UDC 004.056
DOI 10.56525/NGYY6754

RESEARCH ON INTERPRETABLE MACHINE LEARNING AND EXPLAINABLE AI MODELS FOR PREDICTING APPLICANTS CAREER DECISION-MAKING

Orynbassar M., Akberdiyeva M.E.
Yessenov University, Aktau, Kazakhstan
e-mail: maksym1.orynbassar@yu.edu.kz, meruyert1.akberdiyeva@yu.edu.kz

Abstract. The integration of artificial intelligence into human resource management and vocational guidance has catalyzed a paradigm shift from traditional, counselor-driven career advisement to highly complex, data-driven predictive modeling. While sophisticated machine learning architectures—ranging from ensemble decision trees to deep neural networks—demonstrate unprecedented accuracy in mapping psychometric profiles, academic records, and labor market intelligence to optimal career trajectories, their inherent structural opacity poses significant ethical, regulatory, and pedagogical challenges. The "black-box" nature of these predictive engines obscures the fundamental logic behind career recommendations and hiring rankings, thereby risking the exacerbation of historical biases, eroding user trust, and violating emerging international regulatory frameworks. This comprehensive analysis evaluates the theoretical foundations, methodological deployment, and practical implications of Explainable AI (XAI) frameworks, specifically focusing on SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), as critical mechanisms for achieving interpretable machine learning in career decision-making. By systematizing the intersection of established behavioral psychology paradigms—such as Holland's RIASEC model, the Five-Factor Model of personality, and Social Cognitive Career Theory (SCCT)—with advanced algorithmic interpretability, this report demonstrates how XAI transitions predictive models from opaque algorithmic gatekeepers into transparent, developmental tools. The empirical evidence synthesized within this review suggests that while interpretable machine learning enhances predictive fidelity and mitigates subgroup differences, its most profound value lies in fostering human-AI collaboration, auditing algorithmic fairness, and empowering applicants through transparent, data-driven self-efficacy. By shifting the focus from mere predictive accuracy to pedagogically meaningful explanation, organizations and educational institutions can ensure that AI-driven career guidance remains an equitable, reliable, and legally compliant instrument for global workforce development.

Keywords. Explainable Artificial Intelligence (XAI), Interpretable Machine Learning, Career Decision-Making, Vocational Guidance, SHAP, LIME, Psychometric Profiling, Algorithmic Fairness, Predictive Modeling, Human Resource Technology.

Introduction

The trajectory of career decision-making, applicant profiling, and recruitment has undergone a fundamental transformation, evolving from manual heuristics and standardized psychometric assessments to highly sophisticated, automated algorithmic systems. 1 Historically, career prediction relied on traditional approaches that established foundational frameworks for understanding applicant aptitude, vocational interest, and labor market fit. 2 Seminal theoretical perspectives, such as the Person-Environment (P-E) fit theories pioneered in the early twentieth century, dictated that optimal career paths were achieved by matching static individual traits to predefined environmental categories and occupational taxonomies. 3 However, traditional career guidance systems, which heavily utilize static rule-based matching, are increasingly inadequate in addressing the multi-dimensional complexities of the modern global labor market. 4 Such models typically capture broad, average relationships between vocational interests and occupations across general populations but falter significantly when attempting to adapt dynamically to rapid economic fluctuations, evolving technical skill requirements, and highly specific individual psychological nuances.

The advent of machine learning (ML) offers a potent, mathematically rigorous remedy to the limitations of static profiling. By training on massive, diverse datasets encompassing academic records, unstructured natural language resumes, and real-time labor market intelligence, machine learning models identify non-linear patterns and complex correlations that traditional statistical methods fundamentally miss. Algorithms such as eXtreme Gradient Boosting (XGBoost), Random Forests, and deep learning neural networks have demonstrated considerable, quantifiable gains in prediction precision, user alignment, and subsequent job placement outcomes. These systems currently power the core infrastructure of modern talent acquisition, facilitating automated resume screening, behavioral analysis in video interviews, and precision job matching across digital platforms. The development of these technologies has brought unprecedented opportunities to human resource management, becoming the core driving force behind the transformation of the recruitment industry and ushering in a new era of data-driven talent acquisition.

However, the rapid deployment of these advanced predictive systems has unveiled a critical, systemic vulnerability known universally as the "black-box" problem. Explainable Artificial Intelligence (XAI), recognized globally across research communities as a discipline striving to create systems capable of explaining their actions and decisions in a human-comprehensible manner, notes that as models become more complex, their internal decision-making logic becomes increasingly opaque. In high-stakes domains such as career selection, educational tracking, and corporate recruitment, this opacity is untenable. When a deep learning algorithm rejects a highly qualified candidate or recommends an unexpected career transition to a university student, the absolute inability to explain the underlying rationale undermines fundamental fairness, accountability, and user trust. Furthermore, uninterpretable models risk perpetuating or exacerbating systemic, historical biases, as they may silently overweight discriminatory proxies present in the historical training data used to build the algorithms.

Explainable Artificial Intelligence has consequently emerged as a fundamental strategic paradigm to address these severe ethical and operational concerns, effectively bridging the chasm between high-performance machine learning and necessary human interpretability. XAI encompasses a comprehensive suite of computational techniques designed to provide transparent, post-hoc explanations for complex model outputs, thereby revealing the specific weighted factors and feature interactions driving individual algorithmic predictions. In the context of career decision-making, XAI frameworks—most notably SHAP and LIME—enable predictive systems to generate actionable, interpretable feedback tailored to the specific applicant. The conceptualization of XAI has been systematized along nine critical dimensions: credibility, causality, transferability, informativeness, confidence, fairness, accessibility, interactivity, and privacy awareness. Rather than passively receiving a terminal classification or ranking, applicants, students, and human resource professionals are provided with a granular, mathematical breakdown of how specific competencies, psychometric traits, and experiential data influenced the final outcome.

This comprehensive analysis systematically examines the deployment of interpretable machine learning models and post-hoc XAI frameworks within the domain of applicant career prediction. By thoroughly evaluating the mathematical architectures of surrogate models, the integration of established psychometric theories, and the real-world ethical implications surrounding algorithmic fairness, the subsequent sections outline the mechanisms through which artificial intelligence can be optimized for both exceptional predictive accuracy and rigorous, auditable transparency.

Materials and methods. The architecture of a transparent, AI-driven career prediction system requires the seamless, highly engineered integration of behavioral psychology, high-dimensional data engineering, sophisticated machine learning algorithms, and mathematically sound explainability frameworks. The following methodologies detail the theoretical components, data modalities, and computational mechanics utilized in the development and deployment of these predictive systems.

Results and discussion. Predictive machine learning models are inherently constrained by the quality and psychological validity of the features upon which they are trained. In the context of career decision-making, predictive systems synthesize multi-modal signals, heavily leveraging established

psychometric theories to quantify human potential, behavioral tendencies, and occupational preferences.A foundational pillar of career choice modeling is Holland's theory of vocational personalities and work environments, universally recognized as the RIASEC model. This framework categorizes individual occupational interests into six distinct, theoretically ordered domains: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Traditional applications of the RIASEC framework utilized simple rule-based profile matching, calculating basic distance metrics between an applicant's highest scoring domains (high-point codes) and standardized occupational environments mapped within databases such as O*NET. In modern machine learning applications, RIASEC scores are ingested as continuous numerical features rather than discrete typological buckets. Machine learning models augment traditional profiling by effectively managing the complex, multidimensional relationships between the six interest scales and vast arrays of occupational categories, often identifying subtle permutations that human counselors and rule-based systems overlook. Researchers have successfully utilized specialized instruments, such as the 16 PGI-Mini items alongside six RIASEC scale scores, to predict occupational domains with improved accuracy by allowing the algorithm to capture sample-specific relational nuances.

To complement vocational interests, modern career prediction engines extensively integrate the Five-Factor Model of personality, colloquially known as the Big Five: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). Empirical research indicates a clear separation in predictive utility between these frameworks. While RIASEC types accurately predict the specific nature of employment and occupational alignment, Big Five traits are highly indicative of general employment status, overall academic performance, and long-term career success metrics. Large-scale machine learning evaluations highlight deep, meaningful interactions between these parallel frameworks. Extensive meta-analyses illustrate significant correlations between specific traits; for instance, high correlations exist between the Enterprising and Artistic RIASEC profiles and the Extraversion and Openness Big Five traits. Furthermore, predictive algorithms frequently ingest specialized, domain-specific psychological scales depending on the target demographic. In predicting career specialty choices among medical school graduates, models have incorporated the "Sense of Coherence Scale," the "Rosenberg-Self-Esteem-Scale," and measurements of the Dark Triad to achieve highly nuanced clinical performance predictions. Machine learning models utilize these intertwined variables to construct highly personalized, multi-faceted competency and behavioral maps.

The psychological impact of algorithmic recommendations and the design of the user interface are deeply rooted in Social Cognitive Career Theory (SCCT), originally developed by Lent, Brown, and Hackett in 1994. SCCT operates as a framework that seeks to explain the processes involved in career exploration, decision-making, and ultimate success through five interrelated models. The theory emphasizes the development of career interests and decision-making processes, which are heavily mediated by internal self-efficacy and external outcome expectations. The framework posits that successful experiences, mastery, and positive developmental feedback during career exploration are internalized as stable self-beliefs through enhanced self-efficacy. When machine learning models are deployed with explainability features, the system directly intervenes in the SCCT cognitive loop. By transforming an opaque, terminal prediction into a transparent mastery experience, interpretable AI increases the applicant's career self-efficacy, solidifies their core belief in their own capabilities, and directly mitigates the psychological phenomena of career anxiety.

The transition from descriptive psychometrics to predictive machine learning requires robust data pipelines capable of processing highly heterogeneous information. Modern AI-driven career recommendation systems integrate structured psychometric profiling with unstructured academic evidence, such as courses, projects, and certifications, utilizing Natural Language Processing (NLP) to parse resumes and semantic job-text. Furthermore, these systems ingest real-time labor market intelligence signals, including job openings, regional salary percentiles, and sector growth indices, to ensure recommendations are practically viable. Data processing typically involves cleaning and normalizing datasets, often splitting data into 80 percent training and 20 percent testing subsets to validate algorithm performance on unseen instances.

Modality Type	Data Sources	Algorithmic Processing Technique
Psychometric Structured Data	RIASEC assessments, Big Five (OCEAN) inventories, self-efficacy scales.	Ensemble Trees (XGBoost, Random Forest), Support Vector Machines (SVM).
Unstructured Academic/Career Data	Resumes, project portfolios, academic transcripts, MOOC success metrics.	Deep Learning, Natural Language Processing (NLP), Bi-LSTM, BERT.
Labor Market Intelligence (LMI)	O*NET database, job openings, salary percentiles, growth indices.	Semantic matching, multi-label classification architectures.

Ensemble methods, which aggregate the predictions of multiple foundational models to reduce variance and improve robustness, are highly prominent in applicant profiling. Models such as Random Forest and eXtreme Gradient Boosting (XGBoost) iteratively construct decision trees to minimize prediction error. In career pathway predictions based on RIASEC and academic data, XGBoost consistently outperforms conventional baselines and legacy algorithms, delivering superior hit rates and minimizing the Euclidean distances between predicted and optimal career clusters. Research involving multi-modal ensemble models, which seamlessly blend structured psychometric data with unstructured academic evidence, has demonstrated occupational membership prediction accuracy rates exceeding 83%. Despite their robustness and statistical superiority, ensemble trees operate as highly complex "black boxes," requiring external post-hoc techniques for interpretation. To parse vast amounts of unstructured applicant data—such as free-text resumes, complex project portfolios, and dynamic job market descriptions—researchers deploy deep learning architectures. Bi-directional Long Short-Term Memory (Bi-LSTM) networks and advanced Transformer-based models, particularly BERT (Bidirectional Encoder Representations from Transformers), are utilized for automated resume screening, candidate ranking, and skill extraction. Empirical analyses indicate that deep learning frameworks significantly outperform traditional machine learning baselines in complex text-matching tasks. Certain BERT-based models have achieved unprecedented classification accuracies of up to 98% in modeling career satisfaction through multivariate educational characteristics, completely eclipsing the 80-85% accuracy range of legacy logistic regression or support vector machine approaches. However, the extreme dimensionality and millions of parameters inherent in these models render them completely opaque without highly sophisticated explainability interventions. To rectify the deep opacity of ensemble and deep learning models, the field of XAI relies heavily on post-hoc explanation methods. These methods are designed not to alter the underlying complex model or sacrifice its predictive accuracy, but instead to mathematically approximate its behavior to generate human-readable explanations. Explanations generally fall into several categories: visual explanations (such as saliency maps for image data), perturbation-based explanations, knowledge-based explanations utilizing knowledge graphs, and causal explanations employing Structural Causal Models (SCM) or counterfactual reasoning. The two most dominant post-hoc frameworks in career prediction are SHAP and LIME

SHAP represents a unified, theoretically grounded framework that assigns an exact importance value to each feature for a specific prediction, based deeply on cooperative game theory. It calculates the Shapley value, a concept originating in economics, which represents the average marginal contribution of a specific feature across all possible coalitions (combinations) of features present in the dataset. The primary mathematical innovation of SHAP is the representation of the Shapley value explanation as an additive feature attribution method, which essentially functions as a localized linear model. The explanation model g is specified mathematically as:

$$g(\mathbf{z}') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j$$

In this formulation, $\mathbf{z}' \in \{0,1\}^M$ represents a binary vector indicating the presence or absence of a feature within the simplified input space, M is the total number of simplified input features, ϕ_0 is the base expected value of the model output across the entire training dataset, and ϕ_j denotes the calculated feature attribution (Shapley value) for the j-th feature. The SHAP framework is unique in providing robust mathematical guarantees regarding local accuracy, missingness, and consistency. Furthermore, the framework extends to SHAP interaction values, which allow researchers to mathematically isolate the pure interaction effect between two distinct features after accounting for their individual main effects. This generates a highly detailed $M \times M$ interaction matrix per applicant instance, revealing, for example, exactly how a candidate's high extraversion interacts specifically with their programming certifications. Because computing exact Shapley values is exponential in complexity and largely impractical for high-dimensional HR datasets, the framework utilizes optimized estimators such as TreeSHAP (specifically designed for tree-based ensemble models like XGBoost) and KernelSHAP (a model-agnostic, linear LIME-inspired approach utilizing a debiased lasso regression).

While SHAP provides comprehensive feature attribution rooted in game theory, LIME takes a distinct local surrogate approach. Rather than attempting to construct a global explainer, LIME aims to explain single, isolated predictions by training a completely separate, intrinsically interpretable model—such as a simple linear regression or a shallow decision tree—on a newly generated dataset. This new dataset is created by randomly perturbing the original input data and weighting the new samples according to their geometric proximity to the specific applicant instance of interest.The generation of a LIME explanation is defined by optimizing a specific objective function that strictly balances the local fidelity of the surrogate model with its human interpretability (complexity). The mathematical representation is defined as:

$$\xi(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

In this equation, f represents the complex black-box model (e.g., the BERT deep learning network), g is the local interpretable surrogate model selected from a constrained class of interpretable models G, and π_x is a proximity measure that delineates the extent of the local neighborhood around the applicant instance x. The term $L(f, g, \pi_x)$ represents the loss function, which captures the statistical discrepancy between the predictions of the simple surrogate g and the actual black-box model f within the defined neighborhood. Crucially, the term $\Omega(g)$ acts as a regularization penalty that measures and limits the complexity of the surrogate model. For a decision tree, $\Omega(g)$ may limit the maximum tree depth; for linear regression, it may enforce L1 regularization to strictly limit the number of features utilized. This ensures the final explanation remains comprehensible to a human HR professional or applicant, avoiding overwhelming cognitive load.

Conclusion. Artificial intelligence has a significant impact on the development of modern cyber warfare and information conflicts. The use of AI makes it possible to automate data analysis processes, accelerate cyberattacks, and improve digital defense systems.

On the one hand, artificial intelligence increases the efficiency of cybersecurity systems and helps respond quickly to threats. On the other hand, the development of intelligent technologies creates new risks associated with malware, deepfake technologies, and automated cyberattacks.

In the future, the influence of artificial intelligence on cybersecurity will continue to grow. Therefore, governments, international organizations, and IT companies need to improve information protection methods, strengthen international cooperation, and develop unified legal and ethical standards for the use of artificial intelligence.

Results and discussion. The implementation of interpretable machine learning in career decision-making yields profound implications that extend far beyond technical performance metrics.

The synthesis of empirical evidence reveals that Explainable AI acts as a critical operational necessity, impacting predictive efficacy, algorithmic fairness, user psychological well-being, and corporate regulatory compliance.

Conclusions. The transition toward highly sophisticated, data-driven career decision-making represents one of the most critical evolutions in modern talent acquisition, educational tracking, and vocational guidance. While advanced machine learning architectures—particularly multi-modal ensemble models and deep learning Transformer networks—deliver unparalleled, historically unprecedented accuracy in synthesizing complex psychometric profiles and unstructured applicant data, their deployment is structurally flawed without the explicit integration of Explainable Artificial Intelligence. The empirical and theoretical evidence comprehensively demonstrates that frameworks such as SHAP and LIME are not merely supplementary diagnostic tools utilized briefly by data scientists; they are foundational, continuous operational requirements for ensuring ethical, fair, and legally compliant algorithmic behavior in high-stakes environments.

By effectively utilizing the mathematics of cooperative game theory and local surrogate modeling, XAI fundamentally demystifies the "black box," transforming opaque predictive outputs into transparent, highly actionable feedback. In the realm of corporate recruitment, this transparency allows human resource professionals to proactively audit systems for latent historical bias, actively fostering diversity and ensuring strict adherence to rigorous international regulatory standards, such as the EU AI Act. In the vital domain of educational and vocational guidance, the psychological impact of interpretable machine learning is profound. By providing students and job seekers with clear, mathematically grounded justifications for career recommendations, XAI directly engages the cognitive mechanisms of Social Cognitive Career Theory. This actively builds career self-efficacy, aggressively reduces career anxiety, and successfully mitigates the psychological risks of automation bias and the illusion of competence.

Moving forward, the continuous evolution and scaling of career prediction systems will absolutely require deeper, more integrated synergy between behavioral science, psychometrics, and machine learning engineering. Future research and architectural development must focus heavily on enhancing the fidelity of local surrogate models against increasingly complex deep neural networks, ensuring explanations remain perfectly accurate as black-box models expand in dimensionality. Furthermore, the industry must prioritize longitudinal studies to assess the long-term, multi-year impact of AI interventions on vocational identity, while seamlessly embedding fairness constraints directly into the fundamental optimization functions of predictive models. Ultimately, the successful, ethical future of artificial intelligence in career decision-making relies entirely on human-centric architectures, ensuring that algorithms serve to empower and augment human potential through clarity, verifiable equity, and uncompromised algorithmic transparency.

REFERENCES

1. Molnar, C. (2025). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (3rd ed.).

2. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.

3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

4. Lent, R. W., & Brown, S. D. (2019). Social cognitive career theory at 25: Empirical status of the interest, choice, and performance models. Journal of Vocational Behavior, 115, 103340.

5. Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2024). Investigating machine learning's capacity to enhance the prediction of career choices. Personnel Psychology, 77(2), 295-319.

6. Wang, Y., Yang, L., Wu, J., Song, Z., & Shi, L. (2022). Mining campus big data: Prediction of career choice using interpretable machine learning method. Mathematics, 10(8), 1289.

7. Zhang, N., Wang, M., Xu, H., Koenig, N., Hickman, L., Kuruzovich, J., & Ng, V. (2024). Reducing subgroup differences in personnel selection through the application of machine learning. Personnel Psychology, 76(4), 1125-1159.

8. Nakano, S., & Liu, Y. (2025). Interpreting Temporal Shifts in Global Annual Data Using Local Surrogate Models. Mathematics, 13(4), 626.

9. Li, H., Song, J., Xue, M., Zhang, H., & Song, M. (2025). A survey of neural trees: Co-evolving neural networks and decision trees. IEEE Transactions on Neural Networks and Learning Systems, 36(7), 11718-11737.

10. Maddalena, A., & Boccuzzi, G. (2025). The application of ML in post-diploma guidance: A scoping review. Frontiers in Education.

11. Жилкишбаева, Г. С., Утебай, О. Қ., Gadirli, A. V., Қожабай, Қ. Б. (2025). Түсіндірілетін ai (xai) жүйелерді басқарудың қорғалған контурларына интеграциялау. Yessenov Science Journal, 53(4).

ИССЛЕДОВАНИЕ ИНТЕРПРЕТИРУЕМЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ И ОБЪЯСНИМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ ПРОФЕССИОНАЛЬНОГО ВЫБОРА АБИТУРИЕНТОВ

Орынбасар М. А., Ақбердиева М. Е.
Университет Есенова, Актау, Казакстан
e-mail: maksym1.orynbassar@yu.edu.kz, meruyert1.akberdiyeva@yu.edu.kz

Аннотация. Интеграция искусственного интеллекта в управление человеческими ресурсами и профориентацию стала катализатором смены парадигмы: от традиционного карьерного консультирования к высокотехнологичному предиктивному моделированию на основе данных. Хотя современные архитектуры машинного обучения — от ансамблей деревьев решений до глубоких нейронных сетей — демонстрируют беспрецедентную точность в сопоставлении психометрических профилей, академических достижений и аналитики рынка труда с оптимальными карьерными траекториями, их внутренняя структурная непрозрачность создает серьезные этические, нормативные и педагогические проблемы. Природа таких предиктивных систем, подобная «черному ящику», скрывает фундаментальную логику карьерных рекомендаций и рейтингов найма, что чревато усугублением исторических предвзятостей, подрывом доверия пользователей и нарушением новых международных нормативных баз. В данном всестороннем анализе оцениваются теоретические основы, методология внедрения и практические последствия фреймворков объяснимого искусственного интеллекта (Explainable AI, XAI), с особым акцентом на SHapley Additive exPlanations (SHAP) и Local Interpretable Model-agnostic Explanations (LIME), как важнейших механизмов достижения интерпретируемости машинного обучения в принятии карьерных решений.

Систематизируя пересечение устоявшихся парадигм поведенческой психологии — таких как модель RIASEC Холланда, Пятифакторная модель личности и социально-когнитивная теория карьеры (SCCT) — с передовой алгоритмической интерпретируемостью, данный отчет демонстрирует, как XAI превращает предиктивные модели из непрозрачных алгоритмических барьеров в прозрачные инструменты развития. Эмпирические данные, обобщенные в этом обзоре, свидетельствуют о том, что хотя интерпретируемое машинное обучение повышает точность прогнозирования и сглаживает различия между подгруппами, его самая глубокая ценность заключается в содействии сотрудничеству человека и ИИ, аудите алгоритмической справедливости и расширении возможностей соискателей за счет прозрачной самоэффективности на основе данных. Смещая фокус с простой точности прогнозирования на педагогически значимые объяснения, организации и образовательные учреждения могут гарантировать, что карьерная ориентация на основе ИИ останется

справедливым, надежным и юридически правомерным инструментом для развития глобальных трудовых ресурсов.

Ключевые слова*:* объяснимый искусственный интеллект (XAI), интерпретируемое машинное обучение, принятие карьерных решений, профессиональная ориентация, SHAP, LIME, психометрическое профилирование, алгоритмическая справедливость, прогнозное моделирование, технологии управления человеческими ресурсами.

АБИТУРИЕНТТЕРДІҢ КӘСІБИ ТАҢДАУЫН БОЛЖАУҒА АРНАЛҒАН ИНТЕРПРЕТАЦИЯЛАНАТЫН МАШИНАЛЫҚ ОҚЫТУ ЖӘНЕ ТҮСІНДІРМЕЛІ ЖАСАНДЫ ИНТЕЛЛЕКТ МОДЕЛЬДЕРІН ЗЕРТТЕУ

Орынбасар М. А., Ақбердиева М. Е.
Есенов университеті, Ақтау, Қазақстан
e-mail: maksym1.orynbassar@yu.edu.kz, meruyert1.akberdiyeva@yu.edu.kz

Аңдатпа. Жасанды интеллекттің адам ресурстарын басқару және кәсіби бағдар беру жүйесіне интеграциялануы парадигмалық өзгерістің катализаторы болды: дәстүрлі мансаптық кеңес беруден деректерге негізделген жоғары технологиялық болжамдық модельдеуге көшу жүзеге асты. Қазіргі машиналық оқыту архитектуралары — шешім ағаштарының ансамбльдерінен бастап терең нейрондық желілерге дейін — психометриялық профильдерді, академиялық жетістіктерді және еңбек нарығы аналитикасын оңтайлы мансаптық траекториялармен сәйкестендіруде бұрын-соңды болмаған дәлдік көрсеткенімен, олардың ішкі құрылымдық мөлдір еместігі маңызды этикалық, нормативтік және педагогикалық мәселелерді туындатады.

Мұндай «қара жәшік» сипатындағы болжамдық жүйелер мансаптық ұсыныстар мен жұмысқа қабылдау рейтингтерінің негізгі логикасын жасырады, бұл тарихи бейімділіктердің күшеюіне, пайдаланушылар сенімінің төмендеуіне және халықаралық нормативтік базалардың бұзылуына әкелуі мүмкін. Бұл кешенді талдауда Explainable Artificial Intelligence (XAI) фреймворктерінің теориялық негіздері, енгізу әдістемесі және практикалық салдары қарастырылады, әсіресе SHapley Additive exPlanations (SHAP) және Local Interpretable Model-agnostic Explanations (LIME) әдістеріне ерекше назар аударылады, олар мансаптық шешім қабылдауда машиналық оқытудың интерпретациялануын қамтамасыз ететін негізгі механизмдер ретінде бағаланады.

Бекітілген психологиялық парадигмалар — Холландтың RIASEC моделі, тұлғаның Үлкен Бес факторы (Big Five) және әлеуметтік-когнитивтік мансап теориясы (SCCT) — мен заманауи алгоритмдік интерпретацияның тоғысуын жүйелей отырып, бұл зерттеу XAI болжамдық модельдерді мөлдір емес алгоритмдік «қақпалардан» түсінікті даму құралдарына қалай айналдыратынын көрсетеді.

Эмпирикалық деректер интерпретацияланатын машиналық оқытудың болжам дәлдігін арттырып, топтар арасындағы диспропорцияны азайтатынын көрсетеді. Дегенмен, оның ең маңызды құндылығы — адам мен ИИ арасындағы ынтымақтастықты күшейту, алгоритмдік әділдікті аудиттен өткізу және ашық деректерге негізделген өзін-өзі тиімді бағалауды қамтамасыз ету арқылы пайдаланушылардың мүмкіндіктерін кеңейту.

Болжамдық дәлдіктен педагогикалық тұрғыдан маңызды түсіндірмелерге көшу арқылы ұйымдар мен білім беру мекемелері жасанды интеллектке негізделген кәсіби бағдар жүйелерінің әділ, сенімді және құқықтық тұрғыдан негізделген құрал ретінде дамуын қамтамасыз ете алады.

Түйін сөздер: түсіндірмелі жасанды интеллект (XAI), интерпретацияланатын машиналық оқыту, мансаптық шешім қабылдау, кәсіби бағдар беру, SHAP, LIME, психометриялық профильдеу, алгоритмдік әділдік, болжамдық модельдеу, адам ресурстарын басқару технологиялары