***mgr Anna Nastuła***
*Wojskowa Akademia Techniczna w Warszawie*
*anna.nastula@arkadiapolska.com*
*ORCID: https://orcid.org/0000-0002-2061-5067*

# ZAGROŻENIA CYBERNETYCZNE I RYZYKA ZWIĄZANE ZE SZTUCZNĄ INTELIGENCJĄ W ZARZĄDZANIU ZASOBAMI LUDZKIMI

# CYBER THREATS AND AI RISKS IN HUMAN RESOURCES MANAGEMENT

**Streszczenie**

W wyniku cyfrowej transformacji zarządzania zasobami ludzkimi, upowszechnienia pracy zdalnej oraz wdrażania zaawansowanych systemów opartych na sztucznej inteligencji, działy zarządzania zasobami ludzkimi stają się newralgicznym obszarem podatności oraz celem ataków cybernetycznych. W artykule dokonano analizy typologii zagrożeń ukierunkowanych na procesy realizowane przez działy zarządzania zasobami ludzkimi, w tym. phishing, oszustwa rekrutacyjne z zastosowaniem technologii deepfake, zagrożenia wewnętrzne oraz działania dezinformacyjne. Przedstawiono również ramy prawne i etyczne, ze szczególnym uwzględnieniem ogólnego rozporządzenia o ochronie danych osobowych (RODO) oraz AI Act, które nakładają na podmioty obowiązki w zakresie zapewnienia bezpieczeństwa, przejrzystości oraz nadzoru ludzkiego nad systemami zautomatyzowanymi. Celem pracy była identyfikacja i systematyzacja zagrożeń występujących w obszarze zarządzania zasobami ludzkimi, w związku z wykorzystaniem nowych technologii oraz sztucznej inteligencji. W rezultacie opracowano rekomendacje działań o charakterze organizacyjnym

*Słowa kluczowe: cyberbezpieczeństwo, zarządzanie zasobami ludzkimi, zagrożenia w cyberprzestrzeni, sztuczna inteligencja*

**Summary**

As a result of the digital transformation of human resource management, the widespread adoption of remote work, and the implementation of advanced artificial intelligence systems, human resource departments have become a critical point of organizational vulnerability and a target for cyberattacks. This article presents a typological analysis of threats targeting human resource functions, including phishing, recruitment fraud using deepfake technology, insider threats, and disinformation activities. It also outlines the legal and ethical framework, with particular emphasis on the General Data Protection Regulation and AI Act, which impose obligations on entities to ensure security, transparency, and human oversight over automated systems. The study aimed to identify and systematise the threats occurring in the field of human resources management in connection with the use of new technologies and artificial intelligence. Based on the conducted analysis, recommendations for organizational action have been formulated.

*Keywords: cybersecurity, human resource management, cyber threats, artificial intelligence.*

## RESEARCH PROBLEM IDENTIFICATION AND METHODOLOGY

The ongoing digitization of human resource management processes and the growing use of artificial intelligence (AI) tools are leading to significant changes in the structure and functioning of modern entities.The operating environment of enterprises is characterized by dynamic digital transformation, accelerated by global phenomena, including the COVID-19 pandemic. This has led to the widespread adoption of remote and hybrid working models and the implementation of SaaS (Software as a Service) cloud solutions for HR processes.

Digital HR solutions, such as Human Resource Information Systems (HRIS), algorithm-based recruitment platforms, and predictive analytics tools, increase the operational efficiency of HR departments, but at the same time introduce new threats in the areas of cybersecurity, personal data protection and decision-making ethics. Employee data processing takes place in distributed, external and difficult to fully control technological environments.

The importance of the issue under analysis stems from the special role that HR departments play in the organizational structure, collecting, processing and securing, among other things, personal data, information on salaries, periodic evaluations and the health status of employees. The scope of these responsibilities makes them one of the main targets of attacks aimed at sensitive information within an entity. At the same time, the nature of their tasks, which include intensive communication with candidates and staff, processing large volumes of documentation and the need to make decisions within a limited time frame, makes employees of these organizational units vulnerable to social engineering techniques. The ongoing automation using artificial intelligence algorithms in the areas of recruitment, selection and performance management generates a new category of risks. These include not only potential information security breaches, but also algorithmic errors, algorithmic discrimination and a lack of transparency in automated decision-making processes.[1]

The main objective of this paper is to analyze and systematize threats at the intersection of cybersecurity, human resource management and artificial intelligence. Based on the assumption that the digitization and automation of HR processes lead to a significant increase in the attack surface and the entity's exposure to risks, the following research questions were formulated:

1. How does the digitisation of HR processes affect an entity's exposure to cyber threats, and what types of threats are specific to this functional area?

2. What risks to entities arise from the use of advanced technologies, in particular artificial intelligence?

As part of the work, a preliminary literature review was conducted, covering peer-reviewed publications in the fields of human resource management, cybersecurity, new technologies and law. The materials obtained were subjected to critical analysis using theoretical methods such as analysis, synthesis, generalization, and abstraction. The use of analysis made it possible to identify the relevant components of the research problem, while synthesis allowed them to be logically linked into a coherent whole. Generalization provided the basis for formulating general conclusions resulting from observations of recurring phe-

1 Z. Chen, *Ethics and discrimination in artificial intelligence-enabled recruitment practices*, Humanit Soc Sci Commun 10, 567 (2023). https://doi.org/10.1057/s41599-023-02079-x, (accessed 19/06/2025), pp. 6 et seq.

nomena related to the digitization of HR functions and the accompanying threats. The use of abstraction, in turn, allowed for the elimination of features irrelevant to the subject of analysis and the isolation of the constitutive properties of the phenomenon under study, which enabled the theoretical modelling of the relationships between technology and human resource management processes.

## HUMAN RESOURCE DIGITALIZATION AND CYBER RISK EXPOSURE

The last decade has been characterized by a significant acceleration in the digital transformation of human resource management processes. Modern technologies enable the automation of HR tasks, streamline the analysis of large data sets, and allow for the personalization of digital solutions both at the level of the entire entity and in relation to individual employees. The evolution from local databases to integrated cloud-based HR Tech systems, such as HRIS systems and Applicant Tracking Systems (ATS), was a significant turning point in this process.

Modern HRIS platforms integrate payroll data, personnel files, holiday management, periodic assessment results and career development information into a single environment.[2] At the same time, ATS systems have enabled the automation of recruitment processes, and communication platforms such as Microsoft Teams have become an integral part of everyday collaboration, recording huge amounts of formal and informal employee interactions.[3] In addition, the introduction of qualified electronic signatures has revolutionized the circulation of HR documents, transferring sensitive processes entirely to the digital sphere.

An important aspect of modern tools used in HR is their ability to interoperate with various IT systems operating within an entity, resulting in the creation of multi-layered, distributed information processing environments. This type of architecture leads to an expansion of the potential attack surface, understood as the total set of access points through which an unauthorized entity may attempt to gain access to the system or data.[4] It includes not only the technical infrastructure (servers, SaaS applications, employee end devices) but also self-service interfaces, external recruitment modules, electronic signature tools, internal communication channels, and the human factor in the form of HR department employees and candidates, as well as the processes in which they participate. Each of these elements can be a potential vector for data exfiltration or unauthorized modification.

Human resource departments are organizational units that interact extensively with the external environment while also having access to large volumes of personal data. This configuration results in dual exposure: both as direct targets of cyberattacks and as potential, unintentional vectors of organizational vulnerability. Firstly, HR departments should be seen as an area requiring special protection due to the nature of the data processed. Candidate databases contain information such as contact details, employment history and education, which can be used for identity theft or personalized spear phishing attacks. Data on periodic evaluations, corrective actions, and disciplinary proceedings fall into the category of sensitive information, the unauthorized disclosure of which can lead to serious legal, financial,

2 M. Armstrong, *A Handbook of Human Resource Management Practice*, 13th ed., Kogan Page, London 2014, pp. 525-528.
3 A. Branowska, *Proces doboru pracowników w przedsiębiorstwach – przegląd nowoczesnych i tradycyjnych metod selekcji, Organizacja i Zarządzanie*, no 83, 2021, pp. 11-12, https://doi.org/10.21008/j.0239-9415.2021.083.01, (accessed: 20.06.2025).
4 National Institute of Standards and Technology, *Glossary: Attack Surface*, https://csrc.nist.gov/glossary/term/attack_surface (accessed: 20.06.2025).

and reputational consequences for the entity. On the other hand, data on sick leave contains information about health status, which is subject to special protection under applicable law.

On the other hand, HR processes and employees also constitute a vector of vulnerability. In everyday practice, recruiters receive messages from unknown senders (usually candidates) and open attachments sent in response to job advertisements, which may pose a risk of malware infection. HR and payroll departments that handle financial data are particularly vulnerable to Business Email Compromise (BEC) attacks, in which cybercriminals impersonate members of management to extort funds or alter employee bank details.[5]

Another factor exploited by cyber attackers is the specific characteristics of HR departments, such as openness to communication and a high level of responsiveness in their interactions with the outside world. These characteristics can be deliberately exploited in social engineering activities aimed at obtaining information or gaining unauthorized access. As a result, a department whose role is to protect the interests of employees, and the entity may inadvertently contribute to a breach of information security.

In view of the previously discussed interrelations, the process of digitizing human resource management cannot be analyzed solely from the perspective of operational efficiency. It is necessary to simultaneously consider systemic and organizational risks, as well as the perspective of information security, based on security-aware design principles and zero-trust architecture, adapted to the operational specifics of HR departments.

## TYPOLOGY OF CYBER THREATS TARGETING HR FUNCTIONS

Given the above-mentioned interdependencies, a comprehensive identification and classification of cyber threats affecting human resource management is required. The complexity of the technological environment in which HR processes are carried out, combined with the wide range of processed information and intensive interaction with the external environment, results in increased vulnerability to attacks. An analysis of the literature on the subject allows us to identify the basic categories of threats that are particularly relevant to the area of human resource management.

Social engineering, understood as a set of manipulative techniques aimed at persuading the user to take actions leading to the disclosure of confidential information, lowering the level of security or launching malicious code, plays an important role in the context of cyber threats to HR functions.[6]

One of its basic forms is phishing, i.e. sending electronic messages containing malicious content or hyperlinks aimed at obtaining authentication data, installing malicious software or redirecting the user to a fake website.[7] Examples of phishing attack scenarios in the context of HR departments include emails impersonating: job candidates (containing infected CV files), government institutions (e.g. Social Security, tax offices), external HR service providers (e.g. insurance platforms) and members of the board or management with a request to make a confidential transfer or authorize a transaction (so-called whaling or CEO fraud).[8]

5 Federal Bureau of Investigation (FBI), *2022 Internet Crime Report, Internet Crime Complaint Center (IC3)*, Washington D.C. 2023, pp. 10-11.

6 K. Mitnick, W. Simon, *The Art of Deception: Controlling the Human Element of Security*, Wiley, Indianapolis 2002, pp. 16 et seq.

7 C. Banasiński, M. Rojszczak, *Cyberbezpieczeństwo*, Wolters Kluwer, Warszawa 2020, pp.123-126.

8 Warsaw Enterprise Institute, *Odporni na cyberataki. Poziom bezpieczeństwa cyfrowego Polski*, Warszawa: Warsaw Enterprise Institute, 2023, https://wei.org.pl/wp-content/uploads/2023/11/Odporni-na-cyberataki.-Poziom-bezpieczenstwa-cyfrowego-Polski.pdf (accessed: 20.06.2025).

Variations of phishing include smishing and vishing, which use SMS messages and phone calls, respectively, to trick the victim into clicking on a fake link or revealing sensitive data. Spear phishing, on the other hand, relies on carefully crafted, personalized messages, often in the form of fabricated application documents or recruitment enquiries. Such activities are particularly intense during the recruitment and onboarding phases, when the number of interactions with candidates and external entities increases and communication control becomes difficult. Due to the nature of their tasks, HR employees are particularly vulnerable to targeted social engineering attacks. Social engineering attacks require a comprehensive approach to prevention, training and control of access to critical information resources. The effectiveness of social engineering attacks is increasing despite the dissemination of knowledge in this area and the systematic raising of employee awareness. Even people who have been trained multiple times can remain susceptible to manipulation, and some of them regularly repeat the same mistakes. Periods of increased interaction with the external environment, such as recruitment and onboarding, are particularly critical for HR departments. At the same time, there has been an increase in the effectiveness of the collective phishing detection model, which enables the identification of new campaigns in real time with little operational overhead. This model, supported by a transparent incident reporting policy and intuitive reporting tools, can be an important addition to security procedures in human resource management.[9]

Insider threats are also an important category, as they are among the most difficult to detect and the most destructive forms of security incidents. They refer to situations in which the perpetrator of the incident is a person with authorized access to the organization's resources, most often a current or former employee, contractor, business partner or supplier. These types of threats include both intentional and unintentional actions that lead to a breach of confidentiality, integrity or availability of organizational systems and data.[10] Intentional threats include actions aimed at intellectual property theft, infrastructure sabotage, financial fraud and industrial espionage. The perpetrators' motivations may be personal, economic or ideological, and their actions often involve exceeding access rights to gain benefits or cause damage to the entity. Such individuals have detailed knowledge of internal procedures, IT infrastructure and security systems, making it difficult to detect their actions at an early stage.[11]

The second group includes unintentional threats resulting from behavior caused by carelessness, insufficient training or lack of awareness of the potential consequences of actions taken. Examples of such incidents include accidentally sharing data, clicking on a link leading to a phishing website, using unsecured private devices, or unknowingly downloading malware.[12]

Both statistics and case studies highlight the issue of internal threats. According to the 2025 Verizon Data Breach Investigations Report (DBIR), compiled periodically by Verizon based on international reports of information security breaches, insider threats account for approximately 29% of all breaches.[13] Of these, 19% were due to unintentional employee

9 D. Lain, *Phishing in Organizations: Findings from a Large-Scale and Long-Term Study*, IEEE Symposium on Security and Privacy, 2022, https://doi.org/10.1109/SP46214.2022.9833766 (accessed: 21.06.2025).

10 CERT Insider Threat Team, *Unintentional Insider Threats: A Foundational Study*, Carnegie Mellon University, Pittsburgh 2013, pp. 2-6.

11 G. Mazzarolo, A.D. *Jurcut, Insider Threats in Cyber Security: The Enemy Within the Gates*, University College Dublin 2020, pp. 3-6.

12 M. Kont, M. Pihelgas, J. Wojtkowiak, L. Trinberg, A.M. Osula, *Insider Threat Detection Study*, NATO CCDCOE, Tallin 2015, pp. 16-19.

13 Verizon, *Data Breach Investigations Report 2025*, Verizon Enterprise, 2025, https://www.verizon.com/dbir/, (accessed: 20.06.2025)

errors and 8% were due to deliberate misuse of access privileges. In addition, the report indicates that as many as 22% of all security breaches involved credential abuse. Although not all cases of this type are classified as insider threats, as some may result from the acquisition of credentials by third parties, many remain directly linked to the improper actions of individuals with authorized access.

In the HR environment, internal threats occur in particular at three stages of the employee life cycle: during hiring, evaluation and offboarding. In the first case, the risk includes granting excessive privileges to new hires without conducting an adequate level of verification, including behavioral risk analysis and professional background checks[5]. Actions intended to manipulate evaluation outcomes or obtain undue advantage through abuse of formal position constitute a significant threat during the evaluation and promotion process.

At the stage of termination of employment, it is crucial to effectively manage the revocation of access, the recovery of equipment and the securing of information resources to reduce the risk of sabotage or unauthorized transfer of information outside the organization's systems.

Mitigating the risks associated with internal threats requires the use of integrated and multi-layered information security management systems. The literature on the subject points to the need to combine administrative, technical and organizational measures, including the implementation of access controls based on the principle of minimum privileges, the use of behavioural monitoring using UEBA (User and Entity Behavior Analytics), and the implementation of training to raise employee awareness of cybersecurity.[14] The involvement of the HR department in interdisciplinary risk management teams also plays an important role, enabling earlier detection of irregularities and the implementation of preventive and remedial mechanisms at every stage of the employee's life cycle.[15]

The development of generative artificial intelligence technology has led to the emergence of a new category of cyber threats in the employee recruitment phase, related to the use of false digital identities. A particular case in point is recruitment fraud using deepfake technology, based on artificial intelligence algorithms, which enables the generation of realistic but manipulated audiovisual material used to impersonate fictitious candidates or other individuals involved in the recruitment process.[16]

Deepfake poses a serious threat to the integrity of HR processes, especially in the context of remote recruitment, where the lack of physical contact with the candidate makes classic verification methods impossible. Technology enables realistic reproduction of facial expressions, eye movements, gestures and synchronization of sound with lip movements, which makes it very difficult to identify the forgery without the use of specialized detection tools.[17] In such a scenario, a person using a false identity is introduced into the organization and once employed, becomes a potential internal threat, having access to HR systems and information.

The most common attack vectors are recruitment platforms, online interview services, job portals and social media, where fake professional profiles, fabricated CVs and project

14 O. Marwan, *Insider Threats: Detecting and Controlling Malicious Insiders*, [w:] New Threats and Countermeasures in Digital Crime and Cyber Terrorism, IGI Global, Hershey PA 2015, pp. 162-170.

15 Cybersecurity and Infrastructure Security Agency (CISA), *Insider Threat Mitigation Guide*, U.S. Department of Homeland Security, Washington D.C. 2020, pp. 12-19.

16 Y. Hu, J. Huang., M. Zhao., *Deepfake Identity Fraud in Recruitment: Risks and Detection*, Springer, Singapore 2024, pp. 3-4.

17 E. Hydara, M. Kikuchi, T. Ozono, *Empirical Assessment of Deepfake Detection: Advancing Judicial Evidence Verification through Artificial Intelligence*, IEEE Access, IEEE, 2024, https://ieeexplore.ieee.org/iel8/6287639/6514899/10716657.pdf, (accessed: 21.06.2025).

portfolios are disseminated. To reduce the risk, it is necessary to implement mechanisms for additional verification of candidates' identities, including document validation using interactive systems, monitoring of video interviews (with the candidate's consent) and analysis of image consistency in different shots and contexts.

Organizations are not fully prepared to counteract these types of threats, and the procedures used to select and verify candidate data often prove insufficient.[18] An additional problem is the lack of clear legal regulations regarding the use of deepfake technology in the workplace. In most jurisdictions, existing regulations do not consider the specific nature of this type of fraud, which makes it difficult to enforce legal liability against perpetrators.

Given the growing scale of the threats, literature on the subject points to the need for a hybrid approach combining technological, regulatory and educational solutions. It is recommended to implement AI-based detection tools, such as the extended EfficientNet architectures, which enable the detection of anomalies characteristic of algorithmically generated materials.[19] At the same time, systemic preventive measures are important, including improving the digital skills of HR department employees, developing procedures for responding to deepfake incidents, and building organizational resilience to digital fraud throughout the employee lifecycle.

A separate category of threats is disinformation activities targeting the reputation of the entity, in which human resource management plays an important role. The most common include fake job advertisements and fabricated communications purportedly originating from the HR department. The purpose of such activities may be to undermine the credibility of the employer, manipulate public opinion or obtain candidates' personal data.[20]

Another form of attack is fabricated internal communications, disseminated via social media or internal information channels. These may concern, for example, alleged mass redundancies or changes in remuneration policy, leading to escalating anxiety, disruption to the entity's functioning, and a decline in trust in official sources of information. In extreme cases, these activities are used to manipulate the market value of listed companies, as part of a broader information operation.[21] Low levels of information literacy among HR staff increase vulnerability to manipulation and encourage the creation of closed communication loops, which hinders effective response to threats.[22]

## ARTIFICIAL INTELLIGENCE IN HRM: OPPORTUNITIES AND SECURITY RISKS

The use of artificial intelligence in human resource management is one of the most important areas of development in this field. AI systems contribute to increased operational efficiency and shorter decision-making times based on data analysis (data-driven decision making). Artificial intelligence-based solutions can be implemented at every stage of an employee's life cycle. In the recruitment process, algorithms perform automatic pre-selection and scoring of application documents, while chatbots conduct preliminary interviews with

18 M.P.P. Misra, *What's Next for AI: Future Expectations and Predictions, National Institute of Communication Finance*, 2025, https://nicf.gov.in/wp-content/uploads/2025/02/Whats-Next-for-AI-Future-Expectations-and-Predictions.pdf, (accessed: 21.06.2025).

19 W. Hu, G. Wang, Z. Sun and M. Cai, *EfficientNetB4-ES: An Enhanced EfficientNetB4 Model for Deepfake Detection*, 2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS), Yanji, China, 2024,https://ieeexplore.ieee.org/abstract/document/10800337/, (accessed: 21.06.2025)

20 M. Bardin, *Disinformation Threats to Human Capital Management*, Springer, Cham 2025, pp. 27-32.

21 T. Renault, *Market Manipulation and Suspicious Stock Recommendations on social media*, 2023, s.2 l nast., https://arxiv.org/abs/2310.02303 (accessed: 20.06. 2025).

22 D. Bawden, L. Robinson, *The dark side of information: Overload, anxiety and other paradoxes and pathologies*, „Journal of Information Science", vol. 35, no. 2, 2009, pp. 180-191.

candidates. Talent management uses systems that analyze video recordings for body language and tone of voice, as well as predictive models that identify the risk of key employees leaving. Employment planning is supported by forecasting algorithms that enable the entity of the staffing structure according to anticipated organizational needs.

At the same time, the use of artificial intelligence in HR generates new risks, which can be classified into two basic categories: inherent risks resulting from the very nature of the technology, and risks associated with its deliberate misuse.

One of the fundamental problems is the risk of algorithmic bias, resulting from the fact that learning systems are trained on historical data. If this data reflects existing biases in an entity or society regarding, for example, gender, age or ethnic origin, artificial intelligence models may not only reproduce these biases but also intensify them. As a result, the system may automatically lower the ratings of candidates belonging to certain demographic groups, leading to systemic discrimination in recruitment and HR processes.[23]

The limited transparency of decisions made by artificial intelligence models poses a significant challenge in human resource management due to the difficulty of interpreting the mechanisms even for their creators. In this context, legal and ethical issues arise from the inability to justify the reasons for rejecting a candidate or lowering an employee's performance rating, which may violate the principle of procedural fairness and the right to obtain explanations regarding decisions based on automated data processing.[24] At the same time, the limited explainability of models makes it difficult to conduct internal audits and identify the sources of possible HR errors.[25]

Artificial intelligence is becoming a tool that intensifies cybercrime, supporting attack vectors such as generating personalized phishing messages, creating synthetic identities and constructing fake recruitment applications. A new attack vector in the form of so-called prompt injection is also gaining in importance.[26] When the HR department uses chatbots based on large language models (LLMs) to interact with candidates, there is a risk that an attacker could enter a specially crafted query that would violate the model's rules of operation and execute an unauthorised command, including the disclosure of confidential configuration data or information about other candidates.

The implementation of artificial intelligence in human resource management processes reveals a significant conflict between the pursuit of increased efficiency and compliance with ethical standards. The automation of personnel decisions without ensuring adequate control, audit and transparency mechanisms can lead to a loss of trust and a reduction in employee agency. This problem is referred to in the literature as automated HR decision-making in conditions of lack of transparency (black box HR decision-making).[27]

## LEGAL AND ETHICAL CONSIDERATIONS

The introduction of advanced digital technologies, artificial intelligence algorithms and HR process automation poses significant ethical and legal challenges. Entities are required

23 S. O'Neil, *Broń matematycznej zagłady*, Wydawnictwo Naukowe PWN 2017, pp. 152-170.

24 B. Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, „Big Data & Society", t. 3, no 2, 2016, pp. 6-8.

25 Z. Chen, *Ethics and discrimination in artificial intelligence-enabled recruitment practices*. Humanit Soc Sci Commun 10, 567 (2023). https://doi.org/10.1057/s41599-023-02079-x, [dostęp 19.06.2025], pp. 6-7.

26 OWASP Gen AI Security Project, LLM01:2025 Prompt Injection, OWASP, 2025, https://genai.owasp.org/llmrisk/llm01-prompt-injection/, (accessed: 20.06.2025).

27 Z. Chen, *Ethics and discrimination in artificial intelligence-enabled recruitment practices*. Humanit Soc Sci Commun 10, 567 (2023), pp. 6 et seq https://doi.org/10.1057/s41599-023-02079-x, (accessed: 19.06.2025).

to adapt their activities to applicable regulations aimed at protecting fundamental human rights, in particular the right to privacy and personal data protection.

The use of artificial intelligence systems in HR involves the processing of large amounts of personal data, which creates a risk of violating data protection regulations. According to Article 5 of the General Data Protection Regulation (GDPR), processing must be carried out in accordance with the principles of lawfulness, fairness, transparency, minimization, purpose limitation, integrity and confidentiality.[28] This means that HRIS and ATS systems should only collect information necessary to achieve specific purposes, and access to it must be appropriately restricted. In practice, these principles are not always properly implemented in the context of algorithmic candidate selection. Inadequate data security measures can lead to privacy violations and discrimination if algorithms learn from historical biases.

In the context of the use of artificial intelligence in HR processes, restrictions on automated decision-making that may have legal consequences or significantly affect the situation of data subjects are of particular importance.[29] An example of this is the automatic rejection of a candidate by an ATS system without human intervention. The Regulation allows for exceptions to this rule, e.g. when the decision is necessary for the conclusion or performance of a contract but imposes an obligation on the data controller to ensure at least the right to human intervention, to present one's own position and to challenge the decision.[30] Additional requirements for the transparency of decisions made by automated systems are provided for in Article 10 of Directive (EU) 2024/2831 on improving working conditions through digital platforms.[31] This provision obliges employers to use algorithm-based management systems to provide employees and their representatives with access to information on the logic, significance and anticipated consequences of the functioning of these systems. Importantly, the directive extends the information obligation to cases where automated decisions affect employment conditions, task allocation, remuneration or termination of the contract.

The AI Act plays a significant role in regulating the use of artificial intelligence in HR, classifying systems used for human resource management as high risk.[32] This applies to tools supporting the automation of recruitment processes, performance evaluation, career path planning, decisions on promotion or termination of employment, as well as monitoring employee behaviour and performance. Classification as a high-risk system entails a set of stringent obligations for both providers and users, including the implementation and maintenance of a risk management system, the use of high-quality training data to reduce the risk of discrimination, the preparation of detailed technical documentation, the provision of adequate human oversight to mitigate potential harm, and the demonstration of sufficient accuracy, robustness, and cybersecurity.[33]

28 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, pp. 1-88.

29 Art. 22 of the GDPR, see in detail: P. Litwiński, P. Barta, M. Kawecki, *Rozporządzenie w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i swobodnym przepływem takich danych. Komentarz*, C. H. Beck, Warsaw 2018, pp. 430-438; and E. Niezgódka, *Profilowanie a cyberbezpieczeństwo*, in: G. Szpor, A. Gryszczyńska (eds.), Internet. Strategie bezpieczeństwa C. H. Beck, Warsaw 2017, pp. 243-249.

30 Article 29 Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP251rev.01, 6 February 2018, pp. 18-24.

31 Directive (EU) 2024/2831 of the European Parliament and of the Council of 11 April 2024 on improving working conditions in platform work, OJ L 2024/2831, 30.4.2024, pp. 1-32.

32 European Parliament and Council of the European Union, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), OJ L 2024/1689, 12.7.2024, Annex III, point 4.

33 Ibidem, Title III, Chapter 2.

Before implementing artificial intelligence systems in HR processes, it is necessary to carry out a data protection impact assessment in accordance with Article 35 of the GDPR and to ensure that algorithm-based decision-making mechanisms remain explainable (explicability) and that the data subjects are adequately informed about this. These requirements imply the need for close cooperation between HR, IT and legal departments to ensure compliance with applicable regulations. Already at the stage of selecting AI-based technology, it is necessary to implement AI management policies, regular algorithmic audits and training in regulatory compliance and the ethics of automated decision-making processes.

Another issue is the impact of algorithmizing on human rights, including the right to work, equal treatment and privacy protection. Some HR systems may systematically marginalize people with disabilities or of a certain ethnic origin.[34] In this regard, there are calls for collective data management and the strengthening of regulatory control mechanisms at EU and national level.

To summarize this part of the analysis, the implementation of AI-based technologies in HR requires not only compliance with applicable law, but also ethical reflection and practical procedures to ensure fairness, transparency and data protection in personnel decisions.

## ORGANIZATIONAL RECOMMENDATIONS

In response to identified cyber threats, risks associated with the use of artificial intelligence, and regulatory challenges in HR digitalization, it is reasonable to formulate recommendations that can serve as a starting point for in-depth research analyses. From an organizational perspective, it is reasonable to develop integrated procedural, technological and structural mechanisms aimed at strengthening systemic resilience to the risks generated by automation and the digital transformation of HR processes.

It is essential to implement the principle of security by design, which is the foundation for the implementation of HRIS systems, recruitment platforms and AI-based tools. This means that data protection mechanisms must be integrated at the design stage. Of particular importance here are onboarding and offboarding procedures, which should limit the risk of unauthorized access and other internal threats.

The effectiveness of technical security measures is closely dependent on the level of user competence, which is why it is important to systematically develop the knowledge of HR staff in the field of cybersecurity and the ethical use of technology. Training programs should be tailored to the specific tasks performed in this area, considering realistic scenarios of social engineering attacks and the functioning of artificial intelligence-based systems. These activities require close cooperation with IT departments and units responsible for regulatory compliance.

Another recommendation is to improve access policies and strengthen oversight of data lifecycle management and employee permissions. Regular access audits, an effectively implemented offboarding process and a GDPR-compliant data retention policy should be key instruments in limiting the risk of escalation of privileged access and unauthorized data processing.

In the area of implementing AI-based solutions in human resource management, formalizing responsibility for the functioning of these systems is of particular importance.

34 M. Capasso, P. Arora, D. Sharma, C. Tacconi, *On the Right to Work in the Age of Artificial Intelligence: Ethical Safeguards in Algorithmic Human Resource Management*, Business and Human Rights Journal, vol. 9, no. 3, 2024, pp. 348-359, doi:10.1017/bhj.2024.26 (accessed: 23.06.2025).

The establishment of internal supervisory structures, such as AI Governance Committees, responsible for assessing the compliance of implementations with applicable regulations and for the ongoing monitoring of decision-making algorithms, is an important element in ensuring the compliance, transparency and ethics of automated HR processes. The task of such a team would be to develop an internal AI policy, evaluate and approve new tools, supervise compliance with the AI Act, and monitor the operation of implemented systems. An important element of this management model remains human oversight mechanisms for automated decision-making and regular algorithmic audits aimed at identifying systemic errors and biases.

The integrated implementation of the presented solutions is not only an instrument for increasing organizational resilience, but also a starting point for further research on standards for responsible technology management in human resources.

## Conclusion

The objective of the conducted analysis was to identify and systematize threats arising at the intersection of human resource digitalization, cybersecurity, and the application of artificial intelligence. It was assumed that the widespread adoption of automated solutions in the area of human resource management significantly increases the attack surface and organizational exposure to systemic and information-related risks. The analysis clearly indicates that human resource departments, as entities processing some of the most sensitive categories of data, are not only prime targets of cybercrime but also constitute a major vector of organizational vulnerability, particularly in the context of advanced social engineering techniques supported by artificial intelligence algorithms.

These risks go beyond the traditionally understood threats to data confidentiality, integrity and availability, and include legal and ethical consequences such as algorithmic discrimination, lack of transparency in decision-making processes and violations of employee rights.

Effective mitigation of these risks requires the adoption of an integrated, interdisciplinary management model. Constant cooperation between HR, IT and the units responsible for legal compliance and internal audit is essential. The effectiveness of the solutions implemented is closely related to the level of organizational maturity and the entity's ability to translate normative guidelines into specific operational management mechanisms, such as policies, procedures, training models and technology monitoring systems.

## BIBLIOGRAPHY

**Literature**

Armstrong M., *A Handbook of Human Resource Management Practice*, 13th ed., Kogan Page, London 2014.

Banasiński C., Rojszczak M., *Cyberbezpieczeństwo*, Wolters Kluwer, Warszawa 2020.

Bardin M., *Disinformation Threats to Human Capital Management*, Springer, Cham 2025.

Bawden D., Robinson L., *The dark side of information: Overload, anxiety and other paradoxes and pathologies*, Journal of Information Science, vol. 35, nr 2, 2009.

Branowska A., *Proces doboru pracowników w przedsiębiorstwach – przegląd nowoczesnych i tradycyjnych metod selekcji*, Zeszyty Naukowe Politechniki Poznańskiej. Organizacja i Zarządzanie, nr 83/2021, Wydawnictwo Politechniki Poznańskiej, Poznań 2021.

Capasso M., Arora P., Sharma D., Tacconi C., *On the Right to Work in the Age of Artificial Intelligence: Ethical Safeguards in Algorithmic Human Resource Management*, Business and Human Rights Journal, vol. 9, nr 3, Cambridge University Press, Cambridge 2024.

Chen Z., *Ethics and discrimination in artificial intelligence enabled recruitment practices*, Humanities and Social Sciences Communications, vol. 10, Palgrave Macmillan / Springer Nature, 2023.

Hu Y., Huang J., Zhao M., *Deepfake Identity Fraud in Recruitment: Risks and Detection*, Springer, Singapur 2024.

Hydara E., Kikuchi M., Ozono T., *Empirical Assessment of Deepfake Detection: Advancing Judicial Evidence Verification through Artificial Intelligence*, IEEE Access, IEEE, 2024.

Litwiński P., Barta P., Kawecki M., *Regulation on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data – Commentary*, C. H. Beck, Warszawa 2018.

Marwan O., *Insider Threats: Detecting and Controlling Malicious Insiders, w: New Threats and Countermeasures in Digital Crime and Cyber Terrorism*, IGI Global, Hershey PA 2015.

Mazzarolo G., Jurcut A.D., *Insider Threats in Cyber Security: The Enemy Within the Gates*, University College Dublin 2020.

Mittelstadt B. et al., *The ethics of algorithms: Mapping the debate*, Big Data & Society, vol. 3, nr 2, 2016.

Mitnick K., Simon W., *The Art of Deception: Controlling the Human Element of Security*, Wiley, Indianapolis 2002.

Niezgódka E., *Profilowanie a cyberbezpieczeństwo, w: Szpor G., Gryszczyńska A. (red.), Internet. Strategie bezpieczeństwa*, C. H. Beck, Warszawa 2017.

O'Neil S., *Broń matematycznej zagłady*, PWN, Warszawa 2017.

**Legal sources**

Article 29 Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, WP251rev.01, 6 February 2018.

Directive (EU) 2024/2831 of the European Parliament and of the Council of 11 April 2024 on improving working conditions in platform work, OJ L 2024/2831, 30.4.2024.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), OJ L 2024/1689, 12.7.2024.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ

L 119, 4.5.2016.

**Internet sources**

CERT Insider Threat Team, *Unintentional Insider Threats: A Foundational Study*, Carnegie Mellon University (Software Engineering Institute), Technical Note CMU/SEI 2013 TN 022, 2013, https://doi.org/10.1184/R1/6585575.v1.

Cybersecurity and Infrastructure Security Agency (CISA), *Insider Threat Mitigation Guide*, Cybersecurity and Infrastructure Security Agency, Washington D.C. 2020, https://www.cisa.gov/sites/default/files/2022-11/Insider%20Threat%20Mitigation%20Guide_Final_508.pdf.

Kont M., Pihelgas M., Wojtkowiak J., Trinberg L., Osula A. M., *Insider Threat Detection Study*, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn 2015, https://ccdcoe.org/uploads/2018/10/Insider_Threat_Study_CCDCOE.pdf

Lain D., *Phishing in Organizations: Findings from a Large-Scale and Long-Term Study*, IEEE Symposium on Security and Privacy, IEEE, 2022, https://doi.org/10.1109/SP46214.2022.9833766

Misra M.P.P., *What's Next for AI: Future Expectations and Predictions*, National Institute of Communication Finance, New Delhi 2025, https://nicf.gov.in/wp-content/uploads/2025/02/Whats-Next-for-AI-Future-Expectations-and-Predictions.pdf

National Institute of Standards and Technology, *Attack Surface*, w: NIST Glossary, Gaithersburg 2025, https://csrc.nist.gov/glossary/term/attack_surface

OWASP, LLM01:2025 *Prompt Injection*, w: OWASP GenAI Security Project, OWASP, 2025, https://genai.owasp.org/llmrisk/llm01-prompt-injection

Renault T., *Market Manipulation and Suspicious Stock Recommendations on Social Media*, arXiv, 2023, https://arxiv.org/abs/2310.02303

Verizon, *Data Breach Investigations Report* 2025, https://www.verizon.com/dbir/

Warsaw Enterprise Institute, *Odporni na cyberataki. Poziom bezpieczeństwa cyfrowego Polski*, Warsaw Enterprise Institute, Warszawa 2023, https://wei.org.pl/wp-content/uploads/2023/11/Odporni-na-cyberataki.-Poziom-bezpieczenstwa-cyfrowego-Polski.pdf