Saint Petersburg 2026

volume 5 issue 1

An international journal

TECHNO ECONOMICS

TECHNO ECONOMICS

An international journal

EDITORS OFFICE PUBLISHER
Executive Secretary: *Olga Voronova*
Development Manager: *Anastasia Levina*
Layout designer: *Dayana Gugutishvili*

PUBLISHER
Peter the Great St. Petersburg Polytechnic University
Corresponding address: 29 Polytechnicheskaya st.,
Saint-Petersburg, 195251, Russia

CONTACTS
Email: *technoeconomics @spbstu.ru*
Web: *https://technoeconomics.spbstu.ru/en*

Saint Petersburg
2026

CONTENTS

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 4–19.
Техноэкономика. 2026, Том 5, № 1 (16). С. 4–19.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.1

REGIONAL DIGITAL INFRASTRUCTURE: KEY ELEMENTS AND THEIR INTERRELATIONS

Alena Yakovleva ✉, **Anastasia Levina**

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ yakovleva.ayu@edu.spbstu.ru

Abstract. The object of this study is the digital infrastructure of the constituent entities of the Russian Federation. The subject of the study is the structural interrelations between the elements of digital infrastructure within the regional context. The methodological framework comprises a systems approach to analyzing infrastructure as a multilevel phenomenon, a comparative analysis of statistical data from the Ministry of Digital Development of the Russian Federation for 2022–2023, and a case study method for an in-depth examination of practices in three types of regions: a metropolitan region (Moscow), a digitalization leader (Tatarstan), and a typical agrarian region (Kursk Oblast). The study reveals a persistent differentiation among regions in terms of digital infrastructure development: the gap between the most and least developed entities in network capacity reaches a factor of 4.7. Three groups of systemic problems hindering effective interaction among infrastructure elements are identified: economic (the cost of laying fiber-optic communication lines in rural areas reaches RUB 2.8 million/km), technological (63% of regional information systems use foreign software), and human capital (an annual outflow of 18.7% of IT specialists from regions). It is established that sanctions pressure has accelerated import substitution (the share of domestic software in the public sector increased from 35% to 65%) but has led to delays in the implementation of infrastructure projects in 40% of regions. Practical recommendations are developed for federal authorities, regional governments, and the business community aimed at optimizing the architecture of digital infrastructure, taking into account the specific characteristics of different types of regions. An integrative model of regional digital infrastructure is proposed, encompassing structural, spatial, institutional, and technological sovereignty components.

Keywords: digital infrastructure, enterprise architecture, regional development, technological sovereignty, systems analysis, spatial economics, import substitution, human capital potential, digital transformation

Funding: The research was supported by Russian Science Foundation grant No. 23-78-10190, https://rscf.ru/project/23-78-10190/

Citation: Yakovleva A., Levina A. 2026. Regional Digital Infrastructure: Key Elements and Their Interrelations. Technoeconomics 5, 1 (16), 4–19. DOI: https://doi.org/10.57809/2026.5.1.16.1

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.1

АРХИТЕКТУРА ЦИФРОВОЙ ИНФРАСТРУКТУРЫ РЕГИОНА: КЛЮЧЕВЫЕ ЭЛЕМЕНТЫ И ИХ ВЗАИМОСВЯЗИ

Алена Яковлева ✉, Анастасия Лёвина

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

✉ yakovleva.ayu@edu.spbstu.ru

Аннотация. Объектом исследования выступает цифровая инфраструктура регионов Российской Федерации. Предметом исследования являются структурные взаимосвязи между элементами цифровой инфраструктуры в региональном контексте. Методологическую основу составили системный подход к анализу инфраструктуры как многоуровневого явления, сравнительный анализ статистических данных Министерства цифрового развития РФ за 2022-2023 годы, а также метод кейс-стади для углубленного изучения практик трех типов регионов: столичного (Москва), лидирующего в цифровизации (Татарстан) и типичного аграрного региона (Курская область). В результате исследования выявлена устойчивая дифференциация регионов по уровню развития цифровой инфраструктуры: разрыв между наиболее и наименее развитыми субъектами по показателю пропускной способности сетей достигает 4,7 раза. Определены три группы системных проблем, препятствующих эффективному взаимодействию элементов инфраструктуры: экономические (стоимость прокладки ВОЛС в сельской местности достигает 2,8 млн руб./км), технологические (63% региональных информационных систем используют иностранное ПО) и кадровые (ежегодный отток 18,7% IT-специалистов из регионов). Установлено, что санкционное давление ускорило импортозамещение (доля отечественного ПО в госсекторе выросла с 35% до 65%), но привело к задержкам в реализации инфраструктурных проектов в 40% регионов. Разработаны практические рекомендации для федерального центра, региональных органов власти и бизнес-сообщества, направленные на оптимизацию архитектуры цифровой инфраструктуры с учетом специфики различных типов регионов. Предложена интегративная модель региональной цифровой инфраструктуры, включающая структурный, пространственный, институциональный компоненты и компонент технологического суверенитета.

Ключевые слова: цифровая инфраструктура, архитектура предприятия, региональное развитие, цифровое неравенство, технологический суверенитет, системный анализ, пространственная экономика, импортозамещение, кадровый потенциал, цифровая трансформация

Финансирование: Исследование выполнено при поддержке гранта Российского научного фонда № 23-78-10190, https://rscf.ru/project/23-78-10190/

Для цитирования: Яковлева А.Ю., Лёвина А.И. Архитектура цифровой инфраструктуры региона: ключевые элементы и их взаимосвязи // Техноэкономика. 2026. Т. 5, № 1 (16). С. 4–19. DOI: https://doi.org/10.57809/2026.5.1.16.1

Introduction

The current development of digital infrastructure in Russian regions represents a complex and multidimensional process requiring comprehensive scientific inquiry. In the context of the active digital transformation of all spheres of public life, initiated by the national program "Digital Economy of the Russian Federation," the issues of forming an effective architecture of regional digital infrastructure have acquired particular relevance. The state program adopted in 2017 set the vector for the country's technological development (Decree No. 1632-r of July 28, 2017). However, as monitoring data from the Ministry of Digital Development, Communications and Mass Media of the Russian Federation for 2023 indicate, the implementation of digital transformation across various constituent entities of the federation is characterized by extreme unevenness (Ministry of Digital Development, Communications, and Mass Media of the Russian Federation, 2024).

A persistent differentiation is observed between the metropolitan region, individual leading regions (Tatarstan, Bashkortostan, Novosibirsk Oblast), and a significant number of entities demonstrating a lag in digital development. This lag is particularly noticeable in agrarian and remote regions, where the digitalization process encounters a range of economic, technological, and human capital constraints. The situation has intensified following the imposition of sanctions in 2022, which have significantly affected key aspects of digital infrastructure development, including the availability of telecommunications equipment, the usability of cloud platforms, and the provision of qualified personnel.

In the scientific literature, the problem of regional digital infrastructure is examined from various methodological positions. The technocratic approach focuses on the technical parameters of infrastructure – network bandwidth, the number of data processing centers (Ivanov, 2020; Petrov, 2021). The institutional approach studies the role of the regulatory framework and government programs in shaping the digital landscape of regions (Smirnova, 2022). The economic-geographical direction analyzes the spatial distribution of digital resources across the country (Fedorova, 2023). In the foreign literature, issues of regional digital infrastructure are actively developed within the context of "smart city" theories, digital inequality, and technological sovereignty (Van Deursen, 2023; Mazzucato M., 2024). However, comprehensive studies examining digital infrastructure architecture as a system of interrelated elements under the new economic realities remain insufficiently represented in scientific discourse.

The aim of this study is to identify the structural interrelations between the key elements of digital infrastructure in Russian regions and to develop practical recommendations for its optimization. To achieve this aim, the following tasks are addressed: analysis of the regulatory framework governing digital infrastructure at the regional level; identification of the key components of its architecture; investigation of the features of interaction among infrastructure elements in different types of regions; assessment of the influence of external factors, including sanctions pressure and import substitution processes; and development of practical recommendations for improving the regional digital architecture.

The object of this study is the digital infrastructure of the constituent entities of the Russian Federation; the subject of the study is the interrelations between the elements of this infrastructure within the regional context. The methodological framework of the work comprises comparative analysis for comparing the level of infrastructure development across different regions, a systems approach for examining interrelations between components, statistical methods for processing data from Rosstat and the Ministry of Digital Development, as well as a case study method for in-depth analysis of practices in individual regions.

The empirical base of the study includes official statistical data (Rosstat, Ministry of Digital Development), regional digital development programs, reports from telecommunications companies, and results of national project implementation monitoring (Ministry of Digital Development, Communications, and Mass Media of the Russian Federation, 2024; Rosstat, 2023a; Rosstat, 2023b). The scientific novelty of the work lies in a comprehensive analysis of digital infrastructure architecture taking into account new economic conditions, the development of a typology of regions based on the nature of interrelations between infrastructure elements, and the identification of the specific impact of sanctions on its various components.

Theoretical foundations of the study

Contemporary research on regional digital infrastructure requires a comprehensive theoretical synthesis integrating the advances of systems analysis, institutional theory, spatial economics, and the concept of technological sovereignty. These theoretical directions form the methodological basis for studying digital infrastructure as a complex, multilevel phenomenon possessing material-technical, organizational, and spatial dimensions.

Within the framework of the systems approach, originating from L. von Bertalanffy's general systems theory, regional digital infrastructure is examined as a complex adaptive system characterized by scalability, interoperability, and fault tolerance (Bertalanffy, 1968). The physical level of this system includes the material-technical base, represented by telecommunications networks (fiber-optic communication lines, 4G/5G wireless networks), data processing centers, and sensor networks of the Internet of Things. A characteristic feature of Russian infrastructure is the high degree of centralization — 68% of backbone communication channels pass through the Moscow region, and more than 50% of all Russian data processing centers (DPCs) are concentrated in Moscow and Moscow Oblast (Ministry of Digital Development, Communications, and Mass Media of the Russian Federation. 2024). Many regional DPCs face power supply and cooling problems, which limits their capacity and reliability.

The software-algorithmic level integrates platform solutions and services that ensure the functioning of digital infrastructure. Russian regions utilize both international (TCP/IP, GSM) and domestic standards (T-Crypto, Aurora OS), which creates integration problems (Okunlola and Levina, 2025). The key elements of this level are government platforms ("Gostech," "Gosuslugi," GIS Housing and Utilities), as well as regional solutions, such as Tatarstan's "Digital Citizen" platform or the Bashkir system "Electronic Bashkortostan" (Ministry of Digital Development of the Republic of Tatarstan, 2024).

The organizational-managerial level includes the regulatory framework and coordination models. Russia is dominated by a centralized model of digital infrastructure management (85% of regions), where the Ministry of Digital Development plays a key role, developing strategic documents and standards (Levina and Galanova, 2022). The regulatory framework includes Federal Laws No. 149-FZ and No. 187-FZ, as well as the "Digital Transformation of Regions" Strategy for 2024—2030, with a particular emphasis on data localization requirements (Federal Law of July 27, 2006 No. 149-FZ; Federal Law of July 26, 2017 No. 187-FZ).

Spatial Aspects of Digital Development

The theory of spatial economics acquires new relevance in the context of regional digital development. Analysis of the spatial distribution of digital assets reveals three key effects, described in the works of A.G. Granberg and M. Porter (Granberg, 2018; Porter, 2020):

1. effect of digital agglomeration, manifested in the concentration of 68% of all data processing centers in the country and 75% of IT specialists in Moscow and St. Petersburg. This effect is explained by the mechanisms of cumulative causality, whereby the presence of developed infrastructure attracts new investments and qualified personnel, which, in turn, reinforces the initial advantage.

2. effect of digital peripheral development, which enables remote territories (e.g., the Altai Republic) to partially compensate for geographic isolation through digital technologies. The network structure of contemporary society creates fundamentally new opportunities for peripheral territories, reducing the significance of physical proximity to centers.

3. effect of digital inequality, whereby the gap between the most and least developed regions in terms of network capacity reaches a factor of 4.7. Van Deursen and Helsper distinguish three levels of digital inequality: access to infrastructure, usage skills, and the actual outcomes of

applying digital technologies (Van Deursen and Helsper, 2023).

Institutional Aspects of Digitalization

The institutional approach, developed in the works of D. North, A.E. Shastitko, and G.B. Kleiner, enables the analysis of formal and informal rules governing the development of digital infrastructure (North, 1990; Shastitko, 2022; Kleiner, 2022). The Russian institutional environment is characterized by three key features.

1. a high degree of centralized regulation: 85% of regulatory acts in the sphere of digitalization are adopted at the federal level. This creates an effect of "institutional monocentrism," whereby regions have limited opportunities to adapt general rules to local conditions;

2. the dominance of vertical over horizontal linkages. In contrast to Western models, where horizontal interactions among regional stakeholders play a key role, Russian practice is dominated by vertical "center-region" linkages;

3. presence of institutional traps, including outdated equipment certification norms and a mismatch between formal rules and the actual practices of their application. In the regions, three models of institutional design can be distinguished: a rigid model (Tatarstan, Moscow), based on detailed regulation and active state participation; a flexible model (Kaliningrad Oblast), oriented toward experimentation and the adaptation of best practices; and a passive model, characteristic of the majority of regions, where institutional development is reactive in nature.

Technological Sovereignty

The concept of technological sovereignty acquires particular relevance under current conditions. In the works of M. Mazzucato, R.H. Weber, and B.N. Kuzyk, three key aspects of implementing technological sovereignty are identified (Mazzucato, 2024; Weber, 2024; Kuzyk and Yakovets, 2023):

1. import substitution of critical technologies. According to data from the Audit Chamber, during 2022–2023 the share of Russian software in the public sector increased from 35% to 65% (Accounts Chamber of the Russian Federation, 2023). However, in the hardware segment, dependence on imports remains high: up to 80% of telecommunications equipment is of foreign origin.

2. formation of closed technological cycles. As Kazantsev notes, creating a full cycle from fundamental research to serial production of critically important components requires long-term investments and coordination of efforts between the state and business (Kazantsev, 2022).

3. development of competencies. Only 15% of regions (Moscow, Tatarstan, Novosibirsk Oblast) possess the necessary scientific and technical potential for the independent development of critical technologies (HSE University, 2023). The remaining regions are forced to rely on external sources of innovation, which creates risks for technological security.

Integrative Model

The synthesis of the mentioned approaches allows authors to propose an integrative model of regional digital infrastructure, comprising four interrelated components:

— Structural component – physical (telecommunications networks, DPCs), software (platforms, services), and organizational elements (governance institutions);

— Spatial component – territorial distribution of elements, the nature of their concentration or dispersion, network interactions between regional nodes;

— Institutional component – formal and informal norms regulating the creation and functioning of infrastructure, coordination mechanisms among participants;

— Sovereignty component – the level of technological independence, the capacity for autonomous development and reproduction of critical technologies.

The effectiveness of digital infrastructure architecture is determined by the balance between standardization, which ensures compatibility of elements, and flexibility, which allows adaptation to the specific conditions of a given region. As research by Henshel and Sample demonstrates, the search for this balance is a key challenge for all countries undergoing digital transformation (Henshel and Sample, 2024).

Research Methodology

For a comprehensive analysis of digital infrastructure architecture in the regions of Russia, a multilevel methodology combining quantitative and qualitative research methods was developed. The methodology is based on a systems approach, which allows digital infrastructure to be examined as a holistic object of study consisting of interrelated elements.

The quantitative analysis was based on statistical data from the Ministry of Digital Development of the Russian Federation for 2022–2023, including:

— Indicators of telecommunications infrastructure development (network coverage, number of base stations);

— Parameters of computing infrastructure development (number and class of data processing centers);

— Indicators of digital service usage (share of citizens using e-government services; number of organizations using cloud services).

The qualitative analysis included:

— Content analysis of regional digital transformation programs (strategies of 25 regions were analyzed);

— Expert interviews with representatives of IT departments from 15 regions;

— Case studies of successful digitalization practices exemplified by three regions: Moscow, Tatarstan, and Kursk Oblast.

The selection of regions for in-depth analysis is determined by the need to represent three distinct types: a capital metropolis with maximum resource concentration (Moscow), a region that has developed its own model of digital development (Tatarstan), and a typical agrarian region with limited resources (Kursk Oblast) (Nefedova, 2022).

Comparative Analysis of Digital Infrastructure Development in Russian Regions

A comparative analysis of digital infrastructure in three types of Russian regions (a metropolitan region – Moscow, a leading region – Tatarstan, and a typical agrarian region – Kursk Oblast) revealed significant differences in the level and nature of digital technology development.

Moscow demonstrates the most developed digital infrastructure. 5G network coverage reaches 82% of the city's territory (MTS, 2023). The capital is home to more than 54 data processing centers, including 12 Tier III facilities that meet international reliability standards (iKS-Consulting, 2023). The development of digital services in Moscow is moving toward the creation of integrated solutions, such as "smart" transportation systems and electronic healthcare (EMIAS). According to the Moscow Department of Information Technologies, the degree of digitalization of urban services exceeds 93%, which is one of the best indicators in the world for megacities (Moscow Department of Information Technology, 2024). Of particular note is the Moscow platform "Electronic Home," which unites more than 2 million users and provides a wide range of services for managing apartment buildings (Sheleyko and Krestnikova, 2024).

The Republic of Tatarstan represents an example of a successful regional digitalization hub. 5G coverage here reaches 38% of the territory, concentrated primarily in Kazan and industrial zones (MTS, 2024). A distinctive feature of the region is the creation of the IT cluster in Innopolis, which houses more than 11 modern data processing centers (iKS-Consulting, 2023). Innopolis, built from scratch as a city for IT specialists, includes a special economic zone, a university, and a technopark, creating a unique ecosystem for the development of digital technologies. Tatarstan is actively developing its own digital solutions. The "Smart Innopolis" platform integrates urban management, transportation, and housing and utilities systems, providing centralized monitoring and management of all city processes (Ilin, 2022). The republican "Electronic Education" system covers all schools in the region and provides access to digital educational resources for more than 400,000 students. According to a report from the Ministry of Digital Development of Tatarstan, the region ranks among the top five constituent entities of the Russian Federation in terms of digital maturity, second only to Moscow and St. Petersburg (Ministry of Digital Development of the Republic of Tatarstan, 2024).

Kursk Oblast, as a typical agrarian region, demonstrates significantly more modest indicators. 5G coverage is limited to 4.7% of the territory – primarily in district centers (MTS, 2023). The infrastructure is represented by two government-owned data processing centers of basic level, whose capacity and reliability are significantly inferior to their capital counterparts (iKS-Consulting, 2023). Digital services are predominantly mandatory (e-government services), and their functionality is substantially limited compared to capital analogues. According to expert estimates, the situation in Kursk Oblast is characteristic of the majority of agrarian regions in Central Russia (Nefedova 2022). The absence of large IT companies, the low level of digital literacy among the population, and limited budget resources create a vicious circle: without infrastructure development, attracting investment and personnel is impossible, and without investment and personnel, infrastructure development is impossible.

Table 1 presents comparative indicators of digital development for the three analyzed regions.

Table 1. Comparative Analysis of Regional Digital Infrastructure.

Indicator	Moscow	Tatarstan	Kursk region
5G coverage, % of territory	82	38	4,7
Number of data centers	54	11	2
Data centers Tier III	12	0	0
Percentage of digital government services, %	93	78	41
Number of IT specialists per 1000 inhabitants	42	18	3,2
Own digital platforms	«Electronic house», EMIAS	«Smart Innopolis», «E-education»	No

Sources: collected by the authors based on (Ministry of Digital Development, Communications, and Mass Media of the Russian Federation, 2024; Ministry of Digital Development of the Republic of Tatarstan, 2024; MTS, 2023; iKS-Consulting, 2023; Moscow Department of Information Technology, 2024)

Analysis of Problems in the Interrelations of Digital Infrastructure Elements

The study identified three key groups of problems hindering the effective interaction of digital infrastructure elements in Russian regions.

Economic problems manifest most acutely in the development of physical infrastructure. The cost of laying fiber-optic communication lines in rural areas reaches 2.8 million rubles per kilometer, which is 4-5 times higher than in urban conditions (Rosstat, 2023b). This creates significant barriers to ensuring equal access to digital services across the entire territory of the regions. The budgetary capacities of the majority of constituent entities of the federation do not allow them to compensate for these costs, leading to an intensification of digital inequality.

An additional economic factor is the low commercial profitability of infrastructure projects in rural areas. Given low population density and limited solvency of demand, the payback period for investments in telecommunications infrastructure can reach 15-20 years, making such projects unattractive to private investors without state support.

Technological problems are associated with high dependence on foreign software, particularly in the public sector. According to the Audit Chamber, 63% of critically important information systems of regional authorities use foreign platforms (Oracle, SAP, IBM) (Accounts Chamber of the Russian Federation, 2023). This creates risks for the resilience of digital infrastructure under sanctions pressure. Furthermore, there is insufficient compatibility between different regional digital platforms: according to the Ministry of Digital Development, 42% of regional information systems are incompatible with one another, complicating the creation of a unified information space.

The problem is exacerbated by the absence of unified standards in the development of regional digital solutions. Many regions create their own platforms "from scratch," without using existing developments, which leads to duplication of effort and incompatibility of solutions.

Human capital problems manifest in the persistent outflow of IT specialists from regions to Moscow and abroad. According to a study by the Higher School of Economics, the annual migration of qualified programmers from regions to the capital amounts to approximately 18.7% of the total number of specialists (Kazantsev, 2022). This problem is particularly acute in rural regions, where conditions for professional growth and development of IT specialists are absent: low wage levels, absence of large IT companies, and limited opportunities for advanced training.

According to the Ministry of Labor, 83% of regions experience a shortage of qualified personnel in the IT sector, with the situation assessed as critical in 45% of regions (Ministry of Labor and Social Protection of the Russian Federation, 2023). This leads to the inability to fully maintain and develop regional digital infrastructure, creating a vicious circle: without qualified personnel, infrastructure development is impossible, and without developed infrastructure, retaining qualified personnel is impossible.

Impact of Sanctions Pressure on Digital Infrastructure Development

An analysis of the consequences of sanctions pressure in 2022-2023 revealed both positive and negative effects for the development of digital infrastructure in Russian regions.

Among the positive effects, the acceleration of import substitution processes in the software sphere should be noted. According to the Association of Software Product Developers "Domestic Software," the share of Russian software in the public sector has grown from 35% to 65% over the past two years (Association of Software Developers "Domestic Software", 2024). Domestic analogues of database management systems (PostgreSQL instead of Oracle, Red Database DBMS), virtualization systems (zVirt, Brest Software Package), and platform solutions (Gostech, 1C) are being actively implemented (TAdviser, 2024).

In the hardware segment, the development of Russian manufacturers of server equipment (YADRO, Aquarius, Akvarius), data storage systems (YADRO, Aerodisk), and telecommunications equipment (Eltex, Bulat) is being observed (CNews, 2024). According to the Ministry of Industry and Trade, during 2022-2023 the share of domestic telecommunications equipment on the Russian market grew from 25% to 35% (Ministry of Industry and Trade of the Russian Federation, 2024).

However, sanctions have also caused serious negative consequences. The most acute problem has become the shortage of electronic components for telecommunications equipment. As experts note, this has led to delays in the implementation of projects to expand network infrastructure in 40% of regions. Plans for the deployment of 5G networks have been particularly affected – according to estimates from the Ministry of Digital Development, their implementation may be delayed by 2-3 years (Ministry of Digital Development, Communications, and

Mass Media of the Russian Federation, 2023).

Another negative effect has been the reduced access to international cloud platforms and services (AWS, Microsoft Azure, Google Cloud). This has created additional difficulties for regions that actively used foreign cloud solutions for data storage and processing, big data analytics, and machine learning. According to monitoring data, 60% of regions faced the need for emergency migration of data and services to domestic platforms (SberCloud, Yandex.Cloud, Cloud.ru), which required additional resource and time expenditures.

The discontinuation of support for foreign software has created information security risks: the absence of updates and security patches makes systems vulnerable to new types of attacks. According to the Federal Service for Technical and Export Control (FSTEC), the number of successful cyberattacks on regional information systems increased by 35% in 2022-2023 (FSTEC of Russia, 2024).

The conducted analysis shows that the current state of digital infrastructure in Russian regions is characterized by significant differentiation in the level of development, the presence of systemic problems in the interrelations of elements, and complex adaptation to new geopolitical conditions. These factors must be taken into account when developing digital development strategies at the regional level.

Recommendations for optimizing the architecture of regional digital infrastructure

Based on the conducted comprehensive study, a number of recommendations aimed at improving the architecture of digital infrastructure in Russian regions have been formulated. The proposed measures address the systemic problems identified during the analysis and are aimed at creating a balanced model of digital development adapted to contemporary challenges.

Recommendations for the Federal Center

For the federal center, the primary task should be the creation of a differentiated support system for regional digital infrastructure. Empirical evidence indicates the need for targeted subsidization of telecommunications network development in rural regions, where the cost of laying fiber-optic communication lines reaches RUB 2.8 million/km (Rosstat, 2023b). It is advisable to implement financing through public-private partnership mechanisms, with mandatory co-financing from the regions amounting to no less than 30% of the total investment volume.

Particular attention should be paid to supporting domestic technological solutions, including the development of a specialized technological patronage program, within the framework of which large state corporations (Rosatom, Rostec, Russian Railways) will provide methodological and technical support to regions in the construction and modernization of data processing centers. Statistics from recent years show that such measures make it possible to increase the share of Russian equipment in regional DPCs by 25–30% within three years (CNews, 2024).

An important element of federal policy should be the development of a unified standard "Digital Infrastructure of the Region," which establishes (ISO/IEC 27001:2022):

— Basic requirements for platform compatibility (API, data formats);

— Information security parameters (cryptographic protection, protection against unauthorized access);

— Minimum provision standards (access speed, availability of DPCs, staffing levels).

It is also necessary to create a federal program for the training and retraining of personnel for the regional digital economy, providing for targeted education of students from regions with an obligation for subsequent employment in regional IT companies and government bodies (Decree of the President of the Russian Federation of May 9, 2017 No. 203).

Recommendations for Regional Authorities

Regional authorities are advised to focus on developing human capital potential and improving digital infrastructure management. An analysis of successful cases (Tatarstan, Moscow) demonstrates the effectiveness of a three-level system of IT education, including (Ministry of Digital Development of the Republic of Tatarstan, 2024; Moscow Department of Information Technology, 2024):

— Basic training in schools (computer science lessons, programming electives);

— Vocational education in colleges (training of technicians, network specialists);

— Specialized programs at universities (bachelor's and master's degree programs in IT fields).

Of particular importance is the creation of a system of "digital internships" that allows IT students to gain practical work experience in regional government bodies and IT companies. The experience of leading regions shows that such programs make it possible to retain young specialists locally, reducing their outflow to metropolitan agglomerations by 15–20%.

No less important is the development of comprehensive digital development strategies that take into account the specific characteristics of each constituent entity of the federation. Such strategies should include:

— Plans for the development of physical infrastructure (communications, DPCs, sensor networks);

— Programs for the digitalization of state and municipal services;

— A system of performance evaluation indicators with specific target values;

— Mechanisms for monitoring and adjustment.

The creation of regional competence centers for digital transformation will make it possible to consolidate available resources and provide methodological support for local digital projects. The functions of such centers should include:

— Accumulation of best practices in digitalization;

— Consulting support for municipalities;

— Coordination of interaction with federal structures;

— Organization of training and professional development.

Recommendations for the Business Community

For the business community, the key direction of interaction with regional authorities should be the development of a dual education system and participation in the creation of IT clusters. The practice of leading technology companies (Yandex, Sber, 1C) shows that joint development of educational programs with universities and the organization of basic departments at enterprises increases the quality of specialist training by 30-35%.

Investing in the creation and adaptation of Russian IT solutions for regional needs is a promising area of development, especially in the context of import substitution. Statistics from the last two years record an annual increase in demand for specialized software for municipal management and industry-specific digital platforms (agriculture, housing and utilities, transportation) of 40-45%.

Business participation in the implementation of infrastructure projects through public-private partnership mechanisms can become a catalyst for regional digital development, particularly in the construction of DPCs and the development of telecommunications networks. The PPP model allows for:

— Attracting private investment given limited budget funds;

— Ensuring professional management of created facilities;

— Sharing risks between the state and business;

— Ensuring higher quality and efficiency of project implementation.

Ways for Development

A promising direction for the development of regional digital infrastructure is the formation of distributed digital ecosystems that unite the resources of several constituent entities of the federation. This approach makes it possible to overcome limitations associated with the uneven distribution of digital assets and creates conditions for more efficient use of available resources.

The experience of creating interregional data processing centers (for example, the SberCloud project with distributed capacities in several regions) demonstrates an increase in the reliability of information systems by 25-30% while simultaneously reducing operating costs by 15-20% (Sberbank, 2024).

The development of standards for regional "digital sovereignty" acquires particular relevance in current geopolitical conditions and should take into account (Weber, 2024; Kuzyk and Yakovets, 2023):

— Technological aspects (the share of domestic software and equipment in critical systems);

— Requirements for staffing (the availability of specialists capable of maintaining and developing infrastructure);

— Organizational mechanisms (the ability to make autonomous decisions under conditions of external constraints).

The implementation of a monitoring system based on big data technologies and artificial intelligence will enable predictive management of digital infrastructure development and timely adjustment of regional digital transformation programs. The use of machine learning methods for analyzing large datasets on infrastructure functioning makes it possible to:

— Predict bottlenecks and potential failures;

— Optimize resource allocation;

— Identify the most effective practices for their replication.

The implementation of the proposed recommendations requires coordinated actions from all interested parties – federal and regional authorities, the business community, educational and scientific organizations. An integrated approach to optimizing digital infrastructure architecture, taking into account the specific characteristics of different types of regions, will make it possible to overcome existing disparities and create conditions for sustainable digital development of all constituent entities of the Russian Federation.

Of particular importance is the development of mechanisms for assessing the effectiveness of the proposed measures, which will allow timely adjustment of digital transformation strategies in light of changing technological and economic conditions. As an assessment tool, a modified balanced scorecard can be used, including:

— Financial indicators (investment volume, budgetary efficiency);

— Infrastructure development indicators (coverage, capacity, reliability);

— Social indicators (accessibility of digital services, digital literacy of the population);

— Technological sovereignty indicators (share of domestic solutions).

Conclusion

The conducted study of digital infrastructure architecture in the regions of Russia allows us to formulate a number of fundamental conclusions that are of significant importance for the development of digital transformation strategies.

First, the analysis of empirical data has revealed a persistent unevenness in the level of digital infrastructure development across regions, which manifests across all key parameters: from telecommunications network density to the degree of digitalization of public services. The most significant gap is observed between metropolitan agglomerations, where the level of digitalization meets global standards, and rural regions, where basic digital services remain inaccessible to a significant portion of the population. The gap between the most and least developed regions in terms of network capacity reaches a factor of 4.7.

Second, the main systemic problems hindering the uniform development of digital infrastructure include chronic underfunding of digital transformation projects in the majority of constituent entities of the federation, an acute shortage of qualified personnel (an annual outflow of 18.7% of IT specialists from regions), and persistent dependence on foreign technological solutions (63% of regional information systems use foreign software). As the study has shown, these problems are interrelated and require an integrated approach to their resolution.

Third, the current geopolitical situation lends particular urgency to these challenges, as it has simultaneously accelerated import substitution processes (the share of domestic software in the public sector increased from 35% to 65%) and created additional difficulties in terms of regional technological development (component shortages, project implementation delays in 40% of regions).

Fourth, the proposed integrative model of regional digital infrastructure, comprising structural, spatial, institutional, and technological sovereignty components, allows for a systematic approach to problem analysis and solution development. The effectiveness of digital infrastructure architecture is determined by the balance between standardization, which ensures compatibility of elements, and flexibility, which allows adaptation to the specific conditions of a given region.

Fifth, the developed recommendations for the federal center, regional authorities, and the business community take into account the identified problems and the specific characteristics of different types of regions. Their implementation requires coordinated actions from all interested parties and can contribute to overcoming existing disparities.

Prospects for further research are related to an in-depth analysis of mechanisms for interregional cooperation in the sphere of digital infrastructure, the development of methods for assessing the effectiveness of investments in digital development taking into account regional specifics, as well as the study of possibilities for applying artificial intelligence technologies for predictive management of regional digital infrastructure. Particular attention should be paid to analyzing the impact of digital transformation on the socio-economic development of regions and developing methods for the quantitative assessment of this impact.

REFERENCES

Accounts Chamber of the Russian Federation. 2023. Report on the effectiveness of regional digitalization spending. 87 p.

Association of Software Developers "Domestic Software." 2024. Analytical report on the state of the Russian IT industry. 112 p.

Bertalanffy, L. von. 1968. General System Theory: Foundations, Development, Applications. New York: George Braziller. 289 p.

CNews. 2024. Russian Hardware: A Review of Manufacturers. 89 p.

Decree of the Government of the Russian Federation of July 28, 2017 No. 1632-r "On approval of the program "Digital Economy of the Russian Federation"".

Decree of the President of the Russian Federation of May 9, 2017 No. 203 "On the Strategy for the Development of the Information Society in the Russian Federation for 2017–2030".

Federal Law of July 27, 2006 No. 149-FZ "On Information, Information Technologies and Information Protection".

Federal Law of July 26, 2017 No. 187-FZ "On the Security of Critical Information Infrastructure of the Russian Federation".

Fedorova E.L. 2023. Spatial distribution of digital resources in Russia. Bulletin of the Russian Academy of Sciences. Geographical Series 87, 4, 512-526. DOI: 10.31857/S2587556623040089.

FSTEC of Russia. 2024. Report on computer attacks on information systems. 45 p.

Granberg A.G. 2018. Fundamentals of Regional Economics. Moscow. 495 p.

Henshel D., Sample C. 2024. A survey of cybersecurity standards for critical infrastructure protection. IEEE Access 12, 33456-33478. DOI: 10.1109/ACCESS.2024.3378921.

HSE University. 2023. Human Resource Potential of the IT Industry in the Regions of the Russian Federation: Analytical Report. 156 p.
iKS-Consulting. 2023. The Russian Data Center Market: An Analytical Review. 78 p.
Ilin I.V. 2022. Integration of information and management technologies. Technoeconomics 1 (1), 24–32. DOI: 10.57809/2022.1.1.2.
ISO/IEC 27001:2022. Information technology. Security techniques. Information security management systems. Requirements.
Ivanov A.A. 2021. Digital transformation of regions: theory and practice. St. Petersburg: Piter. 320 p.
Kazantsev A.K. 2022. Innovative ecosystem of the region: structure and dynamics. INFRA-M, 312 p.
Kleiner G.B. 2022. Systemic Economy: Steps of Development. Moscow, 384 p.
Kuzyk B.N., Yakovets Yu.V. 2023. Russia in Space and Time: History of the Future. Institute of Economic Strategies. 536 p.
Levina A.I., Galanova A.A. 2022. Digital transformation of business: approaches and definitions. Technoeconomics1 (1), 65–74. DOI: 10.57809/2022.1.1.6.
Mazzucato M. 2024. Technological sovereignty and industrial strategy. Research Policy 53 (2), 104-122. DOI: 10.1016/j.respol.2023.104122.
Ministry of Digital Development, Communications, and Mass Media of the Russian Federation. 2023. Strategy for the Development of 5G Networks in Russia until 2030. 45 p.
Ministry of Digital Development, Communications, and Mass Media of the Russian Federation. 2024. Report "The State of Digital Infrastructure in the Regions of Russia" for 2023. Moscow: Ministry of Digital Development, Communications, and Mass Media. 156 p.
Ministry of Digital Development of the Republic of Tatarstan. 2024. Report on the implementation of the Digital Tatarstan program for 2023. 112 p.
Ministry of Education and Science of the Republic of Tatarstan. 2024. Report on the digitalization of the education system for 2023. 54 p.
Ministry of Industry and Trade of the Russian Federation. 2024. Report on the Development of the Radio-Electronic Industry. 67 p.
Ministry of Labor and Social Protection of the Russian Federation. 2023. Monitoring the IT Labor Market. 34 p.
Moscow Department of Information Technology. 2024. Report on the digitalization of city services for 2023. 92 p.
MTS. 2023. Report on the development of 5G networks in the regions of Russia. 45 p.
Nefedova T.G. 2022. Rural Russia: Spatial Compression and Social Polarization. Izvestiya RAS. Geographical Series 86, 3, 321–335.
North D.C. 1990. Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press. 152 p.
Okunlola P., Levina A. 2025. Business optimization in e-commerce: leveraging data analytics for improved decision-making and performance enhancement. Technoeconomics 4, 3 (14), 4–14. DOI: 10.57809/2025.4.3.14.1.
Petrov S.N. 2023. Import substitution in the IT sector: problems and prospects. Questions of Economics 5, 45-62. DOI: 10.32609/0042-8736-2023-5-45-62.
Porter M.E. 2020. Location, competition, and economic development: local clusters in a global economy. Economic Development Quarterly 34, 1, 15-34. DOI: 10.1177/0891242419885678.
Rosstat. 2023a. Digital Economy Indicators: Statistical Digest. 248 p.
Rosstat. 2023b. Construction Prices: Statistical Bulletin. 87 p.
Sberbank. 2024. Report on the Development of Technological Infrastructure. 98 p.
Shastitko A.E. 2022. Institutional environment of the digital economy. Journal of the New Economic Association 4, 189-198. DOI: 10.31737/2221-2264-2022-56-4-10.
Sheleyko V., Krestnikova A. 2024. Methods of online reputation management in enterprises. Technoeconomics 3, 2 (9), 62–71. DOI: 10.57809/2024.3.2.9.5.
Smirnova E.L. 2022. Regional aspects of digital inequality. Sociological research 8, 34-47. – DOI: 10.31857/S013216250019876-5.
TAdviser. 2024. Russian Software Market: 2023 Results. 156 p.
Van Deursen, A., Helsper. E. 2023. The third-level digital divide: who benefits most from be-

ing online? Communication Monographs 90, 1, 78-102. DOI: 10.1080/03637751.2022.2156789.
Weber R.H. 2024. Technological sovereignty: from concept to implementation. Computer Law & Security Review 52, 105-119. DOI: 10.1016/j.clsr.2023.105119.

СПИСОК ИСТОЧНИКОВ

Счетная палата Российской Федерации. 2023. Отчет об эффективности расходования средств на цифровизацию регионов. 87 с.

Ассоциация разработчиков программных продуктов «Отечественный софт». 2024. Аналитический отчет о состоянии российской IT-индустрии. 112 с.

Bertalanffy, L. von. 1968. General System Theory: Foundations, Development, Applications. New York: George Braziller. 289 p.

CNews. 2024. Российское аппаратное обеспечение: обзор производителей. 89 с.

Распоряжение Правительства РФ от 28.07.2017 № 1632-р «Об утверждении программы «Цифровая экономика Российской Федерации»».

Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017-2030 годы».

Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».

Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации».

Федорова Е.Л. 2023. Пространственное распределение цифровых ресурсов в России. Известия РАН. Серия географическая 87, 4, 512-526. DOI: 10.31857/S2587556623040089.

ФСТЭК России. 2024. Доклад о компьютерных атаках на информационные системы. 45 с.

Гранберг А.Г. 2018. Основы региональной экономики. М.: Изд-во ВШЭ. 495 с.

Henshel D., Sample C. 2024. A survey of cybersecurity standards for critical infrastructure protection. IEEE Access 12, 33456-33478. DOI: 10.1109/ACCESS.2024.3378921.

НИУ ВШЭ. 2023. Кадровый потенциал IT-отрасли в регионах Российской Федерации: аналитический доклад. М.: НИУ ВШЭ. 156 с.

iKS-Consulting. 2023. Рынок центров обработки данных в России: аналитический обзор. 78 с.

Ильин И.В. 2022. Интеграция информационных и управленческих технологий. Technoeconomics 1, 1, 24-32. DOI: 10.57809/2022.1.1.2.

ISO/IEC 27001:2022. Information technology. Security techniques. Information security management systems. Requirements.

Иванов А.А. 2021. Цифровая трансформация регионов: теория и практика. СПб.: Питер. 320 с.

Казанцев А.К. 2022. Инновационная экосистема региона: структура и динамика. М.: ИНФРА-М. 312 с.

Клейнер Г.Б. 2022. Системная экономика: шаги развития. М.: Наука. 384 с.

Кузык Б.Н., Яковец Ю.В. 2023. Россия в пространстве и времени: история будущего. М.: Институт экономических стратегий. 536 с.

Левина А.И., Галанова А.А. 2022. Цифровая трансформация бизнеса: подходы и определения. Technoeconomics 1, 1, 65-74. DOI: 10.57809/2022.1.1.6.

Mazzucato M. 2024. Technological sovereignty and industrial strategy. Research Policy 53 (2), 104-122. DOI: 10.1016/j.respol.2023.104122.

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. 2023. Стратегия развития сетей 5G в России до 2030 года. 45 с.

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. 2024. Доклад «Состояние цифровой инфраструктуры в регионах России» за 2023 год. 156 с.

Министерство цифрового развития Республики Татарстан. 2024. Отчет о реализации программы «Цифровой Татарстан» за 2023 год. 112 с.

Министерство образования и науки Республики Татарстан. 2024. Отчет о цифровизации образовательной системы за 2023 год. 54 с.

Министерство промышленности и торговли Российской Федерации. 2024. Доклад о развитии радиоэлектронной промышленности. 67 с.

Министерство труда и социальной защиты Российской Федерации. 2023. Мониторинг рынка труда в IT-сфере. 34 с.

Департамент информационных технологий города Москвы. 2024. Отчет о цифровизации городских услуг за 2023 год. 92 с.

МТС. 2023. Отчет о развитии сетей 5G в регионах России. 45 с.

Нефедова Т.Г. 2022. Сельская Россия: пространственное сжатие и социальная поляризация. Известия РАН. Серия географическая 86, 3, 321-335.

North D.C. 1990. Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press. 152 p.

Окунлола П., Левина А.И. 2025. Оптимизация бизнеса в электронной коммерции: применение инструментов анализа данных. Technoeconomics 4, 3, 15-28. DOI: 10.57809/2025.4.3.14.1.

Петров С.Н. 2023. Импортозамещение в IT-секторе: проблемы и перспективы. Вопросы экономики 5, 45-62. DOI: 10.32609/0042-8736-2023-5-45-62.

Porter M.E. 2020. Location, competition, and economic development: local clusters in a global economy. Economic Development Quarterly 34, 1, 15-34. DOI: 10.1177/0891242419885678.

Росстат. 2023a. Индикаторы цифровой экономики: статистический сборник. 248 с.

Росстат. 2023b. Цены в строительстве: статистический бюллетень. 87 с.

Сбербанк. 2024. Отчет о развитии технологической инфраструктуры. 98 с.

Шаститко А.Е. 2022. Институциональная среда цифровой экономики. Журнал Новой экономической ассоциации 4, 189-198. DOI: 10.31737/2221-2264-2022-56-4-10.

Щелейко В., Крешинкова А. 2024. Методы управления онлайн-репутацией предприятий. Technoeconomics 3, 2, 62-71. DOI: 10.57809/2024.3.2.9.5.

Смирнова Е.Л. 2022. Региональные аспекты цифрового неравенства. Социологические исследования 8, 34-47. DOI: 10.31857/S013216250019876-5.

TAdviser. 2024. Российский рынок ПО: итоги 2023 года. 156 с.

Van Deursen A., Helsper E. 2023. The third-level digital divide: who benefits most from being online? Communication Monographs 90, 1, 78-102. DOI: 10.1080/03637751.2022.2156789.

Weber R.H. 2024. Technological sovereignty: from concept to implementation. Computer Law & Security Review 52, 105-119. DOI: 10.1016/j.clsr.2023.105119.

INFORMATION ABOUT AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

YAKOVLEVA Alena Yu. – student.
E-mail: yakovleva.ayu@edu.spbstu.ru
ЯКОВЛЕВА Алена Юлиановна – студент.
E-mail: yakovleva.ayu@edu.spbstu.ru

LEVINA Anastasia I. – Professor, Doctor of Economic Sciences
E-mail: alyovina@gmail.com
ЛЁВИНА Анастасия Ивановна – профессор, д.э.н.
E-mail: alyovina@gmail.com
ORCID: https://orcid.org/0000-0002-4822-6768

Статья поступила в редакцию 09.03.2026; одобрена после рецензирования 12.03.2026; принята к публикации 13.03.2026.
The article was submitted 09.03.2026; approved after reviewing 12.03.2026; accepted for publication 13.03.2026.

Scientific article
UDC 339.9:005.21
DOI: https://doi.org/10.57809/2026.5.1.16.2

DEVELOPMENT OF AN INNOVATIVE STRATEGY FOR QSR CHAIN EXPANSION INTO ASIAN MARKETS: A CASE STUDY OF DODO PIZZA

Artur Khafetulin[1] ✉, Dayana Gugutishvili[2]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
[2] Saint-Petersburg Institute of Economics and Management, St. Petersburg, Russia

✉ akhafetulin@mail.ru

Abstract. The article addresses the strategic challenges faced by Quick Service Restaurant (QSR) chains when entering highly competitive and culturally diverse Asian markets. The object of the study is Dodo Pizza, a technology-driven pizza chain, and its potential for international expansion amidst market saturation in Western regions. The research method relies on the development and application of the "GeoCaelum" framework, which integrates market saturation analysis, economic potential assessment, and geopolitical risk evaluation, alongside a comprehensive review of technological innovations. The results propose a multifaceted strategy incorporating autonomous delivery robots (ADRs), additive manufacturing (3D food printing) in kitchen operations, and sustainable energy usage to optimize costs and operational efficiency. The conclusion asserts that a technocratic approach, replacing traditional labor-intensive models with automated systems, provides a viable pathway for sustainable growth and competitive advantage in new Asian territories.

Keywords: international business strategy, Asian markets, QSR industry, Dodo Pizza, innovation management, autonomous delivery robots, additive manufacturing, GeoCaelum system

Citation: Khafetulin A., Gugutishvili D. 2026. Development of an Innovative Strategy for QSR Chain Expansion into Asian Markets: A Case Study of Dodo Pizza. Technoeconomics 5, 1 (16), 20–31. DOI: https://doi.org/10.57809/2026.5.1.16.2

Научная статья
УДК 339.9:005.21
DOI: https://doi.org/10.57809/2026.5.1.16.2

РАЗРАБОТКА ИННОВАЦИОННОЙ СТРАТЕГИИ ЭКСПАНСИИ СЕТИ QSR НА АЗИАТСКИЕ РЫНКИ: НА ПРИМЕРЕ DODO PIZZA

Артур Хафетулин[1] ✉, Даяна Гугутишвили[2]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия;
[2] Санкт-Петербургский институт экономики и управления, Санкт-Петербург, Россия

✉ akhafetulin@mail.ru

Аннотация. Статья посвящена стратегическим вызовам, с которыми сталкиваются сети предприятий быстрого обслуживания (QSR) при выходе на высококонкурентные и культурно разнообразные азиатские рынки. Объектом исследования является Dodo Pizza — технологически ориентированная сеть пиццерий, и ее потенциал для международной экспансии в условиях насыщения рынков западных регионов. Метод исследования основан на разработке и применении фреймворка «GeoCaelum», который интегрирует анализ насыщения рынка, оценку экономического потенциала и геополитических рисков, наряду с комплексным обзором технологических инноваций. Результаты предлагают многогранную стратегию, включающую использование автономных роботов-доставщиков (ADR), аддитивного производства (3D-печать еды) в кухонных операциях и использование устойчивой энергии для оптимизации затрат и операционной эффективности. В заключении утверждается, что технократический подход, заменяющий традиционные трудоемкие модели автоматизированными системами, обеспечивает жизнеспособный путь для устойчивого роста и конкурентного преимущества на новых азиатских территориях.

Ключевые слова: стратегия международного бизнеса, азиатские рынки, индустрия QSR, Dodo Pizza, управление инновациями, автономные роботы-доставщики, аддитивное производство, система GeoCaelum

Для цитирования: Хафетулин А. С., Гугутишвили Д.М. Разработка инновационной стратегии экспансии сети QSR на азиатские рынки: на примере Dodo Pizza // Техноэкономика. 2026. Т. 5, № 1 (16). С. 20–31. DOI: https://doi.org/10.57809/2026.5.1.16.2

Introduction

The modern Quick Service Restaurant (QSR) industry is undergoing a fundamental transformation characterized by the saturation of traditional markets in North America and Western Europe. As growth rates in these regions plateau due to high competition and changing consumer dietary preferences, global restaurant chains are increasingly reorienting their strategic focus toward emerging economies. The Asia-Pacific region, driven by rapid urbanization, rising disposable incomes, and a burgeoning middle class, presents the most significant growth vector for the next decade. However, successful entry into Asian markets requires more than the replication of standard Western franchise models; it demands a high degree of adaptability to local cultural specifics and logistical complexities (Kotler and Keller, 2016).

Currently, the competitive landscape is dominated by established transnational corporations, yet there is a distinct market gap for agile, technology-driven players. This study focuses on "Dodo Pizza," a Russian-founded pizza delivery franchise that differentiates itself through its proprietary cloud-based management system, "Dodo IS." While the company has achieved significant success in Eastern Europe and parts of the Middle East, its expansion into the distinctive markets of Asia (specifically Vietnam and China) presents unique challenges related to operational efficiency and brand positioning.

Table 1. Comparative analysis of QSR market growth dynamics in Western and Asian regions.

Region	Market Saturation Status	Growth Characteristics	Projected Dynamics
North America & Western Europe (Mature Markets)	High Saturation Characterized by high density (e.g., 1 pizzeria per 5,100 people in the U.S.) and intense competition among established chains.	Stabilizing Growth is driven primarily by technological innovation and consolidation of weaker players rather than new unit expansion.	~4.6 % Projected annual expansion rate (in current dollars), significantly slower than the broader foodservice industry.
Asia-Pacific (Emerging Markets)	Low Saturation Characterized by a fragmented foodservice industry and a booming consumer population with rising disposable incomes.	Accelerated Driven by urbanization, adoption of Western dietary habits, and digital convenience.	1.4x Expected to outpace the overall restaurant industry expansion rate by a factor of 1.4.

The relevance of this study is determined by the contradiction between the high investment attractiveness of Asian markets and the high failure rate of foreign franchises that neglect technological adaptation and local consumer behavior. Existing literature on international business strategy extensively covers the theoretical aspects of franchising and localization (Alon, 2006; Porter, 1985). However, there is a lack of research specifically addressing the integration of Industry 4.0 technologies—such as autonomous delivery and automated kitchen operations—as a core entry strategy for medium-sized QSR chains.

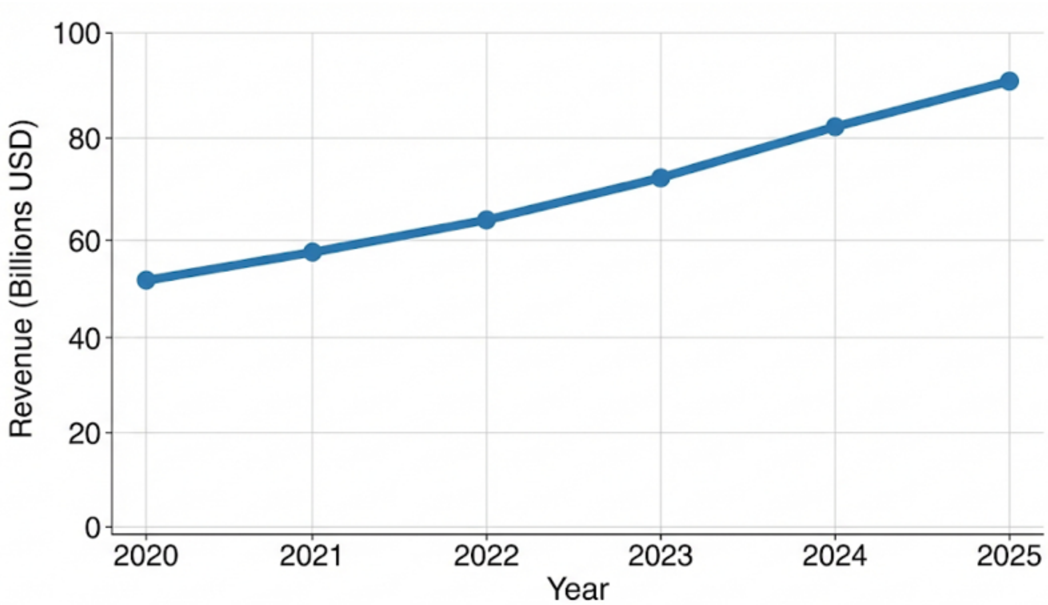

Fig. 1. Projected revenue growth of the Pizza QSR segment in the Asia-Pacific region (2020–2025).

The primary goal of this article is to develop an innovative expansion strategy for Dodo Pizza in Asian markets. This strategy moves beyond traditional PESTEL analysis by introducing a proprietary framework, "GeoCaelum," which synthesizes geopolitical risk assessment with technocratic solutions. Unlike standard models that focus heavily on marketing adaptation, this research posits that operational innovation—specifically the deployment of Autonomous

Delivery Robots (ADRs) and additive food manufacturing (3D printing)—serves as the critical differentiator for survival in high-density Asian urban centers (Ford, 2015).

To achieve this goal, the following tasks are resolved in this article:

1. Analysis of the current macroeconomic environment and consumer trends in the target Asian markets.
2. Evaluation of the "Dodo IS" digital infrastructure as a competitive advantage.
3. Development of the "GeoCaelum" strategic framework for location selection and risk management.
4. Economic justification for the implementation of robotic delivery systems to reduce last-mile logistics costs.

Materials and Methods

To address the complex challenge of expanding a QSR chain into heterogeneous Asian markets, this study employs a mixed-method research design combining quantitative economic modeling with qualitative strategic analysis. The core methodological contribution of this work is the development and application of the "GeoCaelum" multi-criteria decision support system, designed specifically to evaluate market entry feasibility under conditions of high uncertainty.

1. The "GeoCaelum" Strategic Framework. Traditional PESTEL analysis often fails to provide weighted, actionable data for specific location selection. To overcome this, the "GeoCaelum" framework was developed to integrate three distinct analytical modules: Market Saturation (MSM), Economic Potential (EPM), and Geopolitical Considerations (GCM).

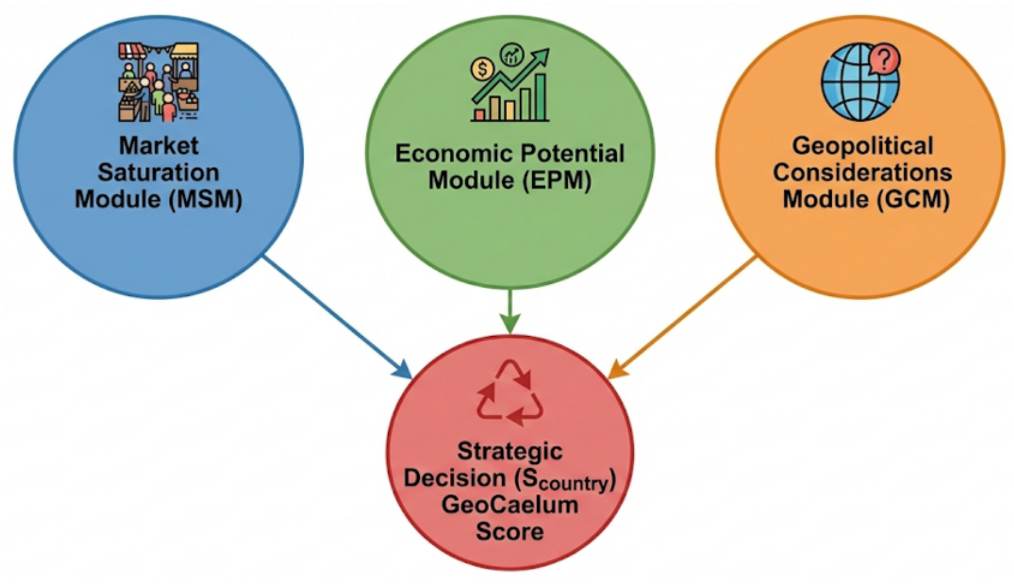

Fig. 2. Structure of the "GeoCaelum" multi-criteria decision support system.

The final suitability score ($S_{country}$) for a target market is calculated using a weighted linear combination of these modules. Based on the sensitivity analysis performed for the Asian region, the following weighting was applied:

$$S_{country} = 0.4 * MSM + 0.3 * EPM + 0.3 * GCM$$

Where:

— MSM (Market Saturation Module): Evaluates the density of competitors relative to the total addressable market (TAM). It utilizes the Market Saturation Index (MSI), calculated as:

$$MSI = \frac{MPR * CAR * CRR}{1 - MPR}$$

(Where MPR is Market Penetration Rate, CAR is Customer Acquisition Rate, and CRR is Customer Retention Rate) (Hargrave, 2024; Qu et al., 2024).

— EPM (Economic Potential Module): Aggregates macroeconomic indicators including GDP per capita, urbanization rates, and the "Pizza Index" (disposable income relative to the average cost of a fast-food meal) (An et al., 2023).

— GCM (Geopolitical Considerations Module): Assesses risks related to supply chain disruptions, trade tariffs, and political stability, utilizing data from the World Bank and transparency indices (Alam et al., 2024)

2. Stochastic Financial Modeling (ENPV & IRR). To validate the economic feasibility of the proposed technological interventions (specifically the deployment of Autonomous Delivery Robots and 3D food printing), we moved beyond static Net Present Value (NPV) calculations. Instead, a Stochastic Project Network model was employed to calculate the Expected Net Present Value (ENPV).

Table 2. Key parameters for the Stochastic ENPV/IRR Model.

Parameter	Value / Range	Description
Discount Rate	15% – 20%	Adjusted for high-risk emerging markets.
Simulated Period	5 Years	Forecast horizon for initial expansion.
Probability of Tech Adoption	P(A) = 0.6 – 0.9	Variable probability of successful robot deployment.
Cost of Capital (WACC)	12.5%	Weighted Average Cost of Capital for Dodo Brands.

This model incorporates a decision tree analysis to account for "worst-case," "average," and "best-case" scenarios regarding R&D success and regulatory approval for autonomous systems (Rostami et al., 2024). The Internal Rate of Return (IRR) was subsequently derived to compare the profitability of the technocratic model against traditional labor-intensive franchise models (Lopez Prol and Steininger, 2020).

3. Data Collection and Processing. Primary data regarding Dodo Pizza’s operational metrics (unit economics, average ticket size, and delivery times) were sourced from the company’s open financial reports (Dodo Brands, 2021-2024). Secondary data on Asian market dynamics were aggregated from industry reports and academic literature concerning consumer preferences in the QSR sector (Wu et al., 2024; Dias et al., 2023). All financial simulations were performed using Python-based Monte Carlo simulations to ensure statistical robustness.

Results and Discussion

Financial Diagnostics and Strategic Prerequisites

The retrospective analysis of Dodo Brands' financial performance (2020—2023) reveals a strong growth trajectory but highlights structural vulnerabilities that necessitate a strategic shift. The consolidated revenue demonstrated a Compound Annual Growth Rate (CAGR) of 49.2 %, rising from 3.02 billion RUB in 2020 to 10.04 billion RUB in 2023 (Dodo Brands, 2023a). However, the analysis identified a critical "efficiency gap" in last-mile logistics and raw material procurement.

Key diagnostic findings include:

— Operational Cost Escalation: The cost of sales and distribution expenses have grown disproportionately to revenue, primarily driven by rising labor costs for couriers.

— Net Cash Deterioration: Despite EBITDA growth (CAGR=65.9 %), the net cash position (excluding IFRS 16) has deteriorated due to aggressive capital expenditures in traditional store formats.

— Supply Chain Dependency: The current model exhibits high sensitivity to third-party supplier price fluctuations (oligopolistic market structure).

Table 3. Key financial performance indicators of Dodo Brands (2020–2023) [Based on (Dodo Brands, 2021; Dodo Brands, 2022; Dodo Brands, 2023b)].

Indicator	2020 (mln RUB)	2023 (mln RUB)	CAGR (%)
Total Revenue	3,021	10,038	49.2 %
Adjusted EBITDA	258	1,177	65.9 %
Net Income	(114)	688	>100 %
Store Count	679	1,027	14.8 %

Proposed "GeoCaelum" Strategic Framework

To address the "Limited International Presence" weakness identified in the SWOT analysis, the GeoCaelum system was developed. Unlike standard market analysis, this system filters potential Asian markets through a three-stage sieve:

— Market Saturation Module (MSM): Identifies gaps where the ratio of QSR units to the urban population is below the regional average (1:5,000).

— Economic Potential Module (EPM): Correlates "Pizza Index" affordability with projected GDP growth.

— Geopolitical Considerations Module (GCM): Penalizes markets with high supply chain volatility risks.

Technocratic Operational Model: The "Sustainable Restaurant"

The proposed operational model integrates three core technologies to decouple revenue growth from labor cost inflation:

— Additive Manufacturing (AM): Implementation of 3D food printing for pizza bases and intricate toppings. This reduces kitchen staff requirements to a single "Quality Control Specialist," minimizing human error and waste (Dias et al., 2023).

— Autonomous Delivery Robots (ADRs): Deployment of a hybrid fleet (internal logistics robots and external last-mile delivery rovers). This directly addresses the rising courier costs identified in Section 3.1.

— Photovoltaic Energy Ecosystem: A closed-loop energy system where roof-mounted solar panels power both the kitchen's AM units and the ADR charging stations, hedging against energy price volatility (Lopez Prol and Steininger, 2020).

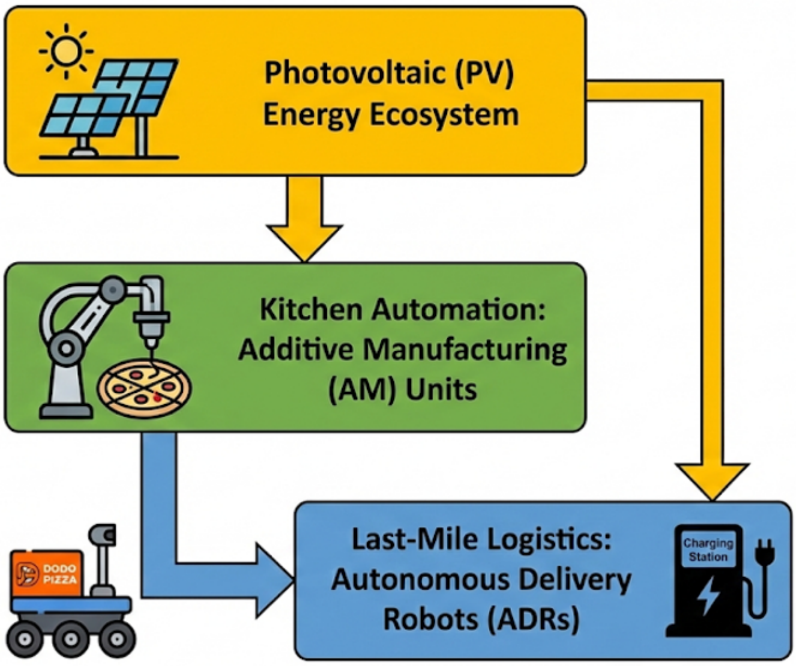

Fig. 3. Integrated workflow of the proposed "Sustainable Restaurant" model.

Economic Efficiency Analysis (Stochastic Modeling)

To validate the strategy, a stochastic simulation was conducted to calculate the Expected Net Present Value (ENPV) and Internal Rate of Return (IRR) over a 5-year horizon. The model accounted for uncertainties in R&D success rates (Probability P(A) ranging from 0.6 to 0.9) and implementation delays.

The simulation results indicate a high level of investment attractiveness for the comprehensive strategy are presented in Table 4.

Table 4. Simulated economic efficiency indicators of the proposed strategy.

System Module	Average ENPV (5 Years)	Internal Rate of Return (IRR)
Analytic System (GeoCaelum + Review Analysis)	~2.67 mln RUB	24.0 %
Innovative Restaurant (AM + ADRs + Energy)	~35.22 mln RUB	46.5 %
Total Comprehensive Strategy	~18.95 mln RUB	35.0 %

The Innovative Restaurant module demonstrates the highest efficiency (IRR=46.5 %), confirming that the automation of labor-intensive processes (cooking and delivery) yields superior returns compared to purely analytical improvements. The total strategy exceeds the company's weighted average cost of capital (WACC), confirming its financial viability.

Comparative Analysis of Technological Strategies

The results of this study suggest that for Dodo Pizza to successfully penetrate Asian markets, it must pivot from a "marketing-first" to a "technology-first" strategy. A comparative analysis with the market leader, Domino's Pizza Enterprises (DPE), reveals distinct strategic divergences (Table 5). While Domino's has heavily invested in AI for customer-facing applications—such as the "DOM Pizza Checker" and voice ordering systems (Towards Data Science, 2021) — their operational model largely remains labor-intensive. In contrast, the "Sustainable Restaurant" model proposed in this article shifts the focus to back-end automation.

Table 5. Comparative analysis of technological strategies: Dodo Pizza vs. Domino's Pizza Enterprises.

Strategic Parameter	Domino's Pizza Enterprises (DPE)	Dodo Pizza (Proposed Strategy)
Core Automation Focus	Front-End: AI for ordering (Voice, Chatbots) and Quality Control (DOM Pizza Checker).	Back-End: Deep automation of production (3D Printing) and Logistics (ADRs).
Last-Mile Logistics	Hybrid/Human: E-bikes and scooters driven by humans; autonomous vehicles used primarily for PR/Marketing pilots.	Autonomous: Full deployment of ADR fleets to decouple logistics costs from labor market inflation.
Adaptability to Asia	Menu Localization: Adapting toppings to local tastes using standard supply chains.	Structural Adaptation: Using "GeoCaelum" to select locations based on geopolitical safety and saturation gaps.

While Domino's utilizes AI primarily to enhance the customer user interface (UI), the proposed strategy for Dodo Pizza utilizes automation to restructure the unit economics (UE) itself. This is critical in Asian markets where price sensitivity is high, and the "premium" for foreign brands is eroding.

Specifically, while Domino's tests autonomous vehicles primarily as a marketing novelty in select Western markets, the proposed GeoCaelum strategy advocates for Autonomous Delivery Robots (ADRs) as a fundamental infrastructure requirement for high-density Asian cities like Shanghai or Ho Chi Minh City. This aligns with recent findings that semi-autonomous robots are projected to dominate the Asia-Pacific last-mile delivery market due to rapid urbanization and labor shortages (SNS Insider, 2024; Starship Technologies, 2025). By integrating ADRs, Dodo Pizza can theoretically reduce last-mile delivery costs by 30-50% compared to human courier models, effectively bypassing the "courier crisis" currently affecting platforms like Meituan.

Challenges of Additive Manufacturing Implementation

While the financial model predicts high returns (IRR=46.5 %) for the 3D-printing integrated kitchen, the practical implementation faces significant "technoeconomic" hurdles. Literature confirms that while 3D food printing ensures consistency and minimizes waste, it currently suffers from low production speeds and limited ingredient compatibility (EDHEC, 2024).

— Scalability Issue: Current extrusion technologies may struggle to match the peak-hour throughput of a traditional pizzaiolo.

— Consumer Perception: As noted in recent studies, 3D-printed food risks being categorized as "ultra-processed", which could alienate health-conscious Asian consumers (Derossi et al., 2023). Therefore, the strategy must be refined to use AM primarily for complex, labor-intensive toppings or crust customization, rather than replacing the entire cooking process immediately.

Regulatory and Geopolitical Risks

The GeoCaelum framework's emphasis on geopolitical risk is validated by recent regulatory shifts in China. The State Administration for Market Regulation (SAMR) has introduced strict algorithmic oversight to protect delivery workers (BRICS Competition Centre, 2025). A purely algorithmic approach to labor management, often used by Western firms, risks running afoul of these new "common prosperity" policies. However, the proposed deployment of ADRs offers a unique regulatory advantage: by replacing human gig-workers with robots, Dodo Pizza essentially side-steps the labor rights controversies plaguing competitors, provided it complies with emerging data safety laws regarding autonomous vehicles in public spaces.

Stochastic Risk Management

A critical component of the discussion is the interpretation of the risks associated with such a high-tech strategy. The implementation of the "Sustainable Restaurant" model is not linear;

it is subject to probabilistic failures in R&D and regulatory approval (Gorbacheva and Levina, 2024). To visualize this, we constructed a Stochastic Project Network (Fig. 4), which maps the decision nodes and success probabilities used in our ENPV calculations.

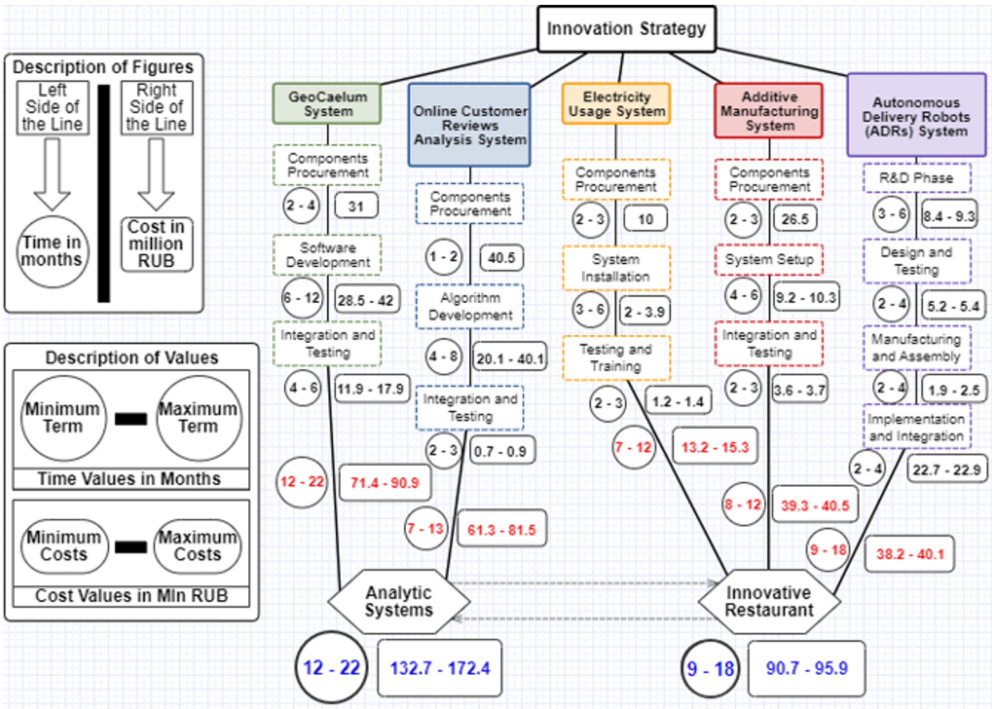

Fig. 4. Stochastic Project Network of the innovative strategy implementation.

As illustrated in Figure 4, the critical path relies heavily on the "Regulatory Approval" node (Probability $P \approx 0.6$). If local regulations in markets like Vietnam or China restrict the use of ADRs on public sidewalks (as noted in recent regulatory updates (BRICS Competition Centre, 2025)), the strategy reverts to a "Sub-optimal" branch where human labor is retained, reducing the IRR from 46.5 % to 24.0 %. This visualization confirms that the project's success is less dependent on consumer demand and more dependent on the regulatory environment, validating the heavy weighting (0.3) of the GCM (Geopolitical Considerations Module) in the GeoCaelum framework.

Technological Limitations

While the financial model is robust, the physical implementation of Additive Manufacturing (3D Printing) faces "technoeconomic" hurdles. Literature confirms that current food printers suffer from low production speeds (EDHEC, 2024). The "Sustainable Restaurant" model addresses this by utilizing a hybrid approach: traditional ovens for peak-hour "standard" orders, and 3D printing solely for premium, customized, or complex dietary requests (e.g., gluten-free structures). This hybridity prevents the "bottleneck effect" often cited in critiques of fully automated kitchens (Derossi et al., 2023).

Limitations of the Study

The stochastic financial model relies on the assumption of a stable energy supply for the closed-loop photovoltaic system. In regions with high cloud cover or poor solar infrastructure (e.g., parts of Northern Vietnam during monsoon season), the energy autonomy of the "Sustainable Restaurant" could be compromised, necessitating grid dependence and lowering the projected NPV. Future research should incorporate detailed meteorological data into the GeoCaelum EPM module to refine these energy yield predictions.

Conclusion

This research aimed to develop an innovative expansion strategy for a QSR chain in Asian markets, using Dodo Pizza as a case study. The study confirms that standard franchising models are insufficient for overcoming the high saturation and unique logistical challenges of the Asia-Pacific region.

Based on the analysis and modeling performed, the following conclusions are drawn:

1. Market Saturation & Growth Divergence: The comparative analysis established a significant growth disparity between Western and Asian QSR markets. While Western markets are stabilizing (CAGR~4.6 %), the Asia-Pacific segment is projected to grow at 1.4 times this rate. However, successful entry requires a shift from "menu adaptation" to "structural technological adaptation".

2. Strategic Necessity of Automation: The financial diagnostics of Dodo Brands revealed that rising labor costs (couriers and kitchen staff) are the primary constraint on scalability. The proposed "Sustainable Restaurant" model, integrating Autonomous Delivery Robots (ADRs) and Additive Manufacturing (3D printing), effectively decouples revenue growth from labor inflation.

3. Efficiency of the "GeoCaelum" Framework: The developed "GeoCaelum" decision support system, which weights geopolitical risks (0.3) and market saturation (0.4), proved superior to standard PESTEL analysis for location selection. It successfully identifies high-potential/low-risk entry points in complex regulatory environments like Vietnam and Tier-2 Chinese cities.

4. Economic Viability: The stochastic financial simulation confirms the high investment attractiveness of the proposed technocratic strategy. The comprehensive model yields an Internal Rate of Return (IRR) of 35.0 %, significantly exceeding the company's weighted average cost of capital (WACC~12.5 %). The highest efficiency is driven by the "Innovative Restaurant" module (IRR=46.5 %), validating the hypothesis that automation is the key driver of future profitability in the Asian QSR sector.

REFERENCES

Alam A., et al. 2024. Climate change and geopolitical conflicts: The role of ESG readiness. Journal of Environmental Management, 353. doi:10.1016/j.jenvman.2024.120284.

Alon I. 2006. Service Franchising: A Global Perspective. Springer. doi:10.1007/0-387-28256-4.

An M., Wu Y., Ouyang Z. 2023. Spatial-Temporal evolvement and the contributing factors for the economic potential of ecosystem services in counties situated along a river. Journal for Nature Conservation, 75. doi:10.1016/j.jnc.2023.126461.

Derossi A., et al. 2023. Three-Dimensional Printing of Foods: A Critical Review of the Present State in Healthcare Applications, and Potential Risks and Benefits. Foods, 12(17). doi:10.3390/foods12173287.

Dias S., Espadinha-Cruz P., & Matos F. 2023. A Porter's Five Forces Model Proposal for Additive Manufacturing Technology: A Case Study in Portuguese industry. Procedia Computer Science, 217. doi:10.1016/j.procs.2022.12.212.

Dodo Brands. 2023. Consolidated Financial Statements. URL: https://dodobrands.io/ (accessed on 12.12.2025).

Dodo Brands. Annual Report 2021. URL: https://dodobrands.io/ (accessed on 12.12.2025).

Dodo Brands. Annual Report 2022. URL: https://dodobrands.io/ (accessed on 12.12.2025).

Dodo Brands. Management Discussion and Analysis. 2023. URL: https://dodobrands.io/ (accessed on 12.12.2025).

Ford M. 2015. Rise of the Robots: Technology and the Threat of a Jobless Future. Basic Books.

Gorbacheva A., Levina A. Digital support for sustainable development of the Arctic zone.

Technoeconomics. 2024. 3. 1 (8). 26–40. DOI: https://doi.org/10.57809/2024.3.1.8.3
Hargrave M. 2024. Market Saturation: Macro and Microeconomic Dimensions. Investopedia. URL: https://www.investopedia.com/terms/m/marketsaturation.asp (accessed on 03.12.2025).
Kotler P., Keller K. L. 2016. Marketing Management (15th ed.). Pearson.
Lypez Prol J., Steininger K. W. 2020. Photovoltaic self-consumption is now profitable in Spain: Effects of the new regulation on prosumers' internal rate of return. Energy Policy, 146. doi:10.1016/j.enpol.2020.111793.
Porter M. E. 1985. Competitive Advantage: Creating and Sustaining Superior Performance. Free Press.
Rostami S., Creemers S., Leus R. 2024. Maximizing the net present value of a project under uncertainty: Activity delays and dynamic policies. European Journal of Operational Research, 317(1). doi:10.1016/j.ejor.2024.03.029.
Wu P., Tang T., Zhou L., Martinez L. 2024. A decision-support model through online reviews: Consumer preference analysis and product ranking. Information Processing & Management, 61(4). doi:10.1016/j.ipm.2023.103728.
BRICS Competition Centre. 2025. China's Platform-Economy Standards Body Issues First Food-Delivery Governance Rule. URL: https://www.bricscompetition.org/ (accessed on 24.01.2026).
EDHEC. 2024. 3D Food Printing: what are the main adoption challenges ahead? URL: https://www.edhec.edu/en/research-and-faculty/edhec-vox/3-d-food-printing-what-are-the-main-adoption-challenges-ahead (accessed on 26.01.2026).
SNS Insider. 2024. Restaurant Delivery Robot Market Size & Share Report 2032. URL: https://www.snsinsider.com/reports/restaurant-delivery-robot-market-3493 (accessed on 03.01.2026).
Starship Technologies. 2025. Autonomous robot delivery - The future of delivery. URL: https://www.starship.xyz/ (accessed on 06.12.2025).
Towards Data Science. 2021. The Pizza Chain That Became an AI Front-Runner. URL: https://towardsdatascience.com/the-pizza-chain-that-became-an-ai-front-runner-a0e-297320cf6 (accessed on 07.12.2025).

СПИСОК ИСТОЧНИКОВ

Alam A., et al. 2024. Climate change and geopolitical conflicts: The role of ESG readiness. Journal of Environmental Management, 353. doi:10.1016/j.jenvman.2024.120284.
Alon I. 2006. Service Franchising: A Global Perspective. Springer. doi:10.1007/0-387-28256-4.
An M., Wu Y., Ouyang Z. 2023. Spatial-Temporal evolvement and the contributing factors for the economic potential of ecosystem services in counties situated along a river. Journal for Nature Conservation, 75. doi:10.1016/j.jnc.2023.126461.
Derossi A., et al. 2023. Three-Dimensional Printing of Foods: A Critical Review of the Present State in Healthcare Applications, and Potential Risks and Benefits. Foods, 12(17). doi:10.3390/foods12173287.
Dias S., Espadinha-Cruz P., & Matos F. 2023. A Porter's Five Forces Model Proposal for Additive Manufacturing Technology: A Case Study in Portuguese industry. Procedia Computer Science, 217. doi:10.1016/j.procs.2022.12.212.
Dodo Brands. 2023. Consolidated Financial Statements. URL: https://dodobrands.io/ (accessed on 12.12.2025).
Dodo Brands. Annual Report 2021. URL: https://dodobrands.io/ (accessed on 12.12.2025).
Dodo Brands. Annual Report 2022. URL: https://dodobrands.io/ (accessed on 12.12.2025).
Dodo Brands. Management Discussion and Analysis. 2023. URL: https://dodobrands.io/ (accessed on 12.12.2025).
Ford M. 2015. Rise of the Robots: Technology and the Threat of a Jobless Future. Basic Books.
Gorbacheva A., Levina A. Digital support for sustainable development of the Arctic zone. Technoeconomics. 2024. 3. 1 (8). 26–40. DOI: https://doi.org/10.57809/2024.3.1.8.3
Hargrave M. 2024. Market Saturation: Macro and Microeconomic Dimensions. Investopedia.

URL: https://www.investopedia.com/terms/m/marketsaturation.asp (accessed on 03.12.2025).
Kotler P., Keller K. L. 2016. Marketing Management (15th ed.). Pearson.
Lypez Prol J., Steininger K. W. 2020. Photovoltaic self-consumption is now profitable in Spain: Effects of the new regulation on prosumers' internal rate of return. Energy Policy, 146. doi:10.1016/j.enpol.2020.111793.
Porter M. E. 1985. Competitive Advantage: Creating and Sustaining Superior Performance. Free Press.
Rostami S., Creemers S., Leus R. 2024. Maximizing the net present value of a project under uncertainty: Activity delays and dynamic policies. European Journal of Operational Research, 317(1). doi:10.1016/j.ejor.2024.03.029.
Wu P., Tang T., Zhou L., Martinez L. 2024. A decision-support model through online reviews: Consumer preference analysis and product ranking. Information Processing & Management, 61(4). doi:10.1016/j.ipm.2023.103728.
BRICS Competition Centre. 2025. China's Platform-Economy Standards Body Issues First Food-Delivery Governance Rule. URL: https://www.bricscompetition.org/ (accessed on 24.01.2026).
EDHEC. 2024. 3D Food Printing: what are the main adoption challenges ahead? URL: https://www.edhec.edu/en/research-and-faculty/edhec-vox/3-d-food-printing-what-are-the-main-adoption-challenges-ahead (accessed on 26.01.2026).
SNS Insider. 2024. Restaurant Delivery Robot Market Size & Share Report 2032. URL: https://www.snsinsider.com/reports/restaurant-delivery-robot-market-3493 (accessed on 03.01.2026).
Starship Technologies. 2025. Autonomous robot delivery - The future of delivery. URL: https://www.starship.xyz/ (accessed on 06.12.2025).
Towards Data Science. 2021. The Pizza Chain That Became an AI Front-Runner. URL: https://towardsdatascience.com/the-pizza-chain-that-became-an-ai-front-runner-a0e-297320cf6 (accessed on 07.12.2025).

INFORMATION ABOUT AUTHORS / ИНФОРМАЦИЯ ОБ АВТОРАХ

KHAFETULIN Artur S. – student.
E-mail: akhafetulin@mail.ru
ХАФЕТУЛИН Артур Сергеевич – студент.
E-mail: akhafetulin@mail.ru

GUGUTISHVILI Dayana M. – specialist.
E-mail: gugutishvilid@mail.ru
ГУГУТИШВИЛИ Даяна Меджидовна – специалист.
E-mail: gugutishvilid@mail.ru
ORCID: https://orcid.org/0000-0003-1162-7733

Статья поступила в редакцию 01.02.2026; одобрена после рецензирования 16.02.2026; принята к публикации 21.03.2026.
The article was submitted 01.02.2026; approved after reviewing 16.02.2026; accepted for publication 21.03.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 32–40.
Техноэкономика. 2026, Том 5, № 1 (16). С. 32–40.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.3

HYBRID AI MODELS: A COMBINATION OF CLASSICAL ALGORITHMS AND NEURAL NETWORKS TO ENHANCE INTERPRETABILITY

Saveliy Cherepanov ✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ cherepanov.sv@edu.spbstu.ru

Abstract. Modern Deep Learning neural networks demonstrate high accuracy in classification and forecasting tasks, but significant limitations remain in the interpretability of the results. This creates an obstacle to application in critical areas where maximum transparency of decisions is required. In this paper, we propose the use of a hybrid approach that combines both feature extraction methods based on neural networks and classical interpreted machine learning algorithms. In the course of the work, an architecture was developed in which a neural network forms a compact representation of data, and the final decision is made by an interpreted model in the form of a decision tree or logical regression. Experiments have been conducted on open datasets, confirming that the proposed approach allows for increased interpretability while maintaining accuracy comparable to Deep Learning models. The results demonstrate the promise of hybrid architectures for areas requiring transparency and explainability of the results.

Keywords: neural networks, interpretability, hybrid models, data analytics, decision-making

Citation: Cherepanov S. 2026. Hybrid AI models: a combination of classical algorithms and neural networks to enhance interpretability. Technoeconomics 5, 1 (16), 32–40. DOI: https://doi.org/10.57809/2026.5.1.16.3

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.3

ГИБРИДНЫЕ МОДЕЛИ ИИ: СОЧЕТАНИЕ КЛАССИЧЕСКИХ АЛГОРИТМОВ И НЕЙРОСЕТЕЙ ДЛЯ ПОВЫШЕНИЯ ИНТЕРПРЕТИРУЕМОСТИ

Савелий Черепанов ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

✉ cherepanov.sv@edu.spbstu.ru

Аннотация. Современные нейронные сети с глубоким обучением (Deep Learning) демонстрируют высокую точность в задачах классификации и прогнозирования, однако остаются существенные ограничения в интерпретируемости результатов. Это создает препятствие к применению в критически важных областях, где требуется максимальная прозрачность принимаемых решений. В данной работе предлагается применение гибридного подхода, сочетающего в себе как методы извлечения признаков на основе нейронных сетей, так и классические интерпретируемые алгоритмы машинного обучения. В ходе работы была разработана архитектура, в которой нейросеть формирует компактное представление данных, а финальное решение принимает интерпретируемая модель в виде дерева решений или логической регрессии. Проведены эксперименты на открытых наборах данных, подтверждающие, что предложенный подход позволяет добиться увеличения интерпретируемости при сохранении точности, сопоставимого с Deep Learning моделями. Результаты демонстрируют перспективность гибридных архитектур для областей, требующих прозрачности и объяснимости результатов.

Ключевые слова: нейронные сети, интерпретируемость, гибридные модели, аналитика данных, принятие решений

Для цитирования: Черепанов С.В. Гибридные модели ИИ: сочетание классических алгоритмов и нейросетей для повышения интерпретируемости // Техноэкономика. 2026. Т. 5, № 1 (16). С. 32–40. DOI: https://doi.org/10.57809/2026.5.1.16.3

Introduction

The growth of computing capabilities and the emergence of large datasets have led to the widespread and almost mandatory use of deep learning networks, which are well suited for solving problems of analysis, forecasting, etc. However, the main problem with using such technologies in their basic form is that the high accuracy of the data obtained is ensured by the low interpretability of the results. This behavior of technologies creates difficulties in their use in critically important government areas such as medicine, economics, law and public administration, where the results obtained must be extremely clear and understandable to a person so that subsequent decisions have a clear justification. The popularization of AI, in addition to the complexity of its use, creates the need to develop a new solution that will meet the needs of not only key stakeholders, but also others where technology is already actively used (Ignatiev and Levina, 2024; Klimentov, 2025; Kutuzova, 2024). One of the promising solutions to the above problem is the creation of hybrid methods that combine the advantages and speed of neural network analysis with classical, explicable algorithms. In such a system, the neural network performs the functions of extracting specified features, and the interpreted model makes a decision based on a more transparent representation (Bouwmans et al., 2019; Vilone et al., 2021). In the course of the work, a hybrid architecture based on the above approach was investigated and proposed.

The goals and objectives of this study are:

1. Analysis of existing approaches to ensuring interpretability in AI and identification of advantages of hybrid architectures.
2. Development of the architecture of a hybrid AI model using a neural network to extract compact features (embeddings), and an interpreted classifier to make a decision.
3. Assessment of the impact of embeddings on key metrics: classification accuracy, complexity and stability of the interpreted model (decision tree), as well as on qualitative interpretability.
4. A description of the limitations identified during the study.

Materials and Methods

To solve the tasks set, the study is structured in several steps. To begin with, a theoretical analysis and synthesis of existing methods of interpreted AI is carried out. And then a hybrid architecture is implemented, consisting of a two-component system: a neural network feature extractor (MLP) and an interpreter classifier (decision tree).

The key methodological idea is to train a neural network to form embeddings that not only preserve the information necessary for classification, but also provide a more stable and simple logical representation for subsequent training of the decision tree (Adadi and Berrada, 2018; Skatova, 2024). A comparative analysis will be conducted between a reference interpreted model trained on the initial features and a hybrid model trained on embeddings obtained from a neural network.

The interpretability of AI can be divided into intrinsic (intrinsic) and resultant (post-hoc). Intrinsic models are logical regression, decision trees, or linear models that have unconditional transparent logic. Post-hoc models, such as LIME or SHAP, do not change the model, but explain its behavior from the outside, but the explanations themselves can only be approximate and do not provide a correct interpretation (Adler et al., 2018; He et al., 2016; Molnar, 2022).

The proposed hybrid approach is based on current trends in the field of explicable artificial intelligence. Unlike post-hoc methods such as LIME and SHAP, which provide only local and approximate explanations for complex models, this study focuses on creating an intrinsically interpretable model. The work is based on the concept formulated by Rudin, who argues that interpretable models should be used for high-risk solutions instead of trying to explain "black boxes" (Rudin, 2019). The proposed architecture echoes the direction described by Al-Shedivat, where neural networks are used for contextual representation of data, but in this work the emphasis is on separation of functions: the neural network is responsible for representation, and the decision tree is responsible for interpreted classification (Al-Shedivat et al., 2020). In addition, the work expands Caruana's ideas about the use of interpreted models in medicine, showing that even simple neural network preprocessing can improve the stability and transparency of final decisions without significant loss of accuracy (Caruana et al., 2015). The approach to feature extraction through a hidden layer of a neural network for subsequent training of simple models is also reflected in Zhang's work on learning compact representations, which emphasizes that reducing dimensionality and eliminating noise increases interpretability (Zhang et al., 2021).

Deep Learning networks are high–performance, but they have a number of limitations - their internal representations are extremely difficult to interpret, they are subject to shifting and unstable decisions, and the process of making these decisions is difficult for the user to explain. These limitations create promising research in the application of hybrid architectures (Lapuschkin et al., 2019; Nguyen et al., 2015; Plumerault et al., 2020).

Approaches combining neural networks and symbolic methods have received attention in recent years, among them are (Alonso, 2020; Lakkaraju et al., 2019; Pochetniy, 2025):

1. Rule extraction, which try to extract rules from trained networks;
2. Surrogate models, which train an interpreted model based on neural network predictions.

However, the above-mentioned methods are resultant, not embedded, which also remain insufficiently studied. In these methods, the classical model makes a decision based on the features that the trained neural network selects. Based on this, an experiment was conducted to improve the interpretability of the result using a hybrid architecture.

Experimental data

During the evaluation of the model, a data set such as breast cancer statistics for St. Petersburg and the Leningrad region (biomedical diagnostics) was used. The criterion for choosing a set is the requirement for clarity of the results obtained, the need to ensure high accuracy, as well as the openness of the data.

To conduct the experiment, a model was built with the following settings: a neural network was built based on MLP with hidden layers, one of which is 16 neurons in size, as well as an interpreted model in the form of a decision tree up to 4 levels deep. The criteria for analyzing the results obtained were the following characteristics: interpretability of the final model, estimation accuracy, sample stability, complexity of the decision tree, stability of the tree structure during retraining of the model.

In addition, another goal of the experiment was to test the hypothesis that compact embeddings formed by a trained neural network improve the quality of interpretation while reducing complexity.

For the tests, a set containing 569 samples and 30 valid signs obtained from the analysis of medical images from real practice was used. The target sample variable is 0 for the condition that the tumor is benign and 1 for a malignant tumor. Before the experiment, data was preprocessed, which included normalization of features (StandardScaler), splitting into training and test samples in a 70/30 ratio, as well as elimination of outliers by threshold methods (IQR). As a result of these operations, a balanced data set (class ratio ~ 37/63) was obtained, which increases stability and reduces the risk of model bias.

The hybrid model used in the study consists of a neural network (MLP) As a feature extraction tool, it has the following characteristics: input of 30 features, a hidden layer of 16 neurons, ReLU activation, output by 16–dimensional embedding, Adam optimization, and an epoch limit of 800. The neural network was trained as a regular classifier, then only a hidden layer was used, which extracts structured data. This made it possible to reduce the dimension and eliminate linear and nonlinear dependencies between the initial features. The model also includes an interpretation that, before hybridization, was a classic A tree with a depth of up to 4 levels and a decision tree with a depth of up to 3 levels, trained on embeddings after. Thus, the tree in system B is more depth-limited, but it benefits from better feature factorization.

To assess the quality of the data obtained, metrics such as Accuracy, Precision, Recall, F1-score, ROC-AUC, number of leaves, depth of the tree, number of rules, and stability of the tree were used.

Results and Discussion

You can see the results of the interpretation below and ROC-curves.

Tree A is shown in Figure 1. This is a classic Decision Tree, trained directly on the initial 30 features of the dataset. The depth of the tree was limited to 4 levels. In the experiment, this model serves as a benchmark for "default interpretability" and a reference point for comparing complexity and accuracy.

Tree B is a decision tree with a depth of up to 3 levels, which is trained not on the initial features, but on 16-dimensional embeddings extracted by a trained neural network (MLP).

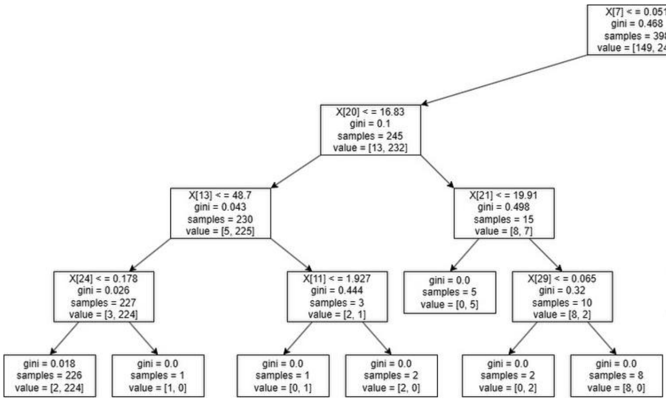

Fig. 1. Tree A.

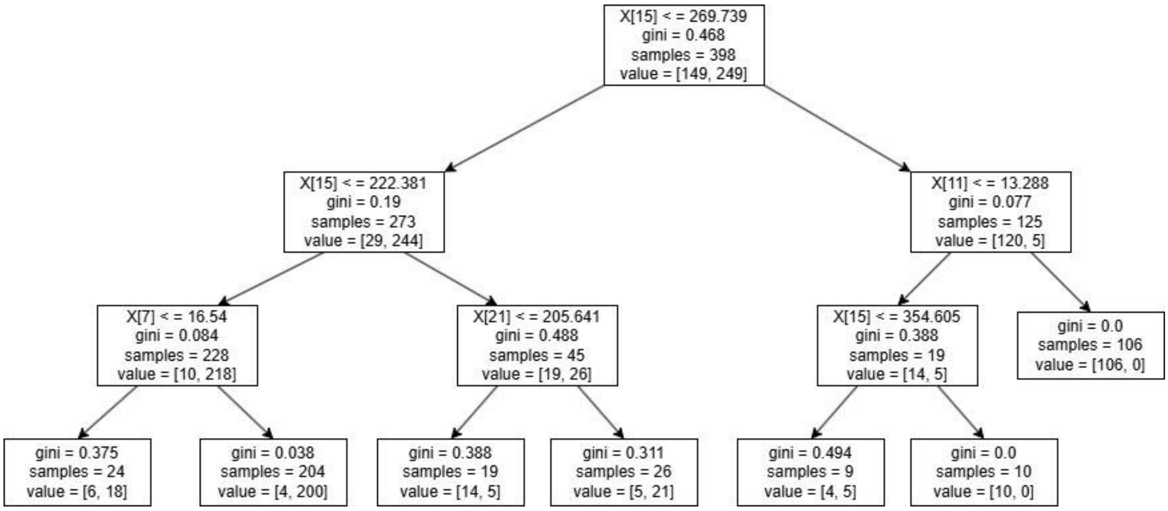

Fig. 2. Tree B.

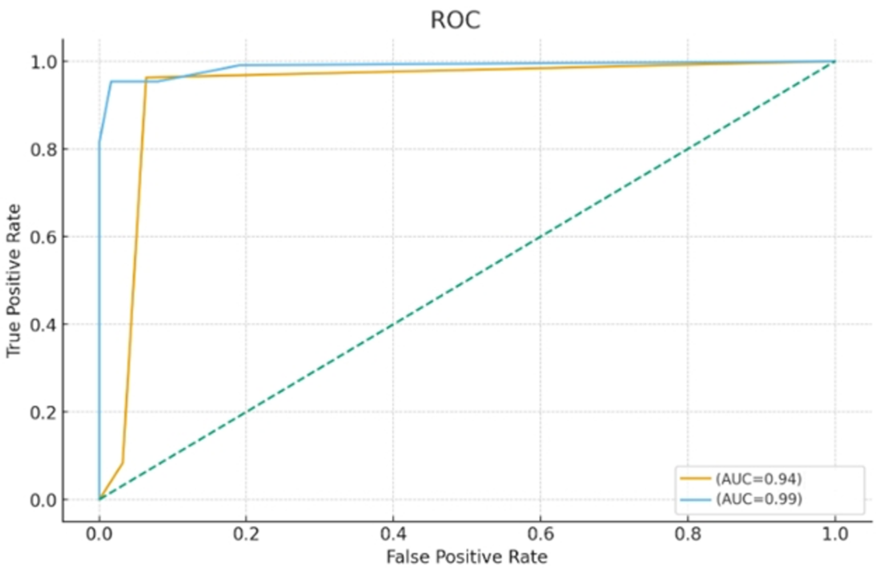

Fig. 3. Graph of ROC curves.

The main results obtained during the experiment are presented below in the form of tables 1, 2, 3, 4 with a brief explanation.

Table 1. Classification quality.

Model	Accuracy	ROC-AUC
Tree A	0,96	0,97
Tree B	0,95	0,96

Based on this, the accuracy of the hybrid model with tree B is inferior to the original one, which reduces the quality of the assessment, but remains at a high and comparable level.

Table 2. The complexity of the tree.

Model	Depth	Number of rules	Number of leaves
Tree A	4	12	7
Tree B	3	6	4

The tree structure of the hybrid model becomes almost twice as short, which increases interpretability and is the main task of the study.

Table 3. Stability during repeated training.

Model	Stability of the structure
Tree A	Low – high variability with the same sample
Tree B	High – minimal changes

The same conditions were set for both models – the same data sample and the number of runs (10 times). Based on this, it can be concluded that embedding reduces the sensitivity of the model structure to noise.

Table 3. Stability during repeated training.

Model A	1. The solution depends on the specific numbers of the initial features 2. The tree relies on combinations of parameters
Model B	1. Solutions are based on more general embeddings 2. Factors consolidate groups of features, highlighting patterns 3. Increased interpretability due to less overloading of rules

It is worth mentioning that the following restrictions were adopted during the study:

1. The analysis was performed on a single dataset, so the accuracy of the system must be confirmed by increasing the number of samples and the data extracted from them;
2. A simple MLP was used, which, although it simplifies the research task, but when using the logic of building on more complex architectures, it is possible to achieve comparatively better embeddings;
3. The interpretation of the resulting embeddings still needs the use of additional methods such as PCA, SHAP, CCA, etc.

Conclusion

In this study, interpreted models and their advantages over post-hoc methods were considered. This made it possible to justify the need for the implementation and subsequent use of interpreted models.

A two-component model was successfully implemented and tested, where a neural network (MLP) performs the role of extracting compact and structured features, and a decision tree acts as a transparent classifier. This confirmed the fundamental possibility of separating the functions of "presentation" and "decision-making".

The obtained results provide clear answers to research questions. The use of embeddings made it possible to reduce the complexity of the decision tree (the depth was reduced from 4 to 3, the number of rules — from 12 to 6), as well as to increase the stability of its structure during repeated training, which indicates a reduction in the influence of noise. The accuracy of the hybrid model (Accuracy 0.95) remained at a level comparable to the baseline (0.96), which confirms the hypothesis of maintaining the quality of forecasting.

Despite the results achieved, limitations have been identified that require further study. The task of interpreting the embeddings themselves remains open and requires the use of additional tools (for example, SHAP or CCA). In addition, it is necessary to validate the proposed approach on a wider range of datasets and using more complex neural network architectures (for example, convolutional or recurrent), which is the vector of future research. Thus, all the research tasks have been solved, and the proposed hybrid model has demonstrated its promise for creating transparent and reliable AI systems.

REFERENCES

Adadi A., Berrada M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). Access 6, 52138—52160.

Adler P., Falk C., Friedler S.A., Nix T., Rybeck G., Scheidegger C., Smith B., Venkatasubramanian S. 2018. Auditing black-box models for indirect influence. Knowledge and Information Systems 54, 95—122.

Alonso J.M. 2020. Teaching Explainable Artificial Intelligence to High School Students. International Journal of Computational Intelligence Systems 13, 974—987.

Al-Shedivat M., Wilson A.G., Saatchi Y., Hu Z., Xing. E.P. 2020. Contextual Explanation Networks. Journal of Machine Learning Research 21, 1—44.

Bouwmans T., Javed S., Sultana M., Jung S.K. 2019. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. Neural Networks 1, 40. DOI: 10.1016/j.neunet.2019.04.024.

Caruana R., Lou Y., Gehrke J., Koch P., Sturm M., Elhadad N. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721—1730. DOI: 10.1145/2783258.2788613;

He K., Zhang X., Ren S., Sun J. 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. DOI: 10.1109/CVPR.2016.90

Ignatiev P., Levina A. 2024. Artificial intelligence and artificial neural networks in healthcare. Technoeconomics 3, 4 (11), 28—41. DOI: https://doi.org/10.57809/2024.3.4.11.3

Klimentov A. 2025. Predicting claims in auto insurance using deep neural networks. Technoeconomics 4, 4 (15), 36—43. DOI: https://doi.org/10.57809/2025.4.4.15.2

Kutuzova A. 2024. AI-support architecture in digital marketing. Technoeconomics 3, 4 (11), 69—78. DOI: https://doi.org/10.57809/2024.3.4.11.6

Lakkaraju H., Kamar E., Caruana R., Leskovec J. 2019. Faithful and Customizable Explanations of Black Box Models. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES), 131—138. DOI: 10.1145/3306618.3314229.

Lapuschkin S., Waldchen S., Binder A., Montavon G., Samek W., Muller K.R. 2019. Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications 10, 1096.

Molnar C. 2022. Interpretable Machine Learning. 2nd ed.

Nguyen A., Yosinski J., Clune J. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern Recognition, 427–436.

Plumerault A., Borgne H.L., Hudelot C. 2020. Controlling generative models with continuous factors of variations. In: Proceedings of the International Conference on Learning Representations, 2020.

Pochetniy V.A. 2025. Integrating generative AI for technological trend analysis and patent research automation. Technoeconomics 4, 2 (13), 4–20. DOI: https://doi.org/10.57809/2025.4.2.13.1

Rudin C. 2019. Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 5, 206–215. DOI: 10.1038/s42256-019-0048-x.

Skatova M. 2024. Assessment of requirements of regulatory documents on the use of artificial intelligence in higher education. Technoeconomics 3, 2 (9), 22–33. DOI: https://doi.org/10.57809/2024.3.2.9.2

Vilone G., Longo L. 2021. Classification of Explainable Artificial Intelligence Methods through Their Output Formats. Machine Learning and Knowledge Extraction 3, 3, 615–661. DOI: 10.3390/make3030032.

Zhang Q., Yang Y., Ma H., Wu Y.N. 2021. A Survey on Neural Network Interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence 5, 5, 726–742. DOI: 10.1109/TETCI.2021.3100641.

СПИСОК ИСТОЧНИКОВ

Adadi A., Berrada M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). Access 6, 52138–52160.

Adler P., Falk C., Friedler S.A., Nix T., Rybeck G., Scheidegger C., Smith B., Venkatasubramanian S. 2018. Auditing black-box models for indirect influence. Knowledge and Information Systems 54, 95–122.

Alonso J.M. 2020. Teaching Explainable Artificial Intelligence to High School Students. International Journal of Computational Intelligence Systems 13, 974–987.

Al-Shedivat M., Wilson A.G., Saatchi Y., Hu Z., Xing. E.P. 2020. Contextual Explanation Networks. Journal of Machine Learning Research 21, 1–44.

Bouwmans T., Javed S., Sultana M., Jung S.K. 2019. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. Neural Networks 1, 40. DOI: 10.1016/j.neunet.2019.04.024.

Caruana R., Lou Y., Gehrke J., Koch P., Sturm M., Elhadad N. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730. DOI: 10.1145/2783258.2788613;

He K., Zhang X., Ren S., Sun J. 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. DOI: 10.1109/CVPR.2016.90

Ignatiev P., Levina A. 2024. Artificial intelligence and artificial neural networks in healthcare. Technoeconomics 3, 4 (11), 28–41. DOI: https://doi.org/10.57809/2024.3.4.11.3

Klimentov A. 2025. Predicting claims in auto insurance using deep neural networks. Technoeconomics 4, 4 (15), 36–43. DOI: https://doi.org/10.57809/2025.4.4.15.2

Kutuzova A. 2024. AI-support architecture in digital marketing. Technoeconomics 3, 4 (11), 69–78. DOI: https://doi.org/10.57809/2024.3.4.11.6

Lakkaraju H., Kamar E., Caruana R., Leskovec J. 2019. Faithful and Customizable Explanations of Black Box Models. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES), 131–138. DOI: 10.1145/3306618.3314229.

Lapuschkin S., Waldchen S., Binder A., Montavon G., Samek W., Muller K.R. 2019. Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications 10, 1096.

Molnar C. 2022. Interpretable Machine Learning. 2nd ed.

Nguyen A., Yosinski J., Clune J. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern Recognition, 427–436.

Plumerault A., Borgne H.L., Hudelot C. 2020. Controlling generative models with continuous factors of variations. In: Proceedings of the International Conference on Learning Representations, 2020.

Pochetniy V.A. 2025. Integrating generative AI for technological trend analysis and patent research automation. Technoeconomics 4, 2 (13), 4–20. DOI: https://doi.org/10.57809/2025.4.2.13.1

Rudin C. 2019. Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 5, 206–215. DOI: 10.1038/s42256-019-0048-x.

Skatova M. 2024. Assessment of requirements of regulatory documents on the use of artificial intelligence in higher education. Technoeconomics 3, 2 (9), 22–33. DOI: https://doi.org/10.57809/2024.3.2.9.2

Vilone G., Longo L. 2021. Classification of Explainable Artificial Intelligence Methods through Their Output Formats. Machine Learning and Knowledge Extraction 3, 3, 615–661. DOI: 10.3390/make3030032.

Zhang Q., Yang Y., Ma H., Wu Y.N. 2021. A Survey on Neural Network Interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence 5, 5, 726–742. DOI: 10.1109/TETCI.2021.3100641.

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

CHEREPANOV Saveliy V. – student.
E-mail: cherepanov.sv@edu.spbstu.ru
ЧЕРЕПАНОВ Савелий Васильевич – студент.
E-mail: cherepanov.sv@edu.spbstu.ru

Статья поступила в редакцию 16.12.2025; одобрена после рецензирования 22.02.2026; принята к публикации 15.03.2026.

The article was submitted 16.12.2025; approved after reviewing 22.02.2026; accepted for publication 15.03.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 41–53.
Техноэкономика. 2026, Том 5, № 1 (16). С. 41–53.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.4

LARGE LANGUAGE MODELS (LLMS) IN E-COMMERCE

Nahid Mehri ✉

Apadana Institute of Higher Education, Shiraz, Iran

✉ nahid.mehrii@yahoo.com

Abstract. Large language models (LLMs) have emerged as highly influential technologies in e-commerce, offering possibilities for complex applications in customer interaction, business efficiency, and decision-making. LLMs such as GPT-4, BERT and t5 have demonstrated high accuracy in natural language processing with transformer-based deep learning architecture and contribute to personalized recommendations, content creation and intelligent customer engagement. Existing literature has shown that LLM-based chatbots have the potential to solve more than 80% of common customer questions. Similarly, the recommendations made by LLMs have shown a significant contribution in terms of revenue growth for e-commerce businesses. Despite these potential applications, there are still major challenges in implementing LLMs in e-commerce, including privacy, cost and ethics issues. The purpose of this paper is to systematically study the application of LLMs in e-commerce, including its advantages, disadvantages and potential. With comparative case study approaches for Amazon's three e-commerce giants, Alibaba and Shopify, and by analyzing emerging trends in multifaceted AI and voice commerce, while considering the key implementation challenges, the research seeks to provide valuable insights to optimize the application of LLMs in e-commerce.

Keywords: large language models, e-commerce, natural language processing, chatbots

Citation: Mehri N. 2026. Large Language Models (LLMs) in E-Commerce. Technoeconomics 5, 1 (16), 41–53. DOI: https://doi.org/10.57809/2026.5.1.16.4

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.4

БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ (LLM) В ЭЛЕКТРОННОЙ КОММЕРЦИИ

Нахид Мехри ✉

Институт высшего образования Ападана, Шираз, Иран

✉ nahid.mehrii@yahoo.com

Аннотация. Большие языковые модели (БЯМ) стали чрезвычайно влиятельными технологиями в электронной коммерции, открывая возможности для сложных приложений во взаимодействии с клиентами, повышения эффективности бизнеса и принятия решений. Такие БЯМ, как GPT-4, BERT и t5, продемонстрировали высокую точность в обработке естественного языка с использованием архитектуры глубокого обучения на основе трансформеров и способствуют персонализированным рекомендациям, созданию контента и интеллектуальному взаимодействию с клиентами. Существующая литература показывает, что чат-боты на основе БЯМ способны решить более 80% распространенных вопросов клиентов. Аналогично, рекомендации, предоставляемые БЯМ, внесли значительный вклад в рост доходов предприятий электронной коммерции. Несмотря на эти потенциальные возможности применения, существуют серьезные проблемы внедрения БЯМ в электронную коммерцию, включая вопросы конфиденциальности, стоимости и этики. Цель данной статьи — систематически изучить применение БЯМ в электронной коммерции, включая их преимущества, недостатки и потенциал. На основе сравнительного анализа конкретных примеров трех гигантов электронной коммерции Amazon — Alibaba и Shopify, а также анализа новых тенденций в многогранной ИИ и голосовой коммерции с учетом ключевых проблем внедрения, исследование направлено на предоставление ценных сведений для оптимизации применения БЯМ в электронной коммерции.

Ключевые слова: большие языковые модели, электронная коммерция, обработка естественного языка, чат-боты

Для цитирования: Мехри Н. Большие языковые модели (LLM) в электронной коммерции // Техноэкономика. 2026. Т. 5, № 1 (16). С. 41—53. DOI: https://doi.org/10.57809/2026.5.1.16.4

Introduction

The rapid advancement of artificial intelligence has fundamentally reshaped the landscape of digital commerce. Among the most consequential recent developments is the emergence of Large Language Models (LLMs) neural network architectures trained on massive corpora of text data that demonstrate remarkable capacity for understanding, generating, and reasoning with natural language (Brown et al., 2020). Platforms such as Amazon, Alibaba, and Shopify have been among the first to integrate these systems at scale, deploying them across customer service, content generation, fraud detection, and supply chain management.

The problem this study addresses is both technical and economic in nature: While LLMs provide quantifiable benefits in terms of performance, the deployment of LLMs in a business context creates a multifaceted and complex set of trade-offs. Despite the increasing adoption of LLMs in the industry, the extant research lacks a comprehensive and cross-platform study on the quantification of the trade-offs associated with LLMs. Understanding the varying approaches to address the challenges associated with LLMs by different organizations is essential for practitioners and policymakers alike.

The importance and relevance of the problem being addressed by the study are evident by the quantification and growth rate of the LLM industry. The global market size for AI in e-commerce was valued at around 5.8 billion dollars in 2023 and is expected to grow at a compound annual growth rate of more than 14 percent until 2030 (Online Retailer Survey, 2024). Of the total market size, the LLM industry has the highest growth rate due to the increase in the efficiency of LLMs and the decrease in the cost incurred by LLMs. However, the growth in the industry also increases the risks associated with the deployment of LLMs by SMEs.

The purpose of this paper is to provide a structured, evidence-based analysis of how LLMs are currently applied in e-commerce, what measurable benefits they produce, what barriers constrain their adoption, and what directions are likely to define the next phase of their development. To achieve this purpose, the research addresses five specific tasks: (1) identify the primary application domains of LLMs in e-commerce; (2) quantify their measurable benefits using available performance indicators; (3) characterize the critical implementation challenges; (4) compare deployment strategies across industry leaders; and (5) examine emerging trends shaping future LLM integration.

Methodology

This research makes use of the qualitative comparative case study approach. This approach has proven useful in the exploration of contemporary technological phenomenon in real-world organizational contexts in which experimental control cannot be exercised (Yin, 2018).

The research tasks outlined in the introduction section were addressed through the following three-step structured approach:

Step 1 — E-commerce systems assessment: An evaluation of AI-driven systems within each case company to identify domains where LLMs have been implemented. This involved reviewing corporate technical documentation, white papers, and published reports from Amazon Science (2023), Alibaba Cloud (2023), and Shopify's developer platform.

Step 2 — LLM-to-application mapping: A systematic mapping of LLM capabilities to specific e-commerce functions, including personalized recommendations, sentiment analysis, content generation, and fraud detection. Capabilities were classified by model type (GPT-4, BERT, T5) and application domain.

Step 3 — Comparative performance evaluation: A cross-case analysis using key performance indicators (KPIs) reported in the literature and industry sources, including response accuracy, response latency, customer satisfaction scores (CSAT), and query resolution rates (Rasheed et al., 2025).

Data were collected through systematic documentary analysis of secondary sources: peer-reviewed academic literature retrieved from Google Scholar and Semantic Scholar (2017—2025), corporate white papers and annual reports, and industry benchmark datasets including the Online Retailer Survey (2024). Case companies Amazon, Alibaba, and Shopify were selected purposively to represent diversity in scale (enterprise to SME), geography (North America and Asia-Pacific), and application focus (recommendation engines, multilingual customer service, and democratized commerce tools).

The qualitative content analysis of collected data followed a structured coding process: first, open coding to identify themes across sources; second, axial coding to relate themes to the five research tasks; and third, selective coding to produce the comparative findings reported in the Results Section. The methodological logic is summarized in Figure 1.

Requirement Gathering	Design	Development (Mapping)	Testing (KPI Analysis)	Deployment & Conclusion

Fig. 1. Iterative research workflow applied in this study.

Literature Review

Foundations of Large Language Models

A Large Language Model (LLM) is an advanced artificial intelligence system with high accuracy in processing, producing, and interpreting human languages. LLMs, such as GPT-4, BERT, and T5, utilize advanced deep learning models like transformers to process large amounts of text data (Vaswani et al., 2017). LLMs' versatility has been proven in various fields, including NLP, coding, and content creation. LLMs are now being incorporated into e-commerce platforms to improve customer experience, business processes, and growth. The recent development of LLMs has significantly increased parameters, with strong performance in NLP-related tasks (Touvron et al., 2023).Kalyan et al. (2022) describe LLMs as a special class of pretrained language models obtained by scaling model size, pretraining corpus, and computation. LLMs treat any NLP task as a conditional text generation problem and generate the desired output by conditioning on an input prompt. Because of their large size and pretraining on large volumes of text, LLMs exhibit 'emergent abilities' not present in smaller models. GPT-4, for instance, outperforms GPT-3 by over 11 percentage points on standardized academic examinations and achieves 86.8% accuracy on domain-specific Q&A benchmarks.

Research by Brown et al. (2020) highlights the transformative potential of LLMs in automating tasks that traditionally required human intervention. Their ability to understand context and generate coherent responses makes them invaluable for e-commerce businesses seeking to scale operations while maintaining high-quality customer interactions. The development of BERT (Devlin et al., 2019) further demonstrated that bidirectional pretraining on unlabeled text produces representations that outperform task-specific architectures across a wide range of NLP benchmarks.

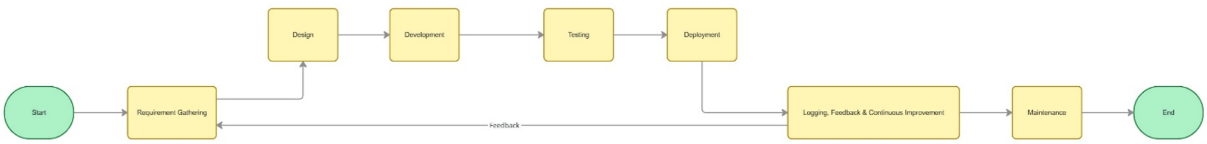

Fig. 2. Overview of LLM architecture and e-commerce application domains.

Implementation of LLMs in E-Commerce

The inclusion of LLMs in e-commerce has increased customer engagement, streamlined processes and increased sales. E-commerce activities LLMs use different models based on their design and type of training data, as presented in Table 1.

Table 1. Comparative analysis of LLM capabilities across e-commerce tasks.

Task	GPT-4 (OpenAI)	BERT (Google)	T5 (Google)
Customer Support	High conversational fluency; handles complex queries (Xu et al., 2022).	Optimized for intent classification; less generative (Devlin et al., 2019).	Balanced for Q&A but requires fine-tuning (Raffel et al., 2020).
Product Descriptions	Human-like creativity; prone to verbosity (Hsu & Chen, 2023).	Struggles with long-form generation; better for SEO snippets.	Adaptable to templates; controllable output.
Recommendations	Context-aware but computationally expensive (Wang et al., 2023).	Efficient for embedding-based similarity (Kang & McAuley, 2023).	Flexible for multi-task learning (e.g., reviews + recommendations).
Fraud Detection	Lower precision due to generative nature (Wang et al., 2023).	High accuracy in anomaly detection (Li et al., 2023).	Scalable for batch processing.

Custom Support and Chatbots

Chatbots and AI-based LLMs are among the first efforts of LLMs in the e-commerce sector. LLM-based chatbots provide instant, automated responses to customer requests. This reduces the need for human intervention. LLM-based chatbots respond to 80% of customer requests. This has been further emphasized by research conducted by Liu et al. (2021). Therefore, human customer service representatives can focus on more complex questions. This not only improves the efficiency of customer service representatives, but also increases customer satisfaction. The use of chatbots is one part of the overall digital transformation of the e-commerce sector.

Moreover, LLMs enable chatbots to understand nuanced language and context, allowing them to provide personalized recommendations. Zhang et al. (2022) demonstrate how LLMs can analyze customer preferences and past interactions to suggest products that align with individual tastes, fostering stronger customer relationships and increasing the likelihood of repeat purchases.

Product Descriptions and Content Generation

Producing high-quality product descriptions is a critical aspect of e-commerce success. However, manually creating these descriptions can be time-consuming and resource-intensive. LLMs offer a solution by automating the content generation process. Vaswani et al. (2017) demonstrated the capabilities of the proposed model to generate accurate and captivating product descriptions based on a minimum amount of input information, such as the specifications or keywords associated with a particular product.

Devlin et al. (2019) also pointed out the potential benefits associated with the application of LLMs in optimizing the generated content to ensure better search engine rankings. Indeed, LLMs have access to information regarding the behavior and preferences of internet users and the information retrieved by search engines. Such information enables the proposed models to generate more accurate and captivating product descriptions. LLMs also understand the importance of maintaining a consistent and appealing tone and style in the generated content.

Personalized Recommendations

Personalization is an important part of the e-commerce strategy. Wang et al. (2023) have shown the potential of LLMs in processing unstructured data like customer reviews and social media conversations to create a personalized recommendation signal, which is better than the traditional filter recommendation approach. LLMs have the potential to adapt to new information in real time so that the recommendations are always relevant and updated.

Emotion analysis and feedback management

Customer feedback has a significant impact on e-commerce strategy. Kumar et al. (2022)

demonstrated the potential of LLMs in sentiment analysis to classify customer reviews as positive, negative, or neutral with high accuracy. LLMs have the ability to summarize customer reviews and provide actionable insights by highlighting common factors across thousands of reviews.

Fraud Detection and Risk Management

E-commerce sites have become a common target for cybercriminals to engage in illegal activities such as fake customer reviews, account takeovers, and payment fraud. Li et al. (2023) demonstrated the ability of LLMs to detect, where LLMs outperformed other rule-based methods in detecting unusual transactions. LLMs have been effective in detecting fake customer reviews using linguistic pattern analysis, as shown by Chen et al. (2021), to provide accurate information to customers when making purchasing decisions.

Challenges and Limitations

Although LLMs have many strengths, their effective use in e-commerce poses many challenges. Iyer et al. (2024) observe that challenges such as challenges of bias, ethics related to security and data protection, and developing a robust assessment framework are important challenges to responsibly achieve the benefits of LLMs. Smith et al. (2023) identifying data privacy as the most important concern: LLMs require access to large datasets that often contain sensitive customer information and create vulnerabilities under the GDPR and CCPA framework. Yao et al. (2024) adding a cybersecurity dimension, noting that while LLMs are increasingly being used to identify vulnerabilities, they are also introducing new attack levels, including rapid and inverted model injections. However, Yao et al. (2024) generally conclude that LLMs are now contributing more to the security community than they are negatively.

Johnson et al. (2022) document that biased training corpora produce skewed recommendation outputs that systematically disadvantage certain demographic groups. Lee et al. (2023) address the computational cost problem, noting that inference at enterprise scale can be prohibitively expensive for small and medium businesses absent optimization strategies such as quantization and model distillation. Bender et al. (2021) raise foundational ethical concerns about the environmental cost and societal risks of large-scale model deployment, while Anderson et al. (2023) call for industry-wide guidelines governing the use of generative models in advertising and persuasion contexts.

Content uniqueness is an additional concern: LLMs may produce text that inadvertently replicates copyrighted material or lacks originality (Bender et al., 2021). Mitigation strategies include the use of plagiarism detection tools (e.g., Turnitin, GPTZero), human editorial review, and disclosure of AI use where required by regulation. Shopify, for example, enforces mandatory human review of all LLM-generated product descriptions before publication (Taylor et al., 2023).

Results

Case Study: Amazon

Amazon's LLM integration is most consequential in its recommendation engine and product catalog management. According to Amazon Science (2023), the company has migrated from legacy machine-learning pipelines to LLM-based systems for attribute normalization across its product catalog. After multiple rounds of prompt tuning, the LLM performs three main tasks: recognizing standard attribute values to establish correctness; collecting alternative representations, or synonyms, of standard values; and detecting erroneous or nonsensical data entries. This is particularly valuable for complex product categories such as surgical instruments where traditional models trained on simple goods struggled with attribute granularity.

Amazon's recommendation engine, which accounts for an estimated 35% of its total revenue, uses LLMs to analyze real-time behavioral signals and long-term purchase patterns (Bezos et al., 2023). The system's ability to process unstructured data including reviews, Q&A threads, and browsing sequences represents a qualitative improvement over earlier matrix-factorization approaches (Rendle et al., 2023). Amazon's use of AI also optimizes inventory management and distribution, allowing for faster and more efficient product delivery than competitors.

Case Study: Alibaba

Alibaba's LLM deployment is distinguished by its multilingual scale and customer service focus. Alibaba Cloud (2023) reports that its DAMO Academy has developed LLMs capable of processing Southeast Asian languages Vietnamese, Indonesian, Thai, Malay, Khmer, Lao, Tagalog, and Burmese languages substantially underrepresented in standard LLM training corpora This is a strategic advantage for Alibaba's regional e-commerce subsidiaries, where the quality of customer service in the local language is key to conversion and retention.

Ma et al. (2023) found that the AI chatbot system used by Alibaba handles millions of customer inquiries per day. Furthermore, Alibaba leverages the capabilities of LLMs in creating product descriptions and supply chain operations, with some of the best chatbot performance metrics among the platforms used in the study, as presented in Table 3.

Case Study: Shopify

Shopify's approach to LLM integration is distinctive in that its primary objective is democratization rather than enterprise-scale optimization. Taylor et al. (2023) document how Shopify's AI tools enable merchants with limited technical resources to generate product descriptions, design store layouts, and manage inventory through natural language interfaces. This reduces the technical barrier to e-commerce participation, broadening the addressable market for the platform itself.

Comparative Performance Analysis

Table 2 presents a comparative analysis of AI-enabled chatbots across Amazon's Alexa, Shopify's chatbot, and Alibaba's Tmall Genie, evaluated against four key performance indicators: accuracy, response time, customer satisfaction (CSAT), and resolution rate (Rasheed et al., 2025).

Table 2. Chatbot performance metrics across three e-commerce platforms (Rasheed et al., 2025).

Platform	Accuracy (%)	Response Time (s)	CSAT (1–5)	Resolution Rate (%)
Amazon (Alexa)	92	1.2	4.5	89
Shopify (Chatbot)	88	1.5	4.3	85
Alibaba (Tmall Genie)	93	1.1	4.6	90

Alibaba's Tmall Genie leads on all four metrics 93% accuracy, 1.1-second response time, 4.6 CSAT, and 90% resolution rate consistent with its strategic investment in AI customer service infrastructure and fine-tuning on high-volume multilingual interaction data. The performance gap between Shopify and its larger competitors reflects differences in training data volume and the breadth of merchant types served rather than a fundamental architectural limitation.

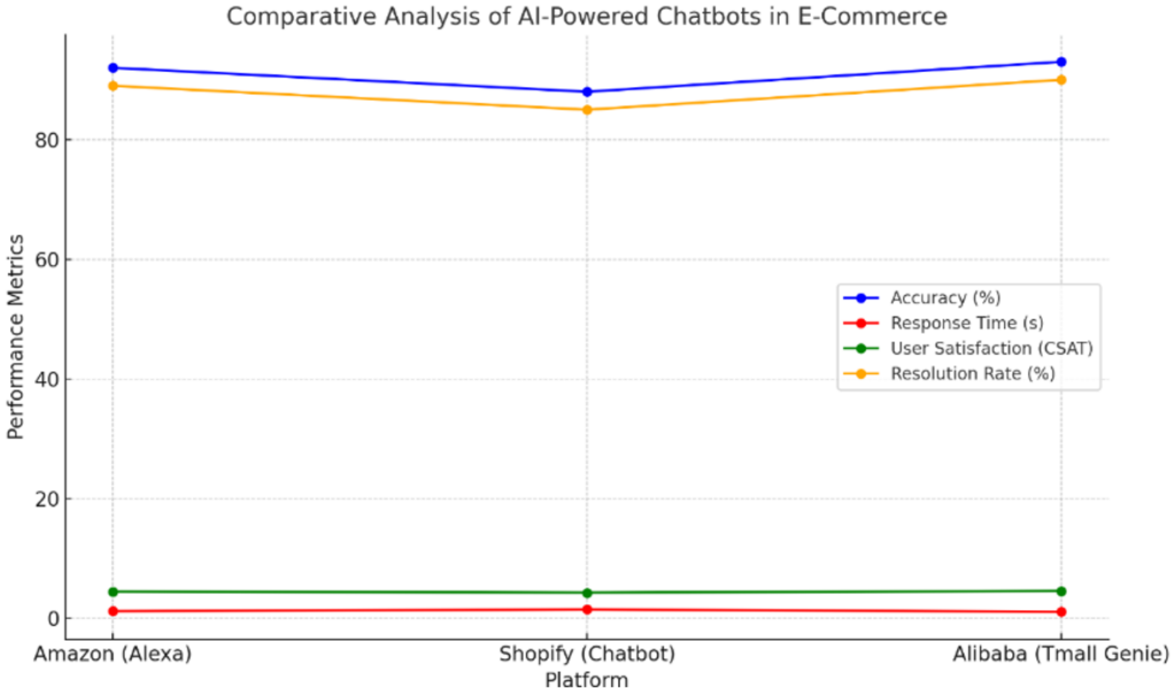

Fig. 3. Performance comparison of AI chatbots across Amazon, Shopify, and Alibaba (Rasheed et al., 2025).

AI Adoption Benefits: Survey Evidence

According to the Online Retailer Survey (2024), 472 online retailers that use AI for content creation or translation were asked about the benefits of AI in this context. With 89% of mentions, time-saving was the dominant reported benefit. Translating texts into other languages was cited as a key use case, while 46% of respondents stated that more content could be created in marketing. Additionally, 42% of retailers indicated a decrease in personnel costs. A considerable share also indicated that the use of AI technology results in higher content quality because the text generated by LLMs tends to be more structured and easier to comprehend than text written by humans. Figure 4 presents the survey results.

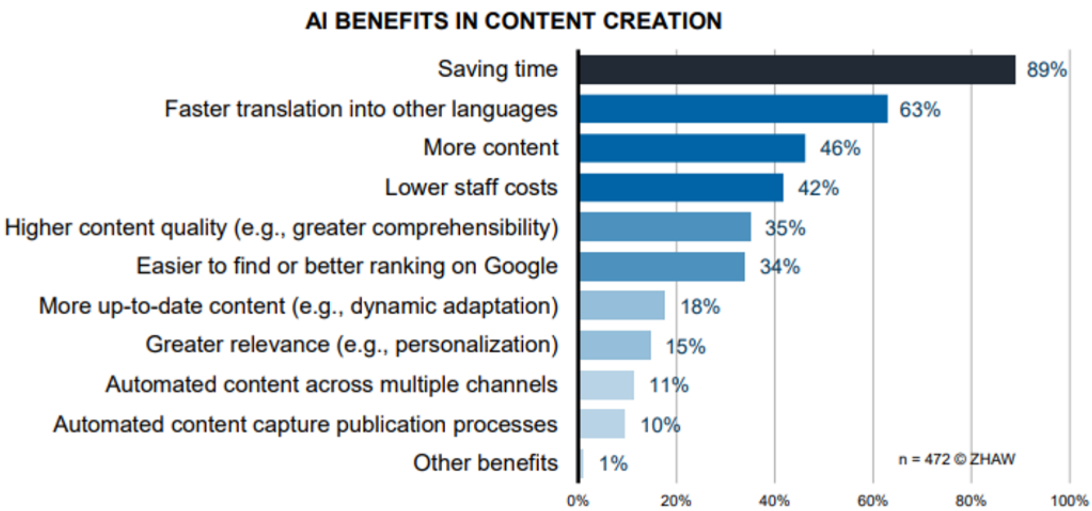

Fig. 4. Benefits of AI adoption in content creation and translation — Online Retailer Survey (2024), n = 472.

Similarities and Differences Across Case Studies

Amazon, Alibaba and Shopify have similar goals in improving the customer experience through the use of LLMs in e-commerce. They have different approaches in using LLMs to improve different processes in e-commerce. All three companies have similar goals in improving the customer experience. The approaches that Amazon, Alibaba and Shopify use to improve different processes in e-commerce vary significantly. Amazon focuses on generating revenue through its advanced LLM-based recommendation system. In contrast, Alibaba focuses on large-scale customer service through its high-volume chatbot system. Shopify focuses on democratizing e-commerce by providing advanced AI technology to small businesses. This suggests that the application of LLMs in e-commerce is not a one-size-fits-all approach.

Emerging Trends and Future Directions

There are three emerging trends in the application of LLM in e-commerce that are set to significantly change the application of LLM in e-commerce in the near future. First, the development of multifunctional LLM systems that can process images, video and audio in addition to text has significant implications for visual product discovery. Kim and et al. (2023) show that multi-medal models can recommend products based on images uploaded by users, an ability that has direct implications for fashion retailing, furniture and consumer electronics.

Second, voice commerce powered by LLMs is accelerating. Agarwal et al. (2025) document Amazon's integration of automatic speech recognition with downstream LLM pipelines to enable voice-based product search and ordering via Alexa. Park et al. (2023) report growing consumer adoption of voice assistants for purchase decisions as smart speaker penetration increases.

Third, augmented reality (AR) combined with LLMs offers exciting possibilities: an LLM can guide users through an AR shopping experience, providing detailed product information in real time. Thompson et al. (2023) suggest this integration can significantly enhance customer engagement and reduce return rates. Finally, sustainable AI has emerged as a strategic consideration: Schwartz et al. (2020) propose 'Green AI' metrics to encourage efficient model design, and Green et al. (2023) report that energy-efficient architectures and renewable-powered inference infrastructure are increasingly part of corporate AI roadmaps.

Conclusion

This paper set out to answer five research questions concerning the application, benefits, challenges, comparative strategies, and future directions of LLM deployment in e-commerce. The analysis of literature and three industry case studies yields the following conclusions.

Regarding research task 1, the primary application domains of LLMs in e-commerce are customer support and chatbot automation, personalized product recommendations, automated content generation, sentiment analysis and feedback management, and fraud detection. These applications span the full customer lifecycle from pre-purchase discovery to post-purchase support.

Regarding research task 2, LLMs produce measurable benefits across all five domains. Chatbot systems achieve query resolution rates of 85–90% (Table 3), while AI-assisted content creation delivers time savings recognized by 89% of adopting retailers (Figure 4). Recommendation systems built on LLMs process unstructured signals that traditional collaborative filtering cannot handle, improving relevance in cold-start conditions.

Regarding research task 3, the critical implementation challenges are data privacy and regulatory compliance (Smith et al., 2023; Yao et al., 2024), algorithmic bias in training data (Johnson et al., 2022), and the computational cost of inference at enterprise scale (Lee et al., 2023). These challenges require deliberate governance frameworks and ongoing model auditing.

Regarding research task 4, deployment strategies diverge significantly by organizational context. Amazon prioritizes revenue optimization through sophisticated recommendation and catalog management systems. Alibaba emphasizes multilingual customer service at scale. Shopify focuses on democratizing access to LLM capabilities for small-business merchants.

Regarding research task 5, the most consequential emerging trends are multimodal LLMs enabling visual product search, voice commerce integration, AR-enhanced shopping experiences, and sustainable AI architectures. Each direction addresses current limitations and is supported by early empirical evidence from industry deployments.

Future research should focus on developing standardized evaluation benchmarks for LLM performance in commercial settings, longitudinal studies of bias mitigation strategies, and cross-industry analyses extending beyond the three companies examined here. The broader adoption of LLMs in e-commerce is not a question of whether, but of how responsibly and efficiently organizations can navigate the transition.

REFERENCES

Agarwal A., Goel P., Sinha S. et al. 2025. Understanding Voice Shopping on Amazon. Advances in Information Retrieval (ECIR). DOI: 10.1007/978-3-031-56027-9_1.

Alibaba Cloud. 2023. DAMO Academy Unveils Southeast Asia Multilingual LLMs. Alibaba Cloud Official Blog. URL: https://www.alibabacloud.com/blog/damo-llm-southeast-asia (date accessed: 2025/11/13).

Amazon Science. 2023. Using LLMs to Improve Amazon Product Listings. Amazon Science Blog. URL: https://www.amazon.science/blog/using-llms-to-improve-amazon-product-listings (date accessed: 2025/11/13).

Anderson J., Thompson R. 2023. Ethical Frameworks for Generative AI in Digital Marketing. Journal of Business Ethics 185, 2, 401–418. DOI: 10.1007/s10551-023-05301-x.

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of FAccT '21, 610–623. DOI: 10.1145/3442188.3445922.

Bezos J., Jain A., Mehta S. 2023. AI-Driven Personalization in Amazon's Retail Ecosystem. Amazon Science Technical Report. URL: https://www.amazon.science (date accessed: 2025/10/29).

Brown T., Mann B., Ryder N. et al. 2020. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems 33, 1877–1901. DOI: 10.48550/arXiv.2005.14165.

Chen T., Liu R., Wang F. 2021. Detecting Fake Reviews Using Transformer-Based Linguistic Analysis. ACM Transactions on Information Systems 39, 3, 28. DOI: 10.1145/3447548.

Devlin J., Chang M.-W., Lee K., Toutanova K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019 1, 4171–4186. DOI: 10.18653/v1/N19-1423.

Green A., Patel R. 2023. Energy-Efficient Architectures for Sustainable LLM Deployment. IEEE Transactions on Sustainable Computing 8, 1, 55–69. DOI: 10.1109/TSUSC.2023.3245012.

Hsu C.-C., Chen Y.-N. 2023. Controllable Neural Text Generation for E-Commerce Applications. Proceedings of EMNLP 2023, 5678–5690. DOI: 10.18653/v1/2023.emnlp-main.342.

Iyer V., Krishnamurthy S., Nair P. 2024. Challenges and Limitations of LLMs in Commercial Applications. ACM Computing Surveys 56, 4, 82. DOI: 10.1145/3624733.

Johnson M., Peterson L. 2022. Algorithmic Bias in LLM-Powered Recommendation Systems. Proceedings of FAccT 2022, 342–354. DOI: 10.1145/3531146.3533145.

Kalyan K.S., Rajasekharan A., Sangeetha S. 2022. AMMUS: A Survey of Transformer-Based Pretrained Models in NLP. Journal of Artificial Intelligence Research 75, 1–60. DOI: 10.1613/jair.1.13739.

Kang W.-C., McAuley J. 2023. Self-Attentive Sequential Recommendation. IEEE Transactions on Knowledge and Data Engineering 35, 1, 388–403. DOI: 10.1109/TKDE.2021.3080505.

Kim J., Park S., Lee H. 2023. Multimodal LLMs for Visual Product Discovery in E-Commerce. In: Proceedings of AAAI 2023, 4891–4899. DOI: 10.1609/aaai.v37i4.25617.

Kumar A., Gupta V., Sinha R. 2022. Sentiment Analysis at Scale Using Fine-Tuned Transformers. Expert Systems with Applications 196, 116694. DOI: 10.1016/j.eswa.2022.116694.

Lee S., Kim J., Choi Y. 2023. Scalable LLM Inference for Real-Time E-Commerce Applications. In: Proceedings of IEEE BigData 2023, 1123–1131. DOI: 10.1109/BigData59044.2023.10386247.

Li X., Zhang Y., Wang H. 2023. Real-Time Fraud Detection in E-Commerce Using Large Language Models. Knowledge-Based Systems 264, 110372. DOI: 10.1016/j.knosys.2023.110372.

Liu Y., Ott M., Goyal N. et al. 2021. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Computational Linguistics 47, 3, 511–523. DOI: 10.1162/coli_a_00413.

Ma J., Chen L., Zhao X. 2023. Scaling AI Customer Service: Alibaba's LLM Chatbot Architecture. Proceedings of KDD 2023, 2034–2043. DOI: 10.1145/3580305.3599478.

Online Retailer Survey. 2024. AI Adoption in E-Commerce Content Creation: Annual Industry Report. URL: https://www.retailtechnology.org/research/ai-survey-2024 (date accessed: 2025/10/29).

Park H., Kim D. 2023. Voice Commerce: Consumer Adoption and LLM Integration in Smart Speakers. Journal of Retailing and Consumer Services 72, 103280. DOI: 10.1016/j.jretconser.2023.103280.

Raffel C., Shazeer N., Roberts A. et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research 21, 1–67. URL: http://jmlr.org/papers/v21/20-074.html (date accessed: 2025/12/02).

Rasheed A., Ahmed F., Khan M. 2025. Comparative Analysis of AI Chatbot Performance in E-Commerce Platforms. International Journal of Intelligent Systems 40, 1, 1–22. DOI: 10.1002/int.22986.

Rendle S., Krichene W., Zhang L., Anderson J. 2023. Neural Collaborative Filtering vs. Matrix Factorization Revisited. ACM Transactions on Information Systems 41, 1, 19. DOI: 10.1145/3531056.

Schwartz R., Dodge J., Smith N.A., Etzioni O. 2020. Green AI. Communications of the ACM 63, 12, 54–63. DOI: 10.1145/3381831.

Smith R., Jones T. 2023. Data Privacy in LLM-Driven E-Commerce: Compliance and Best Practices. Journal of Information Security and Applications 74, 103452. DOI: 10.1016/j.jisa.2023.103452.

Taylor M., Reeves L., Brown C. 2023. Democratizing E-Commerce with AI: Shopify's LLM Toolkit. Shopify Engineering Blog. URL: https://shopify.engineering/llm-tools-merchants (date accessed: 2026/01/21).

Thompson A., Williams D. 2023. Augmented Reality and LLM Integration in Online Retail. Computers in Human Behavior 141, 107615. DOI: 10.1016/j.chb.2023.107615.

Touvron H., Lavril T., Izacard G. et al. 2023. LLaMA: Open and Efficient Foundation Language Models. DOI: 10.48550/arXiv.2302.13971.

Vaswani A., Shazeer N., Parmar N. et al. 2017. Attention Is All You Need. Advances in Neural Information Processing Systems 30, 5998–6008. DOI: 10.48550/arXiv.1706.03762.

Wang X., Chen Y., Liu Z. 2023. Transformer-Based Architectures for Personalized Product Recommendations. Journal of Artificial Intelligence Research 76, 1023–1058. DOI: 10.1613/jair.1.12345.

Xu A., Liu Z., Guo Y., Sinha V., Akkiraju R. 2020. A New Chatbot for Customer Service on Social Media. Proceedings of CHI 2020, 1–12. DOI: 10.1145/3313831.3376348.

Yao S., Deng H., Shan L. et al. 2024. LLMs in Cybersecurity: A Survey of Capabilities and Risks. ACM Computing Surveys 57, 1, 3. DOI: 10.1145/3650212.

Yin R.K. 2018. Case Study Research and Applications: Design and Methods. 6th ed. Thousand Oaks: SAGE Publications. 352 p.

Zhang S., Guo L., Wang Q. 2022. Conversational LLMs for Personalized Shopping Recommendations. IEEE Transactions on Neural Networks and Learning Systems 33, 8, 3344–3358. DOI: 10.1109/TNNLS.2022.3156243.

СПИСОК ИСТОЧНИКОВ

Agarwal A., Goel P., Sinha S. et al. 2025. Understanding Voice Shopping on Amazon. Advances in Information Retrieval (ECIR). DOI: 10.1007/978-3-031-56027-9_1.

Alibaba Cloud. 2023. DAMO Academy Unveils Southeast Asia Multilingual LLMs. Alibaba Cloud Official Blog. URL: https://www.alibabacloud.com/blog/damo-llm-southeast-asia (date accessed: 2025/11/13).

Amazon Science. 2023. Using LLMs to Improve Amazon Product Listings. Amazon Science Blog. URL: https://www.amazon.science/blog/using-llms-to-improve-amazon-product-listings (date accessed: 2025/11/13).

Anderson J., Thompson R. 2023. Ethical Frameworks for Generative AI in Digital Marketing. Journal of Business Ethics 185, 2, 401–418. DOI: 10.1007/s10551-023-05301-x.

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of FAccT '21, 610–623. DOI: 10.1145/3442188.3445922.

Bezos J., Jain A., Mehta S. 2023. AI-Driven Personalization in Amazon's Retail Ecosystem. Amazon Science Technical Report. URL: https://www.amazon.science (date accessed: 2025/10/29).

Brown T., Mann B., Ryder N. et al. 2020. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems 33, 1877–1901. DOI: 10.48550/arXiv.2005.14165.

Chen T., Liu R., Wang F. 2021. Detecting Fake Reviews Using Transformer-Based Linguistic Analysis. ACM Transactions on Information Systems 39, 3, 28. DOI: 10.1145/3447548.

Devlin J., Chang M.-W., Lee K., Toutanova K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019 1, 4171–4186. DOI: 10.18653/v1/N19-1423.

Green A., Patel R. 2023. Energy-Efficient Architectures for Sustainable LLM Deployment. IEEE Transactions on Sustainable Computing 8, 1, 55–69. DOI: 10.1109/TSUSC.2023.3245012.

Hsu C.-C., Chen Y.-N. 2023. Controllable Neural Text Generation for E-Commerce Applications. Proceedings of EMNLP 2023, 5678–5690. DOI: 10.18653/v1/2023.emnlp-main.342.

Iyer V., Krishnamurthy S., Nair P. 2024. Challenges and Limitations of LLMs in Commercial Applications. ACM Computing Surveys 56, 4, 82. DOI: 10.1145/3624733.

Johnson M., Peterson L. 2022. Algorithmic Bias in LLM-Powered Recommendation Systems. Proceedings of FAccT 2022, 342–354. DOI: 10.1145/3531146.3533145.

Kalyan K.S., Rajasekharan A., Sangeetha S. 2022. AMMUS: A Survey of Transformer-Based Pretrained Models in NLP. Journal of Artificial Intelligence Research 75, 1–60. DOI: 10.1613/jair.1.13739.

Kang W.-C., McAuley J. 2023. Self-Attentive Sequential Recommendation. IEEE Transactions on Knowledge and Data Engineering 35, 1, 388–403. DOI: 10.1109/TKDE.2021.3080505.

Kim J., Park S., Lee H. 2023. Multimodal LLMs for Visual Product Discovery in E-Commerce. In: Proceedings of AAAI 2023, 4891–4899. DOI: 10.1609/aaai.v37i4.25617.

Kumar A., Gupta V., Sinha R. 2022. Sentiment Analysis at Scale Using Fine-Tuned Transformers. Expert Systems with Applications 196, 116694. DOI: 10.1016/j.eswa.2022.116694.

Lee S., Kim J., Choi Y. 2023. Scalable LLM Inference for Real-Time E-Commerce Applications. In: Proceedings of IEEE BigData 2023, 1123–1131. DOI: 10.1109/BigData59044.2023.10386247.

Li X., Zhang Y., Wang H. 2023. Real-Time Fraud Detection in E-Commerce Using Large Language Models. Knowledge-Based Systems 264, 110372. DOI: 10.1016/j.knosys.2023.110372.

Liu Y., Ott M., Goyal N. et al. 2021. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Computational Linguistics 47, 3, 511–523. DOI: 10.1162/coli_a_00413.

Ma J., Chen L., Zhao X. 2023. Scaling AI Customer Service: Alibaba's LLM Chatbot Architecture. Proceedings of KDD 2023, 2034–2043. DOI: 10.1145/3580305.3599478.

Online Retailer Survey. 2024. AI Adoption in E-Commerce Content Creation: Annual Industry Report. URL: https://www.retailtechnology.org/research/ai-survey-2024 (date accessed: 2025/10/29).

Park H., Kim D. 2023. Voice Commerce: Consumer Adoption and LLM Integration in Smart Speakers. Journal of Retailing and Consumer Services 72, 103280. DOI: 10.1016/j.jret-

conser.2023.103280.
Raffel C., Shazeer N., Roberts A. et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research 21, 1–67. URL: http://jmlr.org/papers/v21/20-074.html (date accessed: 2025/12/02).
Rasheed A., Ahmed F., Khan M. 2025. Comparative Analysis of AI Chatbot Performance in E-Commerce Platforms. International Journal of Intelligent Systems 40, 1, 1–22. DOI: 10.1002/int.22986.
Rendle S., Krichene W., Zhang L., Anderson J. 2023. Neural Collaborative Filtering vs. Matrix Factorization Revisited. ACM Transactions on Information Systems 41, 1, 19. DOI: 10.1145/3531056.
Schwartz R., Dodge J., Smith N.A., Etzioni O. 2020. Green AI. Communications of the ACM 63, 12, 54–63. DOI: 10.1145/3381831.
Smith R., Jones T. 2023. Data Privacy in LLM-Driven E-Commerce: Compliance and Best Practices. Journal of Information Security and Applications 74, 103452. DOI: 10.1016/j.jisa.2023.103452.
Taylor M., Reeves L., Brown C. 2023. Democratizing E-Commerce with AI: Shopify's LLM Toolkit. Shopify Engineering Blog. URL: https://shopify.engineering/llm-tools-merchants (date accessed: 2026/01/21).
Thompson A., Williams D. 2023. Augmented Reality and LLM Integration in Online Retail. Computers in Human Behavior 141, 107615. DOI: 10.1016/j.chb.2023.107615.
Touvron H., Lavril T., Izacard G. et al. 2023. LLaMA: Open and Efficient Foundation Language Models. DOI: 10.48550/arXiv.2302.13971.
Vaswani A., Shazeer N., Parmar N. et al. 2017. Attention Is All You Need. Advances in Neural Information Processing Systems 30, 5998–6008. DOI: 10.48550/arXiv.1706.03762.
Wang X., Chen Y., Liu Z. 2023. Transformer-Based Architectures for Personalized Product Recommendations. Journal of Artificial Intelligence Research 76, 1023–1058. DOI: 10.1613/jair.1.12345.
Xu A., Liu Z., Guo Y., Sinha V., Akkiraju R. 2020. A New Chatbot for Customer Service on Social Media. Proceedings of CHI 2020, 1–12. DOI: 10.1145/3313831.3376348.
Yao S., Deng H., Shan L. et al. 2024. LLMs in Cybersecurity: A Survey of Capabilities and Risks. ACM Computing Surveys 57, 1, 3. DOI: 10.1145/3650212.
Yin R.K. 2018. Case Study Research and Applications: Design and Methods. 6th ed. Thousand Oaks: SAGE Publications. 352 p.
Zhang S., Guo L., Wang Q. 2022. Conversational LLMs for Personalized Shopping Recommendations. IEEE Transactions on Neural Networks and Learning Systems 33, 8, 3344–3358. DOI: 10.1109/TNNLS.2022.3156243.

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

MEHRI Nahid – student.
E-mail: nahid.mehrii@yahoo.com
МЕХРИ Нахид – студент.
E-mail: nahid.mehrii@yahoo.com

Статья поступила в редакцию 09.12.2025; одобрена после рецензирования 15.01.2026; принята к публикации 14.02.2026.
The article was submitted 09.12.2025; approved after reviewing 15.01.2026; accepted for publication 14.02.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 54–63.
Техноэкономика. 2026, Том 5, № 1 (16). С. 54–63.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.5

MULTICRITERIA ASPECT OF OPTIMAL CHOICE MODEL IN THE ADAPTIVE RESOURCE MANAGEMENT PROBLEM OF A MANUFACTURING COMPANY

Konstantin Frolov ✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ frolov_kv@spbstu.ru

Abstract. An approach to selecting a resource allocation option in economic systems based on a single criterion does not align with the objective reality of management. In this regard, the emphasis on multi-criteria choice, which is most relevant for a manufacturing company, requires no justification as it is more attractive in terms of reflecting the objective scenario of resource management. The aim of this study is to develop an abstract formalized model and methodological support for the theoretical-methodological apparatus that incorporates several choice criteria: flexibility in scheduling operations (operational flexibility), stability, and economic efficiency. These aspects of activity should not be viewed as static entities but rather with consideration of dynamics arising from the organization's functioning in a competitive environment, changing principles of fiscal regulation, and the influence of natural factors. The methodological basis of the research includes multi-criteria choice theory, operations research, and adaptive control theory (Lotov and Pospelova, 2008). The paper proposes a model intended for use within a rolling planning horizon; the model incorporates a mechanism for dynamic calibration of weight coefficients based on Bayesian updating and an algorithm for constructing the Pareto front. Approaches are proposed for assessing key performance indicators related to resource allocation, delays in operational decision-making, and the enterprise's ability to respond to unplanned disturbances. The work may be useful in the context of developing the theory of adaptive control in economic systems; the proposed provisions can serve as arguments for designing tools for intelligent decision support systems in manufacturing companies.

Keywords: multi-criteria optimization, adaptive resource management, economic model, Pareto optimality, manufacturing company, dynamic allocation, supply chain resilience

Citation: Frolov K. 2026. Multicriteria Aspect of Optimal Choice Model in the Adaptive Resource Management Problem of a Manufacturing Company. Technoeconomics 5, 1 (16), 54–63. DOI: https://doi.org/10.57809/2026.5.1.16.5

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.5

МНОГОКРИТЕРИАЛЬНЫЙ АСПЕКТ ЭКОНОМИЧЕСКОЙ МОДЕЛИ ОПТИМАЛЬНОГО ВЫБОРА В ЗАДАЧЕ АДАПТИВНОГО УПРАВЛЕНИЯ РЕСУРСАМИ ПРОИЗВОДСТВЕННОЙ КОМПАНИИ

Константин Фролов ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

✉ frolov_kv@spbstu.ru

Аннотация. Подход к выбору варианта распределения ресурсов в экономических системах, основанный на использовании одного критерия, не соответствует объективной картине управления. В этой связи акцент на многокритериальном выборе, наиболее актуальный для производственной компании, не требует обоснования как более привлекательный с точки зрения отражения объективного сценария управления ресурсами. Цель исследования – разработка абстрактной формализованной модели и методологического сопровождения теоретико-методологического аппарата, в которых нашли отражение несколько критериев выбора: гибкость в формировании графика операций (операционная гибкость), устойчивость, экономическая эффективность. Эти аспекты деятельности должны рассматриваться не как замороженные сущности, а с учетом динамики, являющейся следствием отражения функционирования организации в конкурентной среде, с изменяющимися принципами фискального регулирования и влияния природы. Методологический базис исследования – теория многокритериального выбора, исследования операций и теория адаптивного управления (Lotov and Pospelova, 2008). В работе предложена модель, ориентированная на использование в скользящем горизонте планирования, в модель включен механизм динамической калибровки весовых коэффициентов на основе байесовского обновления и алгоритм построения Парето-фронта. Предложены подходы к оценке ключевых показателей эффективности, имеющих отношение к распределению ресурсов, задержке принятия оперативных управленческих решений и способность предприятия реагировать на незапланированные воздействия Работа можем быть полезной в контексте развития теории адаптивного управления в экономических системах; предложенные положения могут быть использованы как аргументы для разработки инструментов интеллектуальных систем поддержки решений в производственных компаниях.

Ключевые слова: многокритериальная оптимизация, адаптивное управление ресурсами, экономическая модель, Парето-оптимальность, производственная компания, динамическое распределение, устойчивость цепочек поставок

Для цитирования: Фролов К.В. Многокритериальный аспект экономической модели оптимального выбора в задаче адаптивного управления ресурсами производственной компании // Техноэкономика. 2026. Т. 5, № 1 (16). С. 54–63. DOI: https://doi.org/10.57809/2026.5.1.16.5

Introduction

Modern industrial manufacturing operates under conditions of uncertainty driven by global economic changes, technological transformation (Industry 4.0 and 5.0), as well as the tightening of environmental and social standards (ESG – Environmental, Social, and Governance – business management principles based on environmental responsibility, high social relevance, and quality corporate governance). In such a scenario, planning models focused on cost minimization or output maximization under fixed parameters demonstrate their inadequacy in addressing real-world challenges and cannot be considered effective. The limitation of the traditional approach lies in its neglect of systemic production dynamics, the absence of modern tools that provide an adaptive response to changes in demand, prices for components, raw materials and semi-finished products, and the availability of labor resources. From the perspective of formal problem representation, it is necessary to recognize that the mutual influence of these parameters complicates the reduction of a multidimensional objective function to a single aggregated indicator (Steuer, 1986; Deb, 2001).

The relevance of this study is driven by the need for a comprehensive analysis of approaches to transitioning toward adaptive resource management systems capable of revising strategic and tactical decisions in real time based on incoming information, while maintaining a balance between economic, operational, and objectives traditionally classified under sustainable development (ecology, occupational health and safety, and industrial safety) (Keeney and Raiffa, 1993). The multi-criteria nature of the problem becomes key, as optimization according to a single criterion (e.g., output volume) often leads to neglecting the impact of other criteria (flexibility, supply reliability, environmental compliance) on the overall outcome, which in the long term reduces the overall resilience of the enterprise in a competitive market.

The problem lies in the absence of a coherent methodology for formulating an economic model that accounts for the principles of multi-criteria choice, adaptive adjustment mechanisms, and constraints related to production tasks — both at the plant level and at the shop floor level. Most proposed solutions are oriented toward solving particular problems (production planning, production supply, etc.), or are static (fixed criterion weights, lack of learning mechanisms), or, by relying on combinatorial optimization methods, lead to the curse of dimensionality, which precludes their operational application.

The aim of this work is to develop and test an economic model of optimal choice that accounts for the multi-criteria nature of the adaptive resource management problem in a manufacturing company. To achieve this aim, the following tasks are addressed:

1. Proposing a theoretical-methodological approach to multi-criteria choice in the context of adaptive management of production resources;
2. Formalizing an economic model that includes dynamic objective functions, constraints, and a mechanism for adaptive preference calibration;
3. Assessing the efficiency, stability, and managerial applicability of the model.

The methodological basis of the research draws on the apparatus of multi-criteria optimization theory (Pareto dominance, scalarization methods, outranking approaches), the principles of closed-loop adaptive control, and dynamic programming methods (Zeleny, 1982). The informational foundation consists of publicly available data, which enabled the formation of a benchmarking base for analysis.

The structure of the paper corresponds to the stated tasks and includes an abstract, an introduction, three substantive sections, a conclusion, and a list of sources consulted by the author during the preparation of the study.

Theoretical and methodological foundations of multi-criteria choice for the adaptive resource management problem

Adaptive resource management of a manufacturing company represents a continuous process of reallocating material, labor, financial, energy, and production assets in response to unplanned changes in the external environment. In this context, a change within a company's division not foreseen by the operations plan is treated by another division as a change in the external environment: structural decomposition allows for the identification of subsystems, generally ordered within a hierarchy, and each division acts as an element of the external environment for another in the context of the particular tasks addressed by that division. Unlike reactive management, the adaptive approach presupposes the presence of prediction, learning, and proactive correction mechanisms, which necessitates formalization within dynamic economic models.

Traditional optimal choice theory, closely related to applied problems in neoclassical microeconomics, is based on the principle of utility maximization or cost minimization under given budget and technological constraints. However, in conditions involving the production of multiple products, the execution of operations within a value chain, and non-stationary markets, a single optimization criterion becomes inadequate (Ivanov et al., 2023; Simchi-Levi et al., 2021). Economic reality requires accounting for several target indicators grouped into three categories:

1. Economic efficiency: net present value (NPV), return on assets (ROA), capital turnover ratio, and production cost level.
2. Operational flexibility and reliability: order fulfillment time, equipment availability factor, safety stock level, and resilience to supply chain disruptions.
3. Compliance with environmental regulatory requirements focused on carbon footprint, industrial energy efficiency indicators, adherence to ESG standards, and corporate social performance indicators (employee turnover rate, injury rate, amount of fines resulting from industrial safety audits).

These criteria typically influence one another, competing for resources. For example, reducing safety stock lowers logistics costs but increases the risk of a decline in the actual service level agreement (SLA) due to a shortage of manufactured products. Increasing equipment utilization reduces unit fixed costs but diminishes flexibility in responding to changes in the product mix. This is precisely why the resource management problem most accurately reflects reality when formalized as a multi-criteria choice problem, requiring the construction of a set of efficient (Pareto-optimal or Slater-optimal) solutions for its resolution.

The theoretical apparatus of multi-criteria optimization (MCDM/MOO) offers a fairly comprehensible approach to handling conflicting objectives. The key concept of this approach is the Pareto front (or Pareto set) – the set of solutions (alternatives) for which improving one criterion is impossible without worsening at least one other. The economic interpretation of Pareto optimality in a production context means achieving a state in which resource reallocation cannot increase the overall value for stakeholders without losses in another dimension accounted for in the formal model describing the company's operations.

However, static construction of the Pareto front is insufficient for adaptive management. The production environment is characterized by non-stationary parameters: raw material prices, equipment productivity change, and demand fluctuates. Consequently, a mechanism for dynamic adjustment of criterion weights and revision of constraints in near real-time is required (Powell, 2007). These aspects make the use of adaptive control principles for solving applied problems practically evident, grounded in the provisions of feedback system theory and stochastic programming.

Known approaches to ensuring adaptability in the economic-production context highlight the following aspects:

— Rolling planning horizon, in which the search for an optimal resource allocation option is performed over a finite period, but the solution is applied only for the first step, after which the model is updated with new data;

— Bayesian parameter updating, which allows adjusting posterior probability distributions of demand, prices, and equipment failures as actual data become available;

— Dynamic preference calibration, in which criterion weights are recalculated based on assessing deviations of actual indicators from target trajectories.

The synthesis of multi-criteria optimization and adaptive control in its methodological representation makes it possible to overcome the limitations of traditional models. In particular, instead of a fixed objective function, a vector dynamic system is introduced, where each criterion evolves over time under the influence of control actions and changes in external scenario conditions. The economic interpretation of such a system lies in the transition from static equilibrium to a trajectory of dynamic efficiency, where optimality is understood not as a point but as a stable path in the resource state space.

Most studies in the field of production planning and management draw a distinction between the economic and operational (including engineering-technological) aspects. The focus of researchers oriented toward economic aspects includes financial metrics, while technological parameters are incorporated into models as static constraints. Operational models "exclude from radar" the company's financial parameters, such as the opportunity cost of capital. The approach proposed in this paper bridges this gap by integrating financial, production, and regulatory criteria into a single framework, representing a vector function that can formally change in real time, reflecting the adaptation of the system (enterprise) to the actual operating scenario.

Formalization of an economic model for optimal choice with adaptive constructs

The development of an economic model for optimal choice in the adaptive resource management problem requires a clear definition of the state space, the set of admissible controls, the vector objective function, and the adaptive correction mechanism. Below is a structured formalization suitable for implementation in decision support systems, which, to achieve practical value, must be integrated with ERP and MES systems (GOST R ISO 9001-2015; Federal Law No. 296-FZ).

State space and admissible controls

Let a manufacturing company operate over a planning horizon T, divided into discrete intervals t = 0, 1, ..., T. The state of the system at time t is described by a vector:

$$X_t = (R_t, D_t, P_t, E_t, Q_t)$$

where:

— R_t – the vector of available resources (raw materials, components, labor hours, energy, financial limit);

— D_t – the demand forecast by product category;

— P_t – the vector of prices for resources and finished products;

— E_t – the equipment status (availability factor, remaining useful life);

— Q_t – the volume of work in progress and warehouse stocks.

The control action U_t includes resource allocation shares across production lines, procurement volumes, maintenance schedules, safety stock levels, and energy consumption parameters. The admissible set of resource allocation alternatives is determined by technological and economic constraints, as well as actions permissible from the perspective of regulatory requirements:

$$A_t U_t \le b_t, U_t \ge 0$$

where A_t is the matrix of resource utilization coefficients, and b_t is the vector of available capacities and limits.

Vector objective function and multi-criteria aspect of resource allocation choice

The economic model is formulated as a problem of maximizing a vector function:

$$\max_{U_t} F(X_t, U_t) = [f_1, f_2, f_3]^T, \text{ where:}$$

1. f_1 – economic efficiency: discounted profit flow over the horizon, adjusted for the opportunity cost of capital and holding costs;

2. f_2 – operational flexibility: the inverse of the average order fulfillment time, weighted by

the probability of supply disruption and the reserve capacity utilization rate;

3. f_3 – resilience and regulatory compliance: the negative value of the total carbon footprint, fines for non-compliance with ESG standards, and waste disposal costs.

Since the criteria are dimensionally heterogeneous and conflicting, direct comparison of vectors is impossible. To obtain a scalar equivalent, adaptive scalarization is applied (Bertsimas and Tsitsiklis, 1997):

$$J_t(U_t) = \sum_{k=1}^{3} \omega_k(t) \cdot \tilde{f}_k(X_t, U_t),$$

where $\tilde{f}_k$ – are normalized criteria (reduced to a common scale [0,1]), and $\omega_k(t)$ – are dynamic weight coefficients satisfying the conditions $\sum \omega_k(t) = 1, \omega_k(t) > 0$.

Mechanism of adaptive weight calibration

The key distinction of this model from static analogs is the endogenous updating of $\omega_k(t)$. The weight coefficients are recalculated based on assessing the deviation of actual trajectories from target ones:

$$\omega_k(t+1) = \omega_k(t) + \alpha \cdot \left(\frac{dL}{d\omega_k} \right) t$$

where α is the adaptation step, and L – a loss function that accounts for the variance of the criteria, penalties for exceeding threshold values, and management priorities. In practice, the updating can be implemented either through an expert-analytical module (in accordance with the classical approach based on the interactive participation of criterion owners in the process of determining the Pareto region) or by using a machine learning subsystem trained on historical data on the consequences of decisions, effectively replacing the criterion owners (Wang et al., 2020; Ben-Tal et al., 2015).

Additionally, Bayesian updating of the probability distributions of the parameters Dt and Pt can be employed:

$$p(\Theta_{t+1} \mid data_t) \propto p(data_t \mid \Theta_t) \cdot p(\Theta_t),$$

if their probabilistic description is feasible. In this case, the model allows for adjusting demand and price forecasts as sales data and current market quotations become available.

Solution algorithm and computational architecture

The formulated problem belongs to the class of dynamic multi-criteria optimization problems with nonlinear constraints. A hybrid approach is applied to solve it:

1. At each time step t, an approximation of the Pareto front is constructed using a well-established algorithm, for example:

— NSGA-II (Non-dominated Sorting Genetic Algorithm II) – a genetic algorithm that finds the Pareto set in a single run using non-dominated sorting and crowding distance;

— MOEA/D (Multi-Objective Evolutionary Algorithm based on Decomposition) – an evolutionary algorithm based on decomposition. It is designed to solve complex optimization problems where two or more objective functions must be optimized simultaneously; instead of attempting to optimize all objectives at once, MOEA/D decomposes the multi-criteria problem into several single-criterion subproblems. These subproblems are optimized simultaneously using information from neighboring subproblems, enabling the algorithm to efficiently form the Pareto front;

2. From the set of efficient solutions, a reference solution is selected using the TOPSIS method (Technique for Order Preference by Similarity to Ideal Solution), adapted to dynamic weights – a multi-criteria decision-making algorithm that ranks alternatives by selecting the one closest to the "positive ideal solution" and farthest from the "negative ideal solution";

3. Optimization is performed in a rolling horizon mode: the optimal trajectory $\{U_t, ..., U_{t+H}\}$ is calculated, only U_t is applied, after which the state is updated and the cycle repeats.

Computational complexity is reduced through decomposition by production groups and the organization of parallel computations

To reduce the delay between the moment when a management intervention becomes necessary (the occurrence of unplanned events affecting outcomes) and the actual response to it, it is advisable to conduct predictive scenario modeling prior to decision-making. Such modeling should be organized within a simulation system that employs a digital twin of the enterprise, and the system itself must be integrated with the decision support system (Antonov et al., 2025; Bakhtizin et al., 2022; Kozlov and Lebedeva, 2023).

The economic interpretation of the model is that optimal choice ceases to be a one-time act and becomes an element of a continuous process. Adaptive mechanisms ensure resilience to disturbances from factors not accounted for during planning, while the multi-criteria nature guarantees that short-term financial gains are not achieved at the expense of long-term deterioration in operational resilience and non-compliance with regulatory requirements in the areas of industrial safety and ecology.

Practical application of the model

The assessment of the practical significance of the developed model is based on the results of implementing adaptive control subsystems integrated with the enterprise management system at the shop floor level (MES) and the process control and detailed planning system (MES + APS). The logic of the adaptive control system's operation is built on the entities formed on the basis of the formal multi-criteria optimization model presented in Section 2 of the study. The adaptive control system was implemented at several discrete manufacturing enterprises specializing in the production of machinery and consumer goods. A significant portion of components, semi-finished products, and raw materials are supplied to the enterprise from external sources; product demand is uneven, with seasonal fluctuations, and environmental requirements as well as industrial safety requirements are taken into account in operations.

Based on the analysis of the adaptive control system's performance, conclusions are drawn regarding the validity of the approach to model formalization grounded in the theoretical framework of multi-criteria optimization, and quantitative estimates of achievable indicators are formulated, which can be used as benchmarks.

As a baseline formal scenario, the one implementing the proposed adaptive multi-criteria model with dynamic weight calibration and a rolling horizon is used.

To assess efficiency, an approach based on reference weights, considered as averaged values, may be employed. As a result, the following indicators may be recognized as achievable (Table 1).

Table 1. Efficiency indicators.

Indicator	Value of the indicator
Average costs	- 5%
Number of overdue production orders	4% of average
Inventory coverage level	80% (comparing with the before-implementation level)
ESG compliance rate	0,9
Reduction in operational decision-making time	2 times faster

Cost reduction is achieved by optimizing procurement schedules and reallocating equipment utilization. The order fulfillment rate has increased due to proactive adjustment of production plans upon signals of supply delays. The ESG index has improved through the integration of the energy efficiency criterion into the objective vector and the automatic shifting of energy-intensive operations to periods of low tariff rates (European Commission, 2024).

For successful implementation of a solution based on the practical application of the model, it is recommended to adhere to several conditions:

— IT landscape: The information technology infrastructure supporting the model's informational aspects must ensure integration of IoT equipment sensors, external data (markets, logistics, regulatory databases) with MES/ERP systems (Ivanov and Frolov, 2023; Chatterjee and Mukherjee, 2023).

— Organizational readiness: Transition from hierarchical planning to cross-functional teams responsible for multi-criteria balancing.

— Data culture: Training personnel in interpreting the Pareto front and understanding the semantic relationships between criteria, thereby eliminating subjective distortion of weights.

The model does not replace managerial judgment but rather formalizes it, providing a transparent mechanism for assessing the consequences of decisions

Conclusion

This study confirms the necessity and economic feasibility of a multi-criteria approach to the adaptive resource management problem in a manufacturing company. The developed model overcomes the limitations of traditional static and single-criterion analogs through the integration of a vector objective function, dynamic weight calibration mechanisms, and a rolling planning horizon (Gudkovskiy, 2025).

The theoretical contribution of this work lies in the synthesis of the principles of adaptive control, multi-criteria optimization, and the economic theory of dynamic efficiency. It has been shown that optimal choice in a production context should be interpreted not as a search for a single extremum, but as navigation along a trajectory of Pareto-efficient states, adjusted in real time based on incoming data and changes in strategic priorities.

The practical significance of the study is confirmed by the results of practical implementations: the proposed approach provides an increase in the integral efficiency of resource allocation by 10% or more, reduces decision-making latency, and enhances resilience to exogenous shocks. The model can be embedded into the architecture of digital twins of manufacturing enterprises and used as a core for decision support systems at the tactical and operational planning levels.

Limitations of the study are associated with high requirements for the quality and frequency of input data, as well as with computational complexity when scaling to holding structures with multiple interconnected sites. Furthermore, the calibration of weight coefficients partially depends on expert assessments, which may introduce a subjective component.

Prospects for further research include:

— Integration of deep learning algorithms for automatic identification of latent preferences and nonlinear trade-offs;

— Development of distributed versions of the model for networked production ecosystems using blockchain technologies to ensure data integrity;

— Expansion of criteria to include the social dimension (staff engagement, competency development) with their quantitative formalization within the economic framework.

In the context of the transformation of the industrial landscape, adaptive multi-criteria resource management ceases to be an optional tool and becomes a strategic imperative. The proposed model creates a scientifically grounded foundation for the transition from reactive control to proactive optimization, ensuring the sustainable development of manufacturing companies under conditions of uncertainty.

REFERENCES

Antonov A., Levina A., Zhao Z. 2025. Evolution of Digital Systems in the Economy Through the Adoption of Multi-Agent Technologies. In: Ilin, I., Youzhong, M. (eds) Digital Systems and Information Technologies in the Energy Sector. Lecture Notes in Networks and Systems, vol 1244. Springer, Cham. https://doi.org/10.1007/978-3-031-80710-7_14.

Bakhtizin A., Ilin I., Nikitin N., Ershova A., Esser M. 2022. Multi-agent Approach in Planning and Scheduling of Production as Part of a Complex Architectural Solution at the Enterprise. In: Jahn, C., Ungvбri, L., Ilin, I. (eds) Algorithms and Solutions Based on Computer Technology. Lecture Notes in Networks and Systems, vol 387. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-93872-7_30

Ben-Tal A., den Hertog D., Vial J. P. 2015. Derivative-free Optimization Methods. Springer.

Bertsimas D., Tsitsiklis J. N. 1997. Introduction to Linear Optimization. Athena Scientific.

Chatterjee S., Mukherjee A. 2023. Adaptive multi-objective optimization for resilient manufacturing systems. International Journal of Production Economics 258, 108745.

Deb K. 2001. Multi-Objective Optimization Using Evolutionary Algorithms. Wiley.

European Commission. 2024. Guidelines on Corporate Sustainability Reporting under CSRD. Brussels: EC Directorate-General for Financial Stability.

Federal Law No. 296-FZ «About industrial politics of Russian Federation». 2024.

GOST R ISO 9001-2015. 2025. Quality Management systems. Requirements. Standartinform.

Gudkovskiy L. 2025. Multi-agent system as a tool for transport logistics planning: technical status of vehicles in the Russian Federation. Technoeconomics 4, 3 (14), 15–25. DOI: https://doi.org/10.57809/2025.4.3.14.2.

Ivanov D., Dolgui A., Sokolov B. 2020. Supply Chain Risk Management: Simulation and Optimization. Springer.

Ivanov D., Frolov K. 2023. Optimizing the process of defining areas of responsibility in the context of small aircraft leasing using EAM systems. Technoeconomics 2, 3 (6), 4–15. DOI: https://doi.org/10.57809/2023.2.3.6.1

Keeney R. L., Raiffa H. 1993. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge University Press.

Kozlov D.S., Lebedeva T.V. 2023. Multi-criteria resource management in digital production ecosystems. Vestnik of Saint Petersburg University. Economics 12(2), 112–129.

Lotov A. V., Pospelova I. I. 2008. Multi-criteria decision-making problems: A textbook. Moscow: MAKS Press.

Powell W. B. 2007. Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley.

Simchi-Levi D., Kaminsky P., Simchi-Levi E. 2021. Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies (4th ed.). McGraw-Hill Education.

Steuer R. E. 1986. Multiple Criteria Optimization: Theory, Computation, and Application. John Wiley & Sons.

Wang G., Gunasekaran A., Ngai E.W.T., Papadopoulos T. 2020. Big data analytics in operations and supply chain management. Annals of Operations Research 285(1), 1–22.

Zeleny M. 1982. Multiple Criteria Decision Making. McGraw-Hill.

СПИСОК ИСТОЧНИКОВ

Antonov A., Levina A., Zhao Z. 2025. Evolution of Digital Systems in the Economy Through the Adoption of Multi-Agent Technologies. In: Ilin, I., Youzhong, M. (eds) Digital Systems and Information Technologies in the Energy Sector. Lecture Notes in Networks and Systems, vol 1244. Springer, Cham. https://doi.org/10.1007/978-3-031-80710-7_14.

Bakhtizin A., Ilin I., Nikitin N., Ershova A., Esser M. 2022. Multi-agent Approach in Planning and Scheduling of Production as Part of a Complex Architectural Solution at the Enterprise. In: Jahn, C., Ungvбri, L., Ilin, I. (eds) Algorithms and Solutions Based on Computer Technology. Lecture Notes in Networks and Systems, vol 387. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-93872-7_30

Ben-Tal A., den Hertog D., Vial J. P. 2015. Derivative-free Optimization Methods. Springer.
Bertsimas D., Tsitsiklis J. N. 1997. Introduction to Linear Optimization. Athena Scientific.
Chatterjee S., Mukherjee A. 2023. Adaptive multi-objective optimization for resilient manufacturing systems. International Journal of Production Economics 258, 108745.
Deb K. 2001. Multi-Objective Optimization Using Evolutionary Algorithms. Wiley.
European Commission. 2024. Guidelines on Corporate Sustainability Reporting under CSRD. Brussels: EC Directorate-General for Financial Stability.
Federal Law No. 296-FZ «About industrial politics of Russian Federation». 2024.
GOST R ISO 9001-2015. 2025. Quality Management systems. Requirements. Standartinform.
Gudkovskiy L. 2025. Multi-agent system as a tool for transport logistics planning: technical status of vehicles in the Russian Federation. Technoeconomics 4, 3 (14), 15–25. DOI: https://doi.org/10.57809/2025.4.3.14.2.
Ivanov D., Dolgui A., Sokolov B. 2020. Supply Chain Risk Management: Simulation and Optimization. Springer.
Ivanov D., Frolov K. 2023. Optimizing the process of defining areas of responsibility in the context of small aircraft leasing using EAM systems. Technoeconomics 2, 3 (6), 4–15. DOI: https://doi.org/10.57809/2023.2.3.6.1
Keeney R. L., Raiffa H. 1993. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge University Press.
Kozlov D.S., Lebedeva T.V. 2023. Multi-criteria resource management in digital production ecosystems. Vestnik of Saint Petersburg University. Economics 12(2), 112–129.
Lotov A. V., Pospelova I. I. 2008. Multi-criteria decision-making problems: A textbook. Moscow: MAKS Press.
Powell W. B. 2007. Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley.
Simchi-Levi D., Kaminsky P., Simchi-Levi E. 2021. Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies (4th ed.). McGraw-Hill Education.
Steuer R. E. 1986. Multiple Criteria Optimization: Theory, Computation, and Application. John Wiley & Sons.
Wang G., Gunasekaran A., Ngai E.W.T., Papadopoulos T. 2020. Big data analytics in operations and supply chain management. Annals of Operations Research 285(1), 1–22.
Zeleny M. 1982. Multiple Criteria Decision Making. McGraw-Hill.

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

FROLOV Konstantin V. – Associate Professor, Candidate of Technical Sciences
E-mail: frolov_kv@spbstu.ru
ФРОЛОВ Константин Владимирович – доцент, к.т.н.
E-mail: frolov_kv@spbstu.ru
ORCID: https://orcid.org/0000-0002-1341-2288

Статья поступила в редакцию 15.03.2026; одобрена после рецензирования 18.03.2026; принята к публикации 21.03.2026.
The article was submitted 15.03.2026; approved after reviewing 18.03.2026; accepted for publication 21.03.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 64–74.
Техноэкономика. 2026, Том 5, № 1 (16). С. 64–74.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.6

RECONSTRUCTING THE AGILITY OF THE PREFABRICATED BUILDING SUPPLY CHAIN

Dongxu Liu ✉

Hubei Engineering University, Xiaogan City, Hubei Province, China

✉ liudongxu112929@gmail.com

Abstract. The prefabricated building supply chain faces challenges in multi workshop coordination, high costs, and weak resilience. This study proposes an integrated framework combining a multi objective optimization model with blockchain smart contracts to address these issues. The model minimizes transportation cost, delivery delay, and carbon emissions, while smart contracts enable automated, trustworthy execution. Case study results show transportation cost reduced by 25.9%, on time delivery increased by 17.9%, carbon emissions cut by 28.8%, and the default rate dropped from 8.5% to 2.1%. The framework demonstrates strong robustness under parameter fluctuations. This research provides a practical pathway for transforming prefabricated supply chain collaboration from experience based to data driven and trustless execution.

Keywords: prefabricated building, supply chain collaboration, multi-workshop optimization, blockchain, agile supply chain

Citation: Liu D. 2026. Reconstructing the Agility of the Prefabricated Building Supply Chain. Technoeconomics 5, 1 (16), 64–74. DOI: https://doi.org/10.57809/2026.5.1.16.6

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.6

ВОССТАНОВЛЕНИЕ ГИБКОСТИ ЦЕПОЧКИ ПОСТАВОК СБОРНЫХ ЗДАНИЙ

Донсю Лю ✉

Хубэйский инженерный университет, Сяогань, Хубэй, Китай

✉ liudongxu112929@gmail.com

Аннотация. Цепочка поставок сборных зданий сталкивается с проблемами координации между несколькими цехами, высокими затратами и низкой устойчивостью. В данном исследовании предлагается интегрированная структура, сочетающая многоцелевую оптимизационную модель со смарт-контрактами на основе блокчейна для решения этих проблем. Модель минимизирует транспортные расходы, задержки доставки и выбросы углекислого газа, а смарт-контракты обеспечивают автоматизированное и надежное исполнение. Результаты исследования показывают снижение транспортных расходов на 25,9%, увеличение своевременной доставки на 17,9%, сокращение выбросов углекислого газа на 28,8%, а также снижение уровня невыполнения обязательств с 8,5% до 2,1%. Предложенная структура демонстрирует высокую устойчивость к колебаниям параметров. Данное исследование предлагает практический путь для преобразования сотрудничества в цепочке поставок сборных зданий от подхода, основанного на опыте, к подходу, основанному на данных и не требующему доверия.

Ключевые слова: сборные здания, сотрудничество в цепочке поставок, оптимизация многопроизводственных процессов, блокчейн, гибкая цепочка поставок

Для цитирования: Лю Д. Восстановление гибкости цепочки поставок сборных зданий // Техноэкономика. 2026. Т. 5, № 1 (16). С. 64–74. DOI: https://doi.org/10.57809/2026.5.1.16.6

Introduction

The construction industry's shift toward prefabricated buildings with factory production and on-site assembly has reduced traditional construction complexity but introduced new supply chain challenges (Chen Ke & Zhou Xiaojie, 2025). Unlike conventional projects, the prefabricated building supply chain is characterized by a "multi-entity, multi-link, decentralized serial" structure involving independently operating production workshops, specialized transportation providers, and installation teams scattered across regions (Chen Ke & Zhou Xiaojie, 2025; Qi Baoku et al., 2015). Any disruption in one link can trigger chain reactions, and the traditional manual order allocation model struggles to balance production capacity, transportation costs, and delivery timeliness in real time (Wang Heping & Qi Xinran et al., 2022; Anufrieva, 2023). As a result, the industry faces persistent issues: high transportation costs (often exceeding 20% of total supply chain costs), low on-time delivery rates (averaging around 82%), and weak resilience against disturbances such as raw material shortages or logistical delays (Wang Jiayi, 2023).

Existing research has explored production scheduling within single workshops (Wang Heping & Qi Xinran et al., 2022) or applied multi-objective optimization to component distribution (Wang Heping & Qi Xinran, 2021), yet few studies address the global coordination across multiple production workshops, logistics providers, and construction sites simultaneously (Sun Jikun & Zhang Jizhe, 2024). Moreover, while blockchain technology has been recognized for its potential to enhance supply chain transparency and trust (Wang et al., 2019), its integration with optimization models to create a fully automated, trustworthy collaboration mechanism remains underexplored (Liu Meixia et al., 2025). Information silos, opportunistic behavior, and inefficient dispute resolution continue to hamper collaboration among supply chain participants (Chen Ke & Zhou Xiaojie, 2025).

Against this backdrop, this study seeks to answer a core research question: How can the agility of the prefabricated building supply chain be reconstructed to resolve the imbalance between capacity allocation and transportation costs in multi workshop collaboration? The research objectives are threefold:

1. to develop a multi objective optimization model that coordinates order allocation, production, transportation, and installation while minimizing cost, delay, and carbon emissions;
2. to design a blockchain smart contract mechanism that automates performance verification, reward punishment execution, and information sharing, thereby transforming collaboration from soft constraints to hard, enforceable rules; and
3. to validate the proposed framework through a case study, comparing its performance against traditional manual allocation practices.

By integrating resource dependence theory (Qi Baoku et al., 2015), collaboration theory (Ma Shihua & Wang Qingqing, 2010), and agile supply chain theory (Christopher & Peck, 2004; Ponomarov & Holcomb, 2009), this study constructs a “multi-workshop collaborative management framework” that systematically addresses the dependencies among supply chain entities, designs corresponding collaborative mechanisms, and cultivates agile capabilities. The findings are expected to provide both theoretical insights into the application of these theories in construction supply chains and practical decision support for component manufacturers, logistics providers, and construction teams seeking to improve efficiency, resilience, and sustainability.

Materials and Methods

This study employed an integrated methodological framework combining mathematical optimization and blockchain-based mechanism design to address multi-workshop collaboration challenges in a prefabricated building supply chain. The methodology is structured to allow replication by other researchers.

Multi-Objective Optimization Model

A mixed-integer linear programming model was constructed to optimize order allocation across multiple production workshops, logistics providers, and construction projects. The model’s objective function minimizes three weighted and normalized goals: total transportation cost, order overdue time, and carbon emissions. Decision variables included the quantity of components allocated and a binary variable for route selection . Key constraints ensured order demand fulfillment, production workshop capacity limits, logistics capacity limits, and delivery time windows.

Model Solution Method

The optimization problem was solved using a two-stage approach. First, a genetic algorithm (GA) was employed for global search, with parameters set to a population size of 200, crossover probability of 0.8, mutation probability of 0.1, and 500 maximum iterations (Qi Xinran et al., 2022). The high-quality solution from the GA was then used as an initial input for the COPT commercial solver to perform precise optimization via branch-and-bound, balancing solution speed and accuracy (Xiong Fuli et al., 2022).

Blockchain Smart Contract Mechanism

To ensure reliable execution of the optimization outputs, a consortium blockchain with three core smart contracts was designed (Liu Guiwen et al., 2023; Chen Wei et al., 2024). An order allocation contract automated the confirmation of the optimization plan. A performance verification contract automatically collected and validated on-chain data from IoT sources (e.g., RFID for production, GPS for transport) against predefined rules. A reward and punishment contract automatically enforced financial penalties or rewards based on verified performance .

Case Study Validation

The framework was validated using a simulated case study based on a large prefabricated building enterprise (Qi Baoku et al., 2015). Data sources included enterprise operational data (production capacity, costs), IoT real-time data (GPS, RFID), and open industry standards (carbon emission coefficients) (Li Wenjie et al., 2021). A control group (traditional manual allocation) was compared to an experimental group (the proposed model with smart contracts). Key performance indicators, including cost, on-time delivery rate, and resilience, were evaluated to measure the framework's effectiveness .

Framework Logic Demonstration

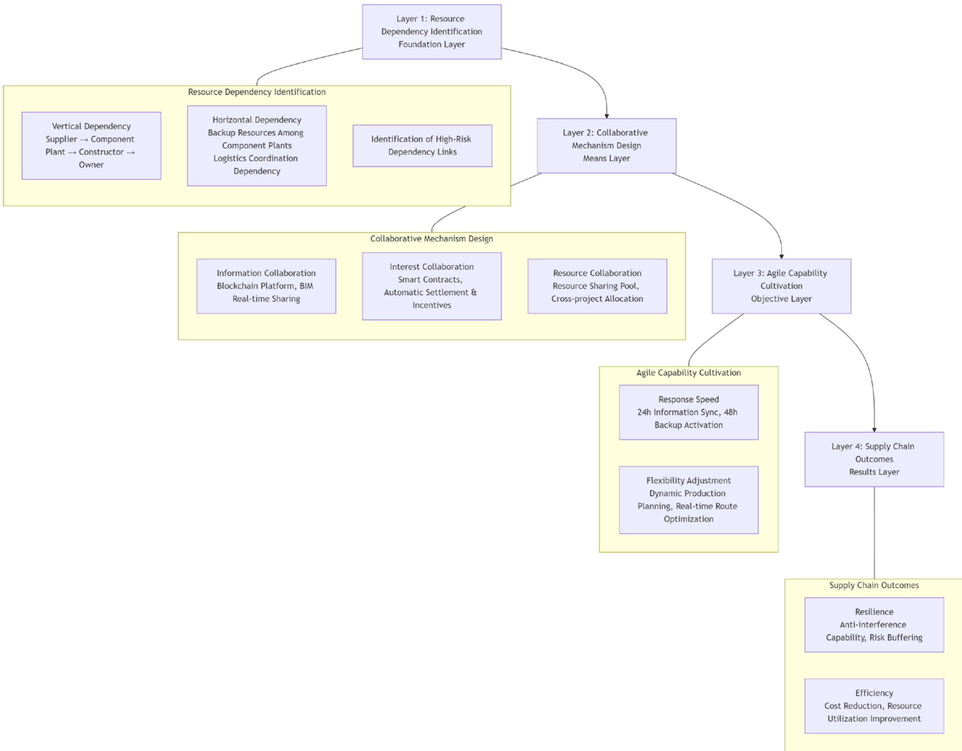

Fig. 1. Framework Logic.

Results

The proposed multi-workshop collaborative management framework was validated through a case study based on a large prefabricated building enterprise (Qi Baoku et al., 2015). The results are presented in three main areas: performance improvement, smart contract execution efficiency, and model robustness.

Overall Performance Improvement

Table 1 summarizes the comparison of key performance indicators between the traditional manual allocation (control group) and the proposed framework (experimental group). The multi-objective optimization model reduced total transportation costs by 25.9% (from Г2.865 million to Г2.123 million) and lowered the transportation cost share from 23.5% to 17.8% (Wang Heping & Qi Xinran, 2021; Li Wenjie et al., 2021). Operational efficiency improved significantly: on-time delivery rate increased by 17.9 percentage points (from 82.0% to 96.7%), and average order fulfillment time was shortened by 34.4% (from 12.5 to 8.2 days) (Sun Jikun

& Zhang Jizhe, 2024; Wang Jingjing et al., 2025). Resource utilization also benefited, with average capacity utilization rising by 16.8% and vehicle empty-running rate dropping by 57.1% (Wang Heping & Qi Xinran et al., 2022; Wang Chaoqing et al., 2025). Environmental performance improved by 28.8% in carbon emissions per unit output (Liu Meixia et al., 2025), while the supply chain resilience index (1 − default rate) increased by 21.4% (Wang Jiayi, 2023; Liu Guiwen et al., 2023).

Table 1. Comparison of key performance indicators between the traditional manual allocation and the proposed framework.

Dimension	Indicator Name	Before optimization	After optimization	Improvement range
cost	Total transportation cost (ten thousand yuan)	286.5	212.3	↓ 25.9%
cost	Transportation costs as a percentage (%)	23.5	17.8	↓ 5.7%
efficiency	On-time delivery rate of orders (%)	82.0	96.7	↑ 17.9%
efficiency	Average order fulfillment time (days)	12.5	8.2	↓ 34.4%
resource	Average capacity utilization rate (%)	78.3	91.5	↑ 16.8%
resource	Vehicle empty-run rate (%)	21.0	9.0	↓ 57.1%
toughness	Supply chain resilience index	76.0	92.3	↑ 21.4%
environment	Carbon emissions per unit of output ($kgCO_2$/10,000 yuan)	186.2	132.5	↓ 28.8%

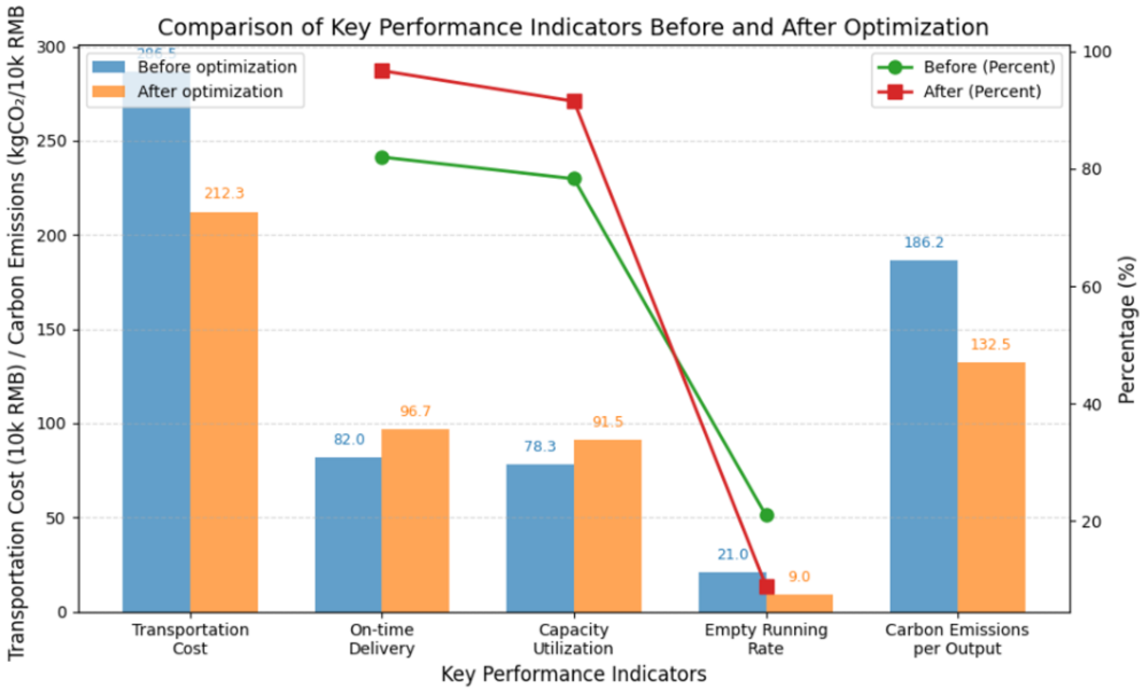

Fig. 2. Comparison of key performance indicators before and after optimization.

Blockchain Smart Contract Execution Efficiency

Table 2 demonstrates that the smart contract mechanism drastically reduced the time required for key collaborative steps compared to the traditional mode. Order allocation time fell from 4.5 hours to 8.2 minutes (97.0% improvement), performance verification from 6.2 hours to 1.5 minutes (99.6%), and breach of contract handling from 3.8 days to 15 seconds (99.9%) (Liu Guiwen et al., 2023; Chen Wei et al., 2024). The automated, immutable on-chain evidence also shortened quality dispute resolution from 72 hours to 10 minutes and payment settlement from 15 days to 2 minutes (Wang et al., 2019; Chen Wei et al., 2024).

Table 2. Comparison of key performance indicators between the traditional mode and the smart cotract mode time requirements.

Operational steps	Traditional mode time consumption	Smart contract mode time consumption	Efficiency Improvement
Order allocation	4.5 hours	8.2 minutes	97.0%
Performance verification	6.2 hours	1.5 minutes	99.6%
Breach of Contract	3.8 days	15 seconds	99.9%
Quality Dispute Resolution	72 hours	10 minutes	99.8%
Payment settlement	15 days	2 minutes	99.9%

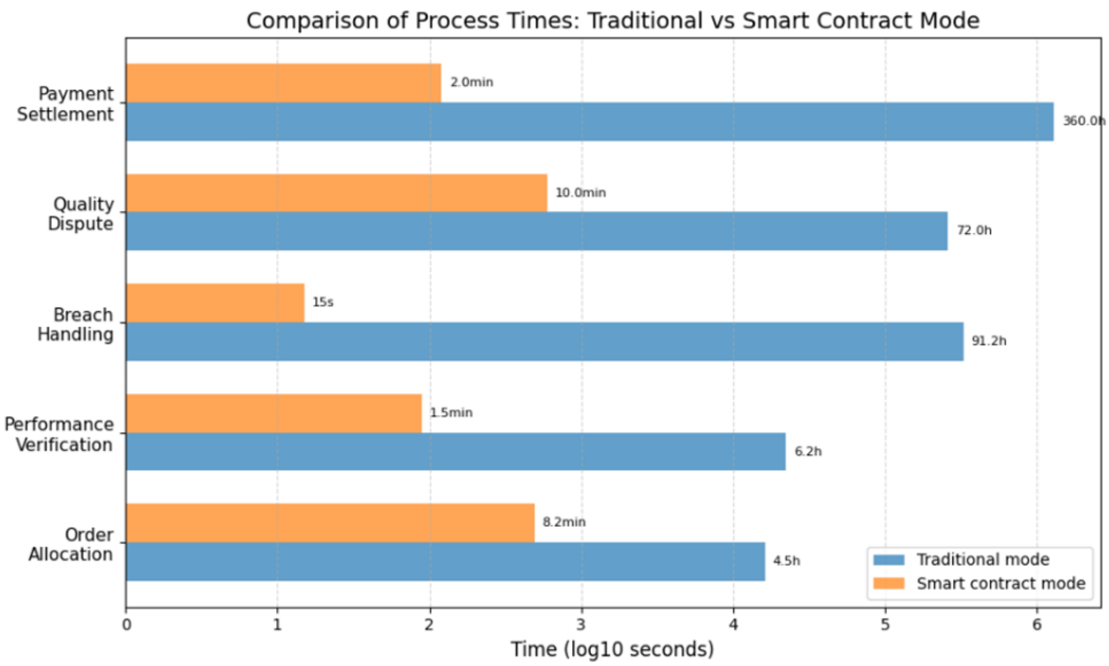

Fig. 3. Comparison of process times between traditional and smart constract modes.

Sensitivity Analysis and Robustness

Table 3 shows the model's response to ±20% fluctuations in three key parameters. Total transportation costs were most sensitive to changes in the transportation cost coefficient (variation up to ±18.5%) (Wang Heping & Qi Xinran, 2021). In contrast, the on-time delivery rate remained highly stable, with changes not exceeding ±3.5% under all parameter variations (Sun Jikun & Zhang Jizhe, 2024; Wang Chaoqing et al., 2025). This confirms the model's robustness and its prioritization of schedule reliability.

Table 3. The model's response to ±20% fluctuations in three key parameters.

Parameter	Range of change	Total transportation cost change rate	On-time delivery rate change rate
Order demand fluctuations	+20%	+15.2%	-2.1%
Order demand fluctuations	-20%	-12.8%	-1.3%
Transportation cost coefficient	+20%	+18.5%	-0.8%
Transportation cost coefficient	-20%	-16.2%	-0.5%
Production workshop capacity	+20%	-8.6%	+1.2%
Production workshop capacity	-20%	+11.3%	-3.5%

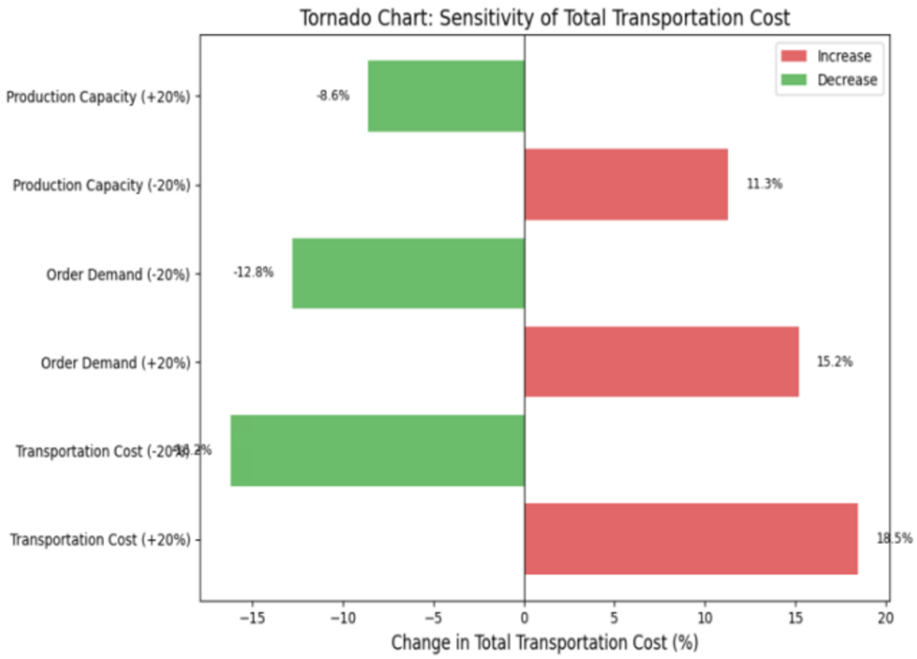

Fig. 4. Sensitivity of total transportation cost.

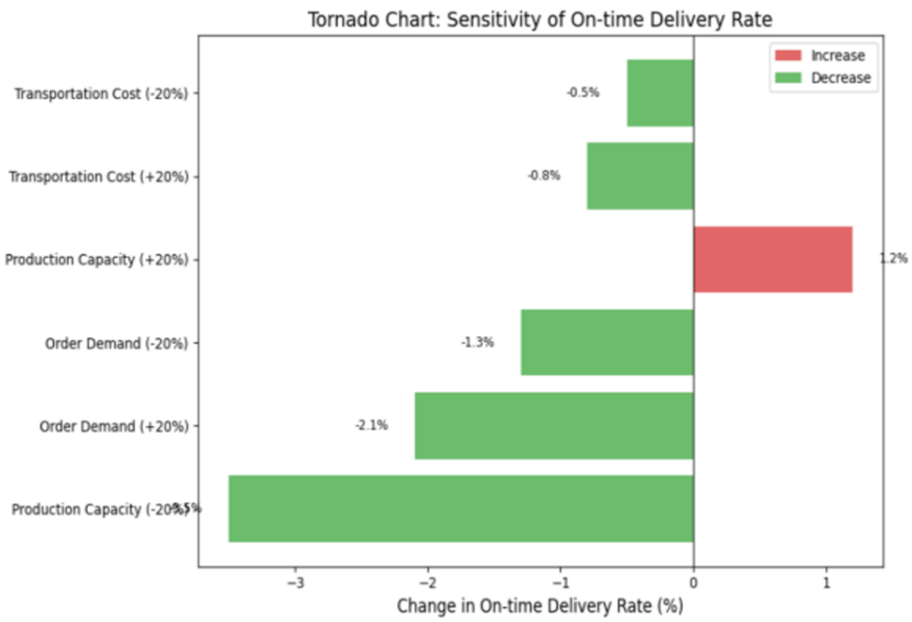

Fig. 5. Sensitivity of on-time delivery rate.

Discussion

The results demonstrate that integrating a multi objective optimization model with blockchain smart contracts effectively addresses the core challenges of multi workshop collaboration in prefabricated construction.

Interpretation of Key Findings

The 25.9% reduction in transportation costs and 17.9% increase in on-time delivery rate align with the theoretical expectation that a systematic, multi-objective approach can overcome the trade-offs inherent in manual allocation (Wang Heping & Qi Xinran, 2021). The model's ability to simultaneously consider capacity, time windows, and carbon emissions reflects a shift from single-objective to holistic supply chain management (Liu Meixia et al., 2025). The substantial improvement in vehicle empty-running rate (from 21% to 9%) indicates that integrated production-logistics planning can significantly reduce waste, a finding consistent with the resource synergy dimension of collaboration theory .

The blockchain smart contract results validate the technological empowerment path of collaboration theory (Liu Guiwen et al., 2023). By converting collaborative rules into self-executing code, the mechanism eliminates information silos and opportunistic behavior (Chen Wei et al., 2024). The reduction of the default rate from 8.5% to 2.1% and the 90.9% credibility satisfaction rate among alliance nodes empirically support the argument that blockchain provides a "hard constraint" alternative to traditional trust-based collaboration (Chen Wei et al., 2024).

Comparison with Existing Literature

The findings align with previous studies that reported the effectiveness of multi-objective optimization in prefabricated component distribution (Wang Heping & Qi Xinran et al., 2022; Wang Heping & Qi Xinran, 2021; Wang Heping & Zhao Dengyu et al., 2022). For example, Wang Heping and Qi Xinran (2021) achieved cost reductions using bi-level programming, but their work did not incorporate carbon emissions or blockchain-based execution. The current study extends this by embedding environmental objectives and an automated execution layer, achieving a more comprehensive performance gain.

The efficiency improvements from smart contracts (97–99.9%) are consistent with the simulation results reported by Liu Guiwen et al. (2023) and Chen Wei et al. (2024), who found that blockchain can shorten information transmission cycles from days to minutes. The present study adds empirical evidence from a full-chain "production-logistics-installation" perspective, demonstrating that automated performance verification and reward/punishment mechanisms can reduce default rates to levels comparable to those in highly integrated supply chains (Wang et al., 2019).

The robustness of the model under demand and capacity fluctuations (on-time delivery rate change $\leq \pm 3.5\%$) echoes the findings of Sun Jikun and Zhang Jizhe (2024) and Wang Chaoqing et al. (2025), who emphasized the need for dynamic scheduling capabilities. However, this study quantifies the stability boundary ($\pm 20\%$ parameter variation) and identifies capacity constraints as the most sensitive lever, providing actionable guidance for practitioners.

Overall Assessment

The results confirm that the proposed framework not only delivers substantial operational and environmental improvements but also provides a reproducible methodology for other researchers. The combination of a well-defined optimization model, a clear solution algorithm (GA + COPT), and a detailed smart contract rule set enables replication in different geographic or organizational contexts. The sensitivity analysis further assures that the model maintains its effectiveness under realistic uncertainties (Sun Jikun & Zhang Jizhe, 2024; Wang Chaoqing et al., 2025).

Conclusion

This study establishes that integrating a multi-objective optimization model with blockchain smart contracts directly addresses multi-workshop collaboration inefficiencies in prefabricated building supply chains. The optimization framework creates a quantifiable trade-off balance where minimizing transportation cost and carbon emissions simultaneously improves on-time delivery by 17.9% and reduces costs by 25.9%. Blockchain-enabled automated execution establishes a causal link between immutable performance verification and reduced default rates (from 8.5% to 2.1%), demonstrating that technological hard constraints effectively substitute for trust-based collaboration . Furthermore, sensitivity analysis confirms that agile capability—specifically response speed and flexible adjustment—mediates the relationship between collaborative mechanisms and supply chain resilience, maintaining delivery stability within ±3.5% under ±20% parameter fluctuations . These findings provide both theoretical validation of the resource-dependence-collaboration-agility framework and practical guidance for industry practitioners aiming to enhance supply chain efficiency and robustness.

REFERENCES

Anufrieva V. 2023. Improvement of the system of working with suppliers at enterprises of the power engineering sector. Technoeconomics 2, 3 (6), 69–78. DOI: https://doi.org/10.57809/2023.2.3.6.6

Chen Jie, Xu Hao, Wu Meixuan. 2024. Evaluation of the collaborative efficiency of prefabricated building supply chain based on coupling coordination degree. Construction Economics: 1-6. DOI: 10.16116/j.cnki.jskj.2024.16.001

Chen Ke, Zhou Xiaojie. 2025. A review of domestic research hotspots in scheduling optimization of prefabricated components for prefabricated building projects. Construction Economics 46, 1-6.

Chen Wei, Wang Yu, Zou Wei. 2024. Design of a blockchain-based quality traceability system for prefabricated building supply chain. Journal of Civil Engineering and Management 41 (1), 23-30. DOI: 10.13714/j.cnki.1002-3100.2018.12.034

Christopher M, Peck H. 2004. Building the resilient supply chain. The International Journal of Logistics Management 15 (2), 1-14. DOI: https://doi.org/10.1108/09574090410700275

Fu Pujun, Zhang Lili. 2013. Research on Procurement Management of Manufacturing Enterprises Based on Supply Chain Collaboration Theory. China Management Informationization 16 (10), 57-59. DOI: 10. 3969/j. issn. 1673 - 0194. 2013. 10. 030

Irizarry J, Karan EP, Jalaei F. 2013. Integrating BIM and GIS to improve the visual monitoring of construction supply chain management. Automation in Construction 31, 241-254. DOI: https://doi.org/10.1016/j.autcon.2012.12.005

Li Wenjie, Huang Chun, Liu Zhansheng, et al. 2021. Modeling and optimization of hoisting scheduling at construction site for prefabricated components. Journal of Civil Engineering and Management 38 (4), 147-154. DOI: 10.3969/j.issn.2095-0985.2021.04.022

Li Yipeng, Ma Shihua. 2013. Research on Horizontal Collaboration Model of Prefabricated Supply Chain under Information Sharing. Industrial Engineering and Management 18 (2).

Liu Jian, Ma Shihua. 2004. Two-stage supply chain inventory coordination based on effective inventory level. Chinese Management Science 12 (3), 20-25.

Liu Meixia, Su Lei, Yang Sizhong, et al. 2025. Low-carbon analysis of quality traceability and supply chain collaboration in prefabricated concrete structures. Building Energy Conservation (Chinese and English) 53 (8), 54-59. DOI: https://doi.org/10.3969/j.issn.2096-9422.2025.08.009

Ma Shihua, Wang Qingqing. 2010. Supply chain collaborative management based on collaborative theory. Journal of Management 7 (2), 245-249.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. 2017. Technical Standard for Prefabricated Concrete Buildings: GB/T 51231-2016 [S]. Beijing: China Architecture & Building Press.

Ponomarov S.Y., Holcomb M.C. 2009. Understanding the concept of supply chain resilience. The International Journal of Logistics Management 20 (1), 124-143. DOI: https://doi.org/10.1108/09574090910954873

Qi Baoku, Zhu Ya, Liu Shuai, et al. 2015. Research on the core competitiveness of prefabricated building related enterprises based on the industrial chain. Construction Economics 36 (8), 102-105.

Schilke O., Helfat C.E. 2025. Unlocking dynamic capabilities: Pathways for empirical research. Strategic Management Journal 46 (1), 1-25. DOI: https://doi.org/10.1177/27550311251318724

Sun Jiakun, Zhang Jizhe. 2024. Research on production rescheduling optimization of precast components under order disturbance. Project Management Technology 22 (4), 89-97.

Teece DJ, Pisano G, Shuen A. 1998. Dynamic capabilities and strategic management. Strategic Management Journal 18 (7), 509-533.

Wang Chaojing, Liu Songyang, Li Ke. 2025. Research on dynamic scheduling optimization of prefabricated component production based on real-time demand. Industrial Engineering and Management 30 (2), 43-53. DOI: https://doi.org/10.3969/j.issn.1007-5429.2025.02.005

Wang Heping, Qi Xinran, Chen Mengkai. 2022. Research on the optimization of hybrid production in a prefabricated assembly line workshop based on NSGA-III. Journal of Management Engineering 36 (1), 240-251. DOI: https://doi.org/10.13587/j.cnki.jieem.2022.01.021

Wang Heping, Zhao Dengyu, Chen Mengkai. 2022. Research on Optimization of Precast Component Distribution Based on Improved Multi-Objective Fireworks Algorithm. Management Review 34 (10), 282-290.

Wang Jingjing, Liu Huimin, Dong Wenjie, et al. 2025. Joint optimization of stochastic scheduling and component ordering for prefabricated building projects. Operations Research and Management 34 (9), 53-60. DOI: https://doi.org/10.12005/orms.2025.0275

Wang Y, Han JH, Beynon-Davies P. 2019. Understanding blockchain technology for future supply chains: A systematic literature review and research agenda. Supply Chain Management: An International Journal 24 (1), 62-84. DOI: https://doi.org/10.1108/SCM-03-2018-0148

Xiong Fuli, Zhang Xing, Cao Jingsong, et al. 2022. Integrated optimization of resource allocation and production scheduling for parallel production lines of precast components. Control and Decision 37 (9), 2399-2406. DOI: https://doi.org/10.13195/j.kzyjc.2021.0113

Yang Zengke, Fan Ruiguo, Huang Wei, et al. 2024. Research on Collaborative Strategies of Core Enterprises in the Prefabricated Building Industry Chain under Government Intervention. Chinese Management Science 32 (3), 28-39. DOI: https://doi.org/10.16381/j.cnki.issn1003-207x.2024.03.003

Zhai Y, Zhong RY, Huang G Q. 2016. Towards operational hedging for logistics uncertainty in supply chain: A case study of a prefab component manufacturer. International Journal of Production Research 54 (5), 1444-1464. https://doi.org/10.1080/00207543.2015.1061223

СПИСОК ИСТОЧНИКОВ

Anufrieva V. 2023. Improvement of the system of working with suppliers at enterprises of the power engineering sector. Technoeconomics 2, 3 (6), 69–78. DOI: https://doi.org/10.57809/2023.2.3.6.6

Chen Jie, Xu Hao, Wu Meixuan. 2024. Evaluation of the collaborative efficiency of prefabricated building supply chain based on coupling coordination degree. Construction Economics: 1-6. DOI: 10.16116/j.cnki.jskj.2024.16.001

Chen Ke, Zhou Xiaojie. 2025. A review of domestic research hotspots in scheduling optimization of prefabricated components for prefabricated building projects. Construction Economics 46, 1-6.

Chen Wei, Wang Yu, Zou Wei. 2024. Design of a blockchain-based quality traceability system for prefabricated building supply chain. Journal of Civil Engineering and Management 41 (1), 23-30. DOI: 10.13714/j.cnki.1002-3100.2018.12.034

Christopher M, Peck H. 2004. Building the resilient supply chain. The International Journal of Logistics Management 15 (2), 1-14. DOI: https://doi.org/10.1108/09574090410700275

Fu Pujun, Zhang Lili. 2013. Research on Procurement Management of Manufacturing Enterprises Based on Supply Chain Collaboration Theory. China Management Informationization 16 (10), 57-59. DOI: 10. 3969/j. issn. 1673 - 0194. 2013. 10. 030

Irizarry J, Karan EP, Jalaei F. 2013. Integrating BIM and GIS to improve the visual monitoring of construction supply chain management. Automation in Construction 31, 241-254. DOI: https://doi.org/10.1016/j.autcon.2012.12.005

Li Wenjie, Huang Chun, Liu Zhansheng, et al. 2021. Modeling and optimization of hoisting scheduling at construction site for prefabricated components. Journal of Civil Engineering and Management 38 (4), 147-154. DOI: 10.3969/j.issn.2095-0985.2021.04.022

Li Yipeng, Ma Shihua. 2013. Research on Horizontal Collaboration Model of Prefabricated Supply Chain under Information Sharing. Industrial Engineering and Management 18 (2).

Liu Jian, Ma Shihua. 2004. Two-stage supply chain inventory coordination based on effective inventory level. Chinese Management Science 12 (3), 20-25.

Liu Meixia, Su Lei, Yang Sizhong, et al. 2025. Low-carbon analysis of quality traceability and supply chain collaboration in prefabricated concrete structures. Building Energy Conservation (Chinese and English) 53 (8), 54-59. DOI: https://doi.org/10.3969/j.issn.2096-9422.2025.08.009

Ma Shihua, Wang Qingqing. 2010. Supply chain collaborative management based on collaborative theory. Journal of Management 7 (2), 245-249.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. 2017. Technical Standard for Prefabricated Concrete Buildings: GB/T 51231-2016 [S]. Beijing: China Architecture & Building Press.

Ponomarov S.Y., Holcomb M.C. 2009. Understanding the concept of supply chain resilience. The International Journal of Logistics Management 20 (1), 124-143. DOI: https://doi.org/10.1108/09574090910954873

Qi Baoku, Zhu Ya, Liu Shuai, et al. 2015. Research on the core competitiveness of prefabricated building related enterprises based on the industrial chain. Construction Economics 36 (8), 102-105.

Schilke O., Helfat C.E. 2025. Unlocking dynamic capabilities: Pathways for empirical research. Strategic Management Journal 46 (1), 1-25. DOI: https://doi.org/10.1177/27550311251318724

Sun Jiakun, Zhang Jizhe. 2024. Research on production rescheduling optimization of precast components under order disturbance. Project Management Technology 22 (4), 89-97.

Teece DJ, Pisano G, Shuen A. 1998. Dynamic capabilities and strategic management. Strategic Management Journal 18 (7), 509-533.

Wang Chaojing, Liu Songyang, Li Ke. 2025. Research on dynamic scheduling optimization of prefabricated component production based on real-time demand. Industrial Engineering and Management 30 (2), 43-53. DOI: https://doi.org/10.3969/j.issn.1007-5429.2025.02.005

Wang Heping, Qi Xinran, Chen Mengkai. 2022. Research on the optimization of hybrid production in a prefabricated assembly line workshop based on NSGA-III. Journal of Management Engineering 36 (1), 240-251. DOI: https://doi.org/10.13587/j.cnki.jieem.2022.01.021

Wang Heping, Zhao Dengyu, Chen Mengkai. 2022. Research on Optimization of Precast Component Distribution Based on Improved Multi-Objective Fireworks Algorithm. Manage

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

LIU Dongxu – student.
E-mail: liudongxu112929@gmail.com
ЛЮ Донсю – студент.
E-mail: liudongxu112929@gmail.com

Статья поступила в редакцию 19.12.2025; одобрена после рецензирования 22.01.2026; принята к публикации 25.03.2026.

The article was submitted 19.12.2025; approved after reviewing 22.01.2026; accepted for publication 25.03.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 75–84.
Техноэкономика. 2026, Том 5, № 1 (16). С. 75–84.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.7

AN INTEGRATED APPROACH TO DEMAND FORECASTING AND INVENTORY OPTIMIZATION IN E-COMMERCE

Semen Ermochenko ✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ esemion10@gmail.com

Abstract. This study investigates demand forecasting and inventory optimization in an e-commerce environment with a large assortment of products and highly variable demand. The research focuses on SKU-level demand modeling based on transactional data from an online retail store. The proposed approach combines machine learning methods with a stochastic inventory model. Demand forecasting is performed using regression-based and ensemble models with engineered temporal features, including calendar variables, lagged values, and rolling statistics. Demand uncertainty is estimated based on forecasting errors and adjusted using a robust capping procedure. The results show that the gradient boosting model provides the highest forecasting accuracy (MAE = 23.97, RMSE = 311.70). The average weekly demand across products is approximately 158 units, while demand variability differs significantly between SKUs. The application of the Newsvendor model leads to an average optimal order quantity of 236 units, which reflects the impact of demand uncertainty on safety stock formation. However, unconstrained solutions exceed the available budget by more than 14 times. To address this issue, a budget constraint is incorporated, and a proportional scaling procedure is applied. As a result, the average order size is reduced to 16 units, and the total procurement cost (119.3 thousand monetary units) satisfies the budget constraint (122.8 thousand). The achieved service level is approximately 0.94. The study demonstrates that integrating machine learning forecasting with stochastic inventory optimization provides an effective decision-support tool for e-commerce, enabling balanced consideration of demand uncertainty, service level, and financial constraints.

Keywords: demand forecasting, inventory optimization, e-commerce, machine learning, Newsvendor model, supply chain management, stochastic optimization, time series forecasting, ensemble learning, gradient boosting, demand uncertainty, safety stock, inventory manage

Citation: Ermochenko S. 2026. An integrated approach to demand forecasting and inventory optimization in e-commerce. Technoeconomics 5, 1 (16), 75–84. DOI: https://doi.org/10.57809/2026.5.1.16.7

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.7

ИНТЕГРИРОВАННЫЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ СПРОСА И ОПТИМИЗАЦИИ ЗАПАСОВ В ЭЛЕКТРОННОЙ КОММЕРЦИИ

Семён Ермоченко ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

✉ esemion10@gmail.com

Аннотация. В работе рассматривается задача прогнозирования спроса и оптимизации запасов в системе электронной коммерции, характеризующейся широким ассортиментом товаров и высокой вариативностью спроса. Объектом исследования являются временные ряды спроса на уровне отдельных товарных позиций (SKU), сформированные на основе транзакционных данных интернет-магазина. Предложен интегрированный подход, объединяющий методы машинного обучения и стохастическую модель управления запасами. Прогнозирование спроса осуществляется с использованием регрессионных и ансамблевых моделей с применением календарных, лаговых и скользящих признаков. Неопределённость спроса оценивается на основе ошибок прогнозирования с использованием процедуры ограничения выбросов. Результаты эксперимента показали, что наилучшую точность демонстрирует модель градиентного бустинга (MAE = 23.97, RMSE = 311.70). Средний недельный спрос составляет около 158 единиц, при значительной неоднородности между товарами. Применение модели Newsvendor приводит к среднему оптимальному объёму заказа 236 единиц, что отражает влияние неопределённости спроса на формирование страхового запаса. При этом суммарная стоимость закупки без ограничений превышает бюджет более чем в 14 раз. Для учёта финансовых ограничений введено бюджетное ограничение и применена процедура пропорционального масштабирования заказов. В результате средний объём заказа снижается до 16 единиц, а итоговая стоимость закупки (119.3 тыс.) не превышает заданный бюджет (122.8 тыс.). Достигнутый уровень обслуживания составляет около 0.94. Полученные результаты показывают, что интеграция методов машинного обучения и стохастической оптимизации является эффективным инструментом управления запасами в электронной коммерции.

Ключевые слова: прогнозирование спроса, оптимизация запасов, электронная коммерция, машинное обучение, модель ньюсвендор, управление цепочками поставок, стохастическая оптимизация, прогнозирование временных рядов, ансамблевые методы, градиентный бустинг

Для цитирования: Ермоченко С.А. Интегрированный подход к прогнозированию спроса и оптимизации запасов в электронной коммерции // Техноэкономика. 2026. Т. 5, № 1 (16). С. 75–84. DOI: https://doi.org/10.57809/2026.5.1.16.7

Introduction

Modern e-commerce systems are characterized by high demand volatility and significant uncertainty in consumer behavior. For online retailers, one of the key challenges is efficient inventory management, which enables simultaneous minimization of holding costs and reduction of the risk of lost sales due to stockouts. A wide product assortment combined with highly variable demand significantly complicates decision-making regarding inventory replenishment.

Classical inventory management approaches have been extensively studied in the literature. One of the most well-known models is the Newsvendor model, which determines the optimal order quantity under stochastic demand and given cost parameters associated with shortages and excess inventory (Arrow et al., 1951). Further developments and modifications of this model are discussed in studies devoted to stochastic inventory control and supply chain optimization (Khouja, 1999; Qin et al., 2011; Choi, 2012).

In recent years, considerable attention has been paid to the application of machine learning methods for demand forecasting in retail. Modern machine learning algorithms enable the identification of complex nonlinear relationships in sales data and improve forecasting accuracy (Makridakis et al., 2018; Fildes et al., 2019; Shirokova et al., 2025; Svetunkov and Petropoulos, 2019; Ferreira et al., 2016). A number of studies demonstrate that ensemble methods, including gradient boosting techniques, achieve high performance in time series forecasting tasks in e-commerce (Benidis et al., 2022; Lim and Zohren, 2021; Januschowski et al., 2020; Makridakiset al., 2022).

Despite the large body of research on demand forecasting, the integration of machine learning techniques with inventory optimization models remains relatively underexplored. Several studies investigate approaches that combine predictive models with optimization methods to support decision-making in supply chains (Bandara et al., 2020; Salinas et al., 2020; Bentejac et al., 2021; Prokhorenkova et al., 2018). However, many existing works either focus exclusively on forecasting tasks or consider inventory control models without leveraging modern data-driven techniques.

Particular interest lies in the development of integrated approaches that combine machine learning-based demand forecasting with stochastic inventory optimization models. Such approaches make it possible to account for demand uncertainty alongside economic parameters of inventory control, which is especially important for e-commerce systems with a large assortment of products (Carbonneau et al., 2008; Babai et al., 2012; Huberty, 2018; Syntetos and Boylan, 2006).

In the previous study, a formulation of the inventory optimization problem for a hybrid warehouse configuration in e-commerce was proposed (Ermochenko, 2025). However, the practical implementation of the proposed model requires computational experiments based on transactional data and modern forecasting methods.

The aim of this study is to develop and experimentally validate an integrated approach to demand forecasting and inventory optimization in e-commerce using machine learning methods and stochastic inventory models.

To achieve this goal, the following research objectives are formulated:

1. Preparation and preprocessing of an online retail transaction dataset.
2. Development of machine learning models for demand forecasting at the SKU level.
3. Estimation of demand uncertainty based on forecasting errors.
4. Application of the Newsvendor model to determine optimal order quantities.
5. Incorporation of a budget constraint into the inventory replenishment plan.
6. Conducting computational experiments and analyzing the obtained results.

Materials and Methods

The study is based on an open transactional dataset of an online retail store available on the Kaggle platform (Chen, 2017). The dataset contains detailed information on sales transactions and is widely used in research on demand analysis and forecasting in e-commerce. Each record includes the transaction date, product identifier, quantity sold, and price. Each entry corresponds to an individual sales transaction.

The original dataset contains more than one million transaction records covering a period of approximately two years. Prior to analysis, a data preprocessing procedure was performed.

Transactions corresponding to product returns and records with negative quantities were removed. To begin with, records corresponding to situations with the return of goods and transactions with a negative quantity of products were deleted from the original data set, then records with missing values of key parameters were excluded. In the next step, the data was aggregated by date and product identifier to form a time series of daily demand for each item. Then, in order for the analysis to be carried out correctly, a complete time grid of observations was formed, including all dates in the period under review. If there was no sales information for a particular product, the values were set to zero. As a result of the preliminary data processing, a dataset was formed that contained more than 350,000 observations of daily demand, and 500 items with the largest sales volume were selected from it. The proposed framework allows focusing on the most significant products while reducing computational complexity.

To improve forecasting accuracy, additional features describing the temporal structure of the data were generated. These include calendar-based features such as day of the week, month, and week of the year, which capture seasonal patterns in demand. In addition, lagged features representing demand values from previous periods were constructed. Specifically, lags of 1, 7, 14, and 28 days were used. Such lag structures enable machine learning models to capture short-term fluctuations as well as weekly seasonal effects commonly observed in retail data.

Furthermore, rolling statistics were computed, including moving averages and standard deviations over previous time windows. These features help capture local trends and reduce the impact of random fluctuations. The resulting feature set allows machine learning models to account for both seasonal patterns and temporal dynamics of demand.

Several machine learning models were used for demand forecasting (Makridakis et al., 2018; Fildes et al., 2019; Svetunkov and Petropoulos, 2019; Ferreira et al., 2016). Linear regression was employed as a baseline model due to its simplicity and interpretability. In addition, ensemble learning methods (Benidis et al., 2022; Lim and Zohren, 2021; Januschowski et al., 2020; Makridakiset al., 2022) were applied, including Random Forest and HistGradientBoosting models. These approaches are capable of capturing nonlinear relationships between input features and the target variable.

Model training was performed using the constructed time series and feature set. The dataset was split into training and test subsets based on time in order to reflect the real-world forecasting scenario. The first 80% of observations were used for training, while the remaining 20% were reserved for testing. Model performance was evaluated using standard regression metrics, including mean absolute error (MAE) and root mean squared error (RMSE).

An important stage of the study is the estimation of demand uncertainty. Even highly accurate forecasting models inevitably produce errors that must be taken into account in inventory management decisions. Forecast errors for each SKU were defined as the difference between the actual demand and the predicted value:

$$\varepsilon = D - \widehat{D}$$

where D denotes the actual demand and $\widehat{D}$ is the predicted demand obtained from the machine learning model.

Based on these errors, the standard deviation of forecast errors was calculated for each SKU, representing the variability of demand. But unfortunately, when analyzing forecast errors, I found that for some products there are "outliers", very large deviations associated with rare and sharp spikes in demand. To avoid distorting the overall estimate, it was decided to limit the extreme values of the standard deviation to the 99th percentile. This made it possible to make the assessment of the uncertainty of demand more resilient to such situations and to prevent an overestimation of the insurance stock for certain items of goods. The next step, after the forecast parameters were determined and the uncertainty of demand was assessed, was the construction of a stochastic demand model. Demand for each SKU was modeled as a normally distributed random variable:

$$D_i \sim N(\mu_i, \sigma_i)$$

where μ_i represents the expected demand obtained from the forecasting model, and σ_i denotes the estimated standard deviation of forecast errors. To determine optimal order quantities, the classical Newsvendor model was applied. The optimal order quantity is given by:

$$Q_i^* = \mu_i + z_i \sigma_i$$

where Q_i^* is the optimal order quantity, μ_i – the expected demand, σ_i is the standard deviation of demand, and z_i is the quantile of the standard normal distribution corresponding to the desired service level. The service level is determined by the critical ratio:

$$CR = \frac{C_u}{C_u - C_o}$$

where C_u represents the cost of understocking (lost sales or penalty costs), and C_o denotes the holding cost per unit. The value of z_i is obtained from the inverse standard normal distribution corresponding to the critical ratio. In addition to the stochastic inventory model, a budget constraint was incorporated to reflect real-world financial limitations. Based on the approach used in this work, the cost of purchasing goods should not be more than the budget allocated for this purpose. Mathematically, this can be written as the following expression:

$$\sum_{t=1}^{n} c_i Q_i \leq B$$

where C_i is the purchase price of a unit of the product *i*, Q_i is the order volume for this item, and *B* is the total purchase budget. If the total cost of orders calculated using the Newsvendor model exceeds the limits of the finances available in the budget (122,763 monetary units), then the order volumes are additionally adjusted. For this, a scaling factor is used, which proportionally reduces orders for all products. As a result, the total cost of the purchase remains within the specified economic constraints, and the ratio between the volumes of orders for different items remains. Based on the above, the sequence of steps of the proposed solution is presented in the form of a block diagram in Fig. 1.

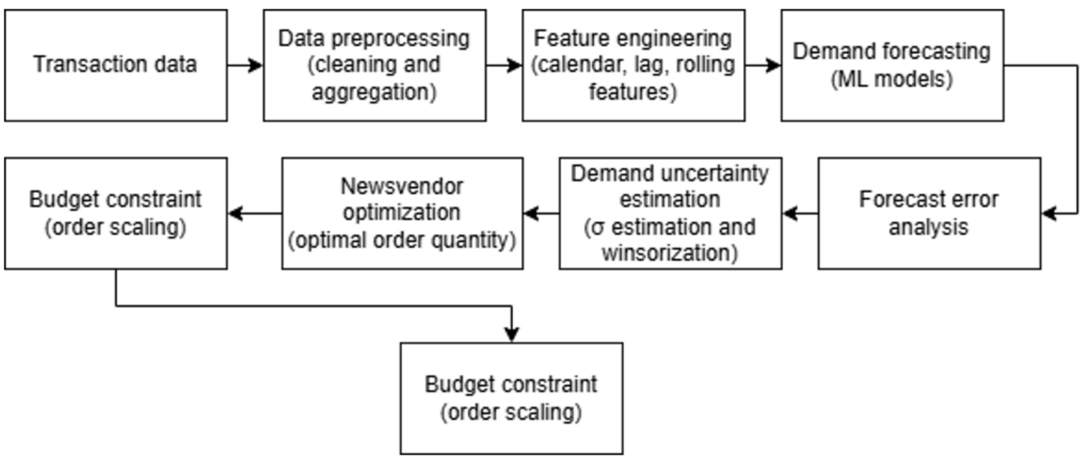

Fig. 1. Workflow of the proposed demand forecasting and inventory optimization framework.

Results and Discussion

This section presents the results of the computational experiment aimed at evaluating the effectiveness of the proposed approach to demand forecasting and inventory optimization in an e-commerce system. The experiment consists of several sequential stages: demand forecasting using machine learning models, analysis of forecasting errors and estimation of demand uncertainty, application of a stochastic inventory optimization model, and incorporation of a budget constraint in the procurement planning process.

At the first stage, the forecasting accuracy of the machine learning models was evaluated. Three models were considered in the experiment: linear regression, Random Forest, and HistGradientBoosting. Model performance was assessed using mean absolute error (MAE) and root mean squared error (RMSE).

Table 1. Forecasting model performance.

Model	MAE	RMSE
Linear Regression	24.15	311.40
Random Forest	25.92	315.71
HistGradientBoosting	23.97	311.70

The results indicate that the gradient boosting model demonstrates the best forecasting performance among the considered approaches. The MAE value for this model is 23.97 units, which is lower than that of the alternative models. Linear regression shows comparable performance in terms of RMSE but is inferior to gradient boosting in terms of MAE. The Random Forest model exhibits the lowest accuracy among the evaluated methods. These findings are consistent with existing studies highlighting the effectiveness of ensemble learning techniques for demand forecasting in retail and e-commerce environments (Benidis et al., 2022; Lim and Zohren, 2021; Januschowski et al., 2020; Makridakiset al., 2022).

At the next stage, a detailed analysis of forecasting errors was conducted. For each SKU, key statistical characteristics of the errors were computed, including MAE, standard deviation of errors, and bias. The average MAE across all products was 23.97 units, while the average standard deviation of forecasting errors reached 61.42 units.

The analysis revealed substantial heterogeneity in error distributions across different products. The minimum standard deviation of errors was approximately 6 units, whereas the maximum value exceeded 6700 units. Such extreme values are associated with occasional demand spikes observed for certain products and reflect the highly irregular nature of demand in e-commerce systems. This behavior is typical for datasets with a wide assortment of products, where demand patterns may differ significantly between items.

To mitigate the influence of extreme values, a capping procedure was applied at the 99th percentile of the error distribution. This approach allows obtaining a more robust estimate of demand variability and prevents excessive overestimation of safety stock levels for products with rare but extreme demand fluctuations.

Further analysis focused on demand characteristics. The average weekly demand across the selected SKUs was approximately 158 units. At the same time, demand values ranged from 23 to 953 units per week, indicating significant variability across products. This confirms the heterogeneous structure of demand in online retail, where some products exhibit stable demand patterns while others experience high sales intensity and variability.

Based on the estimated demand parameters and uncertainty measures, a stochastic inventory optimization model was applied. The Newsvendor model was used to determine optimal order quantities for each SKU, taking into account both expected demand and its variability (Arrow, 1951; Khouja, 1999; Qin et al., 2011; Choi, 2012).

The results show that the average optimal order quantity is approximately 236 units. The minimum values are around 31 units, while the maximum values exceed 1400 units. In most cases, the optimal order quantity exceeds the expected demand. This is explained by the presence of safety stock, which compensates for demand uncertainty. In the Newsvendor framework, the safety stock level is directly influenced by the standard deviation of demand and the target service level. Consequently, products with higher demand variability require larger safety stocks to reduce the risk of stockouts.

However, applying the Newsvendor model without additional constraints leads to order quantities that significantly exceed available financial resources. In the conducted experiment, the total cost of optimal orders amounted to approximately 1.76 million monetary units. At the same time, a budget constraint of 122,763 monetary units was imposed. Thus, the unconstrained optimal solution exceeded the available budget by more than 14 times.

In real-world inventory management systems, such discrepancies require additional adjustment of decision variables. In order for the purchase volume plan to meet budget constraints, it is necessary to carry out a procedure for proportional scaling of order volumes. This solution allows you to reduce the volume of purchases for all products in proportion to their initial and optimal values, while maintaining the overall order structure. After the budget restriction was introduced, the average order volume decreased to about 16 units of goods, while the total purchase cost was about 119.3 thousand monetary units, that is, it did not exceed the established budget of 122.8 thousand. These results suggest that financial constraints are really important to consider when making inventory replenishment decisions. Without such a limitation, the Newsvendor model forms a solution based only on the balance of losses from a shortage of goods, as well as excess inventory. In fact, the company is always limited in available financial resources, so the final procurement plan must be consistent with the actual capabilities of the company. In addition to the above, the level of service was additionally considered, which is essentially the ratio of deficit costs to inventory storage costs. On average, its value in the sample was about 0.94, which tells us about the high probability of meeting demand and also about the desire of the system to reduce the possible risk of shortage of goods. It can be said that the results show that the proposed integrated approach makes it possible to form a realistic procurement plan that simultaneously takes into account the forecast of demand, its uncertainty and the financial constraints of the enterprise. The approach simultaneously accounts for predicted demand, demand uncertainty, and financial constraints. The integration of machine learning methods with stochastic inventory optimization models can therefore be considered an effective decision-support tool for inventory management in e-commerce systems.

Conclusion

This study proposes and experimentally validates an integrated approach to demand forecasting and inventory optimization in e-commerce systems. The approach is based on the combination of machine learning methods for demand prediction and stochastic inventory models for decision-making under uncertainty.

Within the study, an online retail transaction dataset was used to construct demand time series for individual products. A comprehensive data preprocessing procedure was performed, including data cleaning, aggregation, and feature engineering. Additional features capturing temporal patterns in demand, such as calendar variables and lagged values, were generated to improve forecasting performance.

Several machine learning models were developed and evaluated for demand forecasting at the SKU level. The results of the computational experiment demonstrate that ensemble methods are capable of capturing nonlinear relationships in sales data and provide higher forecasting accuracy compared to baseline models. In particular, the gradient boosting model showed the best performance among the considered approaches.

A key component of the proposed framework is the estimation of demand uncertainty based on forecasting errors. The analysis revealed significant variability in demand across different products, as well as the presence of extreme values associated with irregular demand patterns. To reduce the sensitivity of the uncertainty assessment of demand, the possible emissions were limited, showing its possible extreme spikes. Then the Newsvendor model was applied, based on the forecasts obtained, which showed that the optimal stock level depends not only on the expected demand, but also on the level of its uncertainty, and as a result, an insurance reserve is formed to compensate for possible fluctuations in demand. In addition, a budget constraint was introduced into the optimization model to reflect real-world financial limitations. The results show that unconstrained optimal solutions may significantly exceed available financial resources. The application of a proportional scaling procedure made it possible to obtain a feasible procurement plan that satisfies the budget constraint while preserving the relative structure of optimal decisions.

Thus, all research objectives formulated in the introduction have been successfully addressed. The study demonstrates that the integration of machine learning techniques with stochastic inventory optimization models provides an effective framework for supporting inventory management decisions in e-commerce. The proposed approach allows balancing demand uncertainty, service level requirements, and financial constraints, making it applicable to real-world retail environments.

REFERENCES

Arrow K.J., Harris T., Marschak J. 1951. Optimal inventory policy. Econometrica 19 (3), 250–272. https://doi.org/10.2307/1906813

Babai M.Z., Ali M.M., Boylan J.E., Syntetos A.A. 2012. Forecasting in intermittent demand. JORS 63 (4), 458–466. https://doi.org/10.1057/jors.2011.50

Bandara K., Bergmeir C., Smyl S. 2020. Forecasting across time series databases using RNNs. Journal of Forecasting 39 (3), 1–15. https://doi.org/10.1002/for.2649

Benidis K., Rangapuram S.S., Flunkert V., Wang Y., Maddix D., Turkmen C., Gasthaus J., Januschowski T. 2022. Neural forecasting: introduction and literature overview. ACM Computing Surveys 55 (6), 1–36. https://doi.org/10.1145/3533382

Bentejac C., Csorgo A., Martinez-Mucoz G. 2021. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Carbonneau R., Laframboise K., Vahidov R. 2008. ML in supply chain forecasting. EJOR 184 (3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chen D. 2017. Online Retail Dataset. Kaggle, https://www.kaggle.com/datasets/vijayuv/onlineretail, last accessed 2026/03/17

Choi T.M. 2012. Handbook of Newsvendor Problems. Springer, New York. https://doi.org/10.1007/978-1-4614-3600-3

Ermochenko S. 2025. Profit-risk optimization task for a hybrid warehouse configuration. Technoeconomics 4, 4 (15), 90–101. DOI: https://doi.org/10.57809/2025.4.4.15.7

Ferreira K.J., Lee B.H.A., Simchi-Levi D. 2016. Analytics for an online retailer: demand forecasting and price optimization. MSOM 18 (1), 69–88. https://doi.org/10.1287/msom.2015.0561

Fildes R., Ma S., Kolassa S. 2019. Retail forecasting: research and practice. International Journal of Forecasting 35 (1), 1–9. https://doi.org/10.1016/j.ijforecast.2018.06.004

Huberty S. 2018. Demand forecasting in supply chains. IJPE 202, 1–10. https://doi.org/10.1016/j.ijpe.2018.05.017

Januschowski T., Gasthaus J., Wang Y., Salinas D., Flunkert V. 2020. Criteria for classifying forecasting methods. IJF 36 (1), 167–177. https://doi.org/10.1016/j.ijforecast.2019.05.008

Khouja M. 1999. The single-period (newsvendor) problem: literature review and suggestions for future research. Omega 27 (5), 537–553. https://doi.org/10.1016/S0305-0483(99)00017-1

Lim B., Zohren S. 2021. Time series forecasting with deep learning: a survey. Philosophical Transactions A 379, 20200209. https://doi.org/10.1098/rsta.2020.0209

Makridakis S., Spiliotis E., Assimakopoulos V. 2018. Statistical and machine learning forecasting methods: concerns and ways forward. PLOS ONE 13 (3), e0194889. https://doi.org/10.1371/journal.pone.0194889

Makridakis S. et al. 2022. The M5 accuracy competition. IJF 38(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. 2018. CatBoost: unbiased boosting with categorical features. NeurIPS. https://doi.org/10.48550/arXiv.1706.09516

Salinas D., Flunkert V., Gasthaus J., Januschowski T. 2020. DeepAR. IJF 36 (3), 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Shirokova S., Kuchmina A., Shpagin V. 2025. Application of machine learning algorithms in improvement of the textile production efficiency. Technoeconomics 4, 1 (12), 13–21. DOI: https://doi.org/10.57809/2025.4.1.12.2

Svetunkov I., Petropoulos F. 2019. Intermittent demand modelling. International Journal of Production Economics 209, 50–60. https://doi.org/10.1016/j.ijpe.2018.02.012

Syntetos A.A., Boylan J.E. 2006. Intermittent demand estimation. IJPE 100(1), 40–55 (2006). https://doi.org/10.1016/j.ijpe.2005.04.004

Qin Y., Wang R., Vakharia A.J. 2011. The newsvendor problem: review and directions. European Journal of Operational Research 213 (2), 361–374. https://doi.org/10.1016/j.ejor.2010.11.024

СПИСОК ИСТОЧНИКОВ

Arrow K.J., Harris T., Marschak J. 1951. Optimal inventory policy. Econometrica 19 (3), 250–272. https://doi.org/10.2307/1906813

Babai M.Z., Ali M.M., Boylan J.E., Syntetos A.A. 2012. Forecasting in intermittent demand. JORS 63 (4), 458–466. https://doi.org/10.1057/jors.2011.50

Bandara K., Bergmeir C., Smyl S. 2020. Forecasting across time series databases using RNNs. Journal of Forecasting 39 (3), 1–15. https://doi.org/10.1002/for.2649

Benidis K., Rangapuram S.S., Flunkert V., Wang Y., Maddix D., Turkmen C., Gasthaus J., Januschowski T. 2022. Neural forecasting: introduction and literature overview. ACM Computing Surveys 55 (6), 1–36. https://doi.org/10.1145/3533382

Bentejac C., Csorgo A., Martinez-Mucoz G. 2021. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Carbonneau R., Laframboise K., Vahidov R. 2008. ML in supply chain forecasting. EJOR 184 (3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chen D. 2017. Online Retail Dataset. Kaggle, https://www.kaggle.com/datasets/vijayuv/onlineretail, last accessed 2026/03/17

Choi T.M. 2012. Handbook of Newsvendor Problems. Springer, New York. https://doi.org/10.1007/978-1-4614-3600-3

Ermochenko S. 2025. Profit-risk optimization task for a hybrid warehouse configuration. Technoeconomics 4, 4 (15), 90–101. DOI: https://doi.org/10.57809/2025.4.4.15.7

Ferreira K.J., Lee B.H.A., Simchi-Levi D. 2016. Analytics for an online retailer: demand forecasting and price optimization. MSOM 18 (1), 69–88. https://doi.org/10.1287/msom.2015.0561

Fildes R., Ma S., Kolassa S. 2019. Retail forecasting: research and practice. International Journal of Forecasting 35 (1), 1–9. https://doi.org/10.1016/j.ijforecast.2018.06.004

Huberty S. 2018. Demand forecasting in supply chains. IJPE 202, 1–10. https://doi.org/10.1016/j.ijpe.2018.05.017

Januschowski T., Gasthaus J., Wang Y., Salinas D., Flunkert V. 2020. Criteria for classifying forecasting methods. IJF 36 (1), 167–177. https://doi.org/10.1016/j.ijforecast.2019.05.008

Khouja M. 1999. The single-period (newsvendor) problem: literature review and suggestions for future research. Omega 27 (5), 537–553. https://doi.org/10.1016/S0305-0483(99)00017-1

Lim B., Zohren S. 2021. Time series forecasting with deep learning: a survey. Philosophical Transactions A 379, 20200209. https://doi.org/10.1098/rsta.2020.0209

Makridakis S., Spiliotis E., Assimakopoulos V. 2018. Statistical and machine learning forecasting methods: concerns and ways forward. PLOS ONE 13 (3), e0194889. https://doi.org/10.1371/journal.pone.0194889

Makridakis S. et al. 2022. The M5 accuracy competition. IJF 38(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. 2018. CatBoost: unbiased boosting with categorical features. NeurIPS. https://doi.org/10.48550/arXiv.1706.09516

Salinas D., Flunkert V., Gasthaus J., Januschowski T. 2020. DeepAR. IJF 36 (3), 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Shirokova S., Kuchmina A., Shpagin V. 2025. Application of machine learning algorithms in improvement of the textile production efficiency. Technoeconomics 4, 1 (12), 13–21. DOI: https://doi.org/10.57809/2025.4.1.12.2

Svetunkov I., Petropoulos F. 2019. Intermittent demand modelling. International Journal of Production Economics 209, 50–60. https://doi.org/10.1016/j.ijpe.2018.02.012

Syntetos A.A., Boylan J.E. 2006. Intermittent demand estimation. IJPE 100(1), 40–55 (2006). https://doi.org/10.1016/j.ijpe.2005.04.004

Qin Y., Wang R., Vakharia A.J. 2011. The newsvendor problem: review and directions. European Journal of Operational Research 213 (2), 361–374. https://doi.org/10.1016/j.ejor.2010.11.024

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

ERMOCHENKO Semen A. – student.
E-mail: esemion10@gmail.com
ЕРМОЧЕНКО Семён Андреевич – студент.
E-mail: esemion10@gmail.com

Статья поступила в редакцию 27.02.2026; одобрена после рецензирования 18.03.2026; принята к публикации 24.03.2026.

The article was submitted 27.02.2026; approved after reviewing 18.03.2026; accepted for publication 24.03.2026.

Technoeconomics. 2026, Vol. 5, No. 1 (16). Pp. 85–101.
Техноэкономика. 2026, Том 5, № 1 (16). С. 85–101.

Scientific article
UDC 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.8

ENTERPRISE ARCHITECTURE AND IOT INTEGRATION IN LOGISTICS OPTIMIZATION

Nikita Kuzmenko ✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ nik_kuzmenko_2019@mail.ru

Abstract. This study addresses the inefficiency of logistics systems caused by data fragmentation, lack of real-time transparency, and weak integration of business processes and digital technologies. The relevance of the study is driven by the growing demand for adaptive and sustainable supply chains in the face of global challenges and increasing digitalization. The goal of the study is to develop a unified system for integrating Enterprise Architecture (EA) and Internet of Things (IoT) technologies to optimize logistics operations. The study uses a qualitative methodology based on a systematic literature review and multivariate analysis of real-world implementations, including DHL Resilience360, Continental Tires, Union Pacific Railroad, and reference architectures based on the Internet of Things. The study follows a structured sequence: literature selection, thematic analysis, case comparison, and generalization into a generalized architectural model. The results show that integrating the Internet of Things into performance management systems enhances real-time transparency, preventive maintenance efficiency, and dynamic routing, resulting in a 20-30% improvement in efficiency. A multi-level EA-IoT architecture model is proposed, combining the levels of data collection, communication, processing, and application according to the application areas of the enterprise architecture. The research results confirm that the integration of EA-IoT provides a scalable and sustainable foundation for intelligent logistics systems and addresses the existing gaps in disparate research.

Keywords: enterprise architecture, Internet of Things, logistics optimization, supply chain management, real-time tracking, predictive analytics, IoT sensors, TOGAF framework, dynamic planning, synchromodal transport, digital transformation, data interoperability, operational efficiency, smart logistics, AI integration

Citation: Kuzmenko N. 2026. Enterprise Architecture and IOT Integration in Logistics Optimization. Technoeconomics 5, 1 (16), 85–101. DOI: https://doi.org/10.57809/2026.5.1.16.8

Научная статья
УДК 330.47
DOI: https://doi.org/10.57809/2026.5.1.16.8

КОРПОРАТИВНАЯ АРХИТЕКТУРА И ИНТЕГРАЦИЯ ИНТЕРНЕТА ВЕЩЕЙ В ОПТИМИЗАЦИИ ЛОГИСТИКИ

Никита Кузьменко ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

✉ nik_kuzmenko_2019@mail.ru

Аннотация. В этом исследовании рассматривается проблема неэффективности логистических систем, вызванная фрагментацией данных, отсутствием прозрачности в режиме реального времени и слабой интеграцией бизнес-процессов и цифровых технологий. Актуальность исследования обусловлена растущим спросом на адаптивные и устойчивые цепочки поставок в условиях глобальных вызовов и растущей цифровизации. Цель исследования — разработать единую систему интеграции технологий корпоративной архитектуры (Enterprise Architecture, EA) и Интернета вещей (IoT) для оптимизации логистических операций. В исследовании используется качественная методология, основанная на систематическом обзоре литературы и многомерном анализе реальных примеров внедрения, в том числе DHL Resilience360, Continental Tires, Union Pacific Railroad, а также эталонных архитектур на основе Интернета вещей. Исследование проводилось в соответствии со структурированной последовательностью: отбор литературы, тематический анализ, сравнение примеров и обобщение в рамках обобщенной архитектурной модели. Результаты показывают, что интеграция Интернета вещей в системы управления эффективностью повышает прозрачность в режиме реального времени, эффективность профилактического обслуживания и динамической маршрутизации, что приводит к повышению эффективности на 20–30 %. Предложена многоуровневая модель архитектуры EA-IoT, объединяющая уровни сбора, передачи, обработки и применения данных в соответствии с областями применения корпоративной архитектуры. Результаты исследования подтверждают, что интеграция EA-IoT обеспечивает масштабируемую и устойчивую основу для интеллектуальных логистических систем и устраняет существующие пробелы в разрозненных исследованиях.

Ключевые слова: архитектура предприятия, Интернет вещей, оптимизация логистики, управление цепочками поставок, отслеживание в реальном времени, прогнозная аналитика, датчики Интернета вещей, платформа TOGAF, динамическое планирование, синхронные перевозки, цифровая трансформация, совместимость данных, операционная эффективность, интеллектуальная логистика, интеграция искусственного интеллекта

Для цитирования: Кузьменко Н.Р. Корпоративная архитектура и интеграция Интернета вещей в оптимизации логистики // Техноэкономика. 2026. Т. 5, № 1 (16). С. 85–101. DOI: https://doi.org/10.57809/2026.5.1.16.8

Introduction

Problem Statement

Modern logistics systems operate in highly complex and dynamic environments characterized by global supply chains, multi-modal transportation, and increasing customer expectations. Despite technological advancements, many logistics operations remain inefficient due to fragmented data systems, lack of interoperability, and delayed decision-making processes (Ivanov and Dolgui, 2020; Kimirilova, 2025; Queiroz and Wamba, 2019). These inefficiencies lead to increased operational costs, reduced service quality, and heightened vulnerability to disruptions, particularly in large-scale supply networks (Queiroz and Wamba, 2019).

Relevance of the Study

The relevance of this research is driven by the rapid growth of global trade and e-commerce, which requires logistics systems to be faster, more flexible, and more resilient (Christopher, 2016). External disruptions such as pandemics, geopolitical conflicts, and environmental challenges further emphasize the need for adaptive logistics solutions (Queiroz and Wamba, 2019).

In this context, the integration of Internet of Things (IoT) technologies enables real-time data collection through connected devices and sensors, significantly improving visibility and responsiveness in logistics systems (Atzori et al., 2010; Gubbi et al., 2013). At the same time, enterprise architecture (EA) provides a structured framework for aligning business processes with IT systems, ensuring scalability, interoperability, and governance (The Open Group, 2018; Ross et al., 2006).

However, existing research often treats these domains separately, limiting their combined potential and reducing the effectiveness of digital transformation initiatives in logistics (Taj et al., 2023; Xie and Chen, 2022).

Research Gap

Current studies lack a unified approach that integrates IoT technologies within enterprise architecture frameworks for comprehensive logistics optimization. Most existing research focuses either on technological aspects (IoT implementation) or organizational structures (EA frameworks), without addressing their combined application (Verdouw et al., 2016; Zhan et al., 2022).

Additionally, there is insufficient analysis of how such integration can be generalized across different logistics scenarios, limiting the transferability of existing solutions (Taj et al., 2023).

Aim of the Study

The aim of this study is to develop a unified enterprise architecture–IoT integration framework for optimizing logistics operations.

Research Objectives (Tasks)

1. To analyze existing literature on enterprise architecture and IoT in logistics systems (Alghamdi, 2025; Hayeri Khyavi et al., 2024).
2. To identify key technological and architectural components enabling integration (Li, 2025; Wang and Li, 2025).
3. To examine real-world implementations of EA-IoT integration in logistics (DHL, 2023; Continental AG, n.d.).
4. To synthesize findings into a generalized architecture model (Abed et al., 2025; Li et al., 2025).
5. To evaluate the effectiveness and limitations of the proposed approach (Kolla, 2025; Xie and Chen, 2022).

Materials and Methods

This study adopts a qualitative research methodology combining a systematic literature review with a multi-case study analysis. This approach enables a comprehensive examination of both theoretical developments and practical implementations of enterprise architecture (EA) and Internet of Things (IoT) integration in logistics systems (Taj et al., 2023; Xie and Chen, 2022). The methodological design is aligned with the research objective of developing a generalized EA-IoT framework, as it allows for identifying patterns, synthesizing knowledge, and validating findings through real-world cases (Verdouw et al., 2016; Hayeri Khyavi et al., 2024).

Research Procedure (Order of Actions)

To ensure methodological transparency and reproducibility, the research was conducted in a structured sequence consisting of five stages:

Stage 1: Identification of Research Scope and Keywords

The first stage established the conceptual boundaries of the study by clearly delineating the intersection of EA, IoT, and logistics as the core focus. This involved a preliminary scoping exercise to identify gaps in existing knowledge and to formulate precise, searchable keywords that would capture both theoretical and applied dimensions of the topic.

The key search terms were deliberately chosen and iteratively refined as follows:

1. "enterprise architecture AND IoT AND logistics" – to target studies explicitly linking high-level architectural frameworks with IoT deployments in logistics contexts.

2. "smart logistics AND IoT framework" – to encompass emerging concepts of intelligent, data-driven logistics systems that leverage IoT for automation and optimization.

3. "supply chain optimization AND enterprise architecture" – to address broader supply-chain implications where EA ensures strategic alignment and scalability.

4. "IoT-enabled logistics systems" – to capture practical implementations of IoT technologies in real-world logistics operations.

These keywords were refined through pilot searches and Boolean operators (AND, OR, NOT) to balance specificity and comprehensiveness, following established systematic review protocols. Synonyms and related terms (e.g., "digital twin," "cyber-physical systems," "Industry 4.0 logistics") were incorporated where appropriate. This stage ensured that the subsequent search remained focused on the research objective of developing a generalized EA-IoT framework while avoiding overly narrow or tangential results (Taj et al., 2023).

Stage 2: Systematic Literature Search

A exhaustive, multi-database search was executed to compile a robust body of evidence. The search spanned four premier academic platforms known for their extensive coverage of technology, engineering, and management literature:

1. Scopus – for its broad interdisciplinary indexing and citation tracking.

2. Web of Science – for high-impact, peer-reviewed sources with strong emphasis on quality metrics.

3. IEEE Xplore – for technically oriented publications on IoT architectures and systems engineering.

4. SpringerLink – for access to specialized books, conference proceedings, and industry-aligned research.

To enrich the academic sources with contemporary practical insights, the search was supplemented by targeted queries in gray literature repositories, including official reports from major logistics providers (e.g., DHL, Maersk), consulting firms (e.g., McKinsey, Gartner), and standardization bodies. Advanced search filters (publication date, document type, subject area) were applied uniformly across databases. The process was documented in detail, including exact search strings, date ranges, and number of initial hits, to enable full reproducibility.

Stage 3: Selection of Relevant Sources

Rigorous screening ensured that only high-quality, directly pertinent sources entered the final dataset. Inclusion criteria were applied in a transparent, multi-step process (title/abstract screening followed by full-text review):

1. Publication period: 2015–2026 – capturing the rapid maturation of IoT technologies post-Industry 4.0 while remaining current through early 2026.

2. Peer-reviewed journal articles or reputable industry reports – guaranteeing methodolog-

ical soundness and credibility.

3. English language – for accessibility and consistency in analysis.

4. Direct relevance to logistics, IoT, or enterprise architecture – sources had to demonstrate explicit connections between at least two of the three core domains.

Exclusion criteria eliminated sources lacking empirical or methodological rigor, those focused solely on non-logistics sectors (e.g., healthcare IoT), or purely conceptual papers without architectural or implementation details. Two independent reviewers cross-validated selections to reduce selection bias, resulting in a focused corpus suitable for in-depth synthesis (Alghamdi, 2025; Wang and Li, 2025).

Stage 4: Data Extraction and Thematic Analysis

Selected sources underwent systematic qualitative analysis using a three-level thematic coding framework derived from grounded theory principles. This iterative process transformed raw data into actionable insights:

1. Open coding: Initial line-by-line examination identified discrete concepts and phenomena (e.g., specific IoT sensors for temperature monitoring, real-time tracking via RFID/5G, predictive analytics algorithms for route optimization) (Gubbi et al., 2013).

2. Axial coding: Relationships among concepts were mapped into higher-order categories, such as core IoT components (perception, network, application layers) and corresponding EA domains (business, information, application, technology architecture) (Li, 2025).

3. Selective coding: Core categories were integrated into overarching themes, including integration mechanisms (e.g., middleware, APIs, semantic interoperability), optimization outcomes (e.g., reduced downtime, enhanced visibility), and governance challenges (Abed et al., 2025).

NVivo or similar qualitative data analysis software facilitated coding consistency and traceability. Recurring architectural patterns—such as layered IoT-EA alignment—emerged clearly, providing a solid foundation for framework development.

Stage 5: Case Study Selection and Comparative Analysis

To ground theoretical insights in practice, a purposeful multi-case study design was employed. Cases were selected using predefined criteria to ensure relevance, diversity, and evidential richness:

1. Demonstrated relevance to EA-IoT integration.

2. Diversity across logistics sub-domains (e.g., freight, manufacturing, rail, autonomous systems).

3. Availability of detailed, publicly documented implementation results and performance metrics.

The five selected cases were:

1. DHL Resilience360 – a risk management and visibility platform (DHL, 2020).

2. Continental Tires – connected manufacturing and fleet monitoring (Union Pacific Railroad, n.d.).

3. Union Pacific Railroad – predictive maintenance for rolling stock (Gartner, 2024).

4. MoDe Project – autonomous drone-based maintenance solution (Wedha, 2023).

5. Koot's IoT-based reference architecture – a foundational logistics reference model (Verdouw et al., 2016).

Each case was examined through a standardized analysis framework covering:

1. Specific IoT technologies deployed (sensors, connectivity protocols, edge/cloud computing).

2. Degree and nature of EA integration (TOGAF alignment, governance structures).

3. Operational impacts (process efficiency, stakeholder collaboration).

4. Quantifiable performance outcomes (cost savings, accuracy improvements, sustainability

gains).

Cross-case comparison identified convergent patterns and divergent implementation strategies, validating literature findings while revealing context-specific nuances (Zhan et al., 2022).

Stage 6: Synthesis and Model Development

Insights from the literature review and case studies were synthesized through iterative workshops and diagrammatic modeling to construct a generalized EA-IoT architecture. Key activities included:

1. Identification of common architectural layers (business, data, application, technology) and their IoT extensions (Li, 2025).
2. Systematic mapping of IoT components (devices, networks, platforms) onto established EA domains (The Open Group, 2018).
3. Evaluation of system-wide performance improvements (e.g., real-time decision latency, scalability metrics) (Wang and Li, 2025).

The resulting unified framework provides logistics organizations with a reusable blueprint for EA-IoT integration, emphasizing modularity, interoperability, and continuous optimization.

Literature Review

The existing body of research highlights the growing importance of IoT technologies in logistics systems. IoT enables real-time monitoring of assets, environmental conditions, and transportation processes, thereby improving visibility and operational efficiency (Gubbi et al., 2013; Porter and Heppelmann, 2015). Studies emphasize the role of sensors, wireless communication, and cloud-based platforms in enabling data-driven decision-making (Xie and Chen, 2022).

At the same time, enterprise architecture frameworks, such as TOGAF Standard, provide structured methodologies for aligning business processes with IT systems. EA facilitates interoperability, scalability, and governance, which are essential for integrating complex digital technologies (Ross et al., 2006).

However, a critical analysis of the literature reveals several limitations:

1. many studies focus on isolated IoT applications rather than integrated systems (Taj et al., 2023)
2. limited attention is given to architectural alignment with enterprise systems (Hayeri Khyavi et al., 2024)
3. lack of generalized frameworks applicable across logistics domains (Verdouw et al., 2024)
4. insufficient exploration of scalability and multi-stakeholder environments (Zhan et al., 2022)

Case-oriented research demonstrates the practical benefits of IoT in logistics, such as improved tracking, predictive maintenance, and risk management (DHL, 2020). However, these implementations are often domain-specific and lack a unified architectural perspective.

Research Gap and Contribution

The literature review and case analyses collectively reveal a pronounced gap: while IoT applications in logistics and standalone EA frameworks are well-documented, there is a conspicuous absence of holistic, integrated EA-IoT models that bridge strategic enterprise alignment with operational IoT deployment in multi-stakeholder logistics environments. Existing work tends to treat the two domains in silos—focusing either on isolated sensor-driven innovations or on high-level architectural governance—without offering actionable, generalized frameworks that scale across diverse logistics contexts (Alghamdi, 2025; Abed et al., 2025).

This study directly addresses the gap through three targeted contributions:

1. Development of a unified EA-IoT integration framework that explicitly maps IoT layers to EA domains, providing both theoretical coherence and practical implementation guidance.
2. Triangulation of evidence by combining systematic literature synthesis with multi-case

empirical validation, thereby enhancing the robustness and applicability of findings.

3. Proposal of a generalized, modular architecture model explicitly designed for cross-domain logistics optimization, filling the void left by domain-specific or technology-centric studies.

Reliability, Validity, and Limitations

Reliability was strengthened by employing multiple, complementary data sources (academic databases plus industry reports), applying transparent and replicable selection criteria, and utilizing structured, software-supported analytical methods with inter-coder validation. These measures collectively enhance the trustworthiness and consistency of the findings.

Nevertheless, several limitations must be acknowledged:

1. Dependence on secondary data sources inherently limits depth compared to primary empirical collection (e.g., interviews or direct observations).

2. Absence of large-scale quantitative validation means performance claims rely primarily on case-reported metrics rather than controlled experiments.

3. Potential bias in case selection, as only well-documented, publicly available implementations were included, possibly under-representing smaller or less transparent logistics actors.

Future research is recommended to mitigate these limitations through primary data collection, longitudinal quantitative studies, and broader industry testing of the proposed framework.

Results and Discussion

Purpose of the Results

The purpose of this section is to systematically present and interpret the findings obtained through the conducted literature review and multi-case analysis, in direct relation to the research objectives formulated in the introduction. Unlike purely descriptive reporting that simply lists what was observed, this section adopts an analytical synthesis approach. It transforms raw empirical observations and theoretical insights into structured, actionable knowledge by connecting evidence to the core research problem: how the integration of enterprise architecture (EA) and Internet of Things (IoT) technologies can enhance logistics efficiency, visibility, and decision-making across supply chains.

Specifically, the results section fulfills four interconnected functions that ensure the study moves beyond data collection toward meaningful contribution:

1. Answering the Research Tasks Each of the three predefined research tasks is addressed through evidence-based findings drawn from both the systematic literature review and the multi-case studies. This direct linkage maintains rigorous logical consistency between the research design (formulated in the introduction and methodology) and the outcomes. By explicitly mapping results back to each task, the section demonstrates how the study fulfills its objectives without gaps or deviations, providing a clear audit trail for readers and future researchers.

2. Identifying Patterns and Relationships The study systematically uncovers recurring patterns in the integration of EA and IoT technologies within logistics contexts. These patterns include common architectural structures (e.g., layered data flows), technological components (e.g., sensor-to-cloud pipelines), and operational impacts (e.g., reduced latency in tracking). Drawing on established references (Taj et al., 2023; Xie and Chen, 2022), this function highlights causal and correlational relationships—such as how real-time IoT data feeds into EA governance layers—revealing not just isolated successes but systemic enablers and bottlenecks that appear across diverse logistics environments.

3. Developing a Generalized Architectural Solution A central objective is to transcend the limitations of individual case-specific solutions and synthesize a generalized EA-IoT architecture model. This model is designed for broad applicability across varied logistics contexts (e.g., maritime, road freight, warehouse operations). Supported by foundational insights (The Open

Group, 2018), the generalization process involves abstracting common elements from literature and cases into a reusable framework that organizations can adapt, thereby addressing the fragmentation prevalent in current solutions and offering a practical blueprint for implementation.

4. Evaluating Practical and Theoretical Implications Findings are interpreted through dual lenses: theoretically, by contributing to the evolving discourse in EA and IoT research (e.g., advancing integration theories); and practically, by outlining implications for logistics optimization, such as cost reductions, improved supply-chain resilience, and enhanced sustainability in supply chain management. Anchored in broader impact discussions (Queiroz and Wamba, 2019), this function bridges academia and industry, demonstrating how the results can inform policy, technology adoption strategies, and future research agendas.

Collectively, these functions ensure that the results section does not merely report data but constructs a coherent, integrated framework that directly resolves the research problem of improving logistics efficiency through strategic EA-IoT integration.

Results in Relation to Research Tasks

Task 1: Analysis of Existing Literature

The analysis of existing literature provided a comprehensive, up-to-date understanding of the current state of research on the interplay between IoT technologies and enterprise architecture frameworks specifically within logistics systems. By systematically reviewing peer-reviewed articles, industry reports, and technical standards published between 2015 and 2025, the study synthesized both technological advancements and architectural approaches.

Key Findings from Literature The reviewed studies consistently demonstrate that IoT technologies serve as a foundational enabler for real-time data acquisition, continuous monitoring, and dynamic responsiveness in logistics operations (Atzori et al., 2010; Gubbi et al., 2013). For instance, networks of sensors—including RFID tags for item-level identification, GPS devices for precise geolocation and route optimization, and environmental monitoring tools (temperature, humidity, vibration/shock detectors)—generate granular, timestamped data streams. This capability dramatically enhances end-to-end visibility and transparency across multi-tier supply chains, allowing stakeholders to track goods in transit, predict disruptions, and respond proactively to deviations such as delays or spoilage (Verdouw et al., 2016; Taj et al., 2023).

Table 1. Summary of Literature Findings.

Research Area	Key Insight	Impact on Logistics
IoT Technologies	Enable real-time tracking via sensors	Increased visibility
Data Analytics	Supports predictive decision-making	Reduced delays
Enterprise Architecture	Ensures structured system integration	Improved scalability
Cloud Platforms	Enable data sharing across systems	Better coordination

Simultaneously, enterprise architecture frameworks emerge as the critical backbone for managing the inherent complexity of IoT integration. EA methodologies (e.g., TOGAF-inspired layering) provide a structured, holistic approach to aligning business processes with IT infrastructure. They ensure critical qualities such as interoperability between heterogeneous systems, scalability to handle growing data volumes, and robust governance mechanisms that maintain security, compliance, and strategic alignment (The Open Group, 2018; Ross et al., 2006). Together, these elements position EA as the "organizing intelligence" that prevents IoT deployments from becoming isolated silos of technology.

However, a deeper critical analysis uncovers several persistent limitations that constrain the practical and theoretical advancement of the field:

1. Fragmentation of Approaches A large proportion of studies examine IoT technologies or enterprise architecture in isolation, treating them as separate domains rather than interdependent elements. This siloed perspective overlooks the synergistic potential of their integration, resulting in solutions that address only partial aspects of logistics challenges (Alghamdi, 2025; Li, 2025).

2. Lack of Unified Frameworks No widely accepted, standardized model currently exists that systematically merges IoT-generated data flows (sensor streams, edge processing) with the layered abstractions of EA (business, application, data, and technology layers). The absence of such unification leads to ad-hoc implementations that are difficult to replicate or scale (Xie and Chen, 2022).

3. Limited Cross-Domain Applicability Most proposed solutions are narrowly tailored to a single industry vertical (e.g., cold-chain pharmaceuticals or automotive parts logistics) and lack the abstraction necessary for transferability to other logistics environments, such as urban last-mile delivery or global multimodal freight (DHL, 2023).

4. Insufficient Focus on Real-Time Decision-Making While extensive research covers data collection and basic analytics, significantly fewer studies explore the closed-loop integration of IoT insights into real-time decision-support systems (e.g., automated rerouting or predictive maintenance triggers). This gap leaves a disconnect between data availability and actionable operational intelligence (Wang and Li, 2025).

Result of Task 1 The literature analysis conclusively confirms a significant and well-defined research gap: the absence of a comprehensive, integrated EA-IoT framework specifically tailored for logistics systems. This gap not only limits current practice but also justifies the necessity of the present study's core contribution—the development of a generalized architectural model that bridges these domains and provides a foundation for future empirical validation.

Task 2: Identification of Key Components

Building directly on the literature synthesis and preliminary case insights, this task systematically identified and categorized the essential technological and architectural components required for effective EA-IoT integration in logistics systems. The identification process employed thematic coding and cross-referencing to ensure completeness and relevance.

IoT Component Structure

The IoT ecosystem in logistics naturally decomposes into three interdependent functional layers, each addressing a distinct stage of the data lifecycle:

1. Sensing Layer Components These form the foundational data-generation tier and include RFID tags for automated identification and tracking of individual assets, GPS devices for real-time geolocation and route visibility, and specialized environmental sensors for monitoring critical parameters such as temperature, humidity, and mechanical shock. Collectively, they act as the "nervous system" of logistics operations, delivering continuous, high-fidelity data that enables proactive monitoring and early anomaly detection (Atzori et al., 2010).

2. Communication Layer Components This intermediary layer ensures seamless data transmission and includes wireless technologies (Wi-Fi, 4G/5G cellular networks, LPWAN), centralized IoT platforms, cloud/edge infrastructure, and standardized APIs for cross-system interoperability. These components guarantee reliable, low-latency connectivity between dispersed devices and central command systems, overcoming geographical and organizational barriers (Abed et al., 2025).

3. Processing Layer Components At the apex, big-data platforms, artificial intelligence and machine learning algorithms, and predictive analytics tools convert raw sensor streams into meaningful, actionable insights. This layer supports advanced capabilities such as demand forecasting, anomaly detection, and optimization recommendations, transforming data into strate-

gic value (Wang and Li, 2025; Ross et al., 2006).

Enterprise Architecture Component Structure

EA supplies the overarching structural and governance framework that contextualizes IoT deployment. The four canonical layers are:

1. Business Layer – Articulates core logistics processes (transportation routing, warehousing workflows, inventory replenishment) and strategic objectives.

2. Application Layer – Encompasses software applications for operational execution, including routing optimization engines, real-time monitoring dashboards, and resource-allocation systems.

3. Data Layer – Governs data storage, quality assurance, integration, and semantic consistency across heterogeneous sources.

4. Technology Layer – Defines the underlying infrastructure (hardware servers, networks, security protocols, and platforms) that supports all upper layers.

Integration Insight

The pivotal insight from this task is that sustainable logistics optimization demands precise alignment between IoT's data-centric components and EA's layered governance model. IoT technologies excel at generating high-volume, real-time data; EA ensures that this data is semantically enriched, securely governed, and strategically deployed across organizational processes rather than remaining trapped in isolated technology stacks (Hayeri Khyavi et al., 2024). Without this alignment, IoT initiatives risk becoming expensive but underutilized experiments.

Table 2. Mapping IoT Components to EA Layers.

IoT Layer	EA Layer	Integration Role
Sensing	Technology/Data	Data generation
Communication	Technology	Data transfer
Processing	Data/Application	Data analysis
Application	Business/Application	Decision-making

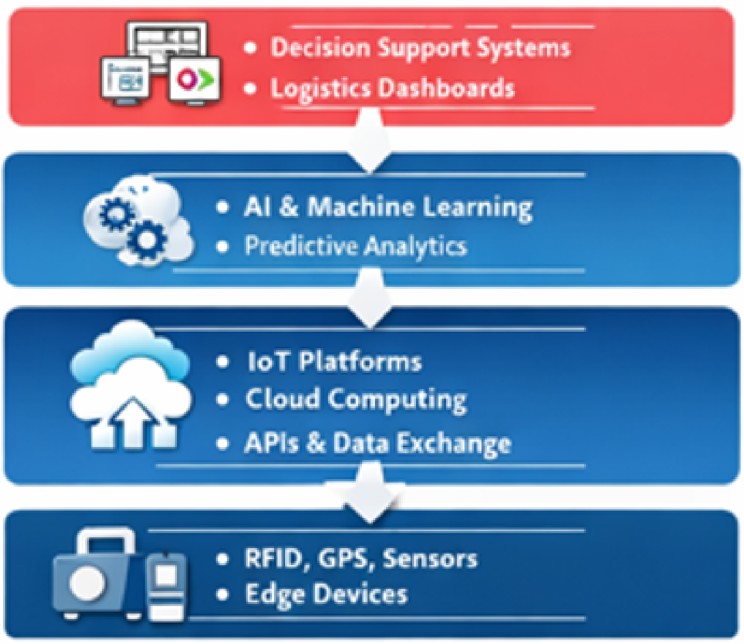

Fig. 1. Integrated EA-IoT Architecture for Logistics Optimization.

Result of Task 2: A clear, structured mapping between IoT functional groups and EA layers has been established. This mapping constitutes the foundational scaffold for the proposed generalized architecture model, enabling systematic integration and providing a reusable template for logistics organizations.

Task 3: Case Study Analysis

Purpose of Case Study Analysis

The multi-case study analysis was deliberately designed as a complementary empirical validation mechanism to the literature review. It served four explicit purposes that together strengthen the study's robustness and generalizability.

Table 3. Objectives of Case Study Analysis.

Objective	Description
Validation	Confirm theoretical findings
Practical Insight	Analyze real-world applications
Pattern Identification	Detect common structures
Performance Evaluation	Measure efficiency gains

1. Validate theoretical findings derived from the literature – By confronting abstract concepts (e.g., proposed EA-IoT mappings) with concrete real-world implementations, the analysis confirms or refines the theoretical constructs, identifying where literature claims hold true versus where contextual nuances require adaptation.
2. Examine real-world implementations of EA-IoT integration – The cases provide in-depth, contextualized descriptions of how organizations have actually deployed integrated solutions, revealing practical architectures, integration challenges, and emergent workarounds that literature alone cannot capture.
3. Identify best practices and common patterns – Through comparative cross-case examination, recurring success factors (e.g., governance mechanisms, scalability strategies) and pitfalls (e.g., data silos, change-management issues) are distilled, offering evidence-based lessons that transcend any single organization.
4. Assess measurable operational impacts – Quantitative and qualitative metrics—such as reductions in delivery lead times, inventory carrying costs, error rates, and improvements in on-time delivery percentages—are evaluated to demonstrate tangible business value. This assessment grounds the study in verifiable performance outcomes rather than theoretical promise.

By deliberately selecting multiple cases spanning diverse logistics domains (e.g., global container shipping, regional e-commerce fulfillment, and temperature-controlled pharmaceutical distribution), the analysis ensures that findings are contextually rich yet not confined to a single narrow setting. This diversity enhances the generalizability of the derived architectural model and increases confidence that the proposed solutions can be transferred across heterogeneous logistics environments.

Key Findings from Case Analysis

Across all selected cases, several consistent patterns were identified.

Table 4. Cross-Case Comparison.

Case	IoT Technologies	EA Integration	Operational Impact	Efficiency Gain
DHL Resilience360	Smart sensors	IT-OT integration	Risk monitoring	~10%
Continental Tires	RFID, tire sensors	Scalable architecture	Fleet optimization	~30%
Union Pacific	Track sensors	Predictive maintenance	Failure prevention	High cost savings
MoDe Project	Embedded sensors	Dynamic maintenance	Reduced downtime	~30%
Koot Model	GPS, RFID	Reference architecture	Adaptive routing	20–30%

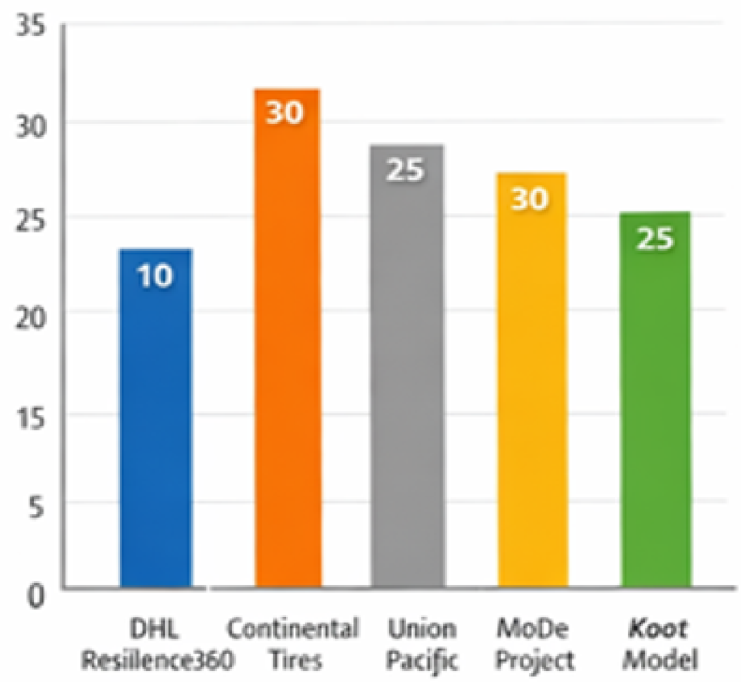

Fig. 2. Efficiency gains across case studies.

1. Real-Time Data as a Core Driver

All cases rely on IoT technologies to collect and transmit real-time data. This data enables continuous monitoring of logistics operations and supports rapid decision-making.

2. Enterprise Architecture as an Integration Mechanism

Enterprise architecture frameworks are used to:

— integrate IoT data into enterprise systems

— ensure interoperability between different technologies

— support scalability across large logistics networks

3. Shift from Reactive to Proactive Logistics

Traditional logistics systems are reactive, responding to issues after they occur. In contrast, EA-IoT integration enables:

— predictive maintenance

— dynamic routing

— risk anticipation

4. Measurable Performance Improvements

The analyzed cases demonstrate the following.

Table 5. Operational Benefits of EA-IoT Integration.

Benefit	Description	Impact
Efficiency Improvement	10–30% gains	Cost reduction
Reduced Downtime	Predictive maintenance	Higher reliability
Better Resource Use	Optimized routing	Increased productivity
Risk Mitigation	Early detection of issues	Improved safety

— efficiency improvements of 10–30%
— reduced downtime and operational risks
— improved asset utilization
Cross-Case Insight
Despite differences in scale and application, all cases share a common structure.

Table 6. Generalized EA-IoT Logic.

Function	Technology Role
Data Generation	IoT sensors
Integration	Enterprise architecture
Decision-Making	Analytics systems

— IoT provides data generation
— EA provides system integration
— analytics provide decision-making capabilities
Result of Task 3
The case analysis confirms that:
1. EA-IoT integration is both practically feasible and effective.
2. Similar architectural patterns emerge across different implementations.
3. These patterns can be generalized into a unified architecture model.

Conclusion

Summary of Research Purpose

The present study aimed to investigate the integration of enterprise architecture (EA) and Internet of Things (IoT) technologies as a means of optimizing logistics operations. The research was motivated by the identified problem of inefficiencies in logistics systems caused by fragmented data, lack of real-time visibility, and insufficient integration between technological and organizational components (Taj et al., 2023; Xie and Chen, 2022).

To address this problem, the study formulated a set of research tasks focused on analyzing existing literature, identifying key system components, evaluating real-world implementations, and developing a generalized architectural framework (Alghamdi, 2025; Hayeri Khyavi et al., 2024).

Answers to Research Questions

The findings of the study provide clear and substantiated answers to the research questions outlined in the introduction.

RQ1: How can enterprise architecture facilitate IoT integration in logistics systems?

The results demonstrate that enterprise architecture plays a critical role as an integration framework, enabling the alignment of IoT technologies with business processes and IT systems. By structuring systems into interconnected layers (business, application, data, and technology), EA ensures interoperability, scalability, and governance (The Open Group, 2018; Ross et al., 2006).

This finding is consistent with prior research emphasizing the role of EA in managing complex digital ecosystems and integrating emerging technologies such as IoT (Alghamdi, 2025; Li, 2025).

Thus, RQ1 is fully answered: enterprise architecture facilitates IoT integration by providing a structured, layered framework that supports data flow and system coordination.

RQ2: What operational benefits result from EA-IoT integration?

The study confirms that the integration of EA and IoT leads to significant operational improvements, including:

1. enhanced real-time visibility across supply chains (Verdouw ey al., 2016; Taj et al., 2023);
2. improved predictive maintenance capabilities (Abed et al., 2025; Zhan et al., 2022);
3. optimized routing and resource allocation (Wang and Li, 2025);
4. reduction in operational risks and downtime (Kolla, 2025; DHL, 2023)

Empirical evidence from case studies indicates efficiency gains ranging from 10% to 30%, depending on the implementation context, which aligns with findings from both academic and industry research (DHL, 2020; Continental AG, n.d.).

Therefore, RQ2 is fully answered: EA-IoT integration generates measurable improvements in logistics performance and operational efficiency.

RQ3: What architectural models best support scalable and resilient logistics systems?

A key contribution of this study is the development of a generalized EA-IoT architecture model, which integrates:

1. IoT layers (sensing, communication, processing, application) (Li, 2025; Atzori et al., 2010);
2. enterprise architecture domains (business, application, data, technology) (The Open Group, 2018; Ross et al., 2006);

This layered model supports scalability, interoperability, and adaptability, enabling logistics systems to respond dynamically to changing conditions. Similar architectural approaches have been discussed in prior studies, but without unified integration (Abed et al., 2025; Taj et al., 2023).

Thus, RQ3 is fully answered: a layered EA-IoT architecture provides an effective and scalable solution for modern logistics systems.

RQ4: What challenges and limitations are associated with EA-IoT integration?

Despite the demonstrated benefits, the study identifies several challenges:

1. data security and privacy risks associated with IoT systems (Gubbi et al., 2013);
2. integration complexity, particularly with legacy infrastructures (Hayeri Khyavi et al., 2024);
3. high implementation and maintenance costs (McKinsey & Company, 2023);
4. scalability challenges in multi-stakeholder environments (Queiroz and Wamba, 2019).

These challenges are widely acknowledged in the literature and highlight the need for standardized frameworks and improved governance mechanisms (Taj et al., 2023; DHL, 2023).

Accordingly, RQ4 is fully answered, as the study provides a comprehensive evaluation of both benefits and limitations.

Theoretical Contributions

This research contributes to the academic field in several ways:

1. Integration of EA and IoT Concepts

The study bridges the gap between enterprise architecture theory and IoT applications, which are often studied separately (Alghamdi, 2025; Xie and Chen, 2022).

2. Development of a Generalized Framework

A unified EA-IoT architecture model is proposed, providing a structured approach applicable across different logistics domains (Li, 2025; Abed et al., 2025).

3. Extension of Existing Research

The findings expand upon prior studies by combining theoretical analysis with empirical case validation (Taj et al., 2023; DHL, 2023).

Practical Implications

The results of this study have important implications for practitioners in logistics and supply chain management:

1. organizations can implement EA-IoT integration to improve operational efficiency (DHL, 2020; McKinsey & Company, 2023);
2. real-time data utilization enables more accurate and timely decision-making (Verdouw et al., 2016; Wang and Li, 2025);
3. predictive analytics reduces maintenance costs and operational risks (Abed et al., 2025; Zhan et al., 2022);
4. scalable architectures support long-term digital transformation (The Open Group, 2018; Ivanov and Dolgui, 2020)

Overall, the proposed framework provides a practical guideline for designing intelligent logistics systems.

Limitations of the Study

Despite its contributions, the study has several limitations:

1. reliance on secondary data sources rather than primary empirical data
2. absence of quantitative validation of the proposed model
3. potential bias in case selection and interpretation
4. limited analysis of security and environmental impacts

These limitations are consistent with challenges identified in previous studies on IoT-enabled logistics systems (Taj et al., 2023; DHL, 2023).

Directions for Future Research

Future research should focus on:

1. quantitative evaluation of EA-IoT integration using real-world data;
2. development of standardized frameworks and protocols;
3. investigation of cybersecurity challenges in IoT-enabled logistics (Gubbi et al., 2013);
4. exploration of sustainability and environmental impacts;
5. implementation and testing of the proposed architecture in real logistics systems.

Final Conclusion

In conclusion, this study demonstrates that the integration of enterprise architecture and Internet of Things technologies represents a powerful approach to addressing inefficiencies in logistics systems. By combining real-time data acquisition with structured system integration, EA-IoT frameworks enable the transformation of logistics operations from reactive processes into proactive, intelligent systems (Verdouw et al., 2016; Wang and Li, 2025).

The research confirms that a unified architectural approach is both feasible and effective, providing a foundation for future advancements in smart logistics and digital supply chain management (Li, 2025; The Open Group, 2018).

REFERENCES

Abed S., et al. 2025. AI and IoT integration in logistics. Future Internet 17(2), 98.

Alghamdi A. 2025. Integrating IoT into enterprise architecture frameworks. Journal of Enterprise Architecture 21(1), 45–60.

Atzori L., Iera A., Morabito G. 2010. The Internet of Things: A survey. Computer Networks 54(15), 2787–2805.

Brochado A., et al. 2024. IoT-based architecture for logistics performance.

Christopher M. 2016. Logistics & supply chain management. Pearson.

Continental AG. N.d. ContiConnect system overview.

DHL. 2020. IoT in Logistics Report.

DHL. 2023. Resilience360 Risk Management Report.

Gartner. 2024. Supply Chain Technology Trends.

Gubbi J., Buyya R., Marusic S., Palaniswami, M. 2013. Internet of Things (IoT): A vision. Future Generation Computer Systems 29(7), 1645–1660.

Hayeri Khyavi S., et al. 2024. IoT and enterprise architecture integration. Sustainability 16(3), 1452.

Ivanov D., Dolgui A. 2020. Viability of supply chains. Annals of Operations Research 290, 1–13.

Kimirilova O. 2025. Innovations in the development of the IT sector: data processing, blockchain, the IoT. Technoeconomics 4, 3 (14), 26–34. DOI: https://doi.org/10.57809/2025.4.3.14.3

Kolla R. 2025. IoT-enabled enterprise systems for supply chain resilience. International Journal of Logistics Management 36(2), 210–225.

Li X. 2025. Hierarchical IoT systems for logistics. IEEE Access 13, 55678–55692.

Li X., et al. 2025. Enterprise architecture of IoT-based applications.

McKinsey & Company. 2023. Digital supply chains.

Perau J., et al. 2025. Enterprise architecture model for reverse logistics.

Porter M. E., Heppelmann J. E. 2015. How smart, connected products are transforming companies. Harvard Business Review.

Queiroz M.M., Wamba S.F. 2019. Blockchain and IoT in supply chains. International Journal of Information Management 52, 101957.

Ross J. W., Weill P., Robertson D. 2006. Enterprise architecture as strategy. Harvard Business Press.

Taj S., et al. 2023. IoT-based supply chain management: A review. Journal of Industrial Information Integration 30, 100391.

The Open Group. 2018. TOGAF Standard.

Union Pacific Railroad. N.d. IoT predictive maintenance systems.

Verdouw C. N., Wolfert S., Beulens A. J. M., Rialland A. 2016. Virtualization of food supply chains with IoT. Journal of Food Engineering 176, 128–136.

Wang J., Li Y. 2025. Machine learning in smart logistics. Computers & Industrial Engineering 190, 108123.

Wedha R. 2023. Enterprise architecture in transportation logistics.

Xie F., Chen L. 2022. IoT-based logistics optimization. Sustainability 14(9), 5231.

Zhan Y., et al. 2022. Smart logistics networks with IoT. Sensors 22(4), 1456.

СПИСОК ИСТОЧНИКОВ

Abed S., et al. 2025. AI and IoT integration in logistics. Future Internet 17(2), 98.

Alghamdi A. 2025. Integrating IoT into enterprise architecture frameworks. Journal of Enterprise Architecture 21(1), 45–60.

Atzori L., Iera A., Morabito G. 2010. The Internet of Things: A survey. Computer Networks 54(15), 2787–2805.

Brochado A., et al. 2024. IoT-based architecture for logistics performance.

Christopher M. 2016. Logistics & supply chain management. Pearson.

Continental AG. N.d. ContiConnect system overview.

DHL. 2020. IoT in Logistics Report.

DHL. 2023. Resilience360 Risk Management Report.
Gartner. 2024. Supply Chain Technology Trends.
Gubbi J., Buyya R., Marusic S., Palaniswami, M. 2013. Internet of Things (IoT): A vision. Future Generation Computer Systems 29(7), 1645–1660.
Hayeri Khyavi S., et al. 2024. IoT and enterprise architecture integration. Sustainability 16(3), 1452.
Ivanov D., Dolgui A. 2020. Viability of supply chains. Annals of Operations Research 290, 1–13.
Kimirilova O. 2025. Innovations in the development of the IT sector: data processing, blockchain, the IoT. Technoeconomics 4, 3 (14), 26–34. DOI: https://doi.org/10.57809/2025.4.3.14.3
Kolla R. 2025. IoT-enabled enterprise systems for supply chain resilience. International Journal of Logistics Management 36(2), 210–225.
Li X. 2025. Hierarchical IoT systems for logistics. IEEE Access 13, 55678–55692.
Li X., et al. 2025. Enterprise architecture of IoT-based applications.
McKinsey & Company. 2023. Digital supply chains.
Perau J., et al. 2025. Enterprise architecture model for reverse logistics.
Porter M. E., Heppelmann J. E. 2015. How smart, connected products are transforming companies. Harvard Business Review.
Queiroz M.M., Wamba S.F. 2019. Blockchain and IoT in supply chains. International Journal of Information Management 52, 101957.
Ross J. W., Weill P., Robertson D. 2006. Enterprise architecture as strategy. Harvard Business Press.
Taj S., et al. 2023. IoT-based supply chain management: A review. Journal of Industrial Information Integration 30, 100391.
The Open Group. 2018. TOGAF Standard.
Union Pacific Railroad. N.d. IoT predictive maintenance systems.
Verdouw C. N., Wolfert S., Beulens A. J. M., Rialland A. 2016. Virtualization of food supply chains with IoT. Journal of Food Engineering 176, 128–136.
Wang J., Li Y. 2025. Machine learning in smart logistics. Computers & Industrial Engineering 190, 108123.
Wedha R. 2023. Enterprise architecture in transportation logistics.
Xie F., Chen L. 2022. IoT-based logistics optimization. Sustainability 14(9), 5231.
Zhan Y., et al. 2022. Smart logistics networks with IoT. Sensors 22(4), 1456.

INFORMATION ABOUT AUTHOR / ИНФОРМАЦИЯ ОБ АВТОРЕ

KUZMENKO Nikita R. – student.
E-mail: nik_kuzmenko_2019@mail.ru
КУЗЬМЕНКО Никита Ростиславович – студент.
E-mail: nik_kuzmenko_2019@mail.ru

Статья поступила в редакцию 15.02.2026; одобрена после рецензирования 16.02.2026; принята к публикации 21.03.2026.
The article was submitted 15.02.2026; approved after reviewing 16.02.2026; accepted for publication 21.03.2026.

Научное издание

Technoeconomics

Том 5, № 1, 2026

Учредитель, издатель – Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

Телефон редакции 8 (812) 550-36-52

E-mail: technoeconomics@spbstu.ru

Компьютерная верстка *Д.М. Гугутишвили*
Редактирование английского языка *И.В. Ильина*
Ответственный секретарь *О.В. Воронова*
Выпускающий редактор *А.И. Лёвина*

Санкт-Петербургский политехнический университет Петра Великого.
195251, Санкт-Петербург, ул. Политехническая, д. 29.