THE ROLE OF ARTIFICIAL INTELLIGENCE IN CRISIS MANAGEMENT AND CYBERSECURITY: CHALLENGES, OPPORTUNITIES AND ETHICAL DILEMMAS

Samed Karovic[1]
[1]Metropolitan University Belgrade, Tadeusa Koscuska 63, Belgrade, Republic of Serbia,
karovic.samed@gmail.com, ORCID ID 0000-0001-5470-4650

Digital Object Identifier (DOI): 10.70995/TIYZ7624
Article Category: Scientific paper
Article Type: Original scientific paper
Received: 28th August 2025
Accepted: 5th September 2025

Abstract: *The development of digital technologies and the increasing frequency of security incidents make artificial intelligence (AI) a key tool in predicting, managing, and responding to crises. This paper analyzes the role of AI in crisis management and cybersecurity, with a focus on threat identification, risk analysis, automated response, and post-incident recovery. In addition to technological capabilities, issues of reliability, accountability, transparency, and ethical frameworks are examined. Methodologically, the paper is based on the analysis of relevant domestic and international literature, regulatory documents, and research reports. This approach demonstrates how AI can enhance organizational resilience, while also highlighting the risks that may arise from inadequate or uncontrolled implementation.*

Keywords: *artificial intelligence; crisis management; cybersecurity; ethics; resilience.*

1. INTRODUCTION

The digitalization of business and public services has significantly increased the exposure of organizations to complex cyber threats. The digital space has become a key environment for economic, political, and social activity, and consequently, the level of risk has grown proportionally. Accordingly, crisis management today is expected to ensure faster and more transparent decision-making.

In this context, artificial intelligence (AI) is increasingly emerging as an indispensable instrument in decision-making processes and the automation of crisis procedures. Its application contributes to incident prediction, risk assessment, and more efficient recovery, while simultaneously raising questions of technical reliability, regulatory governance, and ethical responsibility.

The working hypothesis of this study is based on the assumption that the integration of AI into crisis management and cybersecurity can significantly enhance organizational resilience, reduce response times, and improve the efficiency of reactions to digital threats, while at the same time introducing ethical and regulatory challenges that require systematic resolution.

The aim of this research is to provide an overview of the role of AI in modern crisis management and cybersecurity, analyzing both its technical capabilities and the societal challenges of its application, with particular emphasis on ethical issues and accountability.

2. REVIEW OF PREVIOUS RESEARCH

The development of artificial intelligence (AI) in the field of cybersecurity and crisis management has gained strong research momentum in recent years. Numerous studies

highlight that AI has become a key instrument in early anomaly detection, recognition of complex attack patterns, and optimization of response processes. Its role is particularly evident in environments where the scale and complexity of cyber threats exceed the capacities of traditional protection mechanisms.

Empirical research confirms that the application of AI can significantly accelerate detection and response to incidents compared to traditional rule-based methods. The latest international studies emphasize that the institutional use of AI in the public and security sectors must be based on clearly defined strategies, appropriate regulatory frameworks, and human oversight, in order to achieve a balance between efficiency and accountability (Government of the Republic of Serbia, 2025; UNDP Serbia, 2025; World Economic Forum, 2025).

Contemporary research also shows that the use of predictive analytics and advanced machine learning algorithms can significantly reduce the number of successful attacks and contribute to better resource planning in crisis situations. For example, Fawait, Fakhri, and Muslimah (2024) combine quantitative cyberattack simulations with expert interviews and conclude that AI can achieve a success rate of around 85% in detecting and responding to attacks. However, the authors warn that algorithms have limitations in cases of unexpected incidents, and that excessive reliance on AI without human oversight can increase system vulnerability. This confirms that AI cannot fully replace human judgment, but rather complement it by providing speed and accuracy, while the human factor ensures context and responsibility.

Beyond technical capabilities, a significant body of literature focuses on trust, ethics, and accountability in AI application. Floridi and Cowls (2019) emphasize the need for explainable models and transparent decision-making mechanisms, while Gunning and Aha (2019) develop the concept of "explainable AI" (XAI) as a way to ensure user trust and legitimacy of decisions in crisis situations. González, Pereira, and Martins (2024) further stress that the application of AI in crisis management raises complex legal and ethical issues related to responsibility and transparency of decisions. From a broader societal and theoretical perspective, Zuboff (2019) warns that the phenomenon of "surveillance capitalism" demonstrates how human experience is appropriated as raw material for predictive models, raising questions of power and control over user behavior. It is particularly important to highlight that ethics is no longer a secondary issue, but one of the central research areas in the context of AI application in security.

In domestic literature, Karovic (2024) emphasizes that crisis management must be viewed in continuity with cybersecurity, as digital incidents increasingly threaten business stability and public trust. The author points to the need to strengthen institutional capacities and continuous training of employees, since technological tools cannot fully replace the human factor in decision-making. This confirms the position that technological solutions must be integrated with organizational and regulatory measures in order to ensure sustainability and system resilience.

In conclusion, existing research consistently confirms that AI represents a revolutionary tool in cyber crisis management. Its main advantages lie in predictive analytics, speed of response, and recovery optimization, while challenges include issues of trust, transparency, accountability, and ethical dilemmas. The most recent literature clearly emphasizes that the optimal model of AI application is the synergy of technology and human oversight, supported by clear regulatory and ethical frameworks (Government of the Republic of Serbia, 2025; UNDP Serbia, 2025; World Economic Forum, 2025; Zuboff, 2019).

3. CRISIS MANAGEMENT AND CYBERSECURITY

Crisis management encompasses activities of prevention, preparedness, response, and recovery from events that threaten the survival of organizations, with the aim of minimizing

damage and ensuring business continuity (Karovic, 2024). Digitalization and the growth of cyber threats have significantly expanded the spectrum of challenges, making cybersecurity a key domain of crisis management. It highlights the weaknesses of traditional approaches—slow incident detection, reactive action, and limited analytical capacities. The integration of artificial intelligence enables faster anomaly detection, the creation of predictive models, and the transition from a reactive to an anticipatory framework, thereby reducing the consequences of incidents, strengthening institutional resilience, and improving actor coordination (González, Pereira & Martins, 2024).

Global and national reports confirm that cyber incidents rank among the greatest risks of the modern era, alongside climate and geopolitical crises, which underscores the need for proactive strategies and international cooperation (World Economic Forum, 2025; Government of the Republic of Serbia, 2025; UNDP Serbia, 2025). From a broader societal perspective, Zuboff (2019) warns that surveillance capitalism reshapes the very nature of trust and power, meaning that the development of crisis management must also include ethical dimensions. Therefore, traditional models are no longer sufficient—it is necessary to develop anticipatory and integrated approaches in which artificial intelligence plays a central role, ensuring stability and trust in modern organizations (Karovic, 2024).

4. THE ROLE OF ARTIFICIAL INTELLIGENCE IN CRISIS MANAGEMENT

The role of artificial intelligence (AI) in crisis management goes beyond traditional approaches based on human resources and mechanical procedures. Defined as a set of algorithmic and technological solutions capable of data analysis, learning, and decision-making (Srivastava et al., 2022; González, Pereira & Martins, 2024), AI has become a key element in proactive risk prediction and management. Its application enables rapid data processing, pattern recognition, and reduced response times, thereby strengthening organizational resilience (Karovic, 2024).

Research confirms that AI facilitates the transition from reactive to anticipatory approaches but also highlights limitations—such as susceptibility to evasion attacks and the risk of false positives or false negatives. This underscores the importance of explainable artificial intelligence (XAI), which increases decision transparency and strengthens trust in systems (González, Pereira & Martins, 2024; Srivastava et al., 2022). Global reports (World Economic Forum, 2025) warn that, without ethical and regulatory frameworks, the use of AI may deepen social inequalities, while Zuboff (2019) emphasizes that surveillance capitalism erodes privacy and trust.

The findings confirm that AI brings significant benefits to crisis management but also introduces complex challenges. Therefore, its application must be considered through interconnected technical, ethical, and regulatory dimensions. To provide a clearer overview, the following section presents the key advantages and challenges of applying AI in crisis management (Table 1).

Table 1: Comparative overview of the advantages and challenges of applying AI in crisis management

Advantages of applying AI	Challenges of applying AI
Rapid and accurate detection of anomalies and threats in real time	Susceptibility to “evasion” attacks and data manipulation (Srivastava et al., 2022)
Predictive analytics enables anticipatory action and risk reduction	Risk of false positive and false negative results in detection (Srivastava et al., 2022)

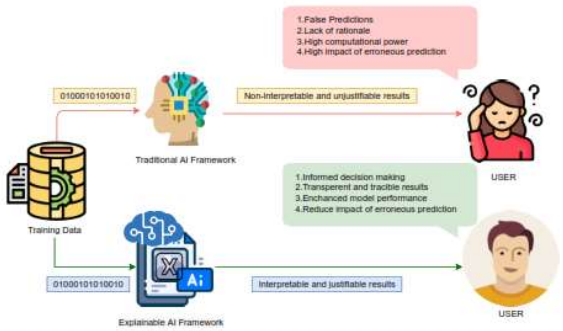
Automation of response and optimization of resources during crises	Lack of transparency and the "black box" nature of algorithms (Floridi & Cowls, 2019)
Strengthening organizational resilience through learning from past incidents	Unclear accountability for decisions (manufacturer, user, regulator) (González, Pereira & Martins, 2024)
Enhancing communication and coordination within crisis teams (NLP and data analysis)	Ethical dilemmas: privacy, bias, and potential misuse of data (Zuboff, 2019; World Economic Forum, 2025)

Source: Author's compilation based on prema Srivastava et al. (2022), Floridi & Cowls (2019), González, Pereira & Martins (2024), World Economic Forum (2025), Zuboff (2019).

The application of artificial intelligence (AI) in crisis management brings significant benefits, such as rapid detection and advanced analytics, but also complex challenges, including vulnerability to manipulation and issues of interpretability. Ethics, transparency, and accountability are as important as technical efficiency, since data misuse or the absence of clear rules can undermine public trust and the legitimacy of decisions (Zuboff, 2019; World Economic Forum, 2025). The optimal model therefore entails a hybrid approach in which technological innovation is combined with human oversight and regulatory frameworks (Srivastava et al., 2022; González, Pereira & Martins, 2024). In this context, tools such as SIEM and SOAR platforms, NLP analytics, and predictive models are particularly significant, making AI an integral part of the continuous process of prevention, response, and recovery, thereby strengthening organizational resilience in a dynamic security environment.

4.1. Technical Aspects and Tools

The technical aspects of applying artificial intelligence in crisis management encompass tools ranging from early warning systems to advanced solutions in the field of cybersecurity. Machine learning and deep learning algorithms enable the prediction of natural disasters, detection of anomalies in network traffic, and identification of sophisticated attacks, while in critical infrastructure monitoring AI has proven to be more efficient than human operators (Srivastava et al., 2022; Karovic, 2024). In the cyber domain, particularly significant are SIEM (Security Information and Event Management) and SOAR (Security Orchestration, Automation and Response) systems, which allow centralized log analysis and automated incident response, as well as NLP (Natural Language Processing) techniques for early detection of crisis situations and disinformation. Predictive models further contribute to risk anticipation and resource mobilization, thereby strengthening institutional resilience (Government of the Republic of Serbia, 2025; UNDP Serbia, 2025).



Slika 1: AI and XAI-user's perspective

Source: Srivastava, A., Gaur, M., Saha, S., Mittal, S., Joshi, A. i Sarkar, S. (2022). XAI for Cybersecurity: State of the Art, Challenges, Open Issues and Future Directions. arXiv preprint arXiv:2206.03585.

Figure 1 illustrates the general workflow of explainable artificial intelligence (XAI) in cybersecurity, encompassing stages from data collection and the application of machine and deep learning algorithms, through explainability mechanisms, to decision-making processes that can also be evaluated by humans. This model demonstrates that the technical aspects of AI go beyond mere automated threat detection, encompassing the development of transparent and verifiable systems integrated into strategic decision-making (Srivastava et al., 2022).

The role of AI cannot be viewed in isolation but rather within the entire crisis management cycle – from prevention and response to recovery. A key element in this process is XAI, which ensures the transparency and interpretability of system decisions, reduces the risk of misjudgments, and strengthens trust in the use of AI technologies in sensitive domains such as cybersecurity and crisis management (González, Pereira & Martins, 2024; Srivastava et al., 2022).

This is precisely why it is important to view the role of AI not in isolation, but in the context of the entire crisis management cycle. Figure 2 shows how AI integrates the phases of prevention, response, and recovery, thereby establishing a continuous flow of crisis management and enhancing organizational resilience.

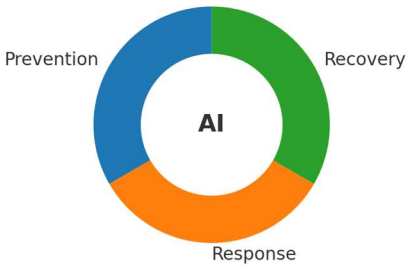


Figure 2: Artificial Intelligence in the Crisis Cycle (Prevention – Response – Recovery)

Note. Adapted from Strategy for the Development of Artificial Intelligence of the Republic of Serbia for the period 2025–2030 (Government of the Republic of Serbia, 2025); Artificial intelligence has the potential to accelerate human development (UNDP Serbia, 2025); and Transforming crisis management with AI: Risks, opportunities and governance frameworks (World Economic Forum, 2025).

The figure illustrates the role of AI in connecting the key phases of crisis management – from prevention and early threat detection, through automated response, to recovery and a return to normal operations. In this way, AI does not operate at a single point but ensures continuity and coordination throughout the entire crisis cycle. The visual representation confirms that technological tools can significantly accelerate response and improve decision-making quality, but also that they require integration with human oversight and clearly defined procedures.

4.2. Responsibility and Reliability

The issue of responsibility and transparency in the application of artificial intelligence is one of the greatest challenges of modern crisis management. When AI makes decisions under conditions of high uncertainty, it remains unclear whether responsibility lies with the system's manufacturer, the user, or the regulator, which can seriously undermine public trust and the

legitimacy of managerial decisions. This is why increasing attention is given to the development of Explainable Artificial Intelligence (XAI), which makes the decision-making process transparent and understandable (Gunning & Aha, 2019). If AI is treated as a "black box," its application may generate additional uncertainty, whereas XAI allows human operators to understand the reasons behind threat classifications and responses, reducing the risk of wrong decisions and strengthening trust in the systems (Srivastava et al., 2022).

Responsibility, however, goes beyond the technical aspect and includes legal and ethical dimensions. As emphasized by González, Pereira, and Martins (2024), it is necessary to develop clear mechanisms that encompass all stakeholders – manufacturers, users, and regulators – in order to prevent misuse and ensure fairness and legal legitimacy. In this way, XAI becomes a bridge between technical efficiency, managerial accountability, and legal sustainability in crisis situations.

4.3. Ethical and Regulatory Aspects

The ethics of using artificial intelligence in crisis management is closely linked to issues of privacy protection, potential misuse, and algorithmic bias. Surveillance systems, while valuable in preventing incidents, can seriously endanger fundamental citizens' rights if implemented without a clear legal and regulatory framework. A particular challenge lies in so-called algorithmic morality, when autonomous systems are forced to make "difficult decisions," raising the question of how to calibrate the values embedded within algorithms.

The application of AI in crisis management also raises the issue of accountability – whether decisions belong to the system's manufacturer, the organization, or the individual supervising the system's operation. Transparency and explainability of algorithms are therefore key prerequisites for trust, with the development of Explainable Artificial Intelligence (XAI) identified as one of the main priorities (Srivastava et al., 2022). Without these principles, the risk of misuse and the erosion of public trust increases.

As González, Pereira, and Martins (2024) emphasize, ethical aspects cannot be separated from legal frameworks, since accountability for algorithmic decisions requires clearly defined regulatory mechanisms. National strategic documents, such as the *Strategy for the Development of AI of the Republic of Serbia 2025–2030*, highlight the importance of ethical principles and human oversight (Government of the Republic of Serbia, 2025), while the World Economic Forum (2025) points out that global cooperation and ethical governance are prerequisites for international trust.

Zuboff (2019) further warns about the phenomenon of surveillance capitalism, where citizens' data are used as "raw material" for predictive models, which in crisis management can lead to commercial and political manipulation. UNDP Serbia (2025) stresses that, if developed in line with ethical guidelines, AI can contribute to both human and institutional development. Therefore, the responsible application of AI requires balancing technical efficiency, ethical principles, and regulatory measures to ensure legitimacy and protect the public interest.

5. DISCUSSION – CHALLENGES AND OPPORTUNITIES

A key question in contemporary crisis management and cybersecurity is how to achieve a balance between the technological capabilities of artificial intelligence and the challenges it generates. Research confirms that AI provides significant advantages in incident detection, automated response, and scenario modeling, thereby strengthening organizational resilience and improving resource planning in crisis situations (Karovic, 2024). This confirms that AI can be a transformative factor in crisis management, particularly in the domain of cybersecurity.

However, the challenges are equally emphasized alongside the opportunities. Srivastava et al. (2022) point out that AI-based systems are vulnerable to "evasion" attacks, which involve situations where an attacker deliberately modifies input data to appear legitimate, thereby deceiving the algorithm into failing to recognize the threat. In this way, AI systems may be misled to classify malicious activities as harmless, leading to false negatives and increasing the risk of serious consequences in crisis conditions. Moreover, false positives or false negatives themselves may result in poor decision-making. Floridi and Cowls (2019) stress that the absence of transparency and clearly defined accountability rules can seriously undermine trust in these technologies. In this context, the concept of Explainable Artificial Intelligence (XAI) is gaining growing importance, as it enables model interpretability and provides the foundation for their legitimate application. Thus, AI cannot be viewed as a fully autonomous solution, but rather as part of a hybrid system in which human oversight continues to play a crucial role (Srivastava et al., 2022).

Future development opportunities are linked to the advancement of explainable AI, the development of regulatory frameworks, and the implementation of standardized ethical guidelines that would enable the legitimate and sustainable application of these technologies (González, Pereira & Martins, 2024; Government of the Republic of Serbia, 2025; UNDP Serbia, 2025; World Economic Forum, 2025). Domestic research also confirms that AI in crisis management should not be regarded solely as a technical solution, but as a strategic resource that strengthens institutional resilience and increases public trust (Karovic, 2024). At this intersection lies the potential to transform crisis management from a reactive discipline into an anticipatory framework, capable of addressing the complex challenges of the digital age.

The results of the analysis confirm the initial hypothesis of the paper that the application of artificial intelligence in crisis management and cybersecurity can significantly contribute to strengthening organizational resilience, accelerating response, and optimizing resource allocation. At the same time, it has been confirmed that ethical and regulatory dilemmas, such as issues of accountability, transparency, and data privacy protection, remain key obstacles to its full legitimate implementation. These findings indicate that the hypothesis can be considered largely confirmed, with an essential emphasis on the importance of a hybrid model that combines technological efficiency with human oversight and a legal framework.

6. CONCLUSION

The application of artificial intelligence (AI) in crisis management and cybersecurity demonstrates that this technology can significantly enhance the resilience of institutions and organizations. Its ability to enable faster anomaly detection, predictive analytics, and automated response shifts crisis management from a reactive to an anticipatory framework. This ensures a higher level of efficiency in preventing and responding to incidents, as well as improved coordination among key actors in crisis situations.

However, the analysis has shown that the use of AI also carries serious challenges: technical vulnerabilities, the risk of false positives and false negatives, unclear rules of accountability, and ethical dilemmas related to privacy and bias. These factors represent key obstacles to sustainable and legitimate implementation.

In conclusion, the most effective model for applying artificial intelligence in crisis management and cybersecurity is based on the synergy of technological innovation and human expertise. Only such an integrated approach, combining technical efficiency with ethical and regulatory responsibility, can ensure sustainable and legitimate use of AI, strengthen institutional resilience, and preserve public trust in the digital age.

The research results thereby confirm the initial hypothesis of the paper, while emphasizing that its full validity can only be achieved through a hybrid model that combines technological efficiency with human oversight and clearly defined legal frameworks.

REFERENCES

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1 (2025-08-01).

Fawait, A., Fakhri, M., & Muslimah, N. (2024). Integrating AI in cybersecurity: Simulation and expert insights. *International Journal of Cybersecurity, 12*(1), 45–63. https://journal.ypidathu.or.id/index.php/multidisciplinary/article/download/1776/1268 (2025-08-01).

González, A., Pereira, A., & Martins, P. (2024). Accountability in artificial intelligence systems: Ethical and legal perspectives in crisis management. *Journal of Ethics in Digital Society, 6*(2), 101–118. https://www.researchgate.net/publication/377180421_Ethics_in_Artificial_Intelligence_an_Approach_to_Cybersecurity (2025-08-10).

Government of the Republic of Serbia. (2025). *Strategy for the development of artificial intelligence of the Republic of Serbia for the period 2025–2030*. https://aiassistant.rs/2025/01/15/strategija-razvoja-ai-u-srbiji-za-period-2025-2030/ (2025-08-12).

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850 (2025-08-12).

Karovic, S. (2024). Crisis management and cyber security: Key strategies for addressing digital threats. *Proceedings of the 10th International Scientific-Professional Conference Security and Crisis Management – Theory and Practice (SeCMan) – International Forum "Safety for the Future" 2024*, 21–28. Belgrade: Regional Association for Security and Crisis Management & S4 GLOSEC Global Security. https://bekmen.rs/wp-content/uploads/2025/03/SeCMan_2024.pdf (2025-08-27).

Srivastava, A., Gaur, M., Saha, S., Mittal, S., Joshi, A., & Sarkar, S. (2022). XAI for cybersecurity: State of the art, challenges, open issues and future directions. *arXiv preprint*. arXiv:2206.03585. https://arxiv.org/abs/2206.03585 (2025-08-05).

UNDP Serbia. (2025, May). *Artificial intelligence has the potential to accelerate human development*. https://www.undp.org/sr/serbia/news/vestacka-inteligencija-ima-potencijal-da-ubrza-ljudski-razvoj (2025-08-27).

World Economic Forum. (2025, January). *Transforming crisis management with AI: Risks, opportunities and governance frameworks*. Geneva: World Economic Forum. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf (2025-08-01).

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. London: Profile Books. https://archive.org/details/zuboff-shoshana.-the-age-of-surveillance-capitalism.-2019 (2025-08-15).