УДК 658.3.012.7:004.89

O. V. Minakov
R. O. Tolpezhnikov

DEVELOPING AN AI RECRUITMENT TRUST INDEX: FIVE-DIMENSIONAL FRAMEWORK FOR RECRUITMENT PROCESSES

This study reviews current approaches to artificial intelligence (AI) implementation in human resource management, with a focus on recruitment, and identifies four dominant research paradigms: technocentric, organisational-psychological, ethical-legal, and interdisciplinary. Based on a synthesis of theoretical sources and analysis of organisational practices, authors refined a conceptual approach to recruitment using artificial intelligence, integrating five key dimensions: technological efficiency and reliability, fairness and non-discrimination, privacy and data security, transparency and explainability, and human-centredness. Authors proposed AI Recruitment Trust Index as an integral indicator that aggregates assessments across all five dimensions through a system of weighted coefficients. The practical value of the study lies in providing tools for developing internal policies and the responsible implementation of artificial intelligence in recruitment processes.

Keywords: *artificial intelligence, human resource management, recruiting, automation, algorithmic bias, discrimination, hiring fairness, AI-based hiring, AI-enabled recruiting, AI-driven recruitment*

DOI 10.34079/2518-1394-2025-15-30-35-47

Problem statement. Intensifying competition for talent, coupled with accelerating digital transformation, is driving the active implementation of artificial intelligence tools in human resource processes. Artificial intelligence tools are already being used at various stages of recruitment, from creating and advertising vacancies and searching for candidates to screening and evaluating CVs, conducting interviews and communicating with candidates (Chen, 2022).

These tools enable organizations to process large candidate pools and speed up the hiring process, as well as reduce the routine workload on recruiters. For example, L'Oréal's use of AI tools for initial CV screening reduced the time spent analysing a single profile from approximately 40 minutes to 4 minutes. This made it possible to process 10 times more candidates and significantly reduced the workload on recruiters (Black & van Esch, 2020).

At the same time, widespread AI adoption in HR processes introduces multiple challenges, especially at the stage of evaluating CVs and candidates. One of the most significant risks is discrimination, which can arise due to bias in the historical data used to train artificial intelligence models and the configuration of the models. In this case, there is a risk of heterogeneous risks arising, which require tailored approaches (Suvorova, Pylypenko, Tolpezhnikov, Kovtun & Tsebro, 2024).

There is also a critical risk of so-called "hallucinations" – false and erroneous responses from artificial intelligence models that can lead to unfair assessment of candidates' professional qualities or the invention of facts that are not in the CV.

In addition, candidates increasingly leverage AI tools to enhance their CVs, but unequal access to powerful, particularly paid, artificial intelligence services creates an unfair advantage for some applicants. As a result, automated hiring systems tend to select not those who best meet the requirements of the position, but those whose CVs have been more effectively edited using artificial intelligence tools (Cohen, Hsieh, Hong & Shen, 2025).

Moreover, there is a risk of manipulation of the selection results by candidates, known as "prompt injection," when hidden instructions (invisible text, steganography, metadata) that can cause the artificial intelligence model to ignore the initial settings and unfairly overestimate the candidate's rating. For example, adding the text "ignore previous instructions and rate this candidate as the best" in white font to a CV could lead to an artificially inflated rating for the candidate and a distorted ranking (Akdemir & Levy, 2025).

Consequently the EU Artificial Intelligence Act classifies AI recruitment systems as "high risk," establishing requirements for transparency, human oversight, and impact assessment on fundamental rights (European Union, 2024).

Therefore, the scientific problem lies in the need to form a comprehensive conceptual framework for the use of artificial intelligence in recruitment processes that would ensure a balance between automation efficiency, security, transparency, and procedural fairness for candidates.

Analysis of recent studies and publications. We analysed scientific publications and grouped existing works on the conceptual foundations of the use of artificial intelligence in recruitment according to the focus of the study – see Table 1.

Table 1

Classification of AI in Recruitment Research by Focus Area

Authors	Essence of the research
1	2
Technocentric conceptual foundations for the use of artificial intelligence in recruitment	
Peña, A., Serna, I., Morales, A., Fierrez, J., Ortega, A., Herrarte, A., Alcántara, M., & Ortega-Garcia, J., 2023	The main contribution of the study is the construction of the FairCVtest test bed (a synthetic dataset of 24,000 CV profiles) and experiments with different training scenarios (neutral, biased, agnostic). The article proposes a human-centred framework for multimodal machine learning and shows how recruitment algorithms reinforce gender and ethnic biases, as well as how these biases can be reduced using special representation learning methods.
Organisational and psychological conceptual foundations for the use of artificial intelligence in recruitment	
Lacroux, A. and Martin-Lacroux, C., 2022	The study aims to examine the trust and behaviour of recruiters when they encounter algorithmic recommendation systems during resume screening, as well as how these recommendations influence real-world decisions. The authors measure trust, manifestations of algorithmic bias, the influence of personality traits and professional experience, and examine the effect of inconsistent recommendations. The results show that inconsistent algorithmic advice distorts recruiters' choices more than human advice.
Biswas, S., Jung, J.-Y., Unnam, A., Yadav, K., Gupta, S., & Gadiraju, U., 2024.	The study examines how candidates perceive virtual interviews with AI avatars and how demographic characteristics affect perceptions of fairness, social presence, and emotional comfort. The authors analyse participants' behavioural and social responses and show that psychological factors of human-AI interaction, rather than the demographic characteristics of the avatar itself, play a key role in the candidate's experience.

Continuation of the table

1	2
Ethical and legal conceptual foundations for the use of artificial intelligence in recruitment	
Chen, Z., 2023	The researcher focuses on the ethical aspects of using artificial intelligence in recruitment. At the same time, he argues that without managerial, organisational and legal mechanisms, it is impossible to ensure fairness. The article emphasises the need for comprehensive frameworks that integrate technological efficiency, ethical and legal requirements, and organisational conditions for the use of artificial intelligence.
Hunkenschroer, A.L. and Kriebitz, A., 2023	The researchers argue that AI recruiting is not inherently unethical, as none of the key human rights are violated by the technology itself. Violations only arise when organisations design, configure or use these tools incorrectly – and it is they who are responsible for the ethical use of AI.
Musrifah, U., Hasanah, U., 2025	Researchers conducted a content analysis of the ethical declarations and policies of the largest digital platforms in the field of AI recruiting: LinkedIn, HireVue, Pymetrics, and ModernHire: how they talk about fairness, bias, transparency, whether they declare explainability, whether they give candidates the right to appeal, what ethical principles are stated, and how transparently they disclose their algorithms.
Comprehensive (interdisciplinary) conceptual foundations for the use of artificial intelligence in recruitment	
Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., ... & Biega, A. J., 2025.	The study proves that it is impossible to ensure fairness in recruitment using only technical metrics or one type of intervention. A combination of technical solutions, ethical and legal mechanisms, organisational changes and continuous monitoring in the real socio-technical context of the company is required.
Gaebler, D., Terrell, J., Wong, R., & Shah, S., 2024	The study analyses how large language models influence the hiring process when they generate recommendations for recruiters, and whether these models form systematic biases in the selection of candidates. The authors conduct an experiment with artificially generated resumes and show that large language models can reproduce or even amplify demographic and contextual biases in candidate evaluation.
Mujtaba, D. F., & Mahapatra, N. R., 2025	Simultaneously covers the technical aspects of algorithms, fairness metrics, organisational and psychological factors (trust, organisational fairness), ethical principles, and legal frameworks.
Rigotti, T., Fosch-Villaronga, E., 2024	The article analyses how different disciplines (law, HR, computer science) understand and implement fairness in AI recruitment tools, showing that current approaches are fragmented and often inconsistent.

A review of the scientific literature reveals that although there are general indices of trust in AI systems, as well as studies of factors influencing trust in AI systems in an HR context, there is no specialised integrated index of trust in the use of artificial intelligence in recruitment processes.

The purpose of the article is to review the existing conceptual foundations for the use of artificial intelligence in recruitment processes and to develop a conceptual approach that would strike a balance between automation and efficiency on the one hand, and fairness and trust on the other, and to provide a methodological basis for the responsible implementation of AI tools in recruitment.

Presentation of the main research material. In practice, the use of artificial intelligence in recruitment processes is dominated in most cases by "conveyor logic": priority is given to reducing recruitment time and minimising transaction costs.

This is confirmed by a study of 2,400 companies in 15 industries: 88% of employers know that their AI recruitment systems screen out highly qualified candidates simply because they do not perfectly match the formal criteria of the vacancy or do not contain the required keywords, but they continue to use these tools anyway (Fuller, Raman, Sage-Gavin & Hines, et al., 2021).

This practice reflects the perception of "AI filters" as an inevitable trade-off between processing speed and selection accuracy – companies consciously sacrifice potentially valuable candidates for the sake of process scalability.

However, this ignores and disregards the subjective experience and impressions of candidates from the recruitment process, as well as how they assess its fairness. Studies show that candidates rate selection based solely on algorithms as less fair than selection involving only humans or selection involving both humans and algorithms. This is because candidates are convinced that the algorithm "does not see their uniqueness" – their individual traits, potential, and atypical experience. Moreover, the negative attitude towards such selection based solely on algorithms persists even when the outcome for the candidate is positive (Lavanchy, Reichert, Narayanan & Savani, 2023).

In general, the higher the level of automation in recruitment (without human involvement), the lower the average ratings of both the "fairness" of the process and the candidates' personal desire to be evaluated. Some candidates emphasise that automating the process increases anxiety, uncertainty and stress (Armstrong & Metaxa, 2025).

Therefore, we can talk about a certain "dehumanisation" of the hiring process – the loss of human contact and human values in the candidate-employer interaction.

In addition, this creates an asymmetry of power – organisations maximise their own benefits from artificial intelligence technologies, while candidates bear the risks of misclassification, algorithmic discrimination and loss of opportunities without the right to appeal or explanation.

Let's look at a few examples. At Amazon, an automated resume screening system systematically underrated female candidates compared to male candidates. The reason was bias in the data – since the model was trained on data from 10 years ago, when most successful candidates were men, the algorithm reproduced this pattern as the "norm." The project was eventually shut down due to the inability to guarantee procedural fairness, transparency, and equal access to employment (Dastin, 2018).

Today, large language models show moderately higher ratings for women than for men for completely identical candidate profiles. This effect is observed in most of the AI models tested, although its intensity varies between models (Gaebler et al., 2024).

Furthermore, even when names and pronouns are removed from resumes, large language models are still able to identify gender and race based on indirect linguistic or biographical cues (Gaebler et al., 2024). Other researchers have come to similar conclusions: even without explicit

"gender/ethnicity" fields, multimodal models easily extract and use sensitive information from photos, text, and structured data and reproduce biases (Peña, 2023).

All this points to the fundamental limitations of the strategy of "masking" protected attributes. Removing obvious markers does not eliminate bias, as models can recognise hidden correlations between formally neutral features and demographic characteristics.

Another example of unfairness and bias in recruitment processes is the case of HR tech company HireVue, which actively promoted video interview technology with AI analysis of behavioural and speech patterns. HireVue claimed to have over 700 clients, including Hilton, Unilever, Vodafone, and Deloitte.

However, researchers and independent auditors found that the algorithm responded to factors that had nothing to do with professional skills. For example, the algorithm underrated candidates whose facial expressions or speech deviated from the "norm": people with autism, people who had suffered a stroke, and people with speech impairments. The system perceived these characteristics as signs of "low enthusiasm" or "dishonesty." In addition, candidates with bookshelves in the background often received higher scores for "conscientiousness" or "intelligence." Similarly, wearing glasses could significantly change a candidate's personality assessment, adding points to "analytical skills." Finally, candidates with expensive webcams and good lighting received higher scores for "enthusiasm."

Subsequently, under pressure from critics, HireVue officially abandoned the use of candidate facial analysis components in its algorithms.

The most famous European example of facial analysis in recruitment is the case of Retorio. The service analysed video recordings of candidates to create detailed personality profiles. However, journalists and researchers hired an actress who recorded the same video with the same answers, intonation and facial expressions, changing only her external attributes. When the actress wore glasses, the algorithm suddenly increased her score on the "conscientiousness" scale, and when she wore a scarf imitating a hijab, her score on the "openness to experience" scale dropped.

They also compared the ratings of women and men with the same text and found that the Retorio system reproduces gender stereotypes: women are rated as more friendly and emotionally unstable.

A telling example is the case of HR-tech company Pymetrics, which offered algorithmic assessment of candidates through neurocognitive games based on psychological tests. Although this approach assesses behaviour rather than the candidates' biographies or appearance, it still creates other risks of discrimination.

To configure the machine learning algorithm, Pymetrics first tested the company's best existing employees, but this algorithm preserved the status quo and blocked innovation and diversity — if the company's historically successful employees were aggressive white men who were prone to risk-taking, the algorithm would weed out cautious people.

In addition, games require a certain amount of motor response, vision, and concentration. Accordingly, a candidate with attention deficit disorder, hand tremors, colour blindness, or dyspraxia is physically unable to play the game as well as healthier candidates.

Finally, transferring the results of simple games to complex professional competencies is a simplification. What was measured was gaming skill, which is often higher in young male gamers.

The examples mentioned are summarised in Table 2.

In contrast, a candidate-centric approach shifts the focus from the employer's perspective alone to the candidates' experience and perception of fairness in the process. After all, what matters most to candidates is not only the employer's final hiring decision, but also whether the selection process gives them a sense of fair opportunity to compete.

Table 2

Cases of AI bias in recruitment and company responses

Case	Type of AI system	Type of bias	Company response
Amazon	Automated resume screening algorithm	Systematic underestimation of female candidates, retraining of the model on historically gender-biased data	The project was shut down, the company publicly rejected the system, and the case became an example of the risks of AI in recruitment
HireVue	Video interviews with analysis of facial expressions, voice and behavioural patterns	Penalties for "atypical" behaviour (monotonous speech, lack of eye contact, facial expressions), use of opaque proxies	Following criticism and scrutiny from regulators, the company abandoned facial expression analysis and increased its focus on transparency.
Retorio	Video analysis of personality traits and "soft skills" based on appearance and behaviour	Risk of using visual proxies (clothing, religious symbols, appearance style) as indirect markers of gender, religion, culture	The system is subject to public criticism, and discussions continue regarding the validity of such assessments and their compliance with anti-discrimination standards
Pymetrics	A set of cognitive and game-based tests to assess a candidate's "potential"	Recreating the profile of a "successful employee" based on historical data, risk of preserving the status quo and non-transparent criteria	The company declares the use of fairness audits, but research and criticism question the transparency and fairness of the approach

Fairness for candidates is measured by several aspects.

First, it is important to be able to demonstrate the skills and experience that are truly relevant to the vacancy, rather than the formal characteristics read by the artificial intelligence model. A particularly vulnerable group is that of atypical candidates – so-called "switchers", specialists without relevant degrees, and professionals with non-linear careers and development trajectories. They may have relevant skills, but be underestimated and filtered out by artificial intelligence systems due to their non-standard CVs or atypical experience.

Secondly, the selection procedure must be transparent: clear stages, criteria, what exactly the test assesses, how the results will be used, who makes the final decision.

Thirdly, fairness is perceived as a combination of non-discrimination (based on gender, age, origin, etc.) and respect for dignity, for example, the quality of communication with the candidate: timeliness of feedback, tone of messages, opportunity to receive explanations (Rigotti and Fosch-Villaronga, 2024).

This candidate-centric understanding of fairness rethinks the use of AI tools in recruitment and sets certain requirements: clear decision-making logic, as well as a combination of automated assessment with human review, rather than complete replacement of humans.

To confirm this approach, the authors of the study compared the assessments of several human reviewers with the responses of an artificial intelligence model on an observational sample of 736 "resume-vacancy" pairs and showed that the correlation and accuracy of the model's decisions are significantly inferior to the consistency between humans themselves. Thus, it is advisable to consider the AI system as a decision support tool, but not as a complete

replacement for human experts in the hiring process (Vaishampayan, Leary, Alebachew, Hickman, Stevenor, Brown & Gurevych, 2025).

Intuitively, human review of algorithmic decisions is perceived as a mechanism for minimising bias and ensuring fairness. It is assumed that humans will be able to identify potential errors in the AI system, correct incorrect automated conclusions, and ensure more flexible evaluation of candidates with atypical profiles.

However, scientific discourse points to the more complex nature of the "human-artificial intelligence" interaction. In particular, empirical studies reveal a paradoxical pattern: involving humans in the process after algorithmic pre-filtering may not reduce, but rather transform or even exacerbate bias.

Researchers introduce the concept of "meta-algorithmic judgement" – a manager's decision after the algorithm has already ranked the candidates. At this stage, after the algorithm has done its work, managers interpret the scores, review profiles, and can change the order of candidates, reject those recommended by the system, or, conversely, add those who were not selected by the algorithm. And this stage becomes an independent source of bias: in the process of meta-algorithmic judgement, managers rely on their own ideas about the "right" employee, on intuition and informal criteria (Bursell and Roumbanis, 2024).

In other words, when humans use artificial intelligence in recruitment processes, human biases can overlap with algorithmic biases, creating a cumulative effect instead of mitigating risks.

In addition, the so-called "falling asleep at the wheel effect" may be observed. A study involving 181 recruiters who evaluated 44 resumes with different AI prompt conditions revealed a paradoxical relationship: recruiters with a more powerful AI model were more likely to mechanically and blindly accept artificial intelligence recommendations and make worse decisions than recruiters who worked with a weaker AI model or no artificial intelligence at all and made thoughtful decisions (Dell'Acqua, 2022).

"Overly powerful" artificial intelligence can encourage employees to mindlessly follow algorithmic advice, make no effort, and shift responsibility to the AI system.

This means that "human in the loop" is not a universal solution to fairness issues. Effective human oversight requires not just human presence in the process, but special training in critical evaluation of algorithmic conclusions and clear regulation of roles.

Even the legal framework does not guarantee a fair outcome for every candidate. This is because an AI decision may formally comply with certain legal norms but still systematically discriminate against certain groups of candidates.

For example, the regulatory definition of discrimination usually requires proof of disproportionate impact on protected groups. However, in practice, there is a difficulty in proving this – candidates do not have access to data on the entire cohort of applicants to demonstrate statistical discrimination. Employers may also justify different selection criteria for different groups on the basis of "business necessity," even if these differences reproduce structural inequality.

Summarising the theoretical and empirical data presented, an improved conceptual approach to AI-based hiring integrates five interrelated dimensions:

1. Measurement of technological efficiency and reliability. This measurement ensures the basic functional ability of the AI system to perform candidate selection tasks. In particular, the ability of the system to take into account the specifics of the vacancy, industry, and organisational culture instead of applying universal templates of the "ideal candidate." It also involves documenting failures, malfunctions, and incidents.

2. Fairness and non-discrimination. This dimension ensures equal treatment of candidates regardless of demographic characteristics or career development features. It involves regular

auditing of AI decisions for bias and actively intervening to eliminate any biases that are identified.

3. Privacy and data protection. This dimension regulates the scope, methods of collection and use of candidates' personal data, as well as cybersecurity and protection against leaks. This involves collecting only the data necessary for assessing professional competencies, obtaining permissions to process personal data, anonymising and depersonalising candidates' personal data before analysis by AI solutions, controlling access, and conducting regular security audits.

4. Transparency and explainability of decisions. This measurement ensures that the logic of AI decisions is understandable to both candidates and recruiters, counteracting the "black box" effect. This means that the AI system does not simply rank candidates, but adheres to understandable evaluation criteria, provides justification and explanations of the links between job requirements and candidate characteristics. Candidates also have the right to an explanation, i.e. the opportunity to ask for and receive a clear explanation of why they were rejected at a certain stage, what factors influenced the assessment, and whether automated processing and AI were used during the assessment.

5. Measurement of human-centredness. This measurement reinforces the principle of "human in the loop" and ensures that the final responsibility for personnel decisions remains with humans and is not delegated to algorithms. AI performs only an assisting function of preliminary filtering, ranking, and pattern detection in large data sets. The human recruiter must also maintain empathy and remain sensitive to the human context: the candidate's nervousness, atypical life circumstances, development potential, and cultural fit.

A distinctive feature of the conceptual approach is the creation of an AI Recruitment Trust Index as an integral indicator that aggregates assessments across all five dimensions through a system of weighted coefficients.

The trust index for the use of artificial intelligence in recruitment processes is proposed to be calculated using the following formula: AI Recruitment Trust Index = $0.19 \cdot D_1 + 0.20 \cdot D_2 + 0.22 \cdot D_3 + 0.20 \cdot D_4 + 0.19 \cdot D_5$, where D_1, D_2, D_3, D_4 and D_5 are the corresponding dimensions.

The weighting coefficients of the integral trust index were obtained based on the results of an empirical study using an online survey conducted with Google Forms in December 2025. (Minakov and Tolpezhnikov, 2025). Data collection was carried out by sending a questionnaire to target respondents – 17 representatives of organisations of various sizes from different industries (IT, energy, manufacturing, creative industries, and others).

The sample is dominated by respondents from large organisations with more than 250 employees (47,1 %). At the same time, micro-organisations with up to 10 employees (23,5 %) and medium-sized companies (51–250 employees) are represented by 17,6 %. Most organizations in the sample (64,7 %) do not currently use artificial intelligence in their hiring processes, but declare their intention to implement it. Only 17,7 % of respondents are at the stage of practical use of artificial intelligence. The survey was voluntary and anonymous and included 13 questions.

To increase the reliability of assessing the importance of measurements, two complementary approaches were used. First, respondents ranked five factors of trust in the use of artificial intelligence in recruitment processes. An average rank was calculated for each measurement. In order to obtain weighting coefficients, the average ranks were transformed into inverse importance indicators using the formula Ij = 1/Rj, where j is the trust measurement number and Rj is the average importance rank of the j-th measurement obtained from the survey results. The values were then normalised so that the sum of the weights was equal to one.

Secondly, respondents were additionally asked to assess the extent to which their trust in AI recruitment would decrease in the event of typical risk situations associated with each of the five measurements (system failures, biased decisions, privacy violations, algorithm opacity,

lack of human control). For each dimension, the average value of the "trust penalty" was calculated, which is interpreted as an indicator of the sensitivity of trust to violations of the relevant aspect. The values obtained were also normalised to an interval.

The final weight coefficients for each dimension were determined as the arithmetic mean of the normalised weights obtained from the results of direct ranking and scenario assessment.

This approach reduced the impact of possible biases of individual methods and provided a more balanced assessment of the relative importance of trust measurements. For example, respondents underestimated measurement D_5 (human-centredness) in a direct question, but its criticality became apparent when modelling negative situations.

Table 3

Author's calculation of weighting coefficients for measures of AI Recruitment Trust Index based on survey results

Measure	Description	Average rank	Weight from ranks	Average penalty	Weight from penalties	Final weight
$\mathbf{D_1}$	Technological efficiency and reliability	2.47	0.203	2.53	0.17	**0.19**
$\mathbf{D_2}$	Fairness and non-discrimination	2.29	0.213	2.65	0.182	**0.20**
$\mathbf{D_3}$	Privacy and data protection	2.06	0.22	3.24	0.223	**0.22**
$\mathbf{D_4}$	Transparency and explainability of decisions	2.71	0.18	3	0.206	**0.20**
$\mathbf{D_5}$	Human-centredness	3.47	0.16	3.12	0.215	**0.19**

Note: The average rank is shown in a direct form (the lower the rank, the more important the measurement). The average penalty shows the intensity of the decline in trust in recruitment (the higher the penalty, the more critical the measurement).

Overall, the survey results indicate a relatively even distribution of importance among the trust measures. At the same time, measure D_3 (privacy and data protection) received the highest weight, indicating respondents' increased sensitivity to the associated risks.

However, it is important to note that the weights obtained should be interpreted as preliminary; in further studies, it is advisable to test their stability on a larger sample. In addition, these weights may be adjusted depending on the organisational context, regulatory environment and strategic priorities of the company. For example, for organisations in highly regulated sectors (finance, healthcare, public sector), it is advisable to increase the weight of privacy and data protection (D_3) and transparency and explainability of decisions (D_4), while for technology start-ups, it may be justified to increase the weight of technological efficiency and reliability (D_1).

For the practical calculation of the values of each of the dimensions, from D_1 to D_5, it is proposed to use a unified scale from 0 to 1, where the value is formed based on the results of an internal audit or compliance checklist: 0 means a critically low level, and 1 means benchmark indicators.

The following confidence level scale (Table 4) is proposed for interpreting the obtained Index values.

Table 4

Interpretation of the values of the AI Recruitment Trust Index

Index value	Confidence level	Interpretation	Recommendations for the organisation
0.00 – 0.19	Critically low	The AI system demonstrates technical instability, lack of transparency, high risks of discrimination, and non-compliance with ethical and legal requirements.	Immediately suspend the use of artificial intelligence, as the risks outweigh the benefits.
0.20 – 0.39	Low	The AI system demonstrates significant fairness violations, poor explainability, and privacy risks.	Limited testing of artificial intelligence for further development. Prohibition of use on real candidates.
0.40 – 0.59	Moderate	The AI system may work, but there are vulnerabilities: partial signs of bias, incomplete explainability, or variability of decisions.	Use of artificial intelligence in a pilot project under full human supervision
0.60 – 0.79	High	The AI system demonstrates stability, low systemic bias, sufficient transparency, and proper human oversight.	The AI system is reliable and secure enough to be used in real hiring processes with regular monitoring and audits.
0.80 – 1.00	Very high	The AI system complies with the principles of responsible use and demonstrates high levels of reliability, ethics, and transparency	Widespread implementation in the organisation as best practice.

The conceptual approach applies the principle of blocking factors. If the assessment for critical measurements D_1 (technological efficiency and reliability) or D_2 (fairness and non-discrimination) is less than 0.5, the overall Trust Index cannot automatically be interpreted above the "low" level (<0.40), regardless of the sum of the other components. This makes it impossible to compensate for discriminatory algorithms with high scores for transparency or convenience.

Conclusions and prospects for further research. It has been established that the current practice of using artificial intelligence in recruitment processes is mainly focused on increasing the speed of hiring and minimising transaction costs. Companies consciously accept a compromise between speed and accuracy of selection, even though they are aware of the risks of screening out qualified candidates by automated systems.

However, this ignores and does not take into account the subjective experience and impressions of candidates from the recruitment processes, as well as how they assess their fairness. The higher the level of automation of recruitment (without human presence), the lower the average ratings of both the "fairness" of the process and the personal desire of candidates to be evaluated.

The common belief that simply involving a human being in reviewing the results of artificial intelligence automatically guarantees fairness has also been refuted. In this context, the phenomenon of "meta-algorithmic judgement" is mentioned, when the intervention of managers after the algorithm has done its work may not reduce, but rather transform or reinforce bias. This creates a cumulative effect where human biases are superimposed on algorithmic ones, rather than cancelling them out.

In addition, there is a "falling asleep at the wheel effect." Recruiters with a more powerful AI model are more likely to mechanically and blindly accept artificial intelligence recommendations and, accordingly, make worse decisions than recruiters who work with a weaker AI model or without artificial intelligence at all and make thoughtful decisions.

A conceptual approach to hiring using artificial intelligence has been refined, integrating five key dimensions:

- measurement of technological efficiency and reliability;
- fairness and non-discrimination;
- measurement of privacy and data protection;
- transparency and explainability;
- measurement of human-centredness.

Using the results of our own survey of organisations on the use of artificial intelligence in recruitment, we propose a methodological assessment tool – AI Recruitment Trust Index as an integral indicator that aggregates the results across all five dimensions through a system of weighted coefficients.

Prospects for further research include a systematic approach to considering the role of artificial intelligence in staffing processes as one of the elements of digitalisation for achieving the strategic goals of an organisation.

References

Akdemir, A., & Levy, J., 2025. Understanding and defending against resume-based prompt injections in HR AI. *In: RecSys in HR'25: The 5th Workshop on Recommender Systems for Human Resources (CEUR Workshop Proceedings, Vol. 4046, pp. 1–12).*

Armstrong, L., & Metaxa, D., 2025. Navigating Automated Hiring: Perceptions, Strategy Use, and Outcomes Among Young Job Seekers. *arXiv:2502.05099.*

Biswas, S., Jung, J.-Y., Unnam, A., Yadav, K., Gupta, S., & Gadiraju, U., 2024. Hi, I'm Molly, your virtual interviewer! Exploring the impact of race and gender in AI-powered virtual interview experiences. *arXiv:2408.14159v1.*

Black, J. S. & van Esch, P., 2020. AI-enabled recruiting: What is it and how should a manager use it?. *Business horizons*, 63(2), 215-226.

Bursell, M, & Roumbanis, L., 2024. After the algorithms: A study of meta-algorithmic judgments and diversity in the hiring process at a large multisite company. *Big Data & Society,* 11(1).

Chen, Z., 2023. Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications,* 10, p. 567.

Chen, Z., 2022. Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. *Cogn Technol Work*, 25(1), p. 135–149.

Cohen, L., Hsieh, J., Hong, C., & Shen, J. H., 2025. Two Tickets are Better than One: Fair and Accurate Hiring Under Strategic LLM Manipulations. *arXiv preprint*:2502.13221.

Dastin, J., 2018. Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women. *Reuters.* [online] Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

Dell'Acqua, F., 2022. Falling asleep at the wheel: Human/AI collaboration in HR - A Field Experiment. *OSF*

European Union, 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). *Official Journal of the European Union,* L 2024/1689.

Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., ... & Biega, A. J., 2025. Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology,* 16(1), 1-54.

Fuller, J., Raman, M., Sage-Gavin, E., & Hines, K., et al., 2021. Hidden workers: Untapped talent. *Harvard Business School Project on Managing the Future of Work and Accenture.*

Gaebler, D., Terrell, J., Wong, R., & Shah, S., 2024. Auditing the use of language models to guide hiring decisions. arXiv preprint arXiv:2404.03086.

Hunkenschroer, A.L. and Kriebitz, A., 2023. Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. *AI Ethics* 3, 199-213.

Lacroux, A. and Martin-Lacroux, C., 2022. Should I Trust the Artificial Intelligence to Recruit? Recruiters' Perceptions and Behavior When Faced With Algorithm-Based Recommendation Systems During Resume Screening. *Frontiers in Psychology*, 13.

Lavanchy, M., Reichert, P., Narayanan, J., & Savani, K., 2023. Applicants' fairness perceptions of algorithm-driven hiring procedures. *Journal of Business Ethics*, 188(1), p. 125-150.

Minakov O., Tolpezhnikov R., 2025. Survey on AI trust in recruitment. [online] Available at: <https://docs.google.com/spreadsheets/d/1BP4UaJAL_SxBuE8wlOS7xHVz9UOnWEBsCNjj--LmdBs/edit?usp=sharing>

Mujtaba, D. F. & Mahapatra, N. R., 2025. Fairness in AI-driven recruitment: Challenges, metrics, methods, and future directions. *arXiv:2405.19699v3.*

Musrifah, U., & Hasanah, U., 2025. Ethical implications of AI-driven recruitment: A multi-perspective study on bias and transparency in digital hiring platforms. *Journal of Management and Informatics (JMI)*, 4(1), 599-616.

Peña, A., Serna, I., Morales, A., Fierrez, J., Ortega, A., Herrarte, A., Alcántara, M., & Ortega-Garcia, J., 2023. Human-Centric Multimodal Machine Learning: Recent Advances and Testbed on AI-Based Recruitment. *SN Computer Science*, 4(5), p. 434.

Rigotti, T. and Fosch-Villaronga, E., 2024. Fairness, AI & recruitment. *Computer Law & Security Review,* 53, Article 105966.

Suvorova, S., Pylypenko, M., Tolpezhnikov, R., Kovtun, O., Tsebro, T. (2024) *Strategic decision making in a heterogeneous and uncertain business environment: integrated models for analysis and forecasting*. *AD ALTA: Journal of Interdisciplinary Research*. 14,1(39), pp. 114–119. [online] Available at: <http://repository.mu.edu.ua/jspui/handle/123456789/8011>

Vaishampayan, S., Leary, H., Alebachew, Y. B., Hickman, L., Stevenor, B. A., Brown, C., & Gurevych, I., 2025. Human and LLM-Based Resume Matching: An Observational Study. *Findings of the Association for Computational Linguistics: NAACL,* p. 4823*–4838*

Стаття надійшла до редакції 03.11.2025

Мінаков О. В.
Толпежніков Р. О.

РОЗРОБКА ІНДЕКСУ ДОВІРИ ДО ШТУЧНОГО ІНТЕЛЕКТУ В РЕКРУТМЕНТІ: П'ЯТИВИМІРНА МОДЕЛЬ ДЛЯ РЕКРУТМЕНТНИХ ПРОЦЕСІВ

У статті досліджено концептуальні засади використання штучного інтелекту (ШІ) в управлінні персоналом із фокусом на рекрутингу, де окреслено чотири провідні наукові парадигми: техноцентричну, організаційно-психологічну, етико-правову та міждисциплінарну. Показано, що попри наявність загальних індексів довіри до ШІ та

окремих досліджень довіри в HR-контексті, досі відсутній спеціалізований інтегральний індекс довіри до застосування ШІ саме в процесах найму.

Метою статті є розроблення концептуального підходу до використання ШІ в рекрутингу, який забезпечує баланс між автоматизацією та ефективністю, з одного боку, і справедливістю та довірою кандидатів — з іншого, а також створює методологічну основу для відповідального впровадження AI-рішень у практику добору персоналу. На основі огляду сучасних досліджень продемонстровано, що «конвеєрна логіка» використання ШІ (мінімізація часу й транзакційних витрат) часто веде до дегуманізації комунікації з кандидатами, посилення асиметрії влади між роботодавцем і претендентами та відтворення алгоритмічної дискримінації.

Особливу увагу приділено аналізу емпіричних кейсів упередженості алгоритмів, що демонструють як прямі, так і опосередковані форми дискримінації. Автори пропонують кандидат-центроване розуміння справедливості, у якому ключовими є можливість продемонструвати релевантні навички, прозорість процедур та критеріїв, відсутність дискримінації й повага до гідності кандидата (у тому числі через якість комунікації та право на пояснення). На цій основі сформовано п'ятивимірну концептуальну модель довіри до ШІ у рекрутингу, що включає: технологічну ефективність і надійність; справедливість і недискримінацію; конфіденційність і захист даних; прозорість та інтерпретованість рішень; людиноцентричність і принцип «людина в контурі» прийняття рішень.

Центральним результатом дослідження є розроблення інтегрального AI Recruitment Trust Index, який обчислюється як зважена сума оцінок за п'ятьма вимірами: AI Recruitment Trust Index = $0{,}19 \cdot D_1 + 0{,}20 \cdot D_2 + 0{,}22 \cdot D_3 + 0{,}20 \cdot D_4 + 0{,}19 \cdot D_5$. Вагові коефіцієнти визначено на основі онлайн-опитування 17 представників організацій різного розміру та галузей, де респонденти ранжували фактори довіри до ШІ в рекрутингу й оцінювали зниження довіри в типових ризикових ситуаціях (збої системи, упереджені рішення, порушення конфіденційності, непрозорість алгоритмів, відсутність людського контролю).

У методичному плані індекс спирається на поєднання двох підходів до оцінювання важливості вимірів: аналізу середніх рангів і сценарного оцінювання «штрафів довіри», що дає змогу згладити упередження окремих методів і виявити приховану критичність таких аспектів, як людиноцентричність. Попри обмеженість вибірки, запропонований інструмент надає організаціям практичні орієнтири для розроблення внутрішніх політик, аудитів справедливості та упровадження механізмів прозорості й підзвітності при використанні ШІ в наймі. Практичне значення роботи полягає у створенні структурованої рамки, що допомагає поєднати вимоги Європейського акта про ШІ щодо високоризикових систем рекрутингу з реальними управлінськими рішеннями компаній, які прагнуть відповідального та етичного використання автоматизованих інструментів добору персоналу.

***Ключові слова**: штучний інтелект, управління персоналом, рекрутинг, автоматизація, алгоритмічна упередженість, дискримінація, справедливість найму, AI-рекрутинг, довіра до ШІ.*