expertise, technical talent management

1. Introduction

Artificial intelligence has become an integral component of modern hardware systems, influencing areas such as power management, performance optimization, predictive control and adaptive system behaviour. As AI models

are increasingly embedded directly into hardware architectures, the boundary between traditional hardware design and intelligent, adaptive systems continues to blur.

The growing autonomy of AI-based hardware systems brings substantial benefits but it also introduces new reliability and security challenges. Unlike conventional deterministic designs, AI-enabled systems may exhibit behaviour that depends on training data, statistical inference and changing operating conditions. As a result, traditional digital verification methodologies are often insufficient to capture all relevant system behaviours, particularly in boundary or stress scenarios.

In this context, the role of human expertise becomes increasingly important. Digital verification engineers and system architects are required to interpret complex verification results, assess system-level risks and make informed decisions in situations where automated tools provide ambiguous or incomplete information. Practical experience and domain knowledge remain essential for distinguishing genuine reliability issues from benign system behaviour.

Technical expertise and organizational factors play a critical role in system reliability. Effective technical talent management, including continuous skill development, mentorship, and retention of experienced engineers, directly affects an organization's ability to manage the risks associated with AI-based hardware systems. From this perspective, talent management should be viewed as an integral part of reliability and security strategy rather than a separate administrative concern.

The contribution of this paper is twofold. First, it highlights the limitations of AI-assisted verification in hardware systems and emphasizes the continuing importance of expert human judgment. Second, it demonstrates how technical talent management and knowledge retention contribute to the overall reliability and security of AI-based hardware systems, supported by an industry-oriented case example.

2. Related Work

Recent research highlights the increasing use of artificial intelligence in hardware design and verification, particularly in areas such as test generation, coverage analysis and prediction of system behaviour. AI-assisted techniques are commonly presented as a means of improving verification efficiency by handling large design spaces and reducing manual effort. Several studies report measurable improvements in verification throughput when machine learning models are integrated into verification workflows (Sculley et al., 2015).

At the same time, the literature recognises important limitations of these approaches. AI-based verification tools depend heavily on the quality and representativeness of training data, which may not fully capture rare or boundary operating conditions. In complex hardware systems, this limitation becomes particularly relevant, as system behaviour may emerge from interactions between multiple components rather than from isolated functional blocks.

Research on the reliability and security of AI-enabled systems emphasises that non-deterministic behaviour introduces new challenges for verification and risk assessment. Recent work stresses that ensuring trustworthy AI in safety-critical systems requires not only robust algorithms, but also effective human oversight and system-level evaluation mechanisms (Zhang et al., 2022). Existing standards and frameworks address robustness and risk management, yet they often focus on algorithmic properties and provide limited guidance for practical verification decision-making in hardware contexts (ISO/IEC 24029-1:2021; ISO/IEC 23894:2023).

Although human-in-the-loop approaches are increasingly discussed, many contributions remain conceptual and do not sufficiently explore the concrete role of expert judgement in hardware verification practice. Moreover, the organisational dimension, including the management of technical talent and knowledge, is frequently treated as a separate issue rather than as an integral part of system reliability. This paper addresses this gap by explicitly linking AI-assisted digital hardware verification with human expertise and technical talent management.

3. Human Expertise in AI-Based Digital Hardware Verification

Despite significant advances in AI-assisted verification tools, human expertise remains a critical factor in ensuring the reliability and security of AI-based hardware systems. Automated tools excel at processing large volumes of data and identifying statistical deviations, but they lack the contextual understanding required to interpret complex system behaviour in real-world operating environments.

In practical verification workflows, engineers are often confronted with situations where AI-generated results are ambiguous or misleading. Certain behaviours may be flagged as critical based on statistical criteria, even though they correspond to expected transient effects during mode transitions or parameter adaptation. Distinguishing between genuine reliability risks and acceptable system behaviour requires deep domain knowledge and familiarity with system architecture.

Human-centred and interactive approaches to AI further underline the importance of expert involvement in interpreting system behaviour, particularly in complex and safety-relevant environments (Holzinger, 2022). Expert verification engineers rely on implicit knowledge developed through long-term engagement with similar systems, enabling them to recognise recurring behavioural patterns and to anticipate potential failure modes that are not explicitly documented.

Decision-making under uncertainty represents another area where human expertise is indispensable. AI-assisted tools may provide probabilistic assessments, but responsibility for final design and verification decisions rests with engineers. Senior professionals play a particularly important role by guiding less experienced team members and by providing system-level judgement that reduces the risk of incorrect escalation or unnecessary design changes.

These observations indicate that AI-assisted verification should be viewed as a complementary mechanism rather than a replacement for human expertise. Reliable and secure AI-based hardware systems emerge from the interaction between advanced digital verification tools and informed human judgement, where each compensates for the limitations of the other.

Expert involvement also contributes to the calibration and continuous refinement of AI-assisted verification tools. Through iterative feedback, engineers help adjust detection thresholds, redefine classifications and improve the alignment between tool outputs and real system behaviour. This interaction reduces the risk of tool overfitting to historical data and supports more reliable verification outcomes over time. Such feedback-driven refinement further highlights the interdependence between automated analysis and expert judgement in complex hardware verification environments.

Table 1 summarises the complementary roles of AI-assisted verification tools and human expertise in hardware verification workflows.

Table 1. Human expertise versus AI-assisted verification in hardware systems

Aspect	AI-assisted verification	Human expertise
Data processing	High-volume, automated	Selective, context-aware
Interpretation of results	Statistical, threshold-based	System-level, experience-based
Handling boundary conditions	Limited by training data	Informed by prior cases
Decision-making	Probabilistic suggestions	Responsibility-driven judgement
Risk assessment	Tool-defined	Contextual and holistic

4. Technical Talent Management as a Reliability Factor

The reliability and security of AI-based hardware systems depend not only on technical architectures and verification methodologies, but also on how organisations manage technical talent and knowledge. In complex engineering environments, expertise represents a critical resource that directly influences system quality and long-term stability.

Technical talent management plays a central role in maintaining verification reliability, particularly in domains where AI-assisted tools are used. Continuous learning enables engineers to adapt to evolving technologies, verification frameworks and emerging risks associated with AI integration. Without systematic skill development, teams risk relying on outdated assumptions or misinterpreting AI-generated verification results.

Knowledge retention and transfer are equally important. Digital hardware verification often involves implicit knowledge that is accumulated through hands-on experience with specific architectures, tools and failure scenarios. When such knowledge is lost due to staff turnover or insufficient documentation, organisations become

vulnerable to repeated errors and increased verification risk. In this sense, the loss of a key expert can represent a single point of failure, comparable to a critical design flaw.

Beyond individual expertise, team composition and continuity significantly influence verification reliability. Stable teams with balanced levels of seniority enable more consistent application of verification practices and reduce variability in decision-making. In contrast, frequent team restructuring or reliance on short-term staffing can weaken collective system understanding, increasing the likelihood of misinterpretation of AI-assisted verification results.

Mentorship structures help mitigate this risk by enabling structured knowledge transfer between senior and less experienced engineers. Through code reviews, joint debugging sessions, and design discussions, senior engineers provide context that cannot be conveyed through formal specifications alone. This process strengthens collective competence and reduces dependency on individual contributors.

From a management perspective, technical talent management should therefore be viewed as a reliability mechanism. Investment in people, knowledge sharing and professional development directly contributes to the organisation's ability to produce reliable and secure AI-based hardware systems. Treating talent management as a strategic component of system reliability aligns organisational practices with the technical demands of AI-enabled hardware development.

5. Industry Case Example

To illustrate the interaction between AI-assisted verification and human expertise, this section presents an anonymised industry-based example from hardware verification practice. The system under consideration includes a hardware module with AI-assisted optimisation of operating parameters under variable load conditions.

During regression testing, an AI-assisted verification tool identified behaviour that deviated from expected statistical patterns and classified it as a potential reliability issue. Based on automated analysis alone, the issue appeared to require immediate escalation and possible design modification.

A detailed review conducted by an experienced digital verification engineer revealed that the observed behaviour occurred exclusively during a transient adjustment phase following a mode transition. By analysing waveform data and system state interactions, the engineer determined that the behaviour did not affect steady-state operation or violate functional requirements.

Based on this expert assessment, verification criteria were refined to distinguish between transient adaptation effects and genuine functional issues. Additional targeted test scenarios were introduced to improve coverage of similar conditions in future regressions. As a result, unnecessary design changes were avoided and confidence in the verification process was strengthened.

This example demonstrates the limitations of relying solely on AI-assisted verification tools in complex hardware environments. While AI tools provide valuable support by highlighting deviations and patterns, human expertise is essential for contextual interpretation and risk assessment. The outcome of the digital verification process depended not on automation alone, but on the informed judgement of an experienced engineer supported by organisational knowledge and established verification practices.

6. Discussion

The analysis presented in this paper confirms that the reliability and security of AI-based hardware systems cannot be evaluated solely through automated verification techniques. While AI-assisted tools significantly improve efficiency and coverage, their outputs remain dependent on training data, statistical thresholds and predefined evaluation criteria. As a result, they may fail to capture system-level implications or misrepresent transient behaviours as critical issues.

The findings from the industry case example support observations in the literature regarding the limitations of fully automated verification in non-deterministic systems. In practice, expert human judgement provides an essential layer of interpretation, enabling engineers to contextualise verification results and make informed decisions under uncertainty. This aligns with human-in-the-loop perspectives yet extends them by demonstrating their concrete impact in hardware verification workflows.

From a management standpoint, the discussion highlights that technical talent management directly influences verification outcomes. Organisations that invest in continuous learning, mentorship, and knowledge retention are

better positioned to mitigate risks introduced by AI-based system behaviour. Treating expertise as an organisational asset rather than an individual attribute reduces dependency on isolated specialists and strengthens overall system resilience.

The results suggest that future verification strategies should emphasise hybrid approaches, where AI-assisted tools and human expertise are deliberately integrated. Such approaches acknowledge the strengths and limitations of both automation and human judgement, providing a more robust foundation for reliable and secure AI-based hardware systems.

This study is subject to several limitations that should be acknowledged. The analysis is based on a single industry-oriented case example, which, although representative of real-world digital verification practice, does not allow for broad generalisation across all AI-based hardware systems. The assessment of human expertise and talent management relies primarily on qualitative insights rather than quantitative performance metrics.

Future research could address these limitations by incorporating multiple case studies across different application domains, as well as by developing measurable indicators for evaluating the impact of technical talent management on system reliability and security. Such extensions would further strengthen the understanding of how human expertise and organisational practices contribute to trustworthy AI-based hardware systems.

7. Conclusion

The increasing integration of artificial intelligence into hardware systems introduces new challenges related to reliability and security. Unlike traditional deterministic designs, AI-based hardware systems exhibit adaptive behaviour that complicates verification and risk assessment processes.

This paper has demonstrated that human expertise remains a critical factor in ensuring system reliability, particularly in complex verification scenarios where AI-assisted tools provide incomplete or ambiguous results. Expert judgement enables contextual interpretation, informed decision-making and effective risk mitigation.

The analysis further demonstrates that technical talent management plays a strategic role in supporting reliable verification processes. Continuous learning, mentorship and knowledge retention contribute directly to organisational capability and system stability. Viewed through this lens, talent management becomes an integral component of reliability and security strategy rather than a purely organisational concern.

Future work may explore formal frameworks for integrating human expertise into AI-assisted verification workflows, as well as organisational models that further strengthen knowledge transfer and long-term reliability in AI-enabled hardware development.

Literature

1. Amershi, S., Weld, D., Vorvoreanu, M., et al. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM.
2. Holzinger, A. (2022). Interactive machine learning for human-centered AI. IEEE Intelligent Systems, 37(3), 26–34.
3. ISO/IEC 23894:2023. Artificial intelligence – Risk management framework.
4. ISO/IEC 24029-1:2021. Artificial intelligence – Assessment of robustness of neural networks – Part 1: Overview.
5. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192–210.
6. Sculley, D., Holt, G., Golovin, D., et al. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems (NeurIPS).
7. Zhang, Y., Li, Y., Wang, X., & Chen, Z. (2022). Trustworthy AI in safety-critical systems: Challenges and perspectives. IEEE Transactions on Artificial Intelligence, 3(4), 238–252.