

Original Scientific Paper/Original naučni rad
Paper Submitted/Rad primljen: 31.12.2025.
Paper Accepted/Rad prihvaćen: 20.01.2026.
DOI: 10.5937/SJEM2600092J

UDC/UDK: 004.8:005.32

Nevidljive pretnje: Gledajući unazad da bismo sa AI krenuli napred – višedimenzionalni uticaji AI na organizacionu bezbednost i ljudsku autonomiju -

Tatjana Jovanović¹,

¹ Belgrade School of Engineering Management, Beopolis University, Belgrade, Serbia,
tatjana.jovanovic@fim.rs

Sažetak rada: Kako sistemi veštačke inteligencije sve više prodiru u najrazličitije sfere ljudske aktivnosti — od zdravstvene zaštite i obrazovanja do zapošljavanja i upravljanja — pojavljuju se novi oblici bezbednosnih rizika, od kojih su mnogi i dalje nedovoljno prepoznati. Ovaj rad istražuje višedimenzionalnu infiltraciju veštačke inteligencije u društveno-organizacione sisteme, analizirajući ne samo tehnološke rizike, već i psihološke, organizacione i epistemološke pretnje koje proizlaze iz nekontrolisane automatizacije i netransparentnog algoritamskog odlučivanja.

Polazeći od praksi upravljanja ljudskim resursima (HRM) u digitalno transformisanim organizacijama, rad identifikuje rane signale zloupotrebe veštačke inteligencije: smanjen uticaj ljudskog prosuđivanja, eroziju profesionalne autonomije, preterano oslanjanje na prediktivne analitike i gubitak poverenja među stakeholderima. Proširujući ove nalaze, studija ocrtava moguće scenarije razvoja — od odgovorne integracije veštačke inteligencije zasnovane na dijalogu sa zainteresovanim stranama, do distopijskih putanja obeleženih dehumanizacijom i društvenom fragmentacijom.

Završni deo rada nudi strateške preporuke za podsticanje odgovorne upotrebe veštačke inteligencije, naglašavajući značaj transdisciplinarnog obrazovanja, kritičke digitalne pismenosti i anticipativnog upravljanja. U oblikovanju narednih generacija stručnjaka, uloga akademskih i istraživačkih institucija postaje ključna za ugradnju etičke refleksivnosti, sistemskog mišljenja i čovekocentričnih vrednosti u samo projektovanje budućnosti koju pokreće veštačka inteligencija.

Ključne reči: veštačka inteligencija, ljudska bezbednost, odgovorna veštačka inteligencija

Invisible Threats: Looking Back to Move Forward with AI – The Multidimensional Impact of AI on Organizational Security and Human Agency –

Abstract in English: As artificial intelligence systems increasingly permeate all domains of human activity — from healthcare and education to employment and governance — new forms of security risks emerge, many of which remain insufficiently recognized. This paper explores the multidimensional infiltration of AI into socio-organizational systems, analyzing not only the technological risks, but also the psychological, organizational, and epistemological threats posed by unchecked automation and opaque algorithmic decision-making.

Drawing from human resource management (HRM) practices in digitally transformed organizations, the paper identifies early warning signals of AI misuse: diminished human agency, erosion of professional autonomy, over-reliance on predictive analytics, and loss of trust among stakeholders. By extending these findings, the study outlines future development scenarios, ranging from responsible integration of AI grounded in stakeholder dialogue to dystopian trajectories marked by dehumanization and social fragmentation.

The final section presents strategic recommendations for promoting responsible AI use, emphasizing the need for transdisciplinary education, critical digital literacy, and anticipatory governance. In shaping the next generation of experts, the role of academia and research institutions becomes vital in embedding ethical reflexivity, systemic thinking, and human-centric values into the very design of AI-driven futures.

Keywords: artificial intelligence, human security, responsible AI

1. Introduction

In recent years, artificial intelligence (AI) has moved from the periphery to the core of socio-organizational systems, subtly yet profoundly transforming how decisions are made, work is organized, and human value is defined. From personalized recruitment algorithms to automated performance evaluations, AI technologies are increasingly embedded in everyday organizational practices—often without transparent oversight or critical reflection. While discussions around AI have primarily focused on efficiency, accuracy, and innovation, growing evidence suggests that deeper, less visible risks are emerging.

These risks extend beyond traditional cybersecurity concerns. They touch on psychological autonomy, professional identity, knowledge legitimacy, and the erosion of trust within complex stakeholder ecosystems. In the pursuit of optimization, organizations may unintentionally displace human judgment, reproduce bias at scale, and weaken the social fabric that sustains cooperation and meaning in work environments.

This paper argues that the invisible threats of AI deserve a central place in the discourse on digital transformation and organizational security. It adopts a multidimensional lens to examine not only technological vulnerabilities but also the epistemological, ethical, and strategic blind spots that arise when AI systems are deployed without adequate stakeholder engagement or anticipatory governance.

Drawing on insights from human resource management (HRM) and critical digital studies, this paper explores:

- early indicators of organizational risk linked to AI,
- future development scenarios—from responsible integration to dystopian drift,
- and strategic recommendations for fostering a human-centered approach to AI deployment.

By looking back at the human foundations of organizational life, we aim to move forward with more reflexivity, caution, and imagination in shaping the AI-enhanced future.

2. Theoretical Framework: Security Beyond Technology

Understanding the invisible risks associated with artificial intelligence requires a departure from purely technological paradigms and a shift toward multidisciplinary frameworks that integrate human, organizational, and ethical dimensions. This section is guided by three interconnected theoretical lenses: the sociotechnical systems theory, which emphasizes the entanglement of technology with social and organizational structures; epistemologies of algorithmic governance, which explore how opaque AI systems reshape authority and knowledge; and a human-centered approach to HRM in the era of digital transformation, which highlights the strategic alignment of AI tools with human development and inclusion.

2.1. Sociotechnical Systems Perspective

Originally formulated in mid-20th century organizational studies, the sociotechnical systems perspective remains highly relevant for analyzing the complex entanglements of technology and human labor. Rather than viewing AI as a neutral tool, this framework conceptualizes it as a dynamic actor embedded within broader social systems. Recent work by Kudina and van de Poel (2024) argues that AI should be understood as part of evolving sociotechnical assemblages that shape—rather than merely support—human decision-making, ethics, and responsibility in real time. In parallel, Salwei et al. (2022) propose a practical framework for applying AI in healthcare through a sociotechnical lens, highlighting how misalignment between technological capabilities and human workflows can generate systemic vulnerabilities.

In the organizational context, this approach emphasizes that the risks posed by AI are not simply a matter of malfunction or misuse, but of structural integration—how technology reconfigures authority, accountability, and labor itself. The interaction between human agency and algorithmic output becomes central to identifying both opportunities and emerging invisible threats.

2.2. Algorithmic Epistemology and Opacity

Critical algorithm studies have revealed that AI systems are not only operational tools but epistemic agents that produce new regimes of knowledge and control. Scholars such as Pasquale (2015) and Ananny and Crawford (2018) demonstrate how algorithmic opacity can obscure accountability, delegitimize traditional expertise, and enable silent shifts in institutional power.

As AI models are trained on massive, often proprietary datasets and refined in black-box architectures, the epistemic authority gradually transfers from human professionals to systems whose logic remains largely inscrutable. This is particularly alarming in sectors such as hiring, education, or finance, where algorithmic decisions bear direct consequences for individuals—without transparent rationale or possibility of appeal. The rise of what Veale and Binns (2017) describe as “fairer machine learning without sensitive data” paradoxically risks reinforcing discrimination through proxy variables and biased optimization targets.

2.3. Human-Centered HRM in the Digital Era

In parallel with the sociotechnical and epistemological shifts, HRM has also undergone significant transformation. The dominant consensus now favors human-centered digitalization, wherein AI tools are deployed not to replace, but to enhance human capacity—provided that their use aligns with ethical principles and organizational inclusivity.

Recent literature, including Armstrong (2024) and Guest (2017), argues for frameworks that prioritize psychological well-being, job satisfaction, and autonomy alongside productivity metrics. The holistic HRM approach, developed in response to Industry 4.0, calls for integrated strategies that blend advanced analytics, personalization, and ethical safeguards. This includes participatory AI implementation, transparent communication with employees, and real-time HR analytics to detect unintended effects such as surveillance fatigue or deskilling.

Ultimately, understanding AI-related risks requires a synthesis of these three lenses: only by combining technical insight, organizational design, and human values can we build resilient systems that resist invisible threats and promote sustainable innovation.

3. Emerging Invisible Threats in AI-Driven Organizations

As artificial intelligence systems become increasingly embedded within organizations, their influence often transcends operational efficiency and begins to reshape fundamental human functions, roles, and relationships. These transformations are not always visible, but they can produce serious long-term consequences for autonomy, judgment, and trust in professional environments.

3.1. Diminished Human Agency

One of the earliest invisible threats lies in the gradual replacement of human judgment with algorithmic decision-making. From HR screening processes to loan approvals, AI models often determine outcomes based on statistical inference, not contextual understanding. While this shift promises efficiency, it reduces individual agency, as decisions are made based on opaque criteria. Ananny and Crawford (2018) argue that even with algorithmic transparency, knowing “how” a decision is made does not guarantee understanding “why” it was made, especially in systems trained on massive datasets with embedded biases. The risk is that human actors may become mere executors of automated recommendations, rather than critical evaluators of them.

3.2. Erosion of Professional Autonomy

In high-stakes domains such as healthcare, education, and human resource management, AI tools increasingly influence or even override professional discretion. This creates a paradox: while these tools are designed to assist decision-making, they can also constrain professionals to follow recommendations produced by opaque and data-driven systems. In the field of medicine, Jotterand and Bosco (2021) describe this tension as a “sword of Damocles” hanging over clinicians — where reliance on AI might reduce diagnostic error, but at the cost of undermining physicians’ epistemic authority and ethical accountability.

Similar dynamics are unfolding in the field of education, where teachers are encouraged to rely on AI-driven learning analytics and adaptive instruction platforms. These tools, although useful, often function as “closed boxes,” limiting pedagogical freedom and sidelining educator expertise. In the realm of human resources, predictive analytics are frequently used to assess employee engagement, performance, or flight risk. However, these models may neglect qualitative, relational, or cultural aspects of workplace life (Veale & Binns, 2017), thereby reducing the HR function to mechanical monitoring rather than human-centered development.

As recent legal developments indicate, AI-based hiring systems are increasingly subject to judicial scrutiny. For instance, a federal court in the United States granted preliminary certification in a landmark case involving

algorithmic bias, highlighting the judiciary's growing readiness to treat automated decision-making systems as potentially discriminatory under employment law (Forman & Wang, 2025). The integration of predictive analytics into HR processes can compromise contextual judgment and professional autonomy—especially when algorithmic outputs are perceived not holistically (Jovanovic, 2025), but as non-negotiable benchmarks rather than advisory tools.

3.3. Opacity and Over-Reliance on Predictive Analytics

AI systems, especially those built on deep learning architectures, often function as “black boxes.” Their inner logic is inaccessible even to their creators, making accountability difficult when errors occur. This opacity creates a new form of organizational vulnerability: over-reliance on predictive analytics without a clear understanding of model limitations. Veale and Binns (2017) note that such systems can perpetuate institutional biases or amplify disparities under the guise of objectivity. When organizations prioritize data-driven predictions over human insight, they risk institutionalizing flawed assumptions, eroding adaptability and innovation.

3.4. Loss of Trust Among Stakeholders

Trust is foundational in any organizational system. When decisions that affect people's lives—such as hiring, promotion, or access to healthcare—are delegated to algorithms, stakeholders may perceive the process as impersonal, unfair, or arbitrary. The absence of explainability fuels suspicion, particularly when affected individuals have no recourse or clarity on how decisions were reached. This can deteriorate workplace climate, reduce employee engagement, and ultimately threaten organizational legitimacy.

A notable example of eroded trust occurred when Amazon discontinued its experimental AI recruiting tool after discovering it systematically downgraded applications from women. The algorithm, trained on historical data from predominantly male-dominated hiring practices, encoded and perpetuated gender bias, favoring resumes that included male-associated terms while penalizing those with indicators of female identity (Dastin, 2018). Although the system was never deployed at scale, internal stakeholders voiced concerns over the fairness and opacity of its decision-making process. The incident not only damaged trust in the AI initiative, but also raised broader questions about ethical governance and bias mitigation in HR analytics.

This case illustrates that trust erosion is not only a theoretical concern but a practical reality, especially when algorithmic opacity intersects with sensitive human decisions.

4. Future Scenarios: Navigating the AI–Security Nexus

As artificial intelligence continues to permeate every layer of organizational life, it becomes increasingly important to anticipate its long-term trajectories. Rather than adopting a binary view—progress versus resistance—this section explores plausible future scenarios that reflect the complex interplay between technological capacity and societal response. Three key pathways are outlined: responsible AI integration, a dystopian trajectory of algorithmic dominance, and institutional anticipatory governance.

4.1. Responsible AI Integration

The responsible integration of AI requires more than just technical safeguards — it calls for inclusive stakeholder participation, ethical foresight, and continuous impact evaluation. This means building systems that are not only efficient but also transparent, accountable, and aligned with social values. Stilgoe, Owen, and Macnaghten (2013) propose a framework for responsible innovation grounded in anticipation, reflexivity, inclusion, and responsiveness — dimensions crucial for aligning AI with democratic principles. In this model, responsibility becomes a collective endeavor, engaging designers, institutions, and the public in shaping technology's direction.

Algorithmic Impact Assessments (AIA), increasingly discussed in legal and policy circles, exemplify this logic. As Selbst (2021) argues, AIA should not be viewed merely as checklists but as institutional processes that foster accountability and visibility of potential harms before deployment. Such tools help translate abstract ethical principles into actionable governance mechanisms, especially in high-risk sectors like employment, education, and healthcare.

4.2. Dystopian Trajectory

Conversely, the unchecked advancement of AI may lead to deeply troubling societal outcomes. Automated decision-making, when detached from ethical deliberation, can evolve into code as governance, where opaque algorithms enforce rules without human oversight or recourse (Pasquale, 2015). In such environments, personalization becomes surveillance, efficiency turns into disposability, and organizational life is depersonalized.

Jotterand and Bosco (2021) raise similar concerns in the medical field, warning that overreliance on AI can erode clinician autonomy and reduce patients to mere datasets. These dynamics can easily be extrapolated to other domains — such as HR, law enforcement, and education — where algorithmic logics risk sidelining empathy, dialogue, and context. Without institutional safeguards, we risk entering a governance model driven by automation rather than deliberation.

4.3. Institutional Role of Anticipatory Governance

To counteract these risks and guide AI development towards socially desirable outcomes, anticipatory governance must become a core institutional function. Kaminski (2021) emphasizes the importance of integrating *ex ante* assessments — particularly under legal regimes like the EU's General Data Protection Regulation (GDPR) — that not only identify privacy risks but also anticipate broader societal consequences of AI deployment.

Anticipatory governance entails proactive scenario planning, interdisciplinary research, and embedding foresight in organizational decision-making. It positions institutions not as reactive regulators, but as forward-looking enablers of innovation that safeguards human dignity, inclusivity, and systemic resilience.

5. Strategic Recommendations

As artificial intelligence becomes embedded in the fabric of organizational decision-making, the need for a strategic, ethically grounded approach grows increasingly urgent. To counteract the invisible threats previously discussed—diminished human agency, erosion of professional autonomy, opacity of systems, and erosion of trust—organizations must adopt proactive, inclusive, and context-sensitive strategies.

First and foremost, human-centered design must guide AI development and implementation. This involves not only technical robustness, but also a commitment to ensuring that AI systems enhance, rather than replace, human judgment. In sectors such as healthcare and education, for instance, algorithms should augment professional expertise rather than dictate decisions. As argued by Jotterand and Bosco (2021), the uncritical integration of AI in high-stakes environments can transform tools of assistance into instruments of dehumanization unless guided by strong ethical oversight.

Second, organizational leaders should institutionalize anticipatory governance mechanisms. These include scenario planning, regular ethical audits, and stakeholder consultation forums. As Stilgoe, Owen, and Macnaghten (2013) have emphasized, responsible innovation requires reflexivity, anticipation, and inclusivity—particularly in the face of uncertain sociotechnical futures. By embedding such practices within strategic management processes, organizations can mitigate risks while enhancing resilience.

Third, the development of critical digital literacy across all hierarchical levels is crucial. Training programs must move beyond basic digital skills to include education on algorithmic bias, data privacy, and the ethical implications of automation. Employees should be equipped not just to use AI tools, but to question and challenge them where necessary. As noted by Ananny and Crawford (2018), transparency in AI is not merely about opening the black box, but also about cultivating a culture of critical inquiry and institutional accountability.

Finally, academic institutions and research bodies must embrace their role in shaping an ethically competent workforce. Curricula should incorporate interdisciplinary approaches to AI—blending technical, philosophical, legal, and organizational perspectives. In this context, the European Skills Agenda for Sustainable Competitiveness, Social Fairness and Resilience (European Commission, 2020) emphasizes the strategic need for lifelong learning and inclusive upskilling pathways, particularly in response to the growing digitalization of work environments. Beyond technical competencies, the EU also highlights the importance of transversal skills—including critical thinking, communication, and adaptability—as outlined in the Key Competences for Lifelong Learning framework (European Commission, 2019). These frameworks underscore the imperative to integrate ethical reflection and human-centered values into AI governance, ensuring that technological advancement is accompanied by social sustainability and professional empowerment across all sectors. Long-term partnerships

between academia, industry, and policymakers can ensure that the next generation of professionals is not only technologically fluent, but also socially responsible and resilient in the face of automation.

Taken together, these recommendations form a blueprint for organizations seeking to navigate the AI-security nexus with foresight and responsibility. In an era where code increasingly governs human activity, maintaining a commitment to democratic values, professional dignity, and institutional transparency becomes not a luxury, but a strategic necessity.

6. Conclusion

As artificial intelligence reshapes the architecture of organizational life, the most pressing challenges are not purely technological, but profoundly human. This paper has examined how AI-driven transformations carry with them invisible threats—subtle shifts that erode professional autonomy, obscure decision-making processes, and diminish trust in institutional systems. These risks, while less visible than traditional cybersecurity breaches, are no less consequential.

Through the lens of human resource management and broader socio-organizational dynamics, the analysis underscores the need for vigilance in how AI is deployed and governed. Responsible AI integration demands more than regulatory compliance—it calls for moral imagination, inclusive dialogue, and a deliberate effort to align technological systems with human values.

Future scenarios outlined in this study suggest that organizations are at a crossroads. One path leads to inclusive, ethically grounded innovation that supports human agency, equity, and long-term resilience. The other risks deepening asymmetries of power, normalizing depersonalized governance, and undermining the legitimacy of institutions.

To move forward wisely, strategic recommendations emphasize anticipatory governance, critical digital literacy, and the vital role of education in shaping reflective practitioners. In doing so, we reclaim space for human judgment, empathy, and ethical reasoning in an increasingly automated world.

Ultimately, the question is not whether AI will transform our organizations—but whether we will allow it to do so without reflection. By looking back at the invisible threats already emerging, we gain the clarity needed to move forward with intelligence, responsibility, and care.

Literature

1. Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
2. Armstrong, M. (2024). *Armstrong's handbook of strategic human resource management* (8th ed.). London: Kogan Page.
3. Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
4. European Commission. (2019). *Key competences for lifelong learning*. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1>
5. European Commission. (2020). *European Skills Agenda for sustainable competitiveness, social fairness and resilience*. <https://ec.europa.eu/social/main.jsp?catId=1223&langId=en>
6. Forman, D. M., & Wang, L. (2025, June 18). *Federal court grants preliminary certification in landmark AI hiring bias case*. CDF Labor Law LLP. Retrieved from <https://www.callaborlaw.com/2025/06/18/federal-court-grants-preliminary-certification-in-landmark-ai-hiring-bias-case/>
7. Guest, D. E. (2017). Human resource management and employee well-being: Towards a new analytic framework. *Human Resource Management Journal*, 27(1), 22–38. <https://doi.org/10.1111/1748-8583.12139>
8. Jotterand, F., & Bosco, C. (2021). Artificial intelligence in medicine: A sword of Damocles? *Journal of Medical Systems*, 46(1), Article 9. <https://doi.org/10.1007/s10916-021-01796-7>
9. Jovanović, T. (2025). *Holističko upravljanje ljudskim resursima u uslovima Četvrtе industrijske revolucije*. Doktorska disertacija, Univerzitet "Union – Nikola Tesla", Beograd.

10. Kaminski, M. E. (2021). Algorithmic impact assessments under the GDPR. *International Data Privacy Law*, 11(2), 125–144. <https://doi.org/10.1093/idpl/ipaa020>
11. Kudina, O., & van de Poel, I. (2024). A sociotechnical system perspective on AI. *Minds and Machines*, 34(3), Article 21. <https://doi.org/10.1007/s11023-024-09680-2>
12. Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.
13. Salwei, M. E., Stevens, K. R., & Pasupathy, K. S. (2022). A sociotechnical systems approach to artificial intelligence integration in health care delivery: A framework for research and design. *Journal of the American Medical Informatics Association*, 29(7), 1204–1213. <https://doi.org/10.1093/jamia/ocac043>
14. Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. *Harvard Journal of Law & Technology*, 35, 117–191. UCLA School of Law, Public Law Research Paper No. 21-25.
15. Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
16. Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 1–17. <https://doi.org/10.1177/2053951717743530>