Pak. j. life soc. Sci. (2026), 24(2): 75-85 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2026-24.2.005

RESEARCH ARTICLE

Analysis of Potential Risks and Ethical Aspects of Using Artificial Intelligence Technologies in Digital Tax Administration

Vadim Zasko[1*], Stanislav Emelyanov[2]

[1,2]Financial University under the Government of the Russian Federation, Moscow, Russia

ARTICLE INFO

Received: AUG 5, 2026

Accepted: AUG 26, 2026

Keywords

Tax policy
Digital economy
Digital tax administration
Information technology
Artificial intelligence
Risks of AI implementation
Ethical risks

***Corresponding Author:**

ABSTRACT

Active implementation of artificial intelligence technologies in tax administration must consider a range of potential risks associated with their use, including legal, technical, informational, and ethical risks. This distinction encompasses both the internal logic of AI application in digital tax administration and the external implications for government interactions. The aim of this study is to analyze the potential risks and ethical aspects of AI application in digital tax administration. This analysis was used to compile a scenario matrix of the practical sensitivity of risks associated with AI application in digital tax administration and formulate corresponding recommendations for mitigating them. This analysis demonstrates that the use of artificial intelligence technologies in digital tax administration cannot be assessed in isolation from a specific use case. The most sensitive cases are those in which an algorithmic signal can influence the intensity of control measures, the legal status of the taxpayer or the processing of a significant amount of tax-relevant data. The applicability of AI should be determined by a combination of four conditions: scenario limitation, preservation of human control in sensitive cases, transparency of service impact and continuous monitoring

INTRODUCTION

The transition toward more intensive use of artificial intelligence technologies in digital tax administration naturally brings to the forefront not the question of the benefits of AI as such, but rather the limits of its permissible application (Yakovleva et al., 2023; Rachkovsky and Yakovlev, 2025). Russian practice already demonstrates a shift from general discussion of AI to its practical implementation: the Federal Tax Service of Russia officially reports the introduction of artificial intelligence technology into electronic business registration and the automated adoption of decisions regarding certain registration procedures.

In the tax context, the key issue is not whether AI is permissible in principle, but which scenarios of its application are acceptable, lawful, manageable, and secure. In digital tax administration, risks are not abstract but scenario-specific: they arise in the routing of taxpayer requests, preliminary document verification, scoring-based prioritization, predictive control, and the automation of analytical audits (Lyutova and Izmailova, 2026).

Within digital tax administration, the risk associated with the use of AI should be understood as the probability that technological support may distort administrative processes, violate data protection requirements, create legal uncertainty, or result in non-transparent impacts on the rights and obligations of the parties involved.

The scenario-based nature of AI application means that risk does not exist independently of a specific function of digital tax administration. Therefore, the analysis should identify the scenarios in which risks arise, determine how they manifest themselves in practice, assess the consequences for both tax authorities and taxpayers, and establish which risk-management measures are appropriate for limiting such risks.

METHODOLOGY AND MATERIALS

The methodological framework of the study comprises a combination of general scientific and specialized research methods. The study employs a mathematical assessment of the integral risk associated with the use of artificial intelligence, based on expert-quantitative evaluations of risk priority across the main AI application scenarios. Within the mathematical risk assessment, the pairwise comparison method is applied using the results of expert pairwise assessments, which makes it possible to determine the relative significance of each parameter within the overall risk structure. The pairwise comparison method is used as the methodological basis for establishing the weighting structure.

Structural and functional analysis methods were used to classify the groups of risks associated with the application of artificial intelligence in tax administration. The scenario-based method made it possible to account for scenario-specific groups of risks related to AI application and, on this basis, to formulate practical recommendations for minimizing risks and security threats arising from the use of AI in digital tax administration.

RESULTS

A substantive analysis of risks in digital tax administration should proceed not from the abstract properties of the technology, but from typical scenarios of its practical application. In the context under consideration, six scenarios are of fundamental importance: routing of taxpayer requests, preliminary document verification, scoring-based prioritization, predictive control, automation of analytical audits, and proactive service-based prevention. Each of these scenarios differs in terms of the nature of the data involved, the permissible degree of automation, the cost of error, and the consequences for both the tax authority and the taxpayer. Risks should be assessed not only as potential threats but also as forms of distortion of the tax administration process (Table 1).

Table 1: Scenario-Based Analysis Of Potential and Practical Risks Associated With the use of Artificial Intelligence in Digital Tax Administration

AI application scenario	Useful function	Potential risk	Practical manifestation	Consequence for the tax authority	Consequence for the taxpayer
Routing of taxpayer requests	Accelerating the distribution of messages and requests	Classification error	A request is misrouted or processed with delay	Increased need for repeated processing and loss of time	Longer response times and unpredictability of interaction
Preliminary document verification	Identifying incomplete information and formal inconsistencies	False error signal or failure to detect an error	Unnecessary return of documents or failure to identify a defect	Reduced quality of preliminary control	Resubmission and additional costs
Scoring-based prioritization	Directing attention toward more significant cases	Bias in criteria and unexplained amplification of risk signals	Overestimation or underestimation of case priority	Distortion in the allocation of control resources	Risk of disproportionate scrutiny of the taxpayer
Predictive control	Forecasting problem areas and deviations	Forecasting error and model drift	False-positive and false-negative signals	Reduced accuracy of case selection and assessment	Risk of unjustified preventive pressure
Automation of analytical audits	Accelerating the comparison of large datasets and identification of inconsistencies	Instability of results when processing heterogeneous data	Increase in false positives and missed discrepancies	Loss of confidence in the analytical framework	Lack of transparency regarding the grounds for increased scrutiny

Source: compiled by the authors.

The scenarios under consideration differ in terms of data type, the permissible degree of automation, and the cost of error. In the scenarios involving the routing of taxpayer requests and proactive service-based prevention, the predominant risks are misclassification and disruption of the predictability of digital communication. In preliminary document verification and automated analytical audits, the major risks are scalable technical errors caused by data heterogeneity and instability in matching rules. In scoring-based prioritization and predictive control, legal and ethical risks become particularly important because the outcome of an algorithmic assessment may influence the intensity of subsequent administrative intervention.

For a systematic representation of the logic of risk management, it is appropriate to use a four-level model in which risks pass through successive stages of identification, assessment, mitigation, and monitoring (Figure 1). The risk-management framework is adapted not to an abstract digital system but to the specific tax administration context, which combines the requirements of legality, data protection, auditability, and service orientation.

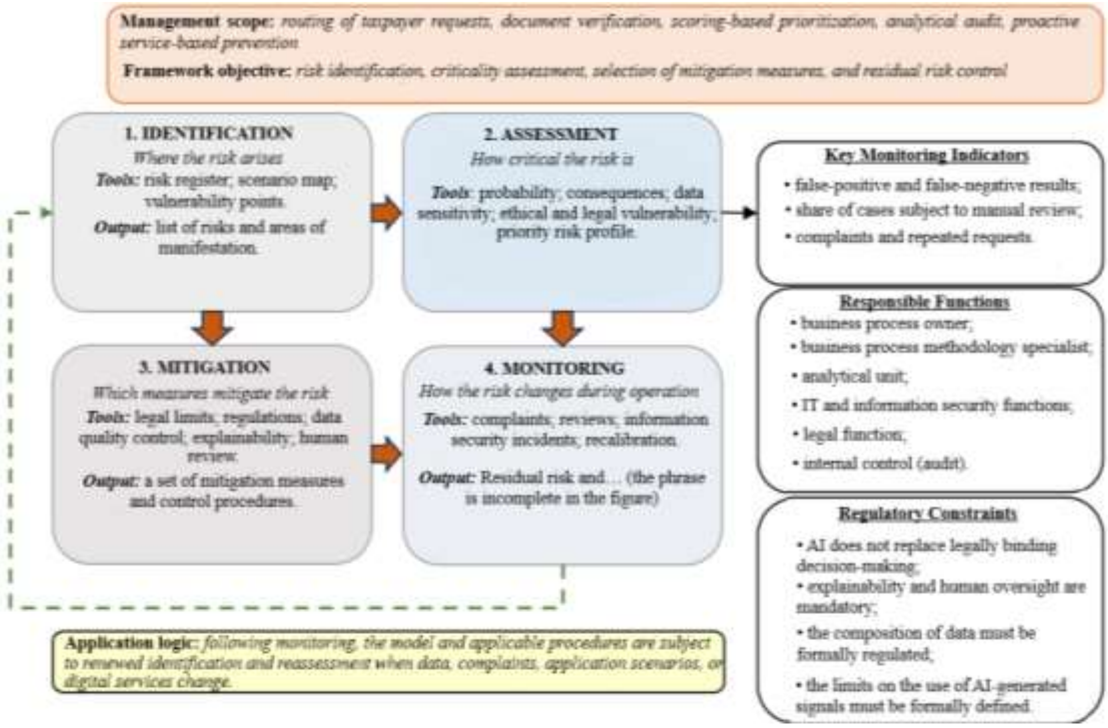

Figure 1: Risk management framework for the use of artificial intelligence in digital tax administration

Source: Compiled by the authors.

The presented framework demonstrates that risk management for the use of artificial intelligence in tax administration is not linear but cyclical in nature. Each level—identification, assessment, mitigation, and monitoring—produces its own management outcome; however, monitoring does not complete the cycle. Instead, it returns the system to repeated risk reassessment when data, operating conditions, complaints, service scenarios, or digital regulations change.

From a practical perspective, the risks associated with the use of artificial intelligence in digital tax administration can be divided into four broad groups: technical risks, information and cyber risks, legal risks, and ethical risks. Technical risks are associated with model errors, data drift, unstable outputs, and improper integration into the digital environment. Information and cyber risks include data leaks, unauthorized access, and compromise of data storage and processing environments. Legal risks arise from non-compliance with legislation governing data, information, and administrative procedures (Tax Code of the Russian Federation (Part One) No. 146-FZ, 1998; Federal Law of, No. 54-FZ, 2003; Federal Law of, No. 149-FZ, 2006; Federal Law No. 152-FZ, 2006; Federal Law No. 63-FZ, 2011). Ethical risks concern lack of explainability, unfair profiling, disproportionate intervention, and excessive reduction of the human role at sensitive stages of the process.

The broad classification of risks is useful as an initial typology (Table 2); however, to address the research objectives, it should be transformed into a managerial risk register in which each risk group is linked to the scenario in which it arises, the most sensitive stage of the process, and the relevant control measures.

Table 2: Risk Register for the Use of Artificial Intelligence in Digital Tax Administration

Risk group	Specific risk	Scenario of occurrence	Most sensitive stage	Possible consequence	Basic control measure
Technical risks	Model error, data drift, unstable classification	Scoring-based prioritization, predictive control, analytical audit	Verification, analytical support	Distortion of priorities and incorrect processing of risk signals	Model calibration, data quality control
Information and cyber risks	Data leakage, unauthorized access, compromise of system integrity	Storage and processing of tax-relevant information	Data transfer, storage, integration	Data compromise and harm to taxpayer rights	Access restrictions, security audits
Legal risks	Insufficient evidentiary reliability of results, disproportionate intervention	Use of AI outputs in administrative procedures	Control response, service communication	Challenges to decisions and reduced legal robustness	Formal definition of the limits of AI use, preservation of human oversight
Ethical risks	Lack of explainability, bias, unfair risk profiling	Communication, scoring, service support	Analytical support, user interaction	Reduced trust, discriminatory effects, lack of transparency	Transparent criteria, explainability, right to review

Source: Compiled by the authors.

Ethical aspects of artificial intelligence use occupy a particularly important place in the context under consideration. In digital tax administration, ethics has a practical meaning: taxpayers should understand that digital services operate predictably; intelligent systems should not generate recommendations that cannot be explained; risk profiling should not result in hidden discrimination; automated support should not replace human review in sensitive cases; and data processing should be proportionate to the purpose of administrative interaction. These requirements are directly related to the legal principles of data and information protection, as well as to the logic of the National Strategy for the Development of Artificial Intelligence, which emphasizes the need for the safe and controlled development of the technology.

For the purposes of tax administration, ethical aspects should be viewed not as external moral constraints but as conditions for the legitimacy of digital interaction (Table 3). In this respect, the tax context directly corresponds to officially established international approaches to the responsible use of AI. UNESCO considers the protection of human rights and human dignity to be a fundamental basis of AI ethics and emphasizes the importance of transparency, fairness, and human oversight.

The presented systematization demonstrates that ethical considerations in tax administration are not an additional external requirement but form part of the very structure of permissible artificial intelligence use. Violations of these principles simultaneously reduce the legal robustness of the procedure, user trust, and the quality of service interaction.

Table 3: Ethical Aspects of the Use of Artificial Intelligence in Digital Tax Administration

Ethical aspect	Where it manifests	Why it is important for tax administration	Risk of violation of participant rights	Mitigation mechanism
Explainability	Service recommendations, scoring signals, analytical conclusions	The user should understand the logic behind the actions of the digital service	Lack of transparency regarding the grounds for action	Mandatory provision of clear explanations
Prohibition of hidden profiling	Scoring-based prioritization, analytical support	Risk of disproportionate attention being directed toward particular groups	Discriminatory effect	Control over profiling criteria
Preservation of the human role	Sensitive cases involving verification and administrative response	AI should not replace legally significant decision-making	Loss of the right to human review	Mandatory verification of complex cases
Proportionality of data processing	Predictive control, analytical audit	The volume of data used should correspond to the purpose of the interaction	Excessive intervention	Regulation of the composition of data
Predictability of the service	Feedback, recommendations, routing	Participants should understand how to use the digital service	Increased distrust and interaction errors	Standards for a clear and understandable user interface

Source: Compiled by the authors.

From a practical perspective, the problem of conflating the technical and legal levels of risk is particularly significant. An algorithm may function correctly from a computational standpoint while still creating risks from the perspective of administrative procedure. Therefore, in the tax context, it is insufficient to limit assessment solely to model accuracy testing; the legal, ethical, and organizational suitability of its use must also be evaluated. It is appropriate to apply an expert-quantitative risk prioritization model in which qualitatively identified parameters are given a formalized expression, making it possible to compare AI application scenarios according to a common analytical framework.

To move from a qualitative classification of risks to their comparable prioritization, a two-tier assessment model can be applied. In the open-data tier, an expert-quantitative assessment of scenario-specific risk sensitivity is formed on the basis of the regulatory framework, official institutional sources, and qualitative comparative analysis. In the internal tier, which can be implemented only if access is available to restricted operational data of the Federal Tax Service of Russia, statistical parameterization of risk is performed using the actual performance results of intelligent services. This distinction makes it possible, on the one hand, to avoid creating an appearance of statistical precision in the absence of restricted operational data and, on the other hand, to establish a mathematically rigorous methodology for subsequent in-depth analysis.

Within the open-data tier, each AI application scenario is assessed according to five parameters: the probability of the practical manifestation of risk, the scale of consequences, the sensitivity of the data involved, the ethical and legal vulnerability of the outcome, and the degree of uncontrollability of the algorithmic signal. To ensure comparability among the parameters, they are normalized, after which the integral priority of scenario-specific risk is calculated.

The proposed analytical framework makes it possible to distinguish between three related but non-equivalent levels of analysis: the initial scenario-specific criticality of risk, the effect of the applied mitigation measures, and the residual risk after the implementation of control measures. Thus, mathematical assessment does not replace qualitative analysis but makes its results comparable and suitable for management purposes.

At the first stage, a scale for expert assessment of risk parameters is introduced. For each AI application scenario, the following are assessed: the probability of risk occurrence, the scale of consequences, data sensitivity, the ethical and legal vulnerability of the outcome, and the degree of uncontrollability of the algorithmic signal. Since these initial assessments are assigned on an ordinal scale from 1 to 5, they are normalized for subsequent calculations.

$$z_{jk} = \frac{x_{jk} - 1}{4} \tag{1}$$

where z_{jk} is the normalized value of the k-th risk parameter for the j-th artificial intelligence application scenario;

x_{jk} is the expert assessment of the k-th parameter on a scale from 1 to 5;

j is the artificial intelligence application scenario;

k is the risk assessment parameter.

Normalization is required so that parameters initially specified using a common expert scale can be incorporated into the integral calculation without introducing artificial bias into the result. Probability, consequences, data sensitivity, ethical and legal vulnerability, and risk uncontrollability are converted to a comparable scale from 0 to 1 and can subsequently be included in the calculation as homogeneous elements of the model.

At the second stage, the integral priority of scenario-specific risk is calculated. Unlike a simple linear sum, the calculation includes an additional interaction term that increases the criticality of a scenario when both the scale of consequences and ethical and legal vulnerability are high.

$$\mathrm{PR}_j = \sum_{k=1}^{5} \omega_k \, z_{jk} + \lambda \sqrt{z_{j2} z_{j4}} \tag{2}$$

where PR_j is the integral risk priority for the j-th artificial intelligence application scenario;

z_{j1} is the normalized probability of risk occurrence;
z_{j2} is the normalized scale of consequences for tax administration and the taxpayer;
z_{j3} is the normalized degree of impact on data and information security;
z_{j4} is the normalized ethical and legal vulnerability of the risk;
z_{j5} is the normalized degree of risk uncontrollability;
ω_k is the significance coefficient of the corresponding risk parameter;
λ is the interaction coefficient between the scale of consequences and ethical and legal vulnerability.

This makes it possible to account for the nonlinear amplification of those scenarios that are simultaneously characterized by a high scale of consequences and high ethical and legal vulnerability.

The parameter weights are not intuitively assigned coefficients but are derived from an expert pairwise comparison procedure that makes it possible to determine the relative significance of each parameter within the overall risk structure. The pairwise comparison method is applied by first constructing a matrix of expert comparisons among the risk parameters and then calculating the final weights on its basis.

$$\omega_k = \frac{\left(\prod_{m=1}^{5} a_{km}\right)^{1/5}}{\sum_{k=1}^{5}\left(\prod_{m=1}^{5} a_{km}\right)^{1/5}} \tag{3}$$

where a_{km} is an element of the pairwise comparison matrix for the significance of risk parameters;
ω_k is the final weight of the -th risk parameter.

This makes it possible to eliminate arbitrariness in determining the weights and link them to a formal expert decision-making procedure. The significance coefficients ω_k reflect the relative priority of the risk parameters within a unified pairwise comparison system. For the purposes of the present study, this procedure is particularly important because, without it, the resulting PR_j indicator could depend not on the actual risk structure but on an arbitrary assignment of weights.

Subsequently, the pairwise comparison method is used primarily as a methodological justification for the weighting structure. The final system of significance coefficients used for the expert-quantitative interpretation of risks is presented.

At the next stage, the influence of mitigation measures must be formalized. For the purposes of tax risk management, it is important to consider not only the initial sensitivity of a scenario but also the extent to which it is reduced after the introduction of legal, organizational, methodological, and technical safeguards; their combined effect can appropriately be calculated as an aggregate effectiveness measure.

$$\mathrm{EM}_j = 1 - \prod_{r=1}^{R_j}\left(1 - e_{jr}\right) \tag{4}$$

where EM_j is the aggregate effectiveness of the set of risk mitigation measures for the j-th scenario;
e_{jr} is the effectiveness of the r-th risk mitigation measure for the j-th scenario;
R_j is the number of measures applied to the j-th scenario.

Risk mitigation measures do not operate as a simple arithmetic sum but rather as a cumulative system of safeguards.

At the final stage, residual risk is determined. Its introduction is important because it makes it possible to directly link the mathematical component to the four-level risk management framework: identification and assessment determine the initial risk priority, mitigation affects the aggregate effectiveness of the safeguards, while monitoring should focus on the residual risk that remains after the application of control measures.

$$\mathrm{OR}_j = \mathrm{PR}_j\left(1 - \mathrm{EM}_j\right) \tag{5}$$

where OR_j is the residual risk for the j-th AI application scenario;
PR_j is the scenario-specific risk priority before the application of safeguards;
EM_j is the aggregate effectiveness of the measures applied.

This provides a direct link between the mathematical component and the risk management framework: risk identification and assessment determine PR_j, mitigation affects EM_j, and monitoring should track OR_j.

Thus, the calculation framework comprises three analytical levels: scenario-specific risk priority, the effect of safeguards, and residual risk after their implementation.

Within the open scenario-based analysis, the resulting PR_j value should be interpreted not as an absolute statistical probability of harm, but as an indicator of the priority of managerial attention to the corresponding artificial intelligence application scenario. The introduction of the EM_j indicator makes it possible to assess not only the initial risk sensitivity but also the extent to which it decreases after safeguards are implemented, whereas the OR_j indicator reflects the residual The numerical values in the Table reflect expert assessments of scenario-specific risk sensitivity on a scale from 1 to 5, where 1 indicates a low level of the corresponding parameter, 2 a moderately low level, 3 a medium level, 4 a high level, and 5 a critical level. The values presented express the relative expert-assessed intensity of risk within the set of scenarios being compared.

The calculated PR_j indicator makes it possible to rank scenarios according to their degree of managerial sensitivity and the need for priority attention by the tax authority. For subsequent qualitative interpretation, the following interval scale is proposed:

– a PR_j, value below 0.60 corresponds to a **moderate** priority level;
– a value from 0.60 to 0.79 corresponds to a **high** priority level;
– a value from 0.80 to 0.99 corresponds to a **very high** priority level;
– a value of 1.00 or higher corresponds to a **critical** priority level.

Expert assessments of the risk parameters for each artificial intelligence application scenario (Table 4) make it possible to obtain a comparable PR_j, indicator, which is subsequently used as an indicator of managerial risk priority.

Table 4: Expert-Quantitative Assessment of Risk Priority for the Main Artificial Intelligence Application Scenarios

Artificial intelligence application scenario	Probability of risk occurrence	Scale of consequences for tax administration and the taxpayer	Degree of impact on data and information security	Ethical and legal vulnerability	Degree of Risk uncontrollability	Integral scenario-specific risk priority	Priority level
Intelligent routing of taxpayer requests	3	3	2	3	2	0.54	Moderate
Preliminary document verification	4	3	3	2	3	0.61	High
Scoring-based prioritization	4	5	4	5	4	1.03	Critical
Predictive control	3	5	4	4	4	0.90	Very high
Automation of analytical audits	4	4	5	3	4	0.86	Very high
Proactive service-based prevention	3	3	2	3	3	0.59	Moderately high

Source: Compiled by the authors.

However, the numerical indicator alone does not reveal the specific nature of the risk, its source, or the control measure required within the practical tax administration environment. Therefore, the calculated prioritization is further translated into a scenario-based matrix of practical sensitivity, in which each scenario is matched with the dominant risk, its source, the most significant consequence, the mandatory control measure, and the responsible management function.

Within the expert-quantitative interpretation adopted in the present study, the following values of the significance coefficients are used: ω{1}=0,20; ω{2}=0,25; ω{3}=0,20; ω{4}=0,20; ω{5}=0,15; λ=0,10. These parameters reflect the increased significance of the scale of consequences while also accounting for the amplification of criticality in cases where high ethical and legal vulnerability is combined with substantial organizational consequences.

The results obtained confirm that the highest risk priority in the tax context is associated with scoring-based prioritization, whereas predictive control and automation of analytical audits are characterized by very high, although somewhat less extreme, levels. Service and communication scenarios demonstrate more moderate integral sensitivity; however, they remain important in terms of the manageability of interaction and taxpayers' trust in digital services.

For applied risk management, risks must be linked to specific AI application scenarios. For this purpose, a scenario-based matrix of practical risk sensitivity is used (Table 5). It demonstrates that the lowest practical sensitivity is characteristic of scenarios involving routing and initial service support, where errors are reversible and do not independently produce legally significant consequences. By contrast, the highest sensitivity is concentrated in scoring-based prioritization and predictive control because, in these scenarios, algorithmic outputs may affect the intensity of subsequent administrative intervention and the actual position of the taxpayer. Preliminary document verification, automation of analytical audits, and proactive service-based prevention occupy an intermediate position: the risks of scalable technical error and non-transparent impact remain highly significant, but they are more amenable to mitigation through formal procedures, verification, and data quality control.

Table 5: Scenario-Based Matrix of Practical Sensitivity of Risks Associated with the use of AI in Digital Tax Administration

Application scenario	Dominant risk	Source of risk	Most significant consequence	Mandatory control measure	Responsible function	Level of practical sensitivity
Intelligent routing of taxpayer requests	Classification error and incorrect routing	Insufficient model training, drift in the subject matter of requests, heterogeneity of texts	Delays in processing requests, repeated rerouting, reduced predictability of interaction	Manual verification of atypical cases, regular model retraining, quality control of request distribution	Process owner, operator, IT team	Moderate
Preliminary document verification	False rejection or failure to detect incompleteness	Instability of verification rules, variability of document fields, differences in incoming data formats	Unnecessary return of documents or failure to identify a defect, resulting in further distortion of the process	Test environment, logging of rejection reasons, sample-based control, periodic review of verification rules	Methodology specialist, IT function, procedure owner	High
Scoring-based prioritization	Priority bias and unjustified case selection	Low interpretability of features, bias in training data, hidden profiling	Disproportionate scrutiny of individual taxpayers and risk of discriminatory effects	Explainability threshold, expert confirmation of sensitive cases, control of profiling criteria	Analyst, owner of the relevant control function	Critical
Predictive control	False-positive or false-negative prediction	Model drift, incomplete or outdated data, instability of predictive relationships	Unjustified preventive pressure or failure to identify a genuinely high-risk situation	Restriction on using predictions as an independent basis for action, mandatory verification of results, model	Analytical unit, IT team, process owner	Critical

				recalibration		
Automation of analytical audits	Non-transparent analytical conclusion and accumulation of false positives	Heterogeneity of datasets, integration errors, difficulty in verifying the model's internal logic	Loss of confidence in the analytical framework and difficulties in challenging or reviewing the result	Explanation log, data quality control, mandatory human review of significant signals	Analyst, IT function, internal control unit	Very high
Proactive service-based prevention	Incorrect recommendation or imposed user pathway	Incorrect user segmentation, oversimplified recommendation logic, insufficient service personalization	Formation of incorrect taxpayer behavior, increased repeated requests, and declining trust in the digital service	Clear distinction between advisory and legally significant information, understandable explanation of recommendation logic, option to decline automated guidance	Service owner, methodology specialist, digital channel operator	High

Source: Compiled by the authors.

From a technical perspective, the most vulnerable scenarios are the automation of analytical audits and preliminary document verification, where a model error can rapidly scale across a large volume of cases. In service interactions, routing and proactive prevention are the most sensitive scenarios because errors directly affect the predictability of communication with taxpayers.

Comprehensive statistical verification of the frequency, stability, and severity of risks within specific operational environments of the Federal Tax Service of Russia requires restricted operational data that are not publicly available. The present study addresses the task of open scenario-based analysis and develops a methodological framework, whereas an in-depth quantitative assessment should be conducted using internal statistical datasets and continuous monitoring of models during operation.

Formula for the internal statistical tier

$$\mathrm{ER}_j = \alpha_1 \frac{\mathrm{FP}_j}{N_j} + \alpha_2 \frac{\mathrm{FN}_j}{N_j} + \alpha_3 \frac{\mathrm{TH}_j}{N_j} + \alpha_4 \frac{C_j}{N_j} + \alpha_5 \frac{\mathrm{IS}_j}{N_j} + \alpha_6 \mathrm{DR}_j \tag{6}$$

where ER_j the empirical risk index for the j-th AI application scenario;

FP_j the number of false-positive results;

FN_j the number of false-negative results;

TH_j the number of cases in which the result was reviewed and overturned following human verification;

C_j the number of complaints and appeals;

IS_j the number of information security incidents;

N_j the total number of cases in which the j-th scenario;

DR_j the data or model drift coefficient;

$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ the significance coefficients of the corresponding empirical risk parameters.

Since the necessary data are not publicly available, the corresponding statistical calculation is not performed in the present study. Nevertheless, the proposed methodology may serve as a basis for subsequent internal mathematical and statistical monitoring of risks associated with the use of artificial intelligence.

Practical risk mitigation measures should be grouped according to their area of impact: legal definition of the boundaries of AI use, organizational control, data quality assurance, information protection, explainability of results, and continuous monitoring. Their purpose is to ensure that AI is used within a controlled and manageable framework. A consolidated set of practical recommendations developed by the authors through adaptation of the official regulatory framework to the needs of tax administration is presented in Table 6.

Table 6: Practical Recommendations for Mitigating Risks and Security Threats Associated with the Use of AI In Digital Tax Administration

Area of measures	Recommendation	Risk mitigated	Scenario for which It Is primarily relevant	Expected result
Legal measures	Establish a list of permissible AI application scenarios and define the limits on the use of AI-generated outputs	Legal and ethical risks	Scoring-based prioritization, predictive control	Reduced risk of arbitrary expansion of algorithmic functions
Organizational measures	Preserve the mandatory role of a specialist in sensitive cases	Ethical and legal risks	Analytical audit, control-related response	Reduced risk of uncontrolled automation
Data quality	Introduce regular verification of the completeness, comparability, and relevance of input data	Technical risks	Preliminary verification, predictive control	Reduction in model errors and false signals
Information protection	Strengthen access control and auditing of actions within the digital environment	Information and cyber risks	All scenarios	Reduced probability of data compromise
Explainability	Establish a requirement to provide a clear explanation of service recommendations	Ethical risks	Routing, proactive service-based prevention	Increased trust and predictability of interaction
Continuous monitoring	Introduce a cycle of repeated assessment of residual risks and revision of applicable procedures	All risk groups	All scenarios	Maintenance of a controlled framework for the use of artificial intelligence

Source: Compiled by the authors.

CONCLUSIONS

The analysis demonstrates that the use of artificial intelligence in digital tax administration is associated with technical, information-related, legal, and ethical risks that should be considered not in the abstract but in relation to specific application scenarios.

It was established that the ethical aspects of AI use are practical in nature and are reflected in requirements concerning explainability, fairness, human oversight, proportionality of automation, and data protection.

It was substantiated that practical risk management should be based on a four-level framework comprising identification, assessment, mitigation, and monitoring. This framework establishes the regulatory boundaries for the subsequent selection of methods and scenarios for AI integration.

ACKNOWLEDGMENTS

The paper was prepared on the research results carried out at the expense of budgetary funds within the framework of the state research assignment to the Financial University under the Government of the Russian Federation, Moscow, Russian Federation.

REFERENCES

Federal Law No. 152-FZ. On personal data. Rossiiskaia Gazeta, State Duma of the Federal Assembly of the Russian Federation, July 27, 2006.

Federal Law No. 63-FZ. On electronic signatures. Rossiiskaia Gazeta, State Duma of the Federal Assembly of the Russian Federation, April 6, 2011.

Federal Law of, No. 149-FZ. On information, information technologies and information protection. Rossiiskaia Gazeta, State Duma of the Federal Assembly of the Russian Federation, July 27, 2006.

Federal Law of, No. 54-FZ. On the use of cash register equipment in settlements in the Russian Federation. Rossiiskaia Gazeta, State Duma of the Federal Assembly of the Russian Federation, May 22, 2003.

Lyutova OI, Izmailova AM. Implementation of a risk-based approach in tax administration using artificial intelligence technology. Legal Sci Pract J Nizhny Novgorod Acad MIA Russia 2026,1:106-10.

Rachkovsky IB, Yakovlev AR. Assessing the effectiveness of digital transformation. Understanding digital transformation. Ekonomika i upravlenie: problemy resheniya 2025,1:214-25.

Tax Code of the Russian Federation (Part One) No. 146-FZ. Rossiiskaia Gazeta, State Duma of the Federal Assembly of the Russian Federation, July 31, 1998.

Yakovleva EA, Vinogradov AN, Aleksandrova LV, Filimonov AP. How artificial intelligence helps transform the digital economy. Russ J Innov Econ 2023,13:707-26. https://doi.org/10.18334/vinec.13.2.117710