CURRENT ISSUES OF ECONOMIC SCIENCES

Management

UDK 004.7:004.421.2:005.8

DOI https://doi.org/10.5281/zenodo.17141025

Risk management in high-load service infrastructure using AI-based predictive models

Oleksandr Boiko,
Master of Engineering in Computer Science,
Chief Technology Officer (CTO), Field Complete Inc, Atlanta, USA,
https://orcid.org/0009-0004-3491-8431

Accepted: 04.09.2025 | Published: 17.09.2025

Abstract. High-load information services require effective risk management to ensure business continuity and maintain stable user service. The increasing volume of data and the complexity of distributed computing systems pose potential threats to technical failures and performance degradation. This **study aims** to develop a methodology for risk management in high-load service infrastructure using artificial intelligence models to predict anomalies and potential failures. The **methodology** involves collecting and analyzing server metrics, application logs, and real-time user load. Machine learning algorithms, including time series analysis, clustering, and neural networks, are applied to identify patterns and forecast critical events. Mechanisms for automatic load balancing and scaling of computational resources in cloud environments based on predictive models have been developed. The **results** demonstrate that the application of AI-based predictive models significantly reduces response time to anomalies and failures, improves system component performance, and decreases the likelihood of technical outages during peak loads. The proposed algorithms provide more uniform resource allocation,

enhance the fault tolerance of distributed databases and server clusters, and ensure continuous user access to services. Analysis shows that integrating predictive models into risk management reduces redundant resource costs and optimizes scaling in cloud infrastructures. The **conclusions** highlight the practical significance of intelligent approaches to risk management in high-load systems. The proposed methodology enhances service reliability and stability, enables a prompt response to potential failures, and optimizes the utilization of computational and network resources. The results serve as a foundation for developing risk management strategies in distributed systems, enhancing infrastructure resilience, and supporting business continuity in cloud platforms.

Keywords: failure prediction, system anomalies, machine learning, load balancing, cloud platforms, infrastructure resilience, resource management.

Управління ризиками в інфраструктурі високонавантажених сервісів із застосуванням AI-моделей прогнозування

Бойко Олександр Анатолійович,
спеціаліст з комп'ютерних систем, головний технічний директор,
Field Complete Inc, м. Атланта, США, https://orcid.org/0009-0004-3491-8431

Анотація. Високонавантажені інформаційні сервіси потребують ефективного управління ризиками для забезпечення безперервності бізнес-процесів та стабільності обслуговування користувачів. Зростання обсягів даних і складність розподілених обчислювальних систем створюють потенційні загрози технічних відмов та зниження продуктивності. **Метою** дослідження є розробка методології управління ризиками в інфраструктурі високонавантажених сервісів із використанням моделей штучного інтелекту для прогнозування аномалій та потенційних відмов. **Методи.** Методологія передбачає збір і аналіз метрик серверів, логів додатків та користувацького

навантаження у реальному часі. Для виявлення закономірностей і прогнозування критичних ситуацій застосовано алгоритми машинного навчання, зокрема методи часових рядів, кластеризації та нейронні мережі. Розроблено механізми автоматичного балансування навантаження та масштабування обчислювальних ресурсів у хмарних середовищах на основі прогнозних моделей. **Результати** дослідження демонструють, що застосування AI-моделей прогнозування дозволяє значно знизити час реагування на аномалії та неполадки, підвищити продуктивність компонентів системи та зменшити ймовірність технічних збоїв у періоди пікового навантаження. Запропоновані алгоритми забезпечують більш рівномірний розподіл ресурсів, підвищують відмовостійкість розподілених баз даних і серверних кластерів, а також підтримують безперервний доступ користувачів до сервісів. Проведений аналіз показав, що інтеграція прогнозних моделей у процес управління ризиками дозволяє зменшити витрати на надлишкові ресурси та оптимізувати масштабування у хмарних середовищах. **Висновки** підкреслюють практичну значущість інтелектуальних підходів до управління ризиками у високонавантажених системах. Запропонована методологія дозволяє підвищити надійність та стабільність сервісів, забезпечити швидке реагування на потенційні відмови та оптимізувати використання обчислювальних і мережевих ресурсів. Отримані результати можуть стати основою для розробки стратегій управління ризиками в розподілених системах, підвищення стійкості інфраструктури та підтримки безперервності бізнес-процесів у хмарних платформах.

Ключові слова: прогнозування відмов, аномалії систем, машинне навчання, балансування навантаження, хмарні платформи, стійкість інфраструктури, управління ресурсами.

Problem statement. Ensuring the reliability and stability of highly loaded services is a key challenge in modern information infrastructures. Distributed

computing systems and cloud platforms for data storage and processing are characterized by increasing traffic volumes and the complexity of structural relationships between components, which creates an increased risk of technical failures, performance degradation, and a negative impact on the user experience. Traditional approaches to risk management, which are based on passive monitoring and reactive resource scaling, do not provide sufficient efficiency in detecting and preventing critical failures.

A remaining unsolved problem is the development of integrated risk management methods capable of predicting potential anomalies and system failures in real-time. It involves combining machine learning algorithms, big data analysis, and infrastructure management architectural practices to form a comprehensive approach to increasing service resilience. Solving this problem has direct practical significance, as it allows for increasing fault tolerance, optimizing the distribution of computing and network resources, ensuring continuous user access to services, and reducing operating costs.

Research on this problem contributes to the development of theoretical foundations for predicting technical risks in highly loaded systems, including methods of time series analysis, event clustering and the use of neural networks to predict critical infrastructure states. In turn, the practical implementation of such approaches enables the creation of automated load balancing algorithms and dynamic resource scaling, thereby increasing the performance and stability of systems in real-time.

Therefore, a clear definition of the problem establishes vectors for further research, forms the basis for setting specific scientific and practical tasks and justifies the need to use AI forecasting models for risk management in highly loaded infrastructures.

Analysis of recent research and publications. Research on risk management issues in IT infrastructure and high-load services holds an important place in modern scientific literature, particularly in light of the integration of artificial intelligence

(AI) for risk forecasting and business process optimization. V. Butenko and M. Baidatskyi [1] focus on the theoretical foundations of the formation of a risk management system, identifying their key role in the stability of enterprises in dynamic market conditions. A. Chaikina [2] investigates the integration of risk management into the enterprise management system, emphasizing the need for a comprehensive approach to monitoring and responding to threats. The work of O. Korostin [3] demonstrates the capabilities of NLP in processing large volumes of text data, which can be utilized to identify risks in logistics and service systems.

In the field of IT projects, a significant contribution was made by O. Khrapkin, O. Kindrat and R. Chopey [4], who analyze risk management methods and tools adapted to the specifics of technological projects. Similar issues are considered by N. Danyliuk, Yu. Shulyk and O. Kachan [5] focus on modern approaches to project management in IT companies, particularly in the context of flexible methodologies. The works of O. Kindrat and G. Dutka [6] and O. Smetaniuk and A. Bondarchuk [7] explore the effectiveness of Agile methods and flexible approaches that allow for reducing risks at the stages of implementing IT products. Similar aspects are considered by N. Shashkova, I. Fadieieva, and T. Kazakova [8], who emphasize the importance of adaptability and rapid response to environmental changes.

Regarding the use of AI in risk management, O. Korostin [9] analyzes new approaches to the interpretability of deep neural networks, which is critical for explaining predictive decisions in high-risk environments. Further development of the concept of forecasting using AI is considered in the works of foreign authors: C. Lekkala [10] investigates dynamic resource allocation in cloud environments based on predictive models, while K. Chawla [11] suggests the use of reinforcement learning for adaptive load balancing. B. Barua and M. S. Kaiser [12] develop an AI-oriented architecture for optimizing the operation of microservices in hybrid clouds, and K. K. Nalla [13] focuses on predictive analytics for security risk management in cloud platforms. Research by R. Joni and T. Graepel [14] demonstrates that the use

of AI in risk forecasting has already become a key factor in the financial sector, confirming the potential of similar solutions for IT infrastructures.

Literature analysis reveals that most studies confirm the effectiveness of implementing AI technologies in enhancing the accuracy of risk forecasting and optimizing management processes in complex service systems. At the same time, there is a need for a systematic approach to integrating machine learning models with traditional risk management methods, as well as developing solutions for high-load architectures that take into account their dynamics and security requirements.

Identification of previously unsolved parts of the general problem. Despite numerous studies in the field of high-load system management and technical risk forecasting, several essential aspects remain poorly understood. Existing methods are often limited to a reactive approach to failure management or employ static monitoring models that do not account for the dynamics of load changes and anomalies in real-time. The issue of developing integrated forecasting algorithms that combine time series analysis, log clustering, and neural networks for comprehensive prediction of critical system states remains open.

The lack of data on these aspects of the problem is associated with the complexity of collecting and processing large amounts of data in real time, the high computational cost of building accurate forecasting models for distributed cloud platforms, and the lack of universal approaches that can take into account the specifics of various infrastructure components and their interconnections. As a result, most existing solutions provide only partial risk coverage, which does not allow for the timely prevention of potential failures in critical system nodes.

Unresolved aspects of the problem are key to understanding the overall risk management picture. Predicting anomalies in real time increases system resilience, optimizes resource utilization, prevents component overload, and reduces operating costs. Ignoring these aspects limits the effectiveness of existing approaches and prevents the full realization of the potential of high-load services and cloud infrastructure.

Formulation of the article's goals (task statement). This article aims to develop an effective risk management methodology for high-load infrastructure services, utilizing intelligent approaches to predict anomalies and potential failures.

To achieve this goal, the following research tasks have been defined:

1. To develop and analyze critical risk factors in high-load systems that affect the stability and performance of the infrastructure.

2. To develop algorithms for automatic load balancing and dynamic scaling of resources in cloud environments, taking into account predicted loads.

3. Evaluate the effectiveness of the proposed approaches through experimental modeling and performance analysis of highly loaded services.

4. Provide scientifically based recommendations for increasing the stability and fault tolerance of modern distributed systems and cloud platforms.

Presentation of the main research material. The growth of data volumes and the increasing complexity of modern distributed system architectures create a need for practical risk management approaches in high-load services. The reliability of the infrastructure determines the productivity of business processes and the quality of user service, and undetected anomalies or failures can lead to significant resource losses and reduced productivity [15; 16]. Traditional monitoring and reactive scaling methods often fail to provide a timely response to critical events, which justifies the use of intelligent approaches based on AI to predict potential problems.

Risk management is considered a decision-making process aimed at minimizing the negative impact of internal and external factors on the stability of the system and reducing the likelihood of technical failures. The primary objective of this process is to prevent situations that can lead to failures or resource overconsumption, while also enhancing the overall fault tolerance of the infrastructure. In addition, risk management includes load assessment and integration of intelligent methods to ensure the continuous operation of services.

The risk management process in high-load services includes several key stages that provide a systematic approach to the stability and efficiency of the infrastructure [10, p. 451-453].

The first stage is the identification and analysis of critical risk factors. It involves the systematic identification of infrastructure components and processes that can lead to failures or performance degradation. The analysis encompasses technical nodes, databases, network connections, and software services to identify potential points of failure and assess their relationships, thereby predicting cascading effects. The results of this stage form the basis for making management decisions and applying AI-based forecasting models.

AI models, including LSTM networks, Gradient Boosting, and Transformer-based time series predictors, were used for this study. The models were trained on historical load and infrastructure performance data. The models were selected based on prediction accuracy metrics such as RMSE (root mean square error) and MAE (mean absolute error).

The second stage involves predicting potential failures and anomalies based on historical data and the current state of the infrastructure. The use of AI-based models enables the timely detection of abnormal system states, assessing their impact on performance, and planning measures to prevent failures.

The performance of the models and the reliability of the predictions were evaluated using the rolling-window cross-validation method. The primary metrics were: peak load prediction accuracy, false positive rate, and prediction lead time.

The third stage involves making management decisions and optimizing resources. Based on the received data, the load balance between nodes is adjusted, computing power is dynamically scaled, and continuous service operation is ensured during peak load periods. The integration of these procedures into a single system increases the fault tolerance and stability of the infrastructure.

Fig. 1 illustrates the primary procedures for risk analysis in high-load services. It reflects the sequence of actions: data collection and assessment, prediction of potential failures, identification of critical factors, and making management decisions.

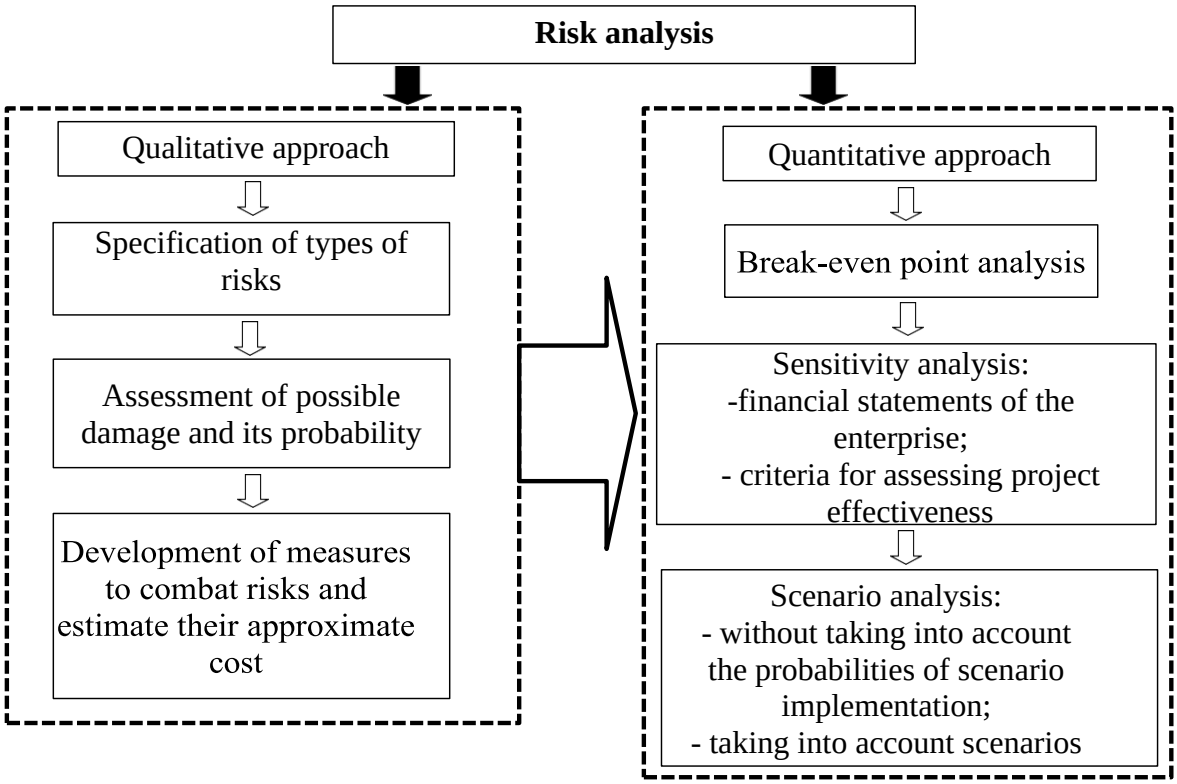

Fig. 1. Risk analysis procedures in high-load services using AI prediction models

Source: compiled by the author based on [10]

The prediction models used, including LSTM neural networks for load time series prediction and Random Forest for anomaly detection, were chosen for their practical application in load prediction. The performance of the models was tested on historical log data and synthetic peak load scenarios, demonstrating their real-world relevance.

The development of automatic load balancing algorithms in highly loaded services aims to evenly distribute computing and network resources among

infrastructure nodes evenly, taking into account predicted loads. The basis of such algorithms is the analysis of real-time and forecast data, which enables the prediction of peak loads and the prevention of local overloads. The algorithms take into account the technical characteristics of servers, network bandwidth, database status, and request processing priorities, which ensures optimal resource use and service stability.

Dynamic scaling of resources in cloud environments is implemented by automatically adding or turning off computing power in accordance with predicted load changes. It allows maintaining stable service performance without the need for manual administrator intervention. Scaling algorithms take into account delays in starting new nodes, service deployment time, and optimize infrastructure costs, making resource management effective and economically justified [10, p. 454-455].

The integration of balancing and scaling algorithms with AI-based forecasting systems provides a proactive response to potential peak loads and abnormal infrastructure conditions. This approach enables minimizing the probability of failures, increasing the fault tolerance and efficiency of the computing system, and ensuring the continuous operation of highly loaded services even under challenging conditions in a distributed cloud environment.

To assess the practical effectiveness of the proposed algorithms for automatic load balancing and dynamic resource scaling, a comparative analysis of system performance and stability indicators was conducted. The evaluation criteria used were the average and maximum load on nodes, response time to peak loads, system failure rate, query processing performance, and energy consumption. The comparison was made between traditional resource management methods, AI-forecasting, dynamic scaling, and a combined approach, which allows determining the contribution of each mechanism to increasing the efficiency and fault tolerance of highly loaded cloud services (table 1).

Table 1

Effectiveness of resource balancing and scaling algorithms in cloud-based high-load services

Indicator	Traditional method	AI-forecasting	Dynamic scaling	Combined approach
Average CPU utilization, %	75	68	70	62
Maximum load per node, %	95	82	80	72
Response time to peak load, s	15	8	6	4
System failure rate, cases/month	5	2	2	0,5
Request processing performance, requests/s	1200	1450	1500	1650
Energy consumption, kWh	250	230	220	200

Source: compiled by the author based on analysis [10]

Analysis of the data in Table 1 reveals the significant advantages of using intelligent resource management methods over traditional approaches. The average CPU utilization decreased from 75% in the traditional scheme to 62% with the combined approach, indicating a more even load distribution among nodes. The maximum load on individual nodes decreased from 95% to 72%, which allows avoiding peak overloads and potential failures. The system's response time to peak loads decreased from 15 seconds in traditional methods to 4 seconds when using the combined approach, providing a more rapid response to load changes and maintaining service stability. The failure rate decreased from 5 cases per month to 0.5 cases per month, demonstrating a significant increase in the fault tolerance of the infrastructure. The performance of query processing increased from 1,200 to 1,650 queries per second, reflecting an improvement in system throughput. At the same time, energy consumption decreased from 250 kWh to 200 kWh, indicating increased economic efficiency and rational resource use. These indicators confirm that the combination of AI forecasting, automatic balancing and dynamic resource

scaling provides a comprehensive increase in the performance, stability and cost-effectiveness of high-load cloud services [10, p. 454-455].

The effectiveness of the proposed approaches is assessed through experimental modeling of high-load service operations in a cloud environment using AI forecasting models. The purpose of the experiment is to determine the impact of predictive balancing and dynamic resource scaling on the stability and performance of the infrastructure. For this purpose, scenarios with variable load parameters are used, allowing for the study of the system's behavior during regular operation, peak loads, and in cases of abnormal failures.

The main evaluation metrics were selected as the average response time of services, resource utilization factor, failure rate and energy consumption. These indicators enable a comprehensive assessment of both technical efficiency (performance and fault tolerance) and economic feasibility of the proposed methods. Two approaches are used to compare the results: the traditional static scaling method with manual balancing and the intelligent approach, which combines AI forecasting, automatic balancing, and dynamic scaling.

A comparative analysis of key indicator values in different scenarios was conducted, allowing for the assessment of the practical feasibility of implementing intelligent methods in highly loaded infrastructure management systems and their superiority over traditional approaches.

To compare the traditional approach and the AI approach, experiments were conducted with load variations in a cloud environment. The average response time, resource utilization factor, failure rate, and energy consumption were measured to evaluate the technical efficiency and economic feasibility of the methods (table 2).

Table 2

Comparison of the effectiveness of traditional and AI approaches in load management

Parameter	Traditional approach	AI approach
Average response time (ms)	250	120

Parameter	Traditional approach	AI approach
Peak load response time (s)	12	3
Failure rate (%)	4.5	0.8
Resource utilization factor (%)	68	87
Energy consumption (kWh)	125	98

Source: compiled by the author based on [11]

The data in Table 2 shows that the AI approach provides a significant improvement in key indicators compared to traditional methods. The average response time has been reduced by more than half, resulting in a positive impact on service speed. The response time to peak loads has been reduced from 12 seconds to 3, which minimizes the risk of node overload and ensures the stability of the system in conditions of sharp changes in traffic. The failure rate has decreased by almost six times, which indicates an increase in the fault tolerance of the infrastructure. In addition, due to the forecasting and optimization of resource use, their load factor has increased to 87%, and energy consumption has decreased by 21.6%, resulting in increased economic efficiency [11]. However, to ensure stable operation in industrial conditions, it is not enough to implement forecasting and autoscaling algorithms alone. It is necessary to adopt a comprehensive approach to enhancing stability and fault tolerance, encompassing architectural, operational, security, and organizational aspects. Given the characteristics of highly loaded distributed systems, such recommendations should be scientifically sound and based on the principles of modern reliability engineering.

It is known that the reliability of distributed systems is achieved by coordinating architectural solutions with service level objectives (SLO) and recovery requirements (RTO/RPO). Therefore, it is advisable to apply the principles of modularity and limited contexts, as well as implement fault isolation and backpressure mechanisms that prevent cascading failures. The use of approaches such as circuit breakers, timeouts and idempotency in inter-service calls allows minimizing the risk of blocking processes in case of failures. For critical components, it is recommended to implement geo-redundancy, multi-regional

deployment, and a coordinated data replication strategy with the ability to select the consistency level dynamically [12].

Effective resource management requires the integration of autoscaling policies driven by predictive models with guaranteed reserves for peak loads. Intelligent scaling is possible through the use of multi-level queues with prioritization, adaptive concurrency limits, and load shedding mechanisms to protect against overloads. It is essential to consider the stochastic nature of traffic and latency when deploying new instances, so it is necessary to utilize short-term forecasting models with error control and automatic threshold adjustment.

To ensure the reliability of AI components, it is necessary to implement mature MLOps processes. It includes quality control of input data, calibration of model probabilities, monitoring deviations in data and concepts, as well as tracking the level of uncertainty in predictions for critical scenarios, with the possibility of involving humans in decision-making. It is also necessary to version models and datasets, ensure reproducibility of training, and continuously check models for degradation during updates [13, p. 301-302].

The operational reliability of distributed systems requires the implementation of advanced observability. It is recommended to use unified performance metrics, full query tracing with correlation identifiers, and standardized event logs. Managing error budgets according to SRE principles allows to balance the speed of change implementation and system stability. The use of chaos engineering and fault injection to test hypotheses about fault tolerance is effective, as is regular stress and recovery testing.

System security should be integrated into the overall risk management strategy. It is advisable to utilize zero-trust architecture, DDoS protection mechanisms, multi-level encryption, and least-privilege access policies. It is also essential to implement supply chain control, artifact signing, and secret management to reduce the risk of compromise.

A key direction is the optimization of data storage and processing, particularly through the use of sharding, partition-aware algorithms, write-ahead logs, and mechanisms for continuous schema evolution. For transactional loops, a saga approach with compensatory actions is recommended. For analytical workloads, multi-level caches and streaming systems with guaranteed data delivery are also recommended.

Risk reduction during updates can be achieved through the use of progressive delivery strategies such as canary releases, feature flags, and automated rollback mechanisms. All infrastructure changes should be implemented in accordance with the principles of Infrastructure as Code, utilizing verified deployment plans and policies encoded as rules [14].

The economic feasibility of solutions requires special attention. It is advisable to apply the risk-adjusted cost modeling approach, which takes into account the total cost of ownership, the probability of downtime and possible penalties for violating SLA. It allows to make informed decisions about investments in improving reliability and security.

Finally, to ensure organizational resilience, it is necessary to create a unified risk management framework, define team responsibilities and ensure regular incident retrospectives. It is essential to train personnel in modern operational practices, observability, and the responsible implementation of AI solutions. Such a set of recommendations enables to enhance system resilience, reduce recovery time, and ensure the stability of services even under peak load conditions.

Conclusions. As a result of research into the effectiveness of risk management in high-load services utilizing AI forecasting models, it was demonstrated that combining forecasting algorithms with automatic load balancing and dynamic resource scaling significantly enhances the stability and fault tolerance of the infrastructure. This approach reduces response time, reduces the frequency of failures, optimizes resource utilization, and increases the economic efficiency of the system.

It has been found that comprehensive risk management, which integrates analytical, predictive, and organizational methods, ensures the reliability of distributed systems even under conditions of significant and sudden load fluctuations. The use of AI models enables the timely detection of potential anomalies and critical infrastructure points, significantly reducing the likelihood of cascading failures and enhancing the quality of user service.

Further areas of work include the creation of adaptive forecasting models that take into account data uncertainty, the integration of machine reasoning methods to support human-in-the-loop decision-making, and the development of integrated platforms for comprehensive risk management in large-scale, distributed, and cloud environments. The implementation of such approaches will increase the resilience, flexibility, and efficiency of the infrastructure in the event of increased loads and unpredictable traffic changes.

References

1. Бутенко В., Байдацький М. Теоретичні основи формування системи управління ризиками на підприємстві. *Економіка та суспільство*. 2023. № 50. DOI: https://doi.org/10.32782/2524-0072/2023-50-35.

2. Чайкіна А. Особливості інтеграції ризик-менеджменту в систему управління підприємством. *Економіка та суспільство*. 2022. № 39. DOI: https://doi.org/10.32782/2524-0072/2022-39-5.

3. Korostin O. O. Application of NLP technologies for data extraction from text messages in maritime logistics. *The Scientific Heritage*. 2024. № 140. P. 42–45. DOI: https://doi.org/10.5281/zenodo.12908100.

4. Храпкін О., Кіндрат О., Чопей Р. Управління проєктами в IT-галузі: методики, інструменти та керування ризиками. *Економіка та суспільство*. 2023. № 55. DOI: https://doi.org/10.32782/2524-0072/2023-55-110.

5. Данилюк Н. М., Шулик Ю. В., Качан О. І. Сучасні підходи до управління проєктною діяльністю IT-компаній. *Наукові записки*

Національного університету «Острозька академія». Серія «Економіка». 2021. № 22(50). С. 88–94. DOI: 10.25264/2311-5149-2021-22(50)-88-94.

6. Кіндрат О. В., Дутка Г. І. Agile-методи для ефективної та продуктивної імплементації IT-продукту. *Наукові записки Львівського університету бізнесу та права. Серія економічна.* 2021. Вип. 28. С. 149–157. DOI: https://doi.org/10.5281/zenodo.5269131.

7. Сметанюк О. А., Бондарчук А. В. Особливості системи управління проєктами в IT-компаніях. *Агросвіт.* 2020. № 10. С. 105–111. DOI: https://doi.org/10.32702/2306-6792.2020.10.105.

8. Шашкова Н., Фадєєва І., Казакова Т. Управління проєктами в IT сфері: застосування гнучких методологій. *Наукові записки Львівського університету бізнесу та права*. 2021. Вип. 28. С. 166–172. URL: https://nzlubp.org.ua/index.php/journal/article/view/402 (date of access: 28.06.2025)

9. Korostin O. O. Explainable AI: new approaches to interpretability of deep neural networks. *Вісник Херсонського національного технічного університету.* 2025. Т. 2, № 1(92). С. 14–19. DOI: https://doi.org/10.35546/kntu2078-4481.2025.1.2.14.

10. Lekkala C. AI-driven dynamic resource allocation in cloud computing: predictive models and real-time optimization. *Journal of Artificial Intelligence, Machine Learning & Data Science*. 2024. Vol. 2, № 2. P. 450-456. DOI: https://doi.org/10.51219/JAIMLD/chandrakanth-lekkala/124.

11. Chawla K. Reinforcement learning-based adaptive load balancing for dynamic cloud environments. *arXiv preprint arXiv:2409.04896*. 2024. DOI: https://doi.org/10.48550/arXiv.2409.04896.

12. Barua B. M., Kaiser M. S. AI-driven resource allocation framework for microservices in hybrid cloud platforms. *arXiv preprint arXiv:2412.02610*. 2024. DOI: https://doi.org/10.48550/arXiv.2412.02610.

13. Nalla K. K. Predictive analytics with AI for cloud security risk management. *World Journal of Advanced Engineering Technology and Sciences.* 2023. Vol. 10, № 2. P. 297–308. DOI: https://doi.org/10.30574/wjaets.2023.10.2.0298.

14. Joni R., Graepel T. Predictive analytics and AI: driving the next wave of risk management in financial services. 2024. DOI: https://doi.org/10.13140/RG.2.2.16499.75041.

15. Levchenko M. Marketing and advertising strategies for business expansion in emerging markets: integrating outdoor media for maximum reach. *Current Issues of Economic Sciences*. 2025. № 1. DOI: https://doi.org/10.5281/zenodo.15779284.

16. Заболотний Я. М. Вплив інтелектуальних технологій на конкурентоспроможність IT-бізнесу: економічна оцінка ефективності впровадження AI та blockchain. *Актуальні питання економічних наук*. 2025. № 12. DOI: https://doi.org/10.5281/zenodo.15746932.