



Artificial intelligence in information security: impact on organisational decision-making

Sérgio Silva

DOI: <https://doi.org/10.47989/ir31262219>

Abstract

Introduction. This study investigates the impact of Artificial Intelligence (AI) on information security and organisational decision-making processes. In the face of the growth of cyber threats and the complexity of digital ecosystems, AI emerges as a resource to strengthen the resilience and effectiveness of defence mechanisms.

Method. Integrative literature review (Scopus, IEEE Xplore, ScienceDirect and SpringerLink) were complemented by technical reports and case studies on the application of AI in Cybersecurity.

Analysis. It focused on three dimensions: operational efficiency, strategic decision support, and human and ethical factors. Benefits, limitations, algorithmic risks, and adoption challenges were examined.

Results. AI improves anomaly detection, reduces response time, decreases false positives, and supports risk-based strategic decisions. However, challenges, such as explainability, trust, accountability, and governance, remain.

Discussion. AI reconfigures decision-making architecture, requiring symbiotic integration between technology, human oversight, and ethical policies.

Conclusion. AI strengthens information security and improves organisational decision-making. However, the optimization of its impact is conditional on the implementation of ethical governance, transparency, and continuous training.

Introduction

The accelerating digital transformation of organisations has expanded the cyber threat landscape, increasing exposure to sophisticated and fast-evolving attacks. Traditional rule-based security mechanisms struggle to cope with the volume and complexity of modern threats, prompting organisations to adopt Artificial Intelligence (AI) techniques for threat detection, prevention, and response (Batool et al., 2025; Sarker, 2024). Machine learning and data-driven approaches enable the automated analysis of large-scale data streams and the identification of anomalous behaviour in near real time.

While existing research has extensively documented the technical benefits of AI in cybersecurity, its influence extends beyond operational performance. AI-based systems increasingly shape how security information is filtered, interpreted, and acted upon, thereby influencing organisational decision-making. From alert prioritization to automated response recommendations, AI becomes an active participant in decision processes, aligning with Human-in-the-Loop (HITL) and bounded rationality perspectives (Khan, 2025; Sushant Kumar et al., 2024).

Despite this potential, organisations face persistent challenges when integrating AI into security decision-making. Limited explainability, concerns about algorithmic bias, and insufficient governance structures can undermine trust and restrict effective use (Mökander et al., 2021). Much of the existing literature remains focused on technical performance, leaving a gap in understanding how AI reshapes organisational decision architectures and governance practices.

This study makes two interrelated contributions. First, it provides an integrative synthesis of recent literature on AI in information security, structuring fragmented research across operational, strategic, and human-ethical dimensions. Second, building upon this synthesis, it develops a conceptual framework for the ethical integration of AI into cybersecurity decision-making. The framework translates dispersed findings into a socio-technical model that connects technical efficiency, decision processes, and governance responsibilities.

Literature review

The application of AI in information security has gained prominence in contemporary literature, driven by the exponential growth of cyber threats, the increasing complexity of digital ecosystems, and the need for faster and more informed organisational decisions. Recent research shows that machine learning (ML) and deep learning (DL) techniques reshape not only threat detection systems, but also strategic decision-making processes in organisational environments (Mohamed, 2023; Ugochukwu Ikechukwu Okoli et al., 2024). Thus, AI is no longer seen only as a technical mechanism for operational support but takes on a broader role in building organisational resilience, risk governance, and defining strategic priorities. This review is structured in three axes: (i) AI-based threat detection, (ii) decision support and risk management, and (iii) trust, explainability, and human factors in decision-making.

AI-based threat detection

The literature presents a robust consensus on the increasing effectiveness of ML models in intrusion detection, malware analysis, and network traffic monitoring. Approaches based on convolutional neural networks (CNNs) and recurrent networks (RNNs) have demonstrated high accuracy in identifying anomalous patterns and suspicious behaviour, allowing near-real-time responses, and reducing exposure to complex attacks (Batool et al., 2025). At the same time, unsupervised learning and reinforcement learning techniques have been investigated to identify zero-day attacks, mitigating the dependence on static signatures and increasing the adaptability of systems in the face of volatile environments (Zoppi et al., 2021).

A recent evolution refers to the incorporation of AI in zero trust architectures, in which continuous validation and adaptive segmentation are reinforced by intelligent algorithms,

expanding the ability to detect suspicious behaviour that escapes traditional mechanisms (Li et al., 2025). In addition, frameworks based on intelligent agents have been proposed for Security Operations Centres (SOCs), allowing automated prioritization of alerts and human decision support, reducing operational overhead, and balancing algorithmic autonomy with human oversight (Ismail et al., 2025; Srinivas et al., 2025).

Despite the advances, the literature identifies significant limitations. Reliance on large volumes of labelled data, which is essential for training trusted models, remains one of the biggest challenges, especially in environments where data privacy is critical (Liang et al., 2022). Another obstacle lies in the interpretability of complex models: deep networks behave like opaque systems, making it difficult to audit and justify automated decisions (Mulita, 2025; Tyagi et al., 2025). In response to these limitations, hybrid proposals emerge that combine AI with traditional threat intelligence, creating systems that are more robust, contextualised, and capable of operating in scenarios of uncertainty (Sarker, 2024; Spyros et al., 2025).

Decision support and risk management

Contemporary literature indicates a progressive transition of AI from a predominantly reactive role to a predictive and strategic role in risk management (Li et al., 2024). AI-based decision support systems can analyse massive streams of data in real-time, correlate dispersed events, and prioritize vulnerabilities based on impact on critical assets. Technologies, such as Security Orchestration, Automation and Response (SOAR) integrate advanced algorithms that automate responses and consolidate alerts, significantly reducing incident response time and improving the allocation of human and technological resources (Ismail et al., 2025). In a complementary way, User and Entity Behaviour Analytics (UEBA) solutions use machine learning to build behavioural profiles and detect deviations that may signal internal or external threats, reinforcing the ability of organisations to anticipate and mitigate risks (Gupta et al., 2024).

The literature also emphasises that the strategic potential of AI depends on organisational factors, including digital maturity, governance infrastructure, and alignment between technology and internal processes. Other studies demonstrate that effective adoption of AI is intrinsically linked to the ability of organisations to integrate new systems into their decision architecture, adjusting policies, workflows, and compliance practices (Aakula & Saini, 2024). At the same time, several authors warn of the risk of over-reliance on automatisms, especially in contexts of high institutional sensitivity, where automated decisions can generate ethical, legal, or reputational implications (Mökander et al., 2021). In this sense, AI should be understood as a support for human intelligence, and not as a substitute, preserving the critical role of decision-makers in the interpretation, validation, and contextualization of algorithmic recommendations (Rodgers et al., 2023).

The literature indicates that AI integration affects cybersecurity decision processes through several interrelated functional mechanisms. First, intelligent automation of alerts and incident handling reduces cognitive load in operational decision-making environments, enabling security analysts to focus on higher-level analytical and strategic tasks rather than routine triage (Batoool et al., 2025; Ismail et al., 2025; Srinivas et al., 2025). Second, AI-driven predictive analytics support risk-based prioritization, by structuring the evaluation of threats and vulnerabilities, which improves the allocation of security resources and strengthens mitigation planning (Li et al., 2024; Gupta et al., 2024). Third, anomaly detection models based on machine learning and deep learning enhance early identification of abnormal system behaviours, shortening response cycles and reducing exposure to rapidly evolving threats (Kumar & Gutierrez, 2025; Muneer et al., 2024; Zoppi et al., 2021). In addition, behavioural profiling approaches, such as User and Entity Behaviour Analytics (UEBA) enable context-sensitive interpretation of user actions, reinforcing insider threat detection and adaptive monitoring capabilities (Gupta et al., 2024; Spyros et al., 2025). Finally, AI-supported recommendation systems in Security Operations Centres introduce

machine-generated options into decision workflows, shifting cybersecurity management from reactive response toward a more anticipatory and preventive posture (Ismail et al., 2025; Srinivas et al., 2025). These functional effects illustrate how AI reshapes not only technical operations but also the structure and tempo of organisational security decision-making.

Studies also show that the impact of AI on strategic decision-making depends on organisational maturity, robust governance, and integration with existing practices (Neiroukh et al., 2024).

Theoretical frameworks for human-AI decision-making

To analytically anchor the discussion on the interaction between AI and organisational decision-making, it is useful to resort to established theoretical frameworks. Two are particularly relevant: the Human-in-the-Loop (HITL) model and the theories of organisational decision making.

The Human-in-the-Loop paradigm, applied to cybersecurity, postulates that the most effective systems are those that combine the speed and processing power of AI with contextual judgment, creativity, and the ethical responsibility of human beings (Sushant Kumar et al., 2024). This model does not see automation as a replacement, but as a symbiosis. AI acts as a cognitive amplifier, filtering out noise, identifying patterns, and suggesting actions, while the human analyst provides the domain *expertise*, understands organisational nuances, and takes ultimate accountability for the decision. This theoretical perspective helps explain why technical AI solutions, even the most precise ones, fail when they are not designed for smooth integration with existing human workflows and capabilities. In addition, organisational decision-making theories offer a lens to understand how AI challenges and transforms decision-making processes (Kaggwa et al., 2023). The classic rational model, which assumes complete information and a single objective of maximization, is insufficient to characterize cybersecurity environments, marked by uncertainty and multiple, sometimes conflicting objectives (Bokhari et al., 2022). AI aligns more closely with the model of Herbert Simon of bounded rationality, where decision-makers *satisfy* rather than optimize. Algorithms can expand these limits by processing more variables than a human could, thus providing a richer informational basis for more effective *satisfaction* (Khan, 2025). However, AI can also be framed in intuitive models, where *machine learning* systems identify subtle patterns and provide *sensitivities* or *alerts* that are not easily justifiable by linear logic, resembling an algorithmic intuition that must nevertheless be validated by human experience (Vouros, 2023).

Trust, explainability and human factors

Explainability emerges as one of the central themes in modern literature. Opaque models hinder transparency, auditing, and validation of results, compromising the confidence of analysts and managers (Arunraju Chinnaraju, 2025). Studies show that the acceptance of intelligent systems depends on the ability of users to understand algorithmic operation, interpret automated decisions, and reconcile technology with existing organisational practices (Vorm & Combs, 2022). The perception of autonomy, the clarity of the results presented, and the adequacy of the interfaces play decisive roles in the adoption of AI (Jocelyn Chew & Achananuparp, 2022).

Ethical and social issues also emerge as crucial elements. The literature highlights concern regarding algorithmic bias, personal data management, and automated decision accountabilities (Kordzadeh et al., 2022). Algorithmic bias poses a tangible threat to the fairness and effectiveness of safety decisions. For example, an insider threat detection system trained predominantly on user behaviour data from a specific technical department (e.g., IT) can learn to associate *normalcy* with nightly work patterns and access to administrative tools. When applied to the entire organisation, this model could generate false positives for employees in finance or marketing departments whose legitimate work patterns diverge from this learned *normal*, leading to unfair investigations, degradation of the work environment, and potential talent loss (Mubeen, 2024). This example illustrates how a biased training dataset can not only reduce technical accuracy but also incur ethical and management risks.

Over time and as organisations integrate AI into critical processes, it becomes imperative to establish robust governance policies, ongoing audit mechanisms, and regular training programs. Recent research suggests that the combination of AI-based decision models and human oversight increases organisational resilience, reduces systemic risks, and promotes greater reliability in decision-making (Zeriuoh & Amara, 2025).

Table 1 systematises the critical factors identified in the literature, detailing their specific impact and proposing concrete risk mitigation strategies. This analytical structure allows visualisation of not only the challenges, such as the opacity of the models (*black box*) or the risks of algorithmic bias, but also the practical ways to overcome them, namely through the implementation of Explainable AI (XAI) or periodic audits. By organising these elements in a relational way, the table acts as a strategic guide to support the design, implementation, and governance of AI systems that are not only efficient, but also transparent, fair, and, therefore, trustworthy.

Factor	Impact	Mitigation strategy
Explainability	Acceptance and auditing	Using XAI, clear dashboards
Autonomy	User engagement	Training and participation in decisions
Algorithmic bias	Unfair decisions	Periodic audit of datasets
Data privacy	Compliance and ethics	Anonymization and encryption

Table 1. Governance and trust risk-mitigation matrix for AI in cybersecurity decision-making. Source: authors' elaboration.

This table focuses specifically on governance, trust, and accountability mechanisms required to manage AI-related risks.

Literature synthesis

In summary, recent scientific production converges on the idea that AI is a transformative element in information security, providing substantial gains in accuracy, automation, and predictive capacity. The theoretical frameworks of HITL and organisational decision-making provide the conceptual lens to understand that its success is intrinsically linked to a symbiotic integration with the human element (Sushant Kumar et al., 2024). However, the same systems that increase efficiency introduce new challenges related to transparency, trust, governance, and accountability, where algorithmic bias stands out as a practical and ethical threat. The impact of AI on organisational decision-making is therefore conditioned by factors, such as digital maturity, organisational culture, ongoing human oversight, and ethical framework (Aakula & Saini, 2024). Despite this convergence on the benefits and challenges, the literature lacks studies that integrate these three dimensions, technical, decision-making, and human, in a single conceptual model, capable of guiding practical implementation in organisations. The present study aims to contribute to fill this gap, by proposing an *integrated framework* that will be detailed in subsequent sections. These gaps motivate the methodological approach described in the following section.

Figure 1 presents a conceptual model that illustrates the multidimensional impact of AI on information security and organisational decision-making. This model integrates the three central axes discussed throughout the review, AI-based threat detection, decision support and risk management, and human and trust factors, demonstrating their dynamic interdependence. The figure illustrates the interdependence between operational AI capabilities, decision-support functions, and human-governance factors. Rather than operating independently, these components form a feedback system in which technical outputs influence decision processes,

which are in turn shaped by trust, oversight, and policy constraints. This visualisation emphasises that cybersecurity effectiveness emerges from alignment among these dimensions rather than from algorithmic performance alone. The gaps and the socio-technical complexity identified in the literature motivate the integrative methodological approach and are presented in the next section.

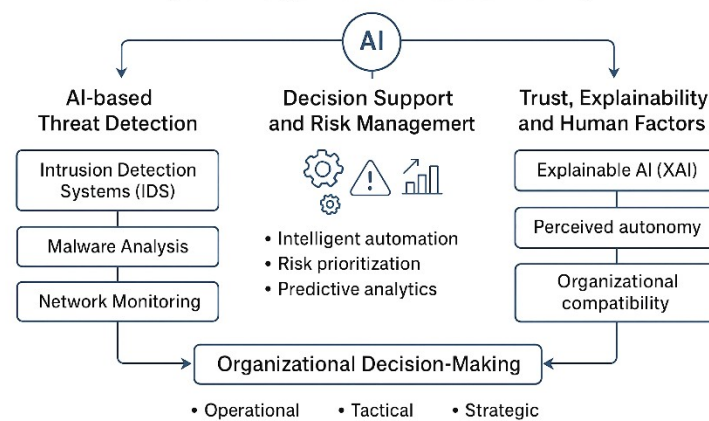


Figure 1. Conceptual model illustrating the impact of AI on information security and organisational decision-making. Source: authors' elaboration.

Method

Type of review

This study adopts an integrative literature review methodology. It was selected because the research problem spans technical, organisational, and ethical domains. Unlike systematic reviews focused on homogeneous empirical designs, integrative reviews enable the synthesis of conceptual, technical, and empirical studies, making them particularly suitable for examining socio-technical phenomena, such as AI-enabled cybersecurity decision-making. This approach allows the combination of performance-oriented research with governance and human-factor perspectives that would otherwise remain analytically disconnected.

Data sources and search strategy

The literature search was conducted across Scopus, IEEE Xplore, ScienceDirect, and SpringerLink. Searches used combinations of the keywords “artificial intelligence,” “cybersecurity,” “information security,” and “decision-making.” The review focused on studies published between 2021 and 2025 to capture recent developments in AI-enabled security.

Inclusion and exclusion criteria

Inclusion criteria:

- Peer-reviewed journal articles and authoritative technical reports
- Published between 2021 and 2025
- Focus on AI applications in information security
- Explicit discussion of organisational decision-making, governance, or human factors

Exclusion criteria:

- Publications prior to 2021
- Purely technical studies without organisational implications

- Non-peer-reviewed opinion or commentary articles

Figure 2 summarizes the literature identification and selection process.

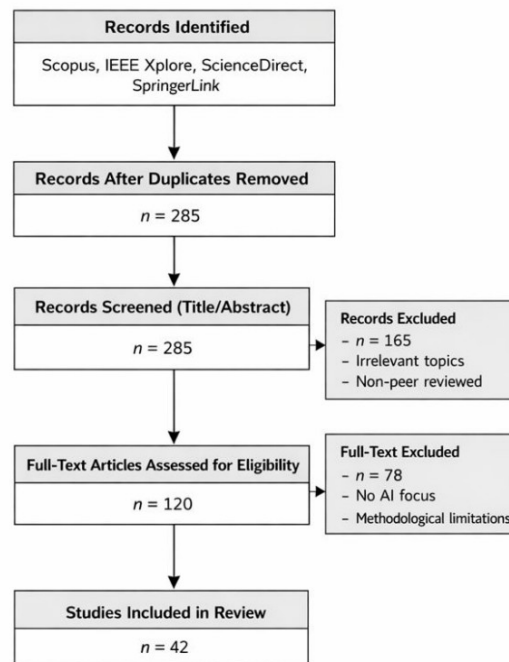


Figure 2. Literature identification and selection process for the integrative review. Source: authors' elaboration.

This figure demonstrates the progressive narrowing of the literature corpus, ensuring that the final sample reflects studies aligned with both technical and organisational dimensions of AI-enabled cybersecurity. The results of the analytical process are presented in the next section.

Analysis

This section synthesises evidence from the reviewed literature, describing the main patterns identified across studies without extending into theoretical or normative interpretation.

Selected studies were analysed using thematic coding. Initial codes were grouped into higher-level themes, resulting in three analytical dimensions: operational efficiency, decision support, and human and ethical factors. Given the heterogeneity of study designs, no formal quality scoring was applied; this limitation is acknowledged in Limitations and future research section.

Operational efficiency

The results show that solutions based on *machine learning* and *deep learning* substantially improve the ability to detect threats and anomalies in corporate networks. Some empirical studies report reductions in response time of up to 40%; however, these findings are context-dependent and derived from heterogeneous operational environments. In addition, the use of supervised algorithms, such as *Random Forest* and *Support Vector Machines*, has been shown to be effective in analysing traffic and preventing intrusions (Kumar & Gutierrez, 2025; Muneer et al., 2024).

Figure 3 presents an analysis of the percentage distribution of the main application areas, based on the literature published between 2021 and 2025. This visualisation enables identification of dominant research priorities of the sector and the scientific community, revealing which areas

have attracted the most attention and intellectual investment. The same figure also serves as a dynamic snapshot of the state of the art, offering an essential macro perspective to guide future technological developments and R&D resource allocation decisions.

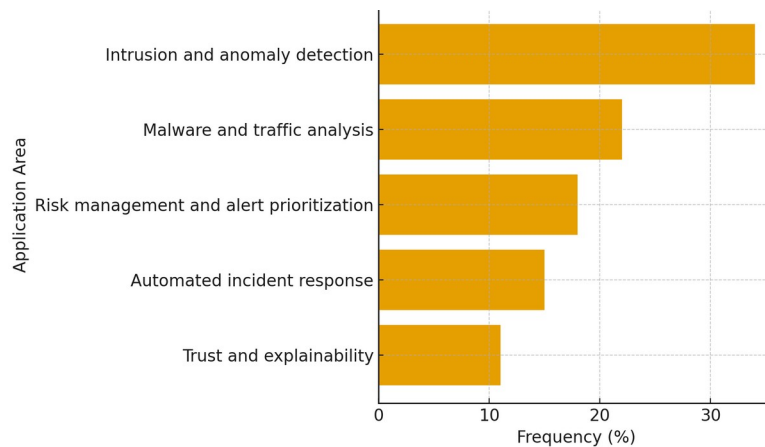


Figure 3. Principal areas of application of AI in Cybersecurity (2021-2025). Source: authors' elaboration.

The predominance observed in the domain of intrusion detection (34%), followed by malware and traffic analysis (22%) and risk management (18%), confirms the focus of the scientific community on operational applications that aim to increase the accuracy and efficiency of cyber defence mechanisms.

Decision support and strategic management

The second dimension is related to the role of AI in decision support and initiative-taking risk management. Tools, such as SOAR and UEBA, are being widely used to correlate alerts and prioritize vulnerabilities (Gupta et al., 2024; Ismail et al., 2025).

The automation of these processes has allowed security managers to shift the focus from reactive response to a preventive and strategic posture. However, some studies warn that blind trust in algorithmic recommendations can lead to biased or non-transparent decisions if the models are not auditable (Landers & Behrend, 2022).

Automated incident management tools, such as SOAR and UEBA, are demonstrating a measurable impact on prioritizing alerts and reducing false positives (ENISA, 2023). Reported improvements (30–40%) reflect specific case settings and should not be interpreted as generalisable performance benchmarks (Batoool et al., 2025).

Human and ethical dimensions

The third dimension highlights the role of human factors, trust, and explainability. The acceptance of AI systems by security professionals relies heavily on the ability to understand the reasons behind system decisions (Vorm & Combs, 2022). The literature emphasises the need for XAI to ensure that decision-makers maintain control and confidence over recommended actions (Arunraju Chinnaraju, 2025).

From an ethical point of view, the collection and processing of sensitive data to train AI models raises privacy concerns and algorithmic bias. The absence of transparency can compromise organisational accountability and generate significant reputational risks (Mökander et al., 2021). The analysis reveals that human and ethical factors are not a separate barrier, but rather the *ground* on which efficiency and decision support are built. As illustrated in Table 1, the mitigation strategy for explainability is the use of XAI. However, the technical implementation of XAI runs into a *trade-off* between precision and explainability: the most complex and accurate models,

such as *deep learning*, are often the opaquest, while the most interpretable models, such as decision trees, can underperform. Thus, the choice of an AI model becomes a strategic decision that weighs the marginal gain in accuracy against the loss of transparency and the consequent erosion of team trust.

Synthesis of findings

In an integrated way, the analysis confirms that AI plays a decisive role in the modernisation of organisational cybersecurity. Its effectiveness, however, depends on the balanced combination of technology and human oversight. The results indicate that success does not lie in unilaterally maximising one of the dimensions (e.g., efficiency), but in actively managing the inherent tensions between them. The organisations that gain the most operational resilience are those that recognise these *trade-offs* and integrate *AI-driven cybersecurity* with ethical governance, algorithmic auditing, and ongoing training, thus giving rise to a decision-making ecosystem, where the speed of AI is tempered by human judgment, its autonomy is counterbalanced by human control, and its complexity is made useful through explainability.

Beyond identifying thematic categories, the analysis revealed that most studies emphasise operational performance, while governance and human oversight dimensions remain comparatively underrepresented. This imbalance suggests that the technical maturity of AI applications is advancing faster than organisational and ethical integration frameworks, reinforcing the need for interdisciplinary approaches.

Results

The thematic coding of the reviewed studies revealed three interdependent dimensions that structure the impact of AI on cybersecurity decision-making.

Operational efficiency

The reviewed literature consistently reports that AI improves intrusion detection accuracy, reduces false positives, and accelerates incident response. Machine learning and deep learning models enable faster identification of anomalous patterns, contributing to measurable efficiency gains (Kumar & Gutierrez, 2025).

Quantitative performance improvements reported in the literature vary substantially across contexts, datasets, and implementation conditions, and, therefore, indicate potential rather than standardized outcomes.

Strategic decision support

AI-based tools, such as Security Orchestration, Automation and Response (SOAR) and User and Entity Behaviour Analytics (UEBA) support decision-making by correlating alerts and prioritizing risks. These systems enable a shift from reactive responses to predictive and risk-based security strategies.

Human and ethical factors

Trust and explainability emerge as critical factors for AI adoption. Blackbox models hinder accountability and user confidence, whereas explainable AI (XAI) techniques improve transparency and acceptance (Arunraju Chinnaraju, 2025). Ethical concerns include data privacy, algorithmic bias, and responsibility for automated decisions.

Framework for ethical AI integration

The proposed framework constitutes the conceptual outcome of the integrative synthesis, translating empirical and theoretical findings into an applied socio-technical model. Based on the synthesis, this study proposes a framework comprising three interdependent pillars: (1) technical efficiency, (2) trust and explainability, and (3) governance and accountability. At the centre lies

contextualised human judgment, supported by continuous feedback between operational outcomes and governance mechanisms. Figure 4 shows a conceptual *framework* for this. The framework operationalises the study's integrative contribution by translating literature synthesis into an implementable decision-governance structure. The findings are interpreted and contextualised in the Discussion section.

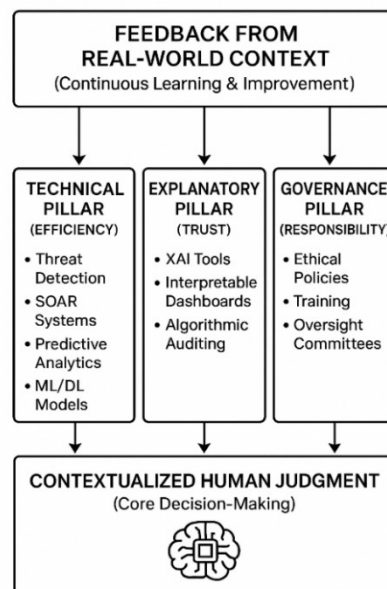


Figure 4. Framework for the ethical integration of AI in cybersecurity decision-making. Source: authors' elaboration.

Discussion

This section interprets the synthesised findings considering organisational decision-making theory, highlighting conceptual implications, trade-offs, and governance challenges. The findings indicate that AI adoption in cybersecurity is not merely a technological transition, but a transformation of decision authority structures. Organisations must, therefore, redesign accountability chains, audit practices, and human oversight mechanisms alongside technical deployment. The discussion highlights that effectiveness depends less on algorithmic sophistication alone and more on alignment between AI systems, decision processes, and governance structures.

A second fundamental trade-off arises here between *autonomy* and *control*. Intelligent automation frees up human resources for strategic tasks, but excessive delegation of authority to systems can have unintended consequences. For example, automating incident responses can contain a threat quickly, but it can also lead to *alert fatigue*, if false positives are not minimised. It can lead to inappropriate automated responses that disrupt legitimate services, causing operational and reputational damage. The ideal balance is not full automation, but rather supervised autonomy, where AI proposes actions, but the execution of high-impact measures requires human validation.

From a theoretical perspective, cybersecurity decision-making can be understood as an extended, bounded rationality process, in which AI expands informational capacity while humans retain interpretive and ethical responsibility. This perspective aligns with HITL principles and emphasises augmentation rather than replacement of human judgment.

Limitations and future research

This review is limited by reliance on published literature, which may underrepresent failed implementations or emerging practices. No formal quality assessment was conducted due to the heterogeneity of study designs. Quantitative findings synthesised from diverse contexts should, therefore, be interpreted cautiously. Future research should employ empirical case studies and longitudinal analyses to validate and refine the proposed framework.

Furthermore, the reliance on published studies may introduce publication bias, as unsuccessful or incomplete AI implementations are less likely to be documented. The rapid evolution of AI technologies means that some findings may become outdated as governance practices mature.

Conclusion

AI significantly enhances information security capabilities and supports organisational decision-making. However, its value depends on transparent, human-centred, and ethically governed integration. AI should be viewed as an enabler of informed judgment, rather than a substitute for human responsibility. The study underscores that sustainable AI adoption in cybersecurity requires simultaneous progress in technical performance, organisational adaptation, and ethical governance. Organisations that address only the technical dimension risk efficiency gains without institutional resilience. When responsibly implemented, AI strengthens both cyber resilience and organisational trust.

About the author

Sérgio Silva, works in the Department of Communication Sciences and Information Technologies, at the University of Maia, 4475-690 Maia, Portugal. They can be contacted at D012196@umaia.pt

References

- Aakula, A., & Saini, V. (2024). The impact of AI on organisational change in digital transformation. In *Internet of Things and Edge Computing Journal*, 4(1), 75–115.
<https://thesciencebrigade.com/iotecj/article/view/465>
- Arunraju Chinnaraju. (2025). Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. *World Journal of Advanced Engineering Technology and Sciences*, 14(3), 170–207.
<https://doi.org/10.30574/wjaets.2025.14.3.0106>
- Batool, F., & Hassnain, S. I. (2025). Neural network-enhanced machine learning applications in cybersecurity for real-time detection of anomalous activities and prevention of unauthorized access in large-scale networks. 1(1). <https://doi.org/10.62762/TNC.2025.920886>
- Bokhari, S., Hamrioui, S., & Aider, M. (2022). Cybersecurity strategy under uncertainties for an IoE environment. *Journal of Network and Computer Applications*, 205, 103426.
<https://doi.org/10.1016/J.JNCA.2022.103426>
- ENISA. (2023). *ENISA threat landscape 2023*. Publications Office of the EU.
<https://op.europa.eu/en/publication-detail/-/publication/3bd053c2-9e1e-11ee-b164-01aa75ed71a1/language-en>

- Gupta, A., Senior, D., & Engineer, S. (2024). Unveiling hidden threats with ML-powered user and entity behavior analytics (UEBA). *Turkish Journal of Computer and Mathematics Education* 15 (1). <https://doi.org/10.61841/turcomat.v15i1.14394>
- Ismail, Kurnia, R., Brata, Z. A., Nelistiani, G. A., Heo, S., Kim, Hyeongon, & Kim, Howon. (2025). Toward robust security orchestration and automated response in security operations centers with a hyper-automation approach using agentic artificial intelligence. *Information (Switzerland)*, 16(5). <https://doi.org/10.3390/info16050365>
- Jocelyn Chew, H. S., & Achananuparp, P. (2022). Perceptions and needs of artificial intelligence in health care to increase adoption: Scoping review. *Journal of Medical Internet Research*, 24(1). <https://doi.org/10.2196/32939>
- Kaggwa, S., Francisa Eleogu, T., Okonkwo, F., Farayola, O. A., Ugomma Uwaoma, P., & Akinoso, A. (2023). AI in decision making: Transforming business strategies. *International Journal of Research and Scientific Innovation (IJRSI)*, 10(12), 423-444, <https://doi.org/10.51244/IJRSI.2023.1012032>
- Khan, S. (2025). From constraints to cognition: Integrating bounded rationality into AI design for realistic decision-making. *The Critical Review of Social Sciences Studies*, 3(3). <https://doi.org/10.59075/n96xaa33>
- Kordzadeh, N., Ghasemaghaei, M., Mikalef, P., Popovic, A., Lundström, J. E., & Conboy, K. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388-409. <https://doi.org/10.1080/0960085X.2021.1927212>
- Kumar, A., & Gutierrez, J. A. (2025). Impact of machine learning on intrusion detection systems for the protection of critical infrastructure. *Information* 2025, 16(7), 515. <https://doi.org/10.3390/INFO16070515>
- Kumar, Sushant, Datta, S., Singh, V., Datta, D., Kumar Singh, S., & Sharma, R. (2024). Applications, challenges, and future directions of human-in-the-loop learning. *IEEE Access*, 12, 75735-75760. <https://doi.org/10.1109/ACCESS.2024.3401547>
- Kumar, Santosh, Sharma, N. K., Sharma, M., & Agrawal, N. (2024). Text extraction from images using Tesseract. In *Deep Learning Techniques for Automation and Industrial Applications* (pp. 1-18). Wiley. <https://doi.org/10.1002/9781394234271.ch1>
- Landers, R. N., & Behrend, T. S. (2022). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, 78(1), 36-49. <https://doi.org/10.1037/AMP0000972>
- Li, H., Yazdi, M., Nedjati, A., Moradi, R., Adumene, S., Dao, U., Moradi, A., Haghighi, A., Obeng, F. E., Huang, C.-G., Kang, H. S., Pirbalouti, R. G., Zarei, E., Dehghan, M., Darvishmotevali, M., Ghasemi, P., Fard, P. S., & Garg, H. (2024). Harnessing AI for project risk management: A paradigm shift. In M. Yazdi (Ed.), *Progressive decision-making tools and applications in project and operation management: Approaches, case studies, multi-criteria decision-making, multi-objective decision-making, decision under uncertainty* (pp. 253-272). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-51719-8_16
- Li, K., Li, C., Yuan, X., Li, S., Zou, S., Ahmed, S. S., Ni, W., Niyato, D., Jamalipour, A., Dressler, F., & Akan, O. B. (2025). Zero-trust foundation models: A new paradigm for secure and collaborative artificial intelligence for Internet of Things. *IEEE Internet of Things Journal*, 12(2), 46269-46293. <https://doi.org/10.1109/JIOT.2025.3603957>

- Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence*, 4(8), 669–677. <https://doi.org/10.1038/s42256-022-00516-1>
- Mohamed, M. (2023). Empowering deep learning based organisational decision making: A survey 2. *Sustainable Machine Intelligence Journal*, 3(5), 1–13. <https://doi.org/10.61185/SMIJ.2023.33105>
- Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, 27(4). <https://doi.org/10.1007/s11948-021-00319-4>
- Mubeen, M. (2024). Zero-trust architecture for cloud-based AI chat applications: Encryption, access control and continuous AI-driven verification. (Master's thesis, Jamk University of Applied Sciences). <http://www.theseus.fi/handle/10024/876662>
- Mulita, R. (2025). Challenging the cybersecurity threats through education programs. In K. Dhoska & E. Spaho (Eds.), *Bridging horizons in artificial intelligence, robotics, cybersecurity, smart cities, and digital economy* (pp. 271–283). Springer Nature Switzerland.
- Muneer, S., Farooq, U., Athar, A., Raza, M. A., Ghazal, T. M., & Sakib, S. (2024). A critical review of artificial intelligence based approaches in intrusion detection: A comprehensive analysis. *Journal of Engineering*, 2024(1), 3909173. <https://doi.org/10.1155/2024/3909173>
- Neiroukh, S., Emeagwali, O. L., & Aljuhmani, H. Y. (2025). Artificial intelligence capability and organisational performance: Unraveling the mediating mechanisms of decision-making processes. *Management Decision*, 63(10), 3501–3532. <https://doi.org/10.1108/MD-10-2023-1946>
- Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, 33(1), 100925. <https://doi.org/10.1016/J.HRMR.2022.100925>
- Sarker, I. H. (2024). Introduction to AI-driven cybersecurity and threat intelligence. In I. H. Sarker (Ed.), *AI-driven cybersecurity and threat intelligence: Cyber automation, intelligent decision-making and explainability* (pp. 3–19). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54497-2_1
- Spyros, A., Koritsas, I., Papoutsis, A., Panagiotou, P., Chatzakou, D., Kavallieros, D., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2025). AI-based holistic framework for cyber threat intelligence management. *IEEE Access*, 13, 20820–20846. <https://doi.org/10.1109/ACCESS.2025.3533084>
- Srinivas, S., Kirk, B., Zendejas, J., Espino, M., Boskovich, M., Bari, A., Dajani, K., & Alzahrani, N. (2025). AI-augmented SOC: A survey of LLMs and agents for security automation. *Journal of Cybersecurity and Privacy*, 5(4), 95. <https://doi.org/10.3390/jcp5040095>
- Tyagi, Himanshu, Goel, P. K., Tyagi, P., & Tyagi, Himani. (2025). AI-driven cybersecurity in industrial automation: Navigating ethical and legal complexities. In *AI-enhanced cybersecurity for industrial automation* (pp. 309–338). IGI Global. <https://doi.org/10.4018/979-8-3373-3241-3.ch016>
- Ugochukwu Ikechukwu Okoli, Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi, & Temitayo Oluwaseun Abrahams. (2024). Machine learning in cybersecurity: A review of threat

detection and defense mechanisms. *World Journal of Advanced Research and Reviews*, 21(1), 2286–2295. <https://doi.org/10.30574/wjarr.2024.21.1.0315>

Vorm, E. S., & Combs, D. J. Y. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (ISTAM). *International Journal of Human-Computer Interaction*, 38(18–20), 1828–1845. <https://doi.org/10.1080/10447318.2022.2070107>

Vouros, G. A. (2023). Explainable deep reinforcement learning: State of the art and challenges. *ACM Computing Surveys*, 55(5). <https://doi.org/10.1145/3527448>

Zerriouh, K., & Amara, M. (2025). AI-driven frameworks for strategic risk management: A systematic review and model for organisational resilience and decision support. *Journal of Intelligent Management Decision*, 4(3), 224–234. <https://doi.org/10.56578/jimd040304>

Zoppi, T., Ceccarelli, A., & Bondavalli, A. (2021). Unsupervised algorithms to detect zero-day attacks: Strategy and application. *IEEE Access*, 9, 90603–90615. <https://doi.org/10.1109/ACCESS.2021.3090957>

Copyright

Authors contributing to *Information Research* agree to publish their articles under a [Creative Commons CC BY-NC 4.0 license](https://creativecommons.org/licenses/by-nc/4.0/), which gives third parties the right to copy and redistribute the material in any medium or format. It also gives third parties the right to remix, transform and build upon the material for any purpose, except commercial, on the condition that clear acknowledgment is given to the author(s) of the work, that a link to the license is provided and that it is made clear if changes have been made to the work. This must be done in a reasonable manner, and must not imply that the licensor endorses the use of the work by third parties. The author(s) retain copyright to the work. You can also read more at: <https://publicera.kb.se/ir/openaccess>