



NJD-iScience

№183

The Norwegian Journal of Development of
the International Science





№183/2026

Norwegian Journal of development of the International Science

ISSN 3453-9875

It was established in November 2016 with support from the Norwegian Academy of Science.

DESCRIPTION

The Norwegian Journal of Development of the International Science, established in 2016 with support from the Norwegian Academy of Science, offers a platform for global researchers, academics, and students. Publishing 24 times annually, it covers current science topics. Submitting your article provides exposure to a diverse audience, enhancing your career. Our commitment to research and SEO ensures timely visibility. Submit to the Norwegian Journal of Development of the International Science today and join esteemed authors.

Editor in chief – Karin Kristiansen (University of Oslo, Norway)

The assistant of the editor in chief – Olof Hansen

- James Smith (University of Birmingham, UK)
- Kristian Nilsen (University Centre in Svalbard, Norway)
- Arne Jensen (Norwegian University of Science and Technology, Norway)
- Sander Svein (University of Tromsø, Norway)
- Lena Meyer (University of Gothenburg, Sweden)
- Hans Rasmussen (University of Southern Denmark, Denmark)
- Chantal Girard (ESC Rennes School of Business, France)
- Ann Claes (University of Groningen, Netherlands)
- Ingrid Karlsen (University of Oslo, Norway)
- Terje Gruterson (Norwegian Institute of Public Health, Norway)
- Sander Langfjord (University Hospital, Norway)
- Fredrik Mardosas (Oslo and Akershus University College, Norway)
- Emil Berger (Ministry of Agriculture and Food, Norway)
- Sofie Olsen (BioFokus, Norway)
- Rolf Ulrich Becker (University of Duisburg-Essen, Germany)
- Lutz Jäncke (University of Zürich, Switzerland)
- Elizabeth Davies (University of Glasgow, UK)
- Chan Jiang (Peking University, China) and other independent experts

1000 copies

Norwegian Journal of development of the International Science

Iduns gate 4A, 0178, Oslo, Norway

email: publish@nor-ijournal.com

site: <https://nor-ijournal.com>

CONTENT

EARTH SCIENCES

Shevchenko I.

MAGNETIC STORMS, MUONS AND HUMAN
HEALTH.....3

ECONOMIC SCIENCES

Baylarov R., Abdullayev K.

DETERMINANTS OF THE SHADOW ECONOMY IN
SELECTED NORTHERN EUROPEAN COUNTRIES: A
MIMIC MODEL APPROACH.....6

Zhumadullaeva D.

TRANSITION FROM A PROFIT MAXIMIZATION MODEL
TO A SUSTAINABLE VALUE CREATION MODEL IN
CORPORATE MANAGEMENT OF STATE-OWNED
ENTERPRISES 18

HISTORICAL SCIENCES

Gasimova N., Heydarova S., Mammadov I.

THE ROLE OF CAMEL CARAVAN ROUTES IN THE
HISTORICAL SYSTEM OF TRADE AND CARGO
TRANSPORTATION OF AZERBAIJAN 23

MATHEMATICAL SCIENCES

**Berdenova B., Orazaly D., Nessipbekov G.,
Bidyut Baran Saha**

MULTIVARIABLE NONLINEAR REGRESSION FOR
ANALYSING ADSORPTION ISOTHERM MODELS - A
CASE STUDY 30

Karimova U., Akhundov R., Abdullayev K.

A NOVEL APPROACH TO APPLYING BAYES' THEOREM
IN MACHINE LEARNING AND PROGRAMMING 38

PEDAGOGICAL SCIENCES

Sagatova Sh., Urunbassarova E., Yunissov Y.

THE DEVELOPMENT OF DIALOGIC COMPETENCE
AMONG STUDENTS IN THE SYSTEM OF
MULTILINGUAL EDUCATION 46

Kolodko T., Kolodko M.

THE INFLUENCE OF EFFECTIVE LANGUAGE LEARNING
STRATEGIES IN SECOND LANGUAGE ACQUISITION ON
FOREIGN LANGUAGE TEACHING TO UNIVERSITY
STUDENTS.....54

PHILOLOGICAL SCIENCES

Guliyev B.

ARABIC CONJUGATIONS IN YUNUS EMRE'S "DİVAN"
(BASED ON THE KARAMAN COPY) 58

PHILOSOPHICAL SCIENCES

Skovronskyi B.

AUTOCOMMUNICATION AS A MODE OF ARTISTIC
SEMIOSIS: THE PAINTING AS A PERFORMATIVE
MODEL 60

PHYSICAL SCIENCES

Abbasov V.

DOUBLE SLIT EXPERIMENT AND WAVE-PARTICLE
DUALITY.....68

SOCIAL SCIENCES

**Shrestha Sofin, Sherpa Pemba, Saru Tilak,
Koirala Ashika**

SERVICE-LED TRANSFORMATION OR PREMATURE
DEINDUSTRIALIZATION? SECTORAL EMPLOYMENT
SHIFTS, STRUCTURAL CHANGE, AND NATIONAL
INCOME IN NEPAL, 2000-2022.....84

TECHNICAL SCIENCES

Shapoval S., Vodka R.

INTEGRATED SOLAR COLLECTORS AS AN ENERGY-
EFFICIENT SOURCE OF ENERGY 93

Kosovan V., Melnyk H., Melnyk V., Melnyk O.

AGENT-ORIENTED ARCHITECTURE FOR CITY
INFRASTRUCTURE ANALYSIS 98

EARTH SCIENCES

MAGNETIC STORMS, MUONS AND HUMAN HEALTH

Shevchenko I.

PhD, Senior Researcher, M. P. Semenenko Institute of Geochemistry Mineralogy and Ore Formation; Kyiv, 03142 Ukraine

<https://doi.org/10.5281/zenodo.21105807>

Abstract

It has recently been discovered that muons produced in the Earth's atmosphere by solar radiation have a strong dynamic influence on hydrolytic processes on Earth. Muons break hydrogen bonds between water molecules, destroy water clusters and increase their hydrolytic activity. This is very important for biological processes and human health as in the absence of muons in the atmosphere water clusters become highly structured and lose their hydrolytic activity. Since muons have an electric charge, their concentration and direction of movement in the Earth's atmosphere strongly depend on the geomagnetic field, the strength and direction of which changes greatly during magnetic storms.

Keywords: hydrolysis; solar influence; muons; magnetic storms; human health

Introduction

The impact of magnetic storms on human health is well known and has been studied in the context of many medical examples. Geomagnetic disturbances during magnetic storms have a particularly strong impact on the cardiovascular system leading to heart attacks and blood pressure fluctuations. The production of melatonin in the human brain is also very sensitive to geomagnetic disturbances which causes headaches and mental disorders, such as depression, sleep disturbances, irritability, decreased concentration and performance. Magnetic storms can also aggravate chronic diseases, as well as reduce reflexes and increase the risk of thrombosis.

However, the mechanism by which magnetic storms affect human health remains unknown, since direct exposure to even a strong magnetic field does not have negative impact on people.

It was logical to assume that unlike local magnets which have a relatively small radius of influence of the magnetic field, changes in the strength and direction of the geomagnetic field on the scale of planet Earth affect living organisms through some intermediate factor that directly influences biological processes.

Discussion

Our research indicates that this intermediate factor is muons, which are constantly produced in the Earth's atmosphere under the influence of the solar radiation.

Unlike the electrons and protons of the solar wind, muons have large penetrating ability. They not only reach the Earth but even penetrate some depth underground. Muons were found to break hydrogen bonds between water molecules, forming free hydroxyl groups. This disrupts water clusters and significantly increases their reactivity in hydrolytic processes. Therefore, with an increase in the formation of muons in the atmosphere under solar influence, hydrolytic reactions are greatly accelerated. This is important for biochemical processes, since in the absence of muons, water clusters are structured to maximum sizes and are deactivated, and hydrolytic reactions are significantly slowed down or stopped. For example, due to fluctuations in solar activity over the 11-year solar cycle and changes in solar influence as the Earth rotates around the Sun and around its axis, the same solution of water in acetonitrile can consist of a small number of large passive clusters, or a large number of small, very active clusters. Therefore, at different times of the year the difference in the rate of hydrolysis at a constant temperature can reach 200 times. This was confirmed by regular measurements of the rate of hydrolysis of triethyl phosphite $[(EtO)_3P]$ in acetonitrile over 6 years 2015-2020, including 5 years of the second half of the 24th solar cycle 2015-2019 and the first year 2020 of the 25th cycle. (Fig.1) [1]

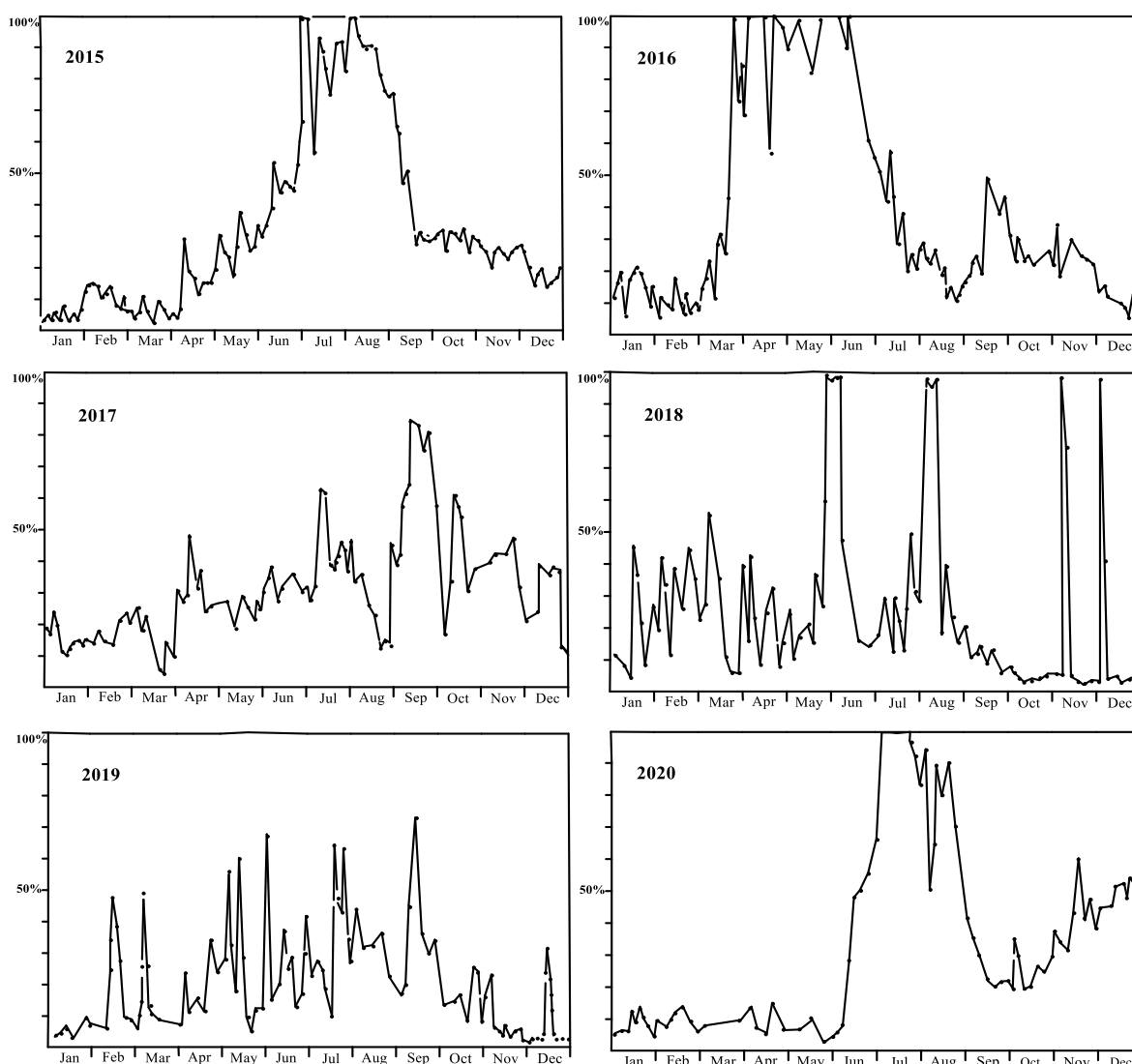


Fig. 1. Variations of the rate of hydrolysis of triethyl phosphite $[(EtO)_3P+H_2O]$ in acetonitrile in 2015-2020 (conversion after 25 minutes of heating at $80^\circ C$). This period includes the second half of the 24-th solar cycle 2015-2019 and the first year 2020 of the 25-th cycle.

By conducting these regular 6-year measurements we noticed that during periods of strong magnetic storms, the rate of triethyl phosphite hydrolysis decreases. This is explained by the fact that the components of the solar wind (electrons and protons) and the muons formed under their influence in the Earth's atmosphere have an electric charge. Therefore, magnetic storms strongly influence the formation of muons, as well as the direction and intensity of their movement in space. This determines the degree of destruction of water clusters by muons, their hydrolytic activity and the rate of hydrolytic processes. This has been well confirmed experimentally. For example, if a solution of water in acetonitrile is protected from muons with a special shield, the water clusters begin to structure to very large sizes and do not react with the added triethyl phosphite. And after removing the shield, hydrolysis immediately begins under the solar influence and accelerates very quickly. [2]

Water is a common structural component of all organs of the human body, so changes in its properties caused by fluctuations in muon influence can manifest themselves in different forms in different organs. This

is exactly what is observed in medical practice. The aggregation of water molecules into large passive clusters in the absence of muons in the atmosphere during geomagnetic storms explains the slowdown of biochemical processes and the increase in blood viscosity leading to thrombosis. This occurs when fluctuations in the interplanetary magnetic field significantly alter the direction of movement of muons in the Earth's magnetic field.

The cessation of muon formation in the atmosphere occurs not only during magnetic storms, but also at night, when the Sun irradiates the opposite side of the Earth. After sunset, the disruption of hydrogen bonds between water molecules stops and the size of water clusters increases again. Therefore, at night the rate of hydrolytic processes in living cells significantly slows down, which promotes the formation of blood clots. This explains why strokes typically occur in the morning, when water clusters are at their largest after a night of solar inactivity, and do not occur in the evening, when water molecules have not yet had time to combine after exposure to muons.

Conclusion

Under the influence of solar radiation, secondary particles muons are constantly formed in the upper layers of the Earth's atmosphere, which have a significant impact on hydrolytic processes on Earth. Muons destroy hydrogen bonds between water molecules in water clusters, forming free hydroxyl groups, which maintains the hydrolytic activity of the clusters at the level necessary for biochemical processes. Since muons have an electric charge, their concentration and the direction of their movement in the Earth's magnetic field are strongly dependent on fluctuations in the strength of the interplanetary magnetic field during magnetic storms. This has a significant impact on the self-organization of water molecules, hydrolytic processes and human

health. Thus, geomagnetic storms indirectly prove that the constant formation of muons in the Earth's atmosphere is a necessary condition for the existence of life on the planet.

References:

1. Shevchenko, I.V., Self-organizations of water molecules over 11-year solar cycle. *Journal of Molecular Liquids* **355**, 118866 (1 June 2022). DOI:10.1016/j.molliq.2022.118866
2. Shevchenko I.V., Influence of muons on hydrolytic processes on Earth. Explanation of the 'Water Memory' mystery. *Norwegian Journal of development of the International Science* **164**, 18-21 (2025). DOI:10.5281/zenodo.17121584

ECONOMIC SCIENCES

DETERMINANTS OF THE SHADOW ECONOMY IN SELECTED NORTHERN EUROPEAN COUNTRIES: A MIMIC MODEL APPROACH

Baylarov R.

Master in International Business

WSB University, Krakow Poland

Abdullayev K.

Doctor of Sciences in Economics, D.Sc. (Economics),

Institute of Economics of the Ministry of Science and Education of the Azerbaijan Republic,

Baku, Azerbaijan

<https://doi.org/10.5281/zenodo.21105809>

Abstract

The shadow economy remains an important challenge for both developed and developing economies, affecting tax revenues, labor market efficiency, and the effectiveness of public policies. This study investigates the determinants of the shadow economy in selected Northern European countries during the period 2013–2023. The analysis covers Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, and Sweden and employs the Multiple Indicators Multiple Causes (MIMIC) model to estimate a latent shadow economy index. Using annual panel data obtained from international databases, the study examines the effects of unemployment, tax revenue, population growth, and economic growth on shadow economic activity. The results indicate that unemployment and population growth have a positive and statistically significant effect on the shadow economy, while tax revenue is negatively associated with informal economic activity. Economic growth does not exhibit a statistically significant impact. The estimated shadow economy index further reveals substantial cross-country differences. Ireland records the highest index values throughout most of the study period, whereas Denmark consistently exhibits the lowest levels. The findings suggest that labor market conditions, demographic dynamics, and institutional effectiveness play a crucial role in shaping shadow economic activity even in highly developed economies. The study contributes to the existing literature by providing a comparative analysis of Northern European countries and extending previous regional research on the determinants of the shadow economy in Europe.

Keywords: National economy, Shadow economy, investment, MIMIC model, Northern Europe, Economic determinants

Introduction

The shadow economy remains one of the most persistent and policy-relevant challenges facing contemporary economies. Broadly defined, it encompasses economic activities that take place outside the scope of official statistics, taxation systems, and government regulation, while nonetheless generating economic value. The prevalence of shadow economic activity has important implications for public revenue collection, the effectiveness of economic policy, labor market performance, and long-term economic development. High levels of informality may reduce tax revenues, distort competitive market conditions, and erode public trust in state institutions. Consequently, identifying the determinants of the shadow economy remains a critical objective for both policymakers and academic researchers.

Although scholarly interest in the shadow economy has expanded substantially over recent decades, the existing literature has predominantly focused on developing economies, large-scale cross-country analyses, or individual country case studies. Evidence from Europe suggests that both the magnitude and determinants of the shadow economy vary considerably across regions [2]. Factors such as taxation, unemployment, institutional quality, economic growth, and demographic change may exert different effects depending on the economic and institutional context. As a result, region-specific analyses are essential for developing a

more comprehensive understanding of the mechanisms underlying shadow economic activity [1].

The Nordic countries constitute a particularly valuable setting for such an investigation. Denmark, Finland, Sweden, Ireland, and the Baltic states are characterized by high income levels, strong institutional frameworks, transparent governance structures, and comprehensive social protection systems. Nevertheless, high institutional quality does not guarantee the complete elimination of informal economic activity. Previous research demonstrates that shadow economic activities persist even in highly developed economies, suggesting that informality is influenced not only by economic development but also by labor market conditions, fiscal institutions, and demographic processes.

In the authors' previous study, the determinants of the shadow economy in selected Southern European countries were examined using the MIMIC model. The findings indicated that unemployment contributed positively to shadow economic activity, whereas population growth exerted a negative effect. Significant cross-country differences were also identified. These results highlighted the importance of examining whether similar relationships hold in other European regions. The present study extends this research agenda by focusing on the Nordic countries and providing a comparative regional perspective [4].

The primary objective of this study is to identify the determinants of the shadow economy in selected

Nordic countries over the period 2013–2023 and to assess cross-country differences in the intensity of shadow economic activity. The analysis includes Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, and Sweden. The shadow economy is conceptualized as an unobservable latent variable and estimated using the Multiple Indicators Multiple Causes (MIMIC) model. Within this framework, unemployment, tax revenues, population growth, and economic growth are incorporated as the principal determinants of shadow economic activity.

This study contributes to the existing literature in several important respects. First, it provides a systematic assessment of shadow economy determinants across the Nordic region within a unified methodological framework. Second, it extends the findings of the author's previous Southern European study, thereby creating a foundation for comparative regional analysis within Europe. Third, it offers empirical evidence on the factors associated with shadow economic activity in countries characterized by high institutional quality and advanced socioeconomic development [5].

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature and identifies the research gap. Section 3 presents the data and methodological framework. Section 4 reports the empirical results. Section 5 discusses the findings and their policy implications. Finally, Section 6 summarizes the main conclusions, acknowledges the limitations of the study, and outlines directions for future research.

Literature review

Conceptual background of the shadow economy

The shadow economy is widely regarded as one of the most complex and pervasive phenomena in contemporary economies. It generally encompasses economic activities that generate value while remaining outside the scope of official statistics, taxation systems, and governmental regulation. Although no universally accepted definition exists, one of the most influential formulations was proposed by Schneider and Enste, who defined the shadow economy as the aggregate of legal value-generating activities that are not recorded within the national accounts system [20].

The literature identifies two principal theoretical perspectives on the emergence of the shadow economy. The first is grounded in rational choice theory, which posits that high tax burdens, administrative constraints, and regulatory compliance costs incentivize economic actors to engage in informal activities [3]. The second stems from institutional theory and emphasizes that the shadow economy is shaped not only by economic incentives but also by broader institutional factors, including trust in public institutions, the quality of governance, and the effectiveness of the legal system [16]. Baylarov et al., in their study, noted that applying the MIMIC model enables a detailed investigation of the current state of the shadow economy in selected Southern European countries [4].

Labor market conditions likewise play a significant role in shaping the shadow economy [17]. High unemployment rates and limited opportunities for for-

mal employment may drive individuals toward informal economic activities [18]. This tendency is particularly evident during periods of economic recession. Moreover, elevated levels of self-employment are frequently regarded as an indirect indicator of informal economic activity [7].

In recent years, increasing scholarly attention has been devoted to the relationship between demographic factors and the shadow economy. Population growth, migration flows, and age structure may affect the balance between formal and informal economic activities [8]. Nevertheless, research in this area remains limited, and the available evidence has yet to produce consistent conclusions. The Nordic countries offer a particularly valuable setting for investigating these dynamics [10]. Despite being characterized by strong institutional quality, efficient tax administration, and comprehensive social protection systems, the shadow economy has not been entirely eradicated within the region. Accordingly, examining the determinants of the shadow economy in the Nordic context is of substantial theoretical and practical relevance [11].

Empirical evidence from previous studies

The empirical literature on the shadow economy has expanded considerably over the past several decades. Early studies primarily concentrated on the effects of taxation and government regulation on informal economic activity. In his seminal analysis of the United States, Tanzi demonstrated that higher tax burdens contribute significantly to the growth of the shadow economy [21]. Subsequent research has provided broader international evidence. In a comprehensive study encompassing 162 countries, Schneider, Buehn, and Montenegro found that the size of the shadow economy is closely linked to tax burdens, unemployment levels, and institutional quality [20]. Similar conclusions were reached by Medina and Schneider in their extensive cross-country investigation [15]. Within the European context, empirical evidence highlights the critical role of institutional differences. Herwartz, Schneider, and Tafenau showed that regional disparities across European countries significantly affect the prevalence of shadow economic activity [9]. Likewise, Buehn and Schneider, applying the MIMIC model to Germany, identified labor market conditions and fiscal factors as the principal determinants of the shadow economy [6]. Research focusing on the Nordic countries indicates that the shadow economy persists despite high levels of institutional quality. Williams and Horodnic argued that tax morale, trust in public institutions, and social norms constitute key determinants of informal economic activity [22]. In the authors' previous study, the MIMIC model was employed to examine selected Southern European countries. The results indicated that unemployment exerted a positive effect on the shadow economy, whereas population growth had a negative effect. Furthermore, Greece and Italy exhibited relatively high latent shadow economy indices, while Cyprus and Malta recorded comparatively lower levels [4].

These findings emphasize the importance of cross-regional comparisons. Whereas labor market vulnerabilities and fiscal challenges appear to be the primary

determinants of the shadow economy in Southern Europe, the magnitude and direction of these effects may differ in the Nordic region. Therefore, a dedicated investigation of the Nordic countries has the potential to make a meaningful contribution to the existing literature.

Research gap and contribution

Although the existing literature offers substantial insights into the determinants of the shadow economy, several important gaps remain unresolved [19]. The first gap concerns regional specificity. The majority of existing studies are based either on large-scale cross-country analyses or on individual country case studies. Consequently, regions characterized by distinctive institutional environments, such as the Nordic countries, have received relatively limited scholarly attention in dedicated empirical investigations.

The second gap relates to the scarcity of comparative regional research. While previous studies have examined Southern European as well as Central and Eastern European countries, only a limited number of studies have analyzed the Nordic countries within a unified analytical framework using the MIMIC model. The third gap concerns the selection of determinants. Although the literature has traditionally focused on variables such as taxation and unemployment, the effects of energy consumption, economic growth, and demographic dynamics on the shadow economy remain underexplored. This study seeks to address these research gaps. Specifically, it presents a comparative regional analysis of Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, and Sweden and employs the MIMIC model to estimate the shadow economy as a latent variable. Moreover, by extending the findings of the author's previous research on Southern Europe, the study provides an opportunity to conduct a systematic comparison between the Nordic and Southern European regions. Such a comparison contributes to a more nuanced understanding of how the determinants of the shadow economy may vary across distinct institutional and socioeconomic contexts [24].

Data and methodology

Data description and sample selection

This study investigates the determinants of the shadow economy in selected Nordic countries over the period 2013–2023. The analysis covers Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, and Sweden. These countries were selected due to their high institutional quality, well-developed social protection systems, and stable economic structures. However, previous research indicates that, despite their advanced levels of economic and institutional development, informal economic activities have not been entirely eradicated.

The empirical analysis is based on a balanced panel dataset covering the period from 2013 to 2023 and comprising 77 observations. The data were obtained from reputable international sources, including the World Bank, Eurostat, the Organisation for Economic Co-operation and Development (OECD), and the International Energy Agency (IEA).

The selected time frame is particularly suitable for analyzing the dynamics of the shadow economy, as it

encompasses both the post-global financial crisis recovery period and the substantial economic disruptions associated with the COVID-19 pandemic. Consequently, the dataset captures significant variations in economic conditions that may influence the scale and determinants of informal economic activity. To address the inherently unobservable nature of the shadow economy, this study employs the MIMIC (Multiple Indicators Multiple Causes) model. The MIMIC framework allows the shadow economy to be estimated as a latent variable through the simultaneous incorporation of multiple causal factors and observable indicators, making it one of the most widely applied approaches in the empirical shadow economy literature.

Variable definitions

In the present study, the shadow economy is modeled as a latent variable that cannot be observed directly and must therefore be inferred from a set of observable causes and indicators within the MIMIC framework.

The causal variables included in the model are defined as follows:

- UNE (Unemployment Rate) – measured as the proportion of unemployed individuals within the working-age population. Higher levels of unemployment are expected to increase participation in informal economic activities, thereby contributing positively to the expansion of the shadow economy.

- TAX (Tax Revenue as a Percentage of GDP) – used as a proxy for the effective tax burden faced by economic agents. A higher tax burden may encourage individuals and businesses to shift economic activities from the formal to the informal sector in order to avoid taxation and regulatory compliance costs.

- POP (Population Growth Rate) – represents the annual rate of population growth. Changes in demographic conditions may affect labor market outcomes and alter the scale of informal economic activity.

- GDP_GROWTH (Annual GDP Growth Rate) – captures the overall performance of the economy. Stronger economic growth is generally expected to reduce the incentives for participation in shadow economic activities by improving employment opportunities and income prospects.

The model includes the following indicator variable:

- SELF (Self-Employment Rate) – measured as the proportion of self-employed individuals within total employment. The self-employment rate is widely recognized in the shadow economy literature as one of the most frequently employed indirect indicators of informal economic activity. Although self-employment does not necessarily imply participation in the shadow economy, elevated levels of self-employment are often associated with a greater likelihood of unreported or partially reported economic activities.

- ENERGY (Energy Consumption) – represents energy consumption and is used as an indicator variable within the MIMIC framework. Previous studies have suggested that energy consumption may capture economic activities that are not fully reflected in official statistics, making it a useful proxy indicator for latent shadow economic activity.

The MIMIC model specification

The MIMIC (Multiple Indicators Multiple Causes) model consists of two complementary components: a structural model and a measurement model. The structural model describes the relationship between the latent shadow economy variable and its underlying causal determinants, whereas the measurement model links the latent construct to observable indicators, thereby enabling its empirical estimation.

The structural component of the model is specified as follows:

$$\eta = \gamma_1 \text{UNE} + \gamma_2 \text{TAX} + \gamma_3 \text{POP} + \gamma_4 \text{GDP_GROWTH} + \zeta$$

where η denotes the latent variable representing the shadow economy; γ_1 , γ_2 , γ_3 , and γ_4 are the structural coefficients to be estimated; and ζ is the disturbance term capturing unexplained variation in the latent construct.

The measurement component is expressed as:

$$y = \lambda \eta + \varepsilon$$

where y represents the observable indicators associated with the shadow economy, λ denotes the factor loadings linking the latent variable to the observed indicators, and ε represents the measurement error terms.

Accordingly, the empirical MIMIC specification employed in this study is given by:

$$\begin{aligned} \text{SHADOW} &\sim \text{SELF} + \text{ENERGY} \\ \text{SHADOW} &\sim \text{UNE} + \text{TAX} + \text{POP} + \text{GDP_GROWTH} \end{aligned}$$

Under this framework, the shadow economy is treated as an unobservable latent variable determined by unemployment, tax revenue, population growth, and economic growth. The latent construct is measured through self-employment and energy consumption indicators. Self-employment is frequently used in the shadow economy literature as a manifestation of informal economic activity, while energy consumption serves as an additional indicator reflecting the intensity of economic activity that may not be fully captured by official statistics. This specification enables the simultaneous

estimation of both the determinants and manifestations of the shadow economy, representing one of the principal methodological advantages of the MIMIC approach in empirical shadow economy research [6].

Estimation procedure

The empirical analysis was conducted using the R statistical computing environment. Prior to model estimation, all variables were standardized to eliminate differences arising from measurement scales and to enhance the comparability of the estimated coefficients and model outputs.

The research was carried out in several sequential stages. First, descriptive statistical analyses were performed to summarize the distributional characteristics of the variables included in the study. Second, correlation analysis was undertaken to examine the strength and direction of the relationships among the variables and to identify potential multicollinearity issues. Third, the MIMIC model was estimated using the lavaan package in R through the Maximum Likelihood (ML) estimation procedure. Finally, the estimated latent shadow economy index was extracted and employed for comparative cross-country analysis across the selected Nordic economies.

The adequacy of the model was evaluated using a range of standard goodness-of-fit statistics. These included the Chi-square (χ^2) statistic, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Collectively, these indicators assess the extent to which the theoretical model reproduces the observed covariance matrix and therefore provide evidence regarding the overall validity and reliability of the specified model. Consistent with the structural equation modeling literature, acceptable model fit is generally indicated by higher CFI and TLI values and lower RMSEA and SRMR values.

Empirical findings

Descriptive statistics

Table 1.

Descriptive statistics of the variables used in the analysis					
VARIABLE	N	MEAN	SD	MIN	MAX
UNE	77	7.34	1.87	4.24	13.73
SELF	77	12.22	2.06	8.14	17.02
TAX	77	35.26	7.95	19.90	48.10
POP	77	0.26	0.51	-1.21	1.24
GDP_GROWTH	77	3.03	4.06	-5.04	25.18
ENERGY	77	19.74	4.58	12.80	31.18

Note: UNE denotes the unemployment rate (% of the labor force), SELF represents the self-employment rate (% of total employment), TAX refers to tax revenue (% of GDP), POP indicates annual population growth (%), GDP_GROWTH measures annual GDP growth (%), and ENERGY represents energy consumption. SD denotes the standard deviation.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data for the period 2013–2023 [23].

Table 1 presents descriptive statistics on the variables used in the study. The results show that the average unemployment rate was 7.34%, with a minimum of 4.24% and a maximum of 13.73%. The average self-employment rate was determined to be 12.22%, indicating that a certain part of the labor market in the selected Northern European countries operates outside of traditional wage employment.

The average tax revenue indicator was 35.26%, with significant variability observed between observations. The average population growth rate was 0.26%, with negative demographic dynamics also recorded in some countries. The average value for the economic growth indicator was 3.03%. At the same time, the dif-

ference between the minimum and maximum values reflects the cyclical nature of economic activity and, in particular, the effects of the pandemic period.

The average energy consumption was 19.74 units. The relatively high standard deviation indicates that energy use varies across countries. Overall, the observed

variation across variables suggests that there is sufficient statistical diversity to evaluate the MIMIC model.

Correlation analysis

Table 2.

Correlation matrix						
VARIABLE	UNE	SELF	TAX	POP	GDP_GROWTH	ENERGY
UNE	1.000	0.457	-0.163	-0.324	0.121	-0.347
SELF	0.457	1.000	-0.694	0.032	0.366	-0.424
TAX	-0.163	-0.694	1.000	0.133	-0.439	0.535
POP	-0.324	0.032	0.133	1.000	0.225	0.016
GDP_GROWTH	0.121	0.366	-0.439	0.225	1.000	-0.310
ENERGY	-0.347	-0.424	0.535	0.016	-0.310	1.000

Note: UNE denotes unemployment, SELF refers to self-employment, TAX represents tax revenue, POP indicates population growth, GDP_GROWTH denotes annual GDP growth, and ENERGY represents energy consumption. Correlation coefficients are reported using Pearson's correlation measure.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023) [23].

Table 2 presents the correlations between the variables. The results show that there is a positive relationship between the unemployment rate (UNE) and self-employment (SELF) ($r = 0.457$). This result can be explained by the fact that individuals seek alternative sources of income during periods of limited employment opportunities in the labor market. On the other hand, there is a negative correlation between unemployment and population growth ($r = -0.324$), which suggests that higher demographic dynamics may be associated with relatively more favorable labor market performance.

A strong negative correlation was found between self-employment and tax revenues ($r = -0.694$). This result supports the possibility that a certain proportion of

economic activity remains outside the formal sector in conditions of high self-employment. In addition, a moderate positive correlation was observed between tax revenues and energy consumption ($r = 0.535$), indicating that higher economic activity is accompanied by an increase in both fiscal revenues and energy use.

Although the correlation coefficients generally confirm the existence of some relationships between the variables, none of them reach a very high level. This indicates that there is no serious multicollinearity problem and that the simultaneous use of the variables in the MIMIC model is statistically acceptable.

MIMIC estimation results

Table 3.

MIMIC model estimation results			
PATH	STANDARDIZED COEFFICIENT	Z-VALUE	P-VALUE
SHADOW $\leftarrow$ UNE	0.442	6.204	<0.001
SHADOW $\leftarrow$ TAX	-0.681	-8.950	<0.001
SHADOW $\leftarrow$ POP	0.277	3.707	<0.001
SHADOW $\leftarrow$ GDP_GROWTH	-0.049	-0.625	0.532
SELF $\leftarrow$ SHADOW	1.000	—	—

Model fit statistics

INDICATOR	VALUE
CFI	1.000
TLI	1.000
RMSEA	0.000
SRMR	0.000
AIC	143.116
BIC	154.835
R ² (SHADOW)	0.666

Note: Standardized coefficients are reported. SHADOW denotes the latent shadow economy index.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023) [23].

Table 3 presents the MIMIC model estimation results for Northern European countries. The findings reveal a positive and statistically significant association

between the unemployment rate and the shadow economy indicator ($\beta = 0.442$, $p < 0.001$). This result suggests that increasing labor market instability may foster the expansion of informal economic activities.

A negative and highly statistically significant relationship is observed between the tax burden and the shadow economy ($\beta = -0.681$, $p < 0.001$). This finding is particularly noteworthy given the institutional characteristics of Northern European countries. Higher tax revenues are likely accompanied by more effective public institutions, robust regulatory and monitoring mechanisms, and stronger tax compliance, all of which contribute to curbing informal economic activity.

Population growth is found to have a positive effect on the shadow economy ($\beta = 0.277$, $p < 0.001$). This result may be explained by demographic dynamics

that increase the number of new entrants to the labor market and, in certain circumstances, encourage the expansion of informal employment. By contrast, the impact of economic growth is not statistically significant ($\beta = -0.049$, $p = 0.532$), suggesting that economic growth per se is not a primary determinant of changes in the shadow economy across the selected countries.

The model's coefficient of determination ($R^2 = 0.666$) indicates that approximately 66.6% of the variation in the latent shadow economy indicator is explained by the determinants incorporated into the model. Taken together, the findings suggest that the shadow economy in Northern European countries is predominantly influenced by labor market conditions, fiscal institutions, and demographic developments.

Country-level shadow economy index results

Table 4.

RANK	COUNTRY	MEAN SHADOW INDEX
1	Ireland	1.490
2	Finland	0.576
3	Latvia	0.564
4	Estonia	-0.019
5	Lithuania	-0.146
6	Sweden	-0.701
7	Denmark	-1.760

Note: Higher index values indicate a relatively greater estimated propensity toward shadow economic activity within the sample. The index is a relative measure derived from the MIMIC model and should not be interpreted as a direct percentage of GDP.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023) [23].

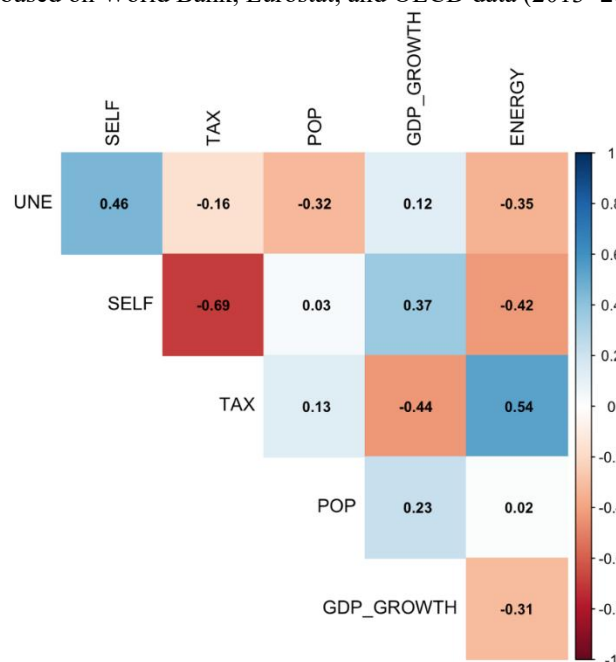


Figure 1. Estimated Shadow Economy Index Trends in Northern European Countries (2013–2023)

Note: The figure presents annual shadow economy index scores derived from the MIMIC model. Higher values indicate a relatively greater propensity toward shadow economic activity within the sample, while lower values indicate a relatively smaller propensity. The index represents a relative measure and should not be interpreted as a direct percentage of GDP.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023) [23].

Table 4 and Figure 1 present the average levels and temporal dynamics of the shadow economy index estimated using the MIMIC model across Northern European countries. The results indicate that Ireland exhibited the highest average index value (1.490) among the countries included in the sample. Finland (0.576) and Latvia (0.564) also recorded relatively high index levels. In contrast, Denmark (-1.760) and Sweden (-0.701) displayed the lowest values. When interpreting these findings, it is important to recognize that the MIMIC approach does not provide a direct measure of the absolute size of the shadow economy. Rather, it generates a latent index that captures the relative position of each country within the sample. Consequently, a higher index value implies that a country is more strongly affected by the determinants associated with shadow economic activity.

The temporal dynamics reveal several notable trends. In Ireland, the index declined gradually from 2.326 in 2013 to 0.905 in 2023. This downward trajectory may reflect improvements in labor market perfor-

mance, enhanced fiscal stability, and strengthened institutional oversight mechanisms. Although Finland maintained positive index values throughout the study period, some fluctuations were observed over time.

By contrast, the index remained in negative territory throughout the entire observation period in both Denmark and Sweden. In Denmark, the persistently low index values may be attributed to the country's high institutional quality, effective tax administration, and extensive social protection system, all of which help to constrain informal economic activity. Similar patterns are evident in Sweden.

Taken together, the results demonstrate that the relative prevalence of the shadow economy differs considerably across Northern European countries. These variations are largely associated with the effectiveness of fiscal institutions, labor market characteristics, and demographic developments. The findings are also consistent with the existing literature, which highlights the pivotal role of institutional factors in shaping the evolution of the shadow economy across the region.

Table 5.

Annual Shadow Economy Index by Country (2013–2023)

YEAR	DENMARK	ESTONIA	FINLAND	IRELAND	LATVIA	LITHUANIA	SWEDEN
2013	-1.495	0.131	0.663	2.326	0.668	-0.214	-0.399
2014	-1.571	0.086	0.399	2.223	0.692	-0.054	-0.522
2015	-1.622	0.060	1.000	2.058	0.779	0.027	-0.559
2016	-1.660	0.041	0.764	1.944	0.825	0.058	-0.689
2017	-1.877	0.071	0.386	1.562	0.485	-0.101	-0.753
2018	-1.975	-0.004	0.317	1.356	0.476	-0.266	-0.821
2019	-1.877	-0.075	0.466	1.057	0.383	-0.270	-0.758
2020	-1.789	-0.033	0.592	1.117	0.619	-0.221	-0.722
2021	-1.641	-0.105	0.826	0.944	0.437	-0.285	-0.761
2022	-1.915	-0.237	0.604	0.875	0.374	-0.137	-0.854
2023	-1.966	-0.141	0.319	0.905	0.471	-0.142	-0.877

Note: Values represent annual shadow economy index scores estimated using the MIMIC model. Higher values indicate a relatively greater propensity toward shadow economic activity within the sample. The index is a relative latent measure and should not be interpreted as a direct percentage of GDP.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023).

Discussion

Interpretation of determinant effects

The results of the MIMIC model indicate that unemployment, tax revenues, and population growth are statistically significant determinants of the shadow economy in the selected Nordic countries. These findings suggest that, despite the region's high institutional quality and advanced socioeconomic development, shadow economic activity remains responsive to a number of underlying structural factors.

Unemployment exerts a positive and statistically significant effect on the shadow economy ($\beta = 0.442$, $p < 0.001$). This finding implies that reduced opportunities within the formal labor market encourage individuals to seek alternative sources of income, thereby increasing participation in informal economic activities. The result is consistent with the theoretical expectations of the shadow economy literature and aligns with the findings of Schneider and Enste (2000) and Medina and Schneider (2019). Even in highly developed economies, labor market uncertainty may incentivize participation in informal economic arrangements as a means of income generation and risk mitigation.

Tax revenues were found to have a negative and highly statistically significant effect on the shadow economy ($\beta = -0.681$, $p < 0.001$). At first glance, this result appears counterintuitive, as higher taxation is often expected to encourage informal economic activity. However, within the Nordic institutional context, higher tax revenues may serve as an indicator of effective public administration, strong tax enforcement mechanisms, and high levels of tax compliance. Consequently, greater tax revenue collection is likely to reflect stronger institutional capacity and a reduced prevalence of shadow economic activity. This finding provides empirical support for institutional theory, highlighting the importance of effective governance and state capacity in constraining the shadow economy.

Population growth also exhibits a positive and statistically significant relationship with the shadow economy ($\beta = 0.277$, $p < 0.001$). This result suggests that demographic expansion may increase the number of individuals entering the labor market, thereby creating conditions under which informal employment opportunities can emerge or expand. In particular, migration inflows and the incomplete integration of newly arrived

populations into the formal economy may contribute to higher levels of informal economic activity. Nevertheless, the empirical evidence regarding this relationship remains inconclusive, and its magnitude and direction may vary across countries depending on institutional and labor market conditions.

In contrast, economic growth does not exert a statistically significant effect on the shadow economy ($\beta = -0.049$, $p = 0.532$). This finding indicates that variations in economic growth alone do not adequately explain changes in shadow economic activity within the selected Nordic countries. Rather, the results suggest that institutional quality, labor market conditions, and demographic dynamics play a more influential role in shaping the shadow economy than short-term fluctuations in economic performance.

The model demonstrates substantial explanatory power. The estimated coefficient of determination (R^2) equals 0.666, indicating that approximately 66.6% of the variation in the latent shadow economy variable is

explained by the determinants included in the model. Moreover, the goodness-of-fit statistics indicate an excellent model fit (CFI = 1.000, TLI = 1.000, RMSEA = 0.000, and SRMR = 0.000), suggesting a very high degree of consistency between the theoretical specification and the observed data.

The findings identify labor market conditions, the effectiveness of fiscal institutions, and demographic processes as the principal determinants of the shadow economy in the Nordic region. More importantly, they reinforce the argument that the shadow economy should be understood not merely as an economic phenomenon but as a multidimensional process shaped by the interaction of economic, institutional, and social factors. These results contribute to the growing body of evidence emphasizing the central role of institutional effectiveness in limiting informal economic activity, even in highly developed economies.

Cross-country comparative discussion

Shadow Economy Index Trends in Northern Europe (2013–2023)

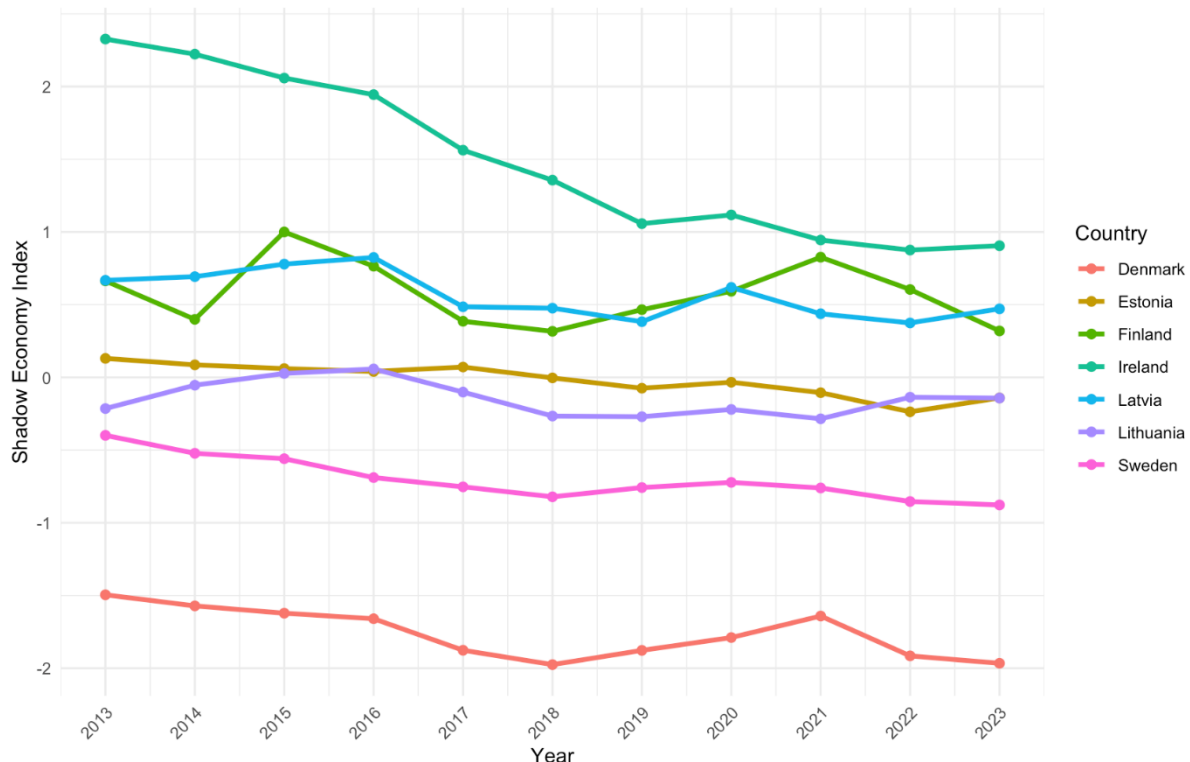


Figure 2. Shadow Economy Index Trends in Northern European Countries (2013–2023)

Note: The figure illustrates annual shadow economy index values estimated through the MIMIC model. Positive values indicate a relatively higher propensity toward shadow economic activity within the sample, whereas negative values indicate a relatively lower propensity.

Source: Authors' calculations based on World Bank, Eurostat, and OECD data (2013–2023) [23].

The results reveal considerable heterogeneity in the estimated latent shadow economy index across the selected Nordic countries. Finland also exhibited positive index values over the study period. The index reached its peak at approximately 1.0 in 2015 and subsequently followed a general downward trajectory, declining to around 0.3 by 2023. This pattern suggests that the influence of the underlying determinants associated with shadow economic activity weakened over time, resulting in a gradual reduction in the estimated latent index. Latvia displayed relatively stable positive index

values throughout the observation period. Although its trajectory occasionally mirrored that of Finland, the index generally fluctuated within a relatively narrow range of 0.4 to 0.8. Such stability may indicate that the structural determinants of shadow economic activity remained broadly unchanged during the study period. Estonia and Lithuania were characterized by index values clustered around zero. Both countries experienced only limited fluctuations, with the estimated indices remaining close to a neutral level for most of the period. This

similarity may reflect the comparable institutional environments, economic structures, and historical development paths shared by the Baltic economies. In contrast, Sweden and Denmark consistently recorded negative index values. Denmark exhibited the lowest estimated index among all countries in the sample, reaching approximately -2.0 in 2023. This finding suggests that Denmark had the lowest relative intensity of shadow economic activity during the study period. Sweden displayed a similar pattern, although its index values remained consistently higher than those observed for Denmark.

Ireland's relatively high latent index values may be attributable to the country's comparatively elevated levels of unemployment and self-employment, both of which are theoretically associated with greater participation in informal economic activities. Nevertheless, the latent nature of the MIMIC methodology must be carefully considered when interpreting these findings. The estimated index should not be understood as a direct measure of the size of the shadow economy. Rather, it reflects a relative latent position derived from the relationships between the selected causal determinants and observable indicators included in the model. Consequently, the index is more appropriately interpreted as a comparative measure of shadow economy intensity across countries rather than as an estimate of its absolute magnitude. Ireland consistently recorded

the highest shadow economy index throughout the study period. The index declined steadily from approximately 2.3 in 2013 to around 0.9 in 2023, indicating a gradual reduction in the relative intensity of shadow economic activity. This downward trend may reflect improvements in institutional effectiveness, regulatory enforcement, and overall economic conditions. Nevertheless, Ireland maintained the highest relative position among all countries in the sample throughout the entire observation period. Given the latent nature of the MIMIC index, this finding should be interpreted as a relative ranking rather than as evidence of the largest absolute shadow economy.

Overall, the figure demonstrates a broad decline in the estimated latent shadow economy index across most countries between 2013 and 2023. This trend may be associated with continued improvements in institutional quality, enhanced tax administration, and more efficient labor market performance across the region. At the same time, the persistence of cross-country differences indicates that national institutional arrangements, labor market structures, and demographic conditions continue to influence the development of shadow economic activity. These findings further support the argument that the shadow economy is shaped by a combination of economic, institutional, and social factors rather than by economic conditions alone.

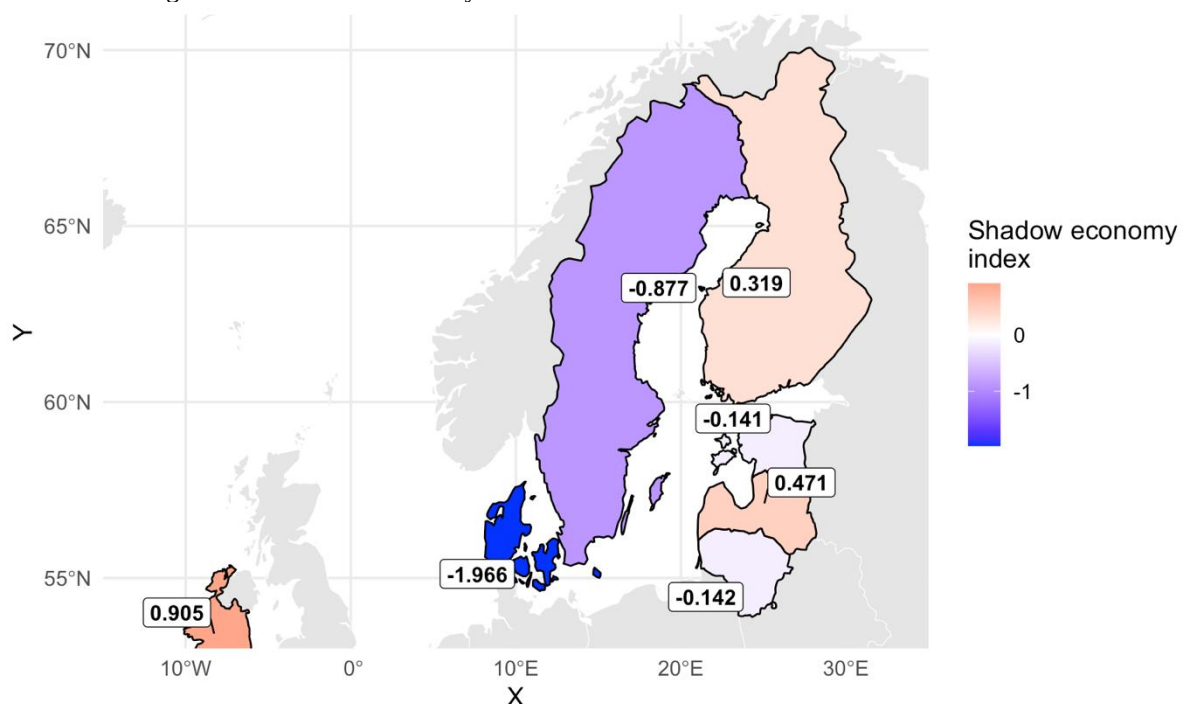


Figure 3. Estimated Shadow Economy in Northern Europe (2023)

Source: Authors' calculations based on World Bank, Eurostat, OECD, and International Energy Agency (IEA) databases (2013–2023).

Note: The map presents estimated shadow economy index values for selected Northern European countries in 2023 derived from the MIMIC model. Positive values indicate a relatively higher propensity toward shadow economic activity within the sample, whereas negative values indicate comparatively lower levels. The shadow economy index is a latent measure and should not be interpreted as the actual share of the shadow economy in GDP. Among the selected countries, Ireland exhibited the highest estimated shadow economy index (0.905), while Denmark recorded the lowest value (-1.966) in 2023.

Figure 3 presents the geographical distribution of the estimated latent shadow economy index across the selected Northern European countries in 2023. The spatial pattern reveals substantial cross-country variation in the relative intensity of shadow economic activity, highlighting the continued importance of national institutional and economic characteristics in shaping informal economic behavior. Ireland (0.905) recorded the highest latent shadow economy index among the countries included in the analysis. This finding is consistent with the temporal trends discussed earlier and indicates that Ireland maintained the highest relative position throughout most of the study period. Latvia (0.471) and Finland (0.319) also exhibited positive index values, suggesting that the determinants associated with shadow economic activity exert a comparatively stronger influence in these countries.

In contrast, Estonia (-0.141) and Lithuania (-0.142) reported index values close to zero, positioning them near the midpoint of the regional distribution. The similarity of these outcomes may reflect the comparable institutional environments and economic structures shared by the Baltic states. Sweden (-0.877) displayed a lower index value, indicating a comparatively weaker latent propensity toward shadow economic activity.

The lowest index value was observed in Denmark (-1.966). This result is consistent with Denmark's reputation for strong institutional quality, highly effective

tax administration, extensive digitalization of public services, and a well-developed social welfare system. These characteristics may contribute to reducing incentives for participation in informal economic activities and strengthening compliance within the formal economy. The map demonstrates that although the shadow economy appears to be relatively limited across Northern Europe, meaningful differences persist among countries. These differences suggest that institutional effectiveness, fiscal capacity, labor market conditions, and demographic factors continue to influence the relative intensity of shadow economic activity within the region.

Furthermore, the spatial distribution shown in Figure 3 corroborates the patterns identified in the preceding trend analysis. Countries such as Ireland and Finland remain associated with relatively higher latent index values, whereas Denmark and Sweden continue to occupy the lower end of the distribution. Importantly, these findings reinforce the argument that the shadow economy is not a homogeneous phenomenon within Northern Europe. Rather, its manifestation varies according to country-specific economic structures, institutional arrangements, and policy environments. As with all MIMIC-based estimates, the reported values should be interpreted as relative latent indices rather than direct measures of the actual size of the shadow economy.

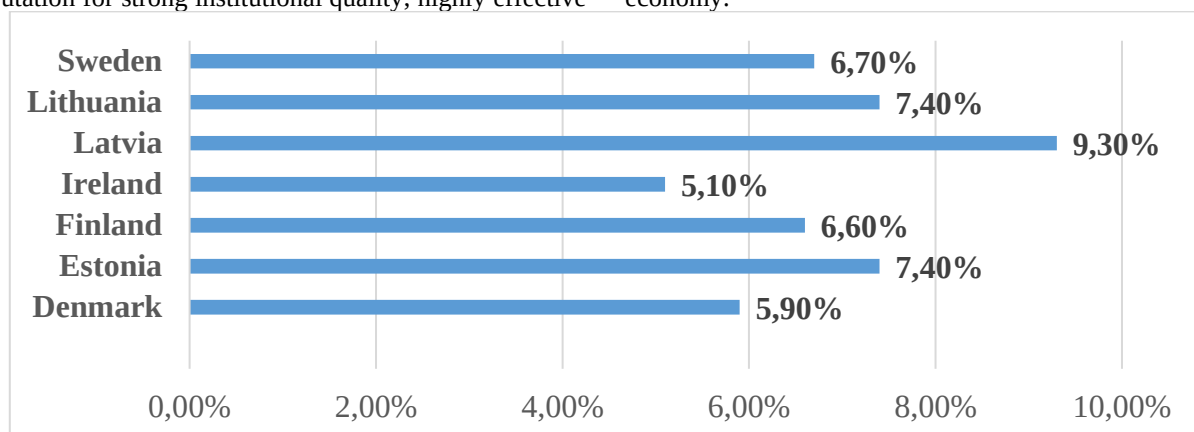


Figure 4. Shadow economy size (% of GDP) in selected Northern European countries (2023)

The conducted analysis shows that the share of the shadow economy in GDP is increasing in some Northern European countries. For example, this figure was 9.3% in Latvia, 7.4% in Lithuania, and 6.7% in Sweden. Furthermore, the share of the shadow economy accounted for 6.6% in Finland and 5.1% in Ireland. In general, it is essential to implement continuous measures at the state level to eliminate the shadow economy (Figure 4).

Source: <https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/insights/tax/documents/ey-gl-shadow-economy-report-02-2025.pdf>

Policy implications

The findings of this study generate several important policy implications for efforts aimed at reducing the shadow economy in the Nordic countries. First, the positive and statistically significant effect of unemployment on the shadow economy highlights the central role of labor market policies. Measures designed to expand formal employment opportunities, facilitate the integration of young people into the labor market, and reduce long-term unemployment are likely to contribute to a reduction in informal economic activity. By mitigating labor market uncertainty and improving em-

ployment prospects, such policies can reduce the incentives that drive individuals toward participation in the shadow economy [12].

Second, the negative relationship between tax revenues and the shadow economy underscores the importance of effective tax administration and institutional trust [13]. This finding suggests that the prevalence of informal economic activity depends less on the nominal level of taxation and more on the transparency, fairness, and efficiency of the tax system. Consequently, policies aimed at strengthening digital tax administration, expanding electronic reporting systems, improving taxpayer services, and enhancing regulatory

transparency may play a critical role in reducing the shadow economy. More broadly, the results reinforce the view that institutional effectiveness constitutes one of the most powerful mechanisms for limiting informality.

The positive effect of population growth further indicates that demographic dynamics should be incorporated into policy design [14]. Migration flows and the successful integration of new labor market entrants into the formal economy represent important policy challenges. In particular, improving access to legal employment opportunities, vocational training programs, and labor market integration initiatives may help reduce the likelihood that newly arrived or economically vulnerable groups engage in informal activities.

The findings also demonstrate that economic growth alone does not automatically lead to a reduction in the shadow economy. As a result, policy strategies that rely exclusively on economic expansion are unlikely to be sufficient. Sustainable reductions in shadow economic activity require economic growth to be accompanied by institutional reforms, active labor market policies, and continued improvements in tax administration and regulatory enforcement.

Overall, the results suggest that the most effective strategy for reducing the shadow economy in the Nordic region involves the simultaneous strengthening of labor market performance, institutional capacity, and tax compliance. Such an integrated policy approach can contribute not only to the containment of informal economic activity but also to greater economic transparency, improved fiscal sustainability, and enhanced public trust in state institutions. These findings support the broader institutional perspective that long-term reductions in the shadow economy are achieved not solely through economic development, but through the creation of effective, transparent, and trusted public institutions.

Conclusion

This study investigated the determinants of the shadow economy in selected Nordic countries over the period 2013–2023 using the MIMIC (Multiple Indicators Multiple Causes) model. The analysis covered Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, and Sweden, with the shadow economy conceptualized and estimated as a latent variable.

The empirical findings indicate that unemployment and population growth exert positive and statistically significant effects on the shadow economy. These results suggest that labor market vulnerabilities and demographic dynamics may create conditions conducive to the expansion of informal economic activity. By contrast, tax revenues were found to have a negative and statistically significant association with the shadow economy. This finding highlights the importance of effective tax administration, institutional capacity, and public trust in limiting informality. Economic growth, however, did not exhibit a statistically significant effect, suggesting that economic performance alone is insufficient to explain variations in shadow economic activity within the Nordic context.

The cross-country analysis revealed substantial heterogeneity among the selected countries. Ireland

consistently recorded the highest latent shadow economy index throughout most of the study period, whereas Denmark and Sweden exhibited comparatively low index values. The Baltic countries—Estonia, Latvia, and Lithuania—displayed relatively stable and moderate index levels. More broadly, the declining trend observed in the latent shadow economy index across most countries between 2013 and 2023 may reflect improvements in institutional quality, enhanced tax administration, and stronger labor market performance.

This study contributes to the existing literature in several important respects. First, it provides a systematic analysis of the determinants of the shadow economy in the Nordic region within a unified methodological framework. Second, it complements and extends the author's previous research on Southern Europe, thereby facilitating a comparative assessment of the factors shaping the shadow economy across different European regions. Third, it demonstrates that unemployment, fiscal variables, and demographic dynamics remain important determinants of shadow economic activity even in highly developed economies characterized by strong institutional environments.

Several limitations of the study should be acknowledged. The analysis is limited to seven Nordic countries, and the range of explanatory variables was constrained by data availability. Furthermore, because the MIMIC approach estimates the shadow economy as a latent index rather than as a direct measure, the results should be interpreted as relative indicators of shadow economy intensity and not as estimates of the actual share of the shadow economy in GDP.

Future research may benefit from extending the temporal coverage of the dataset and incorporating additional explanatory variables, such as institutional quality, corruption, digitalization, tax morale, and governance effectiveness. Moreover, comparative analyses integrating Northern, Southern, and Eastern European countries within a common modeling framework could provide deeper insights into the regional determinants and dynamics of the shadow economy.

In conclusion, although the shadow economy appears to be relatively limited in the Nordic region, labor market conditions, fiscal institutions, and demographic processes continue to play a significant role in shaping its development. Consequently, policies aimed at reducing the shadow economy should extend beyond the promotion of economic growth and place greater emphasis on strengthening institutional quality, improving governance effectiveness, and expanding opportunities for formal employment. Such an integrated policy approach is likely to be more effective in fostering economic transparency, enhancing fiscal sustainability, and reducing the incentives for participation in informal economic activities.

References:

1. Abdullayev, K., Tkachenko, A., Metreveli, S., Maziashvili, N., Bichai, V. Strategies for Enhancing Global Economic Resilience: A Focus on International Financial Structures and Their Impact. – *Economic*

Studies (Ikonomicheki Izsledvania), 2025. 34(6), pp. 3–20.

2. Abdullayev, K., Yusifov, E., Tazabekova, G., Tkachenko, A., & Mezheryskyi, D. International Trade and Its Implementation: Studying The Impact on The Economic Development of Azerbaijan, Kyrgyzstan, and Ukraine. *International Journal of Accounting and Economics Studies*, 2025. 12(2), 347–360. <https://doi.org/10.14419/9302vg44>

3. Allingham, M. G., & Sandmo, A. Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1972. 1(3–4), 323–338. [https://doi.org/10.1016/0047-2727\(72\)90010-2](https://doi.org/10.1016/0047-2727(72)90010-2)

4. Baylarov R., Karimova U. & Abdullayev K. Determinants of the Shadow Economy In Selected Southern European (EU) Countries: A MIMIC Model Approach. *Norwegian Journal of Development of the International Science*, 2026. 181. <https://doi.org/10.5281/zenodo.20448070>

5. Bollen, K. A. *Structural Equations with Latent Variables*. New York: John Wiley & Sons. 1989. <https://doi.org/10.1002/9781118619179>

6. Buehn, A., & Schneider, F. Shadow economies around the world: Novel insights, accepted knowledge, and new estimates. *International Tax and Public Finance*, 2012. 19(1), 139–171. <https://doi.org/10.1007/s10797-011-9187-7>

7. Dell'Anno, R., & Schneider, F. A complex approach to estimate shadow economy: The structural equation modelling approach. In M. Faggini & T. Lux (Eds.), *Coping with the Complexity of Economics*, 2009. (pp. 111–130). Springer. https://doi.org/10.1007/978-88-470-1083-3_7

8. Eurostat. Eurostat Database. 2025. <https://ec.europa.eu/eurostat>

9. Herwartz, H., Schneider, F., & Tafenau, E. One share fits all? Regional variations in the extent of the shadow economy in Europe. *Regional Studies*, 2015. 49(9), 1575–1587. <https://doi.org/10.1080/00343404.2013.848034>

10. International Energy Agency. IEA Data and Statistics. 2025. <https://www.iea.org/data-and-statistics>

11. Jöreskog, K. G., & Goldberger, A. S. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 1975. 70(351a), 631–639. <https://doi.org/10.1080/01621459.1975.10482485>

12. Korsunska, M., Butorina, V., Abdullayev, K., Kravtsov, Y., Ustyomenko, L. The role of Creative Po-

tential in the Project Management Process for the Implementation of the Company's Strategies. *Review of Economics and Finance*, 2022. 20(1), pp. 255–262. <https://doi.org/10.55365/1923.x2022.20.30>

13. Lackó, M. Hidden economy—An unknown quantity? Comparative analysis of hidden economies in transition countries, 1989–95. *Economics of Transition*, 2000. 8(1), 117–149. <https://doi.org/10.1111/1468-0351.00038>

14. Mammadov Y., Abdullayev K., Karimova U., Pironti V., Baylarov R. Main Directions of Ensuring Fiscal and Social Justice in Ensuring Sustainable Development of Countries. *European Journal of Economics and Management Sciences*, No 2. 2026. <https://doi.org/10.29013/EJEMS-26-2-90-94>

15. Medina, L., & Schneider, F. Shedding light on the shadow economy: A global database and the interaction with the official one. *IMF Working Paper*, 2019. No. 19/278. <https://doi.org/10.5089/9781513520698.001>

16. North, D. C. *Institutions, institutional change and economic performance*. 1990. Cambridge University Press.

17. OECD. OECD Data Explorer. 2025. <https://data.oecd.org>

18. Rosseel, Y. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 2012. 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>

19. Schneider, F., & Enste, D. H. Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 2000. 38(1), 77–114. <https://doi.org/10.1257/jel.38.1.77>

20. Schneider, F., Buehn, A., & Montenegro, C. E. Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007. *World Bank Policy Research Working Paper*, 2010. No. 5356. <https://doi.org/10.1596/1813-9450-5356>

21. Tanzi, V. The underground economy in the United States: Annual estimates, 1930–1980. *IMF Staff Papers*, 1983. 30(2), 283–305. <https://doi.org/10.2307/3866786>

22. Williams, C. C., & Horodnic, I. A. An institutional theory of the informal economy: Some lessons from the United Kingdom. *International Journal of Social Economics*, 2016. 43(7), 722–738. <https://doi.org/10.1108/IJSE-12-2014-0256>

23. World Bank. World Development Indicators. 2025. <https://databank.worldbank.org/source/world-development-indicators>

TRANSITION FROM A PROFIT MAXIMIZATION MODEL TO A SUSTAINABLE VALUE CREATION MODEL IN CORPORATE MANAGEMENT OF STATE-OWNED ENTERPRISES**Zhumadullaeva D.***Tashkent State University of Economics***ПЕРЕХОД ОТ МОДЕЛИ МАКСИМИЗАЦИИ ПРИБЫЛИ К МОДЕЛИ УСТОЙЧИВОГО СОЗДАНИЯ ЦЕННОСТИ В КОРПОРАТИВНОМ УПРАВЛЕНИИ ПРЕДПРИЯТИЙ С ГОСУДАРСТВЕННЫМ УЧАСТИЕМ****Жумадуллаева Д.Ш.***Ташкентский государственный экономический университет*<https://doi.org/10.5281/zenodo.21105811>**Abstract**

The article examines the limits of the profit maximization model in corporate governance of state-owned enterprises. The author argues that state-owned enterprises require a broader governance framework because the state acts not only as a shareholder, but also as a regulator and representative of public interests. The proposed approach is based on sustainable value creation, where ESG indicators become part of strategic KPI, managerial motivation and board performance assessment.

Аннотация

В статье исследуется ограниченность традиционной модели корпоративного управления, ориентированной на максимизацию прибыли, применительно к предприятиям с государственным участием. Обосновано, что специфика таких предприятий определяется двойственной ролью государства как собственника и регулятора, наличием социальных обязательств, сниженной рыночной дисциплиной и повышенными требованиями к прозрачности. На основе анализа агентской теории, stakeholder-подхода и международных принципов корпоративного управления предложен переход к модели устойчивого создания ценности. В данной модели ESG-показатели рассматриваются не как элемент внешней отчетности, а как часть стратегического управления, системы KPI, мотивации менеджмента и оценки деятельности наблюдательного совета. Авторский подход позволяет согласовать финансовую эффективность, общественные интересы и долгосрочную устойчивость предприятий с государственным участием.

Keywords: corporate governance, state-owned enterprises, ESG, sustainable value creation, agency conflict, state ownership, strategic KPI, supervisory board.

Ключевые слова: корпоративное управление, предприятия с государственным участием, ESG, устойчивое создание ценности, агентский конфликт, государственная собственность, стратегические KPI, наблюдательный совет.

Введение

Корпоративное управление на предприятиях с государственным участием не может оцениваться теми же критериями, которые применяются к частным акционерным обществам. Частная компания в большей степени ориентирована на доходность капитала, рост стоимости бизнеса и удовлетворение интересов акционеров. Предприятие с участием государства действует в иной системе координат: оно должно обеспечивать финансовую отдачу, поддерживать отраслевую стабильность, участвовать в реализации государственных программ и сохранять социальную функцию.

Эта двойственность делает традиционную модель корпоративного управления методологически неполной. Если результативность компании измеряется только прибылью, рентабельностью капитала или дивидендными выплатами, за пределами оценки остаются прозрачность, общественная подотчетность, качество контроля, социальные эффекты и долгосрочные риски. Для предприятий с

государственной долей эти элементы не являются второстепенными. Они определяют устойчивость бизнеса и доверие со стороны государства, инвесторов, кредиторов и общества.

Цель статьи состоит в теоретическом обосновании перехода от модели максимизации прибыли к модели устойчивого создания ценности в корпоративном управлении предприятиями с государственным участием. В рамках данной цели решаются следующие задачи: раскрыть ограничения традиционной модели, показать специфику агентских отношений в предприятиях с государственной долей. А также, обосновать роль ESG-показателей как инструмента стратегического управления и предложить концептуальную схему перехода к устойчивому созданию ценности.

Методы исследования

Теоретическую основу исследования составляют агентская теория М. Дженсена и У. Меклинга, stakeholder-подход Р. Фримена, исследования А.

Шлейфера и Р. Вишни по корпоративному управлению и защите инвесторов, а также принципы корпоративного управления ОЭСР.

Методологическая база статьи включает сравнительный анализ, институциональный анализ, системный подход и метод концептуального моделирования. Сравнительный анализ использован для сопоставления традиционной модели максимизации прибыли и предлагаемой модели устойчивого создания ценности. Институциональный анализ применен для раскрытия двойственной роли государства как собственника и регулятора. Системный подход позволил рассмотреть корпоративное управление как совокупность взаимосвязанных элементов. К ним относятся собственность, наблюдательный совет, менеджмент, раскрытие информации, ESG-показатели и механизмы подотчетности.

Материалами исследования послужили научные труды по корпоративному управлению, документы ОЭСР, международные подходы к ESG-отчетности, а также положения национальной практики корпоративного управления предприятий с государственным участием.

В статье используется авторская логика анализа. Сначала выявляются ограничения традиционной модели, затем раскрывается специфика предприятий с государственной долей, после чего обосновывается переход к модели устойчивого создания ценности.

Результаты

Традиционная модель корпоративного управления исходит из того, что основной целью компании является максимизация прибыли и увеличение стоимости для акционеров. В качестве ключевых индикаторов используются чистая прибыль, ROE, ROA, дивиденды, рыночная капитализация и рост стоимости акций. Такая система оценки понятна для частного бизнеса, поскольку собственник вкладывает капитал и ожидает финансовую отдачу.

Для предприятий с государственным участием эта модель оказывается неполной. Её ограниченность проявляется в нескольких аспектах.

Первое ограничение связано с множественностью целей. Государственное предприятие может участвовать в реализации государственных программ, поддерживать занятость, обеспечивать доступность услуг, финансировать приоритетные отрасли или выполнять социальные обязательства. Эти задачи не всегда дают прямой коммерческий эффект, но имеют значение для экономики и общества.

Второе ограничение связано с размытостью ответственности. В частной компании ответственность распределяется между акционерами, советом директоров и менеджментом. В государственном предприятии структура более сложная - государственный орган, наблюдательный совет, исполнительный орган, отраслевой регулятор и общественные ожидания. В результате финансовый показатель не всегда показывает, где именно возникла управленческая проблема.

Третье ограничение проявляется в слабой рыночной дисциплине. Государственные предприятия часто имеют доступ к поддержке, гарантиям или административному ресурсу. Это снижает давление рынка и ослабляет стимулы к повышению прозрачности, снижению издержек и повышению качества управления.

Четвертое ограничение касается социальных и экологических обязательств. В традиционной модели они воспринимаются как внешняя нагрузка. В модели устойчивого создания ценности они должны рассматриваться как часть долгосрочной стратегии.

Пятое ограничение связано с инвестиционной привлекательностью. Инвестор оценивает не только прибыль, но и качество раскрытия информации, независимость наблюдательного совета, предсказуемость корпоративных решений, управление рисками и деловую этику. Компания может демонстрировать прибыль, но оставаться непрозрачной и институционально слабой. В таком случае инвестиционный риск сохраняется высоким.

Классическая агентская теория рассматривает конфликт между собственником и менеджером. В предприятиях с государственным участием этот конфликт приобретает более сложную форму.

Первый уровень конфликта возникает между государством-собственником и менеджментом. Государство как акционер заинтересовано в сохранности и эффективном использовании государственных активов. Менеджмент может быть ориентирован на собственные интересы, формальное исполнение решений или краткосрочные показатели.

Второй уровень конфликта возникает между государством-регулятором и обществом как конечным бенефициаром. Государство может ожидать от предприятия прибыли и дивидендов. Однако одновременно использовать его для достижения социальных целей поддержки занятости, льготного кредитования, развития инфраструктуры или стабилизации отдельных отраслей.

Эта двойственность усложняет систему корпоративного управления. В частной компании корпоративное управление в основном направлено на защиту интересов акционеров и контроль за действиями менеджмента.

В государственном предприятии круг интересов шире. Здесь важно обеспечить баланс между интересами государства, менеджмента, инвесторов и общества.

ESG как элемент стратегического управления

ESG-подход позволяет перевести экологические, социальные и управленческие параметры из зоны формальной отчетности в сферу стратегического управления. Для предприятий с государственным участием это особенно значимо, поскольку многие их функции изначально связаны с общественными интересами.

Экологический блок ESG отражает энергоэффективность, зеленое финансирование, экологические риски и воздействие на природные ресурсы.

Социальный блок включает доступность услуг, защиту клиентов, развитие персонала, поддержку социально значимых проектов и вклад в развитие регионов.

Управленческий блок охватывает прозрачность, состав наблюдательного совета, внутренний аудит, антикоррупционную политику, управление конфликтом интересов и раскрытие информации.

Для предприятий с государственной долей особую значимость имеет именно управленческий компонент ESG. Он связан с ключевыми проблемами таких предприятий: ограниченной независимостью наблюдательных советов, недостаточной прозрачностью, слабой рыночной дисциплиной и административным влиянием.

Принципиальное значение имеет не само наличие ESG-отчета, а включение ESG-показателей в

систему стратегических KPI. Ответственность фиксирует результат, но не всегда меняет поведение менеджмента. KPI влияет на управленческие решения, распределение ответственности и мотивацию.

Интеграция ESG в стратегические KPI означает, что экологические, социальные и управленческие показатели становятся частью оценки деятельности менеджмента и органов управления. Например, для банка с государственным участием такими показателями могут быть уровень раскрытия информации, доля независимых директоров, наличие комитетов при наблюдательном совете, качество внутреннего аудита, доля устойчивого финансирования, цифровая доступность услуг и наличие политики управления конфликтом интересов.

Таблица 1

Механизм интеграции ESG-факторов в стратегические KPI¹

Параметр	Модель максимизации прибыли	Модель устойчивого создания ценности (предлагаемая)
Главная цель	Максимизация прибыли	Устойчивое развитие и баланс интересов
Решение агентского конфликта	Механизмы контроля (независимые директора)	ESG-KPI, транспарентность, социальная ответственность
Роль государства	Краткосрочное политическое вмешательство	Долгосрочный стратегический собственник
ESG-факторы	Обязательная отчетность, как ограничение	Стратегические KPI, увязанные с системой вознаграждения
Критерии успеха	ROE, ROA, дивиденды	ROE + социальные и экологические индикаторы

При таком подходе ESG становится не внешним требованием, а внутренним механизмом управления. Менеджмент получает стимул не только увеличивать прибыль, но и улучшать прозрачность, снижать риски, развивать устойчивые продукты, повышать качество раскрытия информации и укреплять доверие инвесторов.

Обсуждение

Переход к модели устойчивого создания ценности имеет теоретическое и прикладное значение.

С теоретической точки зрения предложенный подход расширяет классическую агентскую модель. Для предприятий с государственным участием недостаточно анализировать только конфликт между собственником и менеджментом. Необходимо учитывать конфликт между различными ролями государства: собственника, регулятора и представителя общественных интересов.

Stakeholder-подход позволяет дополнить агентскую теорию. Он показывает, что компания не может быть сведена только к интересам акционеров. Для предприятий с государственной долей этот тезис особенно значим, поскольку их деятельность затрагивает работников, клиентов, кредиторов, государственные органы, местные сообщества и общество.

Принципы ОЭСР усиливают институциональную сторону анализа. Они подчеркивают значение прозрачности, профессионального выполнения государством функции собственника, ответственности

органов управления и разграничения роли собственника и регулятора. Эти положения формируют основу для перехода от административного контроля к стратегическому управлению государственными активами.

С практической точки зрения ESG-показатели позволяют сделать социальные, экологические и управленческие цели измеримыми. Если они остаются декларативными, их влияние на управление ограничено. Если они включаются в KPI, они становятся частью механизма принятия решений.

Для банков с государственным участием такая модель имеет прямое значение. Банки не только формируют прибыль, но и участвуют в финансировании приоритетных отраслей, поддержке населения и реализации государственных программ. Поэтому их эффективность должна оцениваться не только через финансовые показатели, но и через прозрачность, качество корпоративного контроля, уровень раскрытия информации, ESG-интеграцию и доверие инвесторов.

При этом внедрение ESG-показателей в KPI не должно превращаться в формальное расширение отчетности. Оно требует изменения управленческой логики. Наблюдательный совет должен оценивать не только финансовые результаты, но и устойчивость стратегии. Менеджмент должен отвечать не только за прибыль, но и за качество раскрытия информации, внутренний контроль, управление рисками и социальные эффекты.

¹ Составлено автором на основе проведенных исследований

В условиях Узбекистана предложенный подход может использоваться при совершенствовании корпоративного управления предприятий с государственной долей, особенно в банковском секторе. Он позволяет связать национальные реформы

корпоративного управления с международными принципами ОЭСР и ESG-повесткой, не копируя зарубежные модели механически, а адаптируя их к особенностям государственной собственности.

Концептуальная модель перехода от максимизации прибыли к устойчивому созданию ценности



Рис. 1. Концептуальная модель перехода от максимизации прибыли к устойчивому созданию ценности²

Предложенный подход имеет несколько научных и прикладных результатов.

С научной точки зрения он расширяет классическое понимание корпоративного управления на предприятиях с государственным участием. Агентский конфликт рассматривается не только как проблема отношений между собственником и менеджером, но и как конфликт между государством в разных институциональных ролях. Это позволяет точнее объяснить причины низкой эффективности, слабой прозрачности и сложности оценки результатов государственных предприятий.

С методологической точки зрения модель устойчивого создания ценности соединяет три исследовательские линии: агентскую теорию, stakeholder-подход и ESG-концепцию. Их объединение позволяет перейти от описательной характеристики проблем к построению управленческой модели.

С практической точки зрения предложенная модель даёт возможность сформировать систему KPI, в которой финансовые и нефинансовые показатели не противоречат друг другу. Для предприятий с государственным участием это означает переход от формального исполнения требований к управлению, основанному на измеримых результатах.

Для банков с государственным участием применение модели может повысить качество стратегического планирования, усилить прозрачность, укрепить наблюдательные советы, снизить агент-

ские издержки и повысить инвестиционную привлекательность. При этом ESG-показатели становятся не декоративной частью годового отчета, а инструментом управления стоимостью и устойчивостью.

Заключение

Традиционная модель максимизации прибыли сохраняет значение для оценки финансовой результативности, но не раскрывает всей сложности корпоративного управления предприятий с государственным участием. Государственная собственность формирует особую систему интересов, где прибыль сочетается с социальными задачами, регуляторными функциями и общественной ответственностью.

Переход к модели устойчивого создания ценности позволяет устранить главный недостаток традиционного подхода — одностороннюю ориентацию на финансовый результат. В предлагаемой модели ESG-показатели становятся частью стратегических KPI, системы мотивации менеджмента и оценки деятельности наблюдательного совета. Это создает основу для согласования интересов государства, менеджмента, инвесторов и общества.

Для банков с государственным участием такая модель имеет прикладное значение, поскольку позволяет связать финансовую эффективность с прозрачностью, рыночной дисциплиной, внутренним контролем и общественной подотчетностью. Авторский подход показывает, что устойчивое создание ценности не заменяет прибыль, а формирует

² Составлено автором

условия для её долгосрочного и институционально устойчивого роста.

Список литературы:

1. Jensen M.C., Meckling W.H. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure // *Journal of Financial Economics*. 1976. Vol. 3. No. 4. P. 305–360. DOI: 10.1016/0304-405X(76)90026-X.
2. Shleifer A., Vishny R.W. A Survey of Corporate Governance // *The Journal of Finance*. 1997. Vol. 52. No. 2. P. 737–783. DOI: 10.1111/j.1540-6261.1997.tb04820.x.
3. Freeman R.E. *Strategic Management: A Stakeholder Approach*. Boston: Pitman, 1984. 276 p.
4. OECD. *G20/OECD Principles of Corporate Governance 2023*. Paris: OECD Publishing, 2023. DOI: 10.1787/ed750b30-en.
5. OECD. *OECD Guidelines on Corporate Governance of State-Owned Enterprises 2024*. Paris: OECD Publishing, 2024.
6. Tricker B. *Corporate Governance: Principles, Policies, and Practices*. 3rd ed. Oxford: Oxford University Press, 2015. 550 p.
7. Cadbury A. *Report of the Committee on the Financial Aspects of Corporate Governance*. London: Gee and Co. Ltd., 1992. 90 p.
8. La Porta R., Lopez-de-Silanes F., Shleifer A., Vishny R. Investor Protection and Corporate Governance // *Journal of Financial Economics*. 2000. Vol. 58. No. 1–2. P. 3–27. DOI: 10.1016/S0304-405X(00)00065-9.
9. Hart O. Corporate Governance: Some Theory and Implications // *The Economic Journal*. 1995. Vol. 105. No. 430. P. 678–689. DOI: 10.2307/2235027.
10. Global Reporting Initiative. *GRI Standards*. Amsterdam: GRI, latest edition.
11. Кодекс корпоративного управления Республики Узбекистан. Утвержден протоколом заседания Комиссии по повышению эффективности деятельности акционерных обществ и совершенствованию системы корпоративного управления от 31 декабря 2015 г. № 9.
12. Закон Республики Узбекистан «Об акционерных обществах и защите прав акционеров» от 6 мая 2014 г. № ЗРУ-370.
13. Постановление Кабинета Министров Республики Узбекистан от 29 марта 2021 г. № 166 «Об утверждении Стратегии управления и реформирования предприятий с государственным участием на 2021–2025 годы».
14. Указ Президента Республики Узбекистан от 28 января 2022 г. № УП-60 «О Стратегии развития Нового Узбекистана на 2022–2026 годы».
15. Указ Президента Республики Узбекистан от 27 октября 2020 г. № УП-6096 «О мерах по ускоренному реформированию предприятий с участием государства и приватизации государственных активов».

HISTORICAL SCIENCES

THE ROLE OF CAMEL CARAVAN ROUTES IN THE HISTORICAL SYSTEM OF TRADE AND CARGO TRANSPORTATION OF AZERBAIJAN

Gasimova N.

Baku State University,

Baku, Azerbaijan

Heydarova S.

PhD in History

Baku State University,

Baku, Azerbaijan

Mammadov I.

Baku State University,

Baku, Azerbaijan

<https://doi.org/10.5281/zenodo.21105813>

Abstract

This article examines the significant role of caravan routes operated by camel caravans during the ancient and medieval periods in developing trade and freight transport networks, expanding economic exchange, and fostering interregional integration. The study explores the structure of transport and trade networks established along caravan routes, the logistical advantages of camel transport, and its function in ensuring the resilience of long-distance trade relations. Furthermore, it highlights the impact of this system not only on economic development but also on the expansion of cultural and civilizational interactions. The primary objective of this research is to determine the functional significance of camel freight transport within the historical caravan transport system of Azerbaijan and to provide a scholarly rationale for its position in the regional trade system.

Keywords: Azerbaijan, ancient and medieval periods, transport, camel caravan routes, trade, freight transport.

Introduction. Azerbaijan's geographical position at the crossroads of Asia and Europe has, since ancient times, created highly favourable conditions for its development as a major transit hub for trade and freight transport, thereby driving the demand for transport infrastructure. Caravan routes represent one of the oldest communication and trade networks, having played a pivotal role in shaping economic, cultural, and political relations throughout human history. These arteries served not only for the transportation of commodities but also acted as the primary channels for transferring knowledge, technology, religious concepts, and cultural values between diverse civilisations. Established in deep antiquity, this networked system spanned the vast geographical expanse of Eurasia, functioning as a strategic bridge connecting East and West. Notably, within the framework of the Silk Road system, caravan routes acquired a more structured and international character, intensifying trade and cultural ties between China, Central Asia, the Middle East, and Europe. The cities and outposts situated along these thoroughfares evolved not only into major economic hubs but also into spaces of intercultural convergence. The territory of Azerbaijan was an integral part of this extensive caravan network. The region's advantageous geographical location at the intersection of the East–West and North–South trade routes fostered the dynamic development of local transport and logistical infrastructure. Consequently, urban centres such as Baku, Shamakhi, Ganja, and Tabriz emerged as key transit points on international trade routes.

The study of caravan routes holds significant theoretical importance not only for reconstructing historical trade systems but also for understanding the dynamics of regional socio-economic development, intercultural interaction processes, and geopolitical transformations. Within this context, a comprehensive investigation into these issues is highly relevant to both historiography and cultural studies. Researching the historical, geographical, and functional characteristics of caravan routes is fundamental to a scholarly comprehension of regional economic development dynamics, the transformation of trade networks, and cultural integration processes. This subject remains a pertinent area of study at the intersection of historiography, the history of intercultural relations, and economic history.

The primary objective of this study is to conduct a comprehensive analysis of the historical development dynamics of caravan routes, their economic and cultural functions, and the strategic position of the territory of Azerbaijan within international trade networks. The work aims to conceptualise caravan routes not merely as commercial paths, but as multifunctional systems that played a pivotal role in shaping inter-civilisational ties. To achieve this objective, the following tasks have been defined: to investigate the genesis and formative stages of caravan routes; to analyse the structural features of the Silk Road system; to identify the main caravan routes that traversed the territory of Azerbaijan; to evaluate the cultural and social consequences of trade relations; and to determine the role of caravan roads in regional and international integration.

Alongside this, structural-functional analysis provides a scholarly evaluation of the economic, political,

and cultural functions performed by caravan routes within society. The source base for this research comprises archaeological materials, historical chronicles, notes and diaries of medieval travellers, as well as contemporary research literature.

Domestic animals played a vital role in the economic activities of Azerbaijan, serving as a key means of freight transportation. For centuries, Azerbaijanis and their ancestors utilised horses, donkeys, mules, camels, as well as oxen and buffaloes as riding and pack animals.

Based on archaeological materials, it has been established that the inception of camel utilisation in the South Caucasus dates back to the second half of the 2nd millennium BCE (27). The results of archaeological excavations conducted in 1896 in the village of Qarabulag indicate that by the beginning of the 1st millennium BCE, the camel was already being actively exploited as a means of transport and riding in Azerbaijan (9). This thesis is further corroborated by the works of B. B. Pitrovsky (17). In particular, materials dating to the 1st–2nd centuries CE indicate that camels were the primary mode of transport for delivering Caspian fish and other trade commodities to distant countries (21). Medieval sources also note the use of camels as a priority means of freight transportation (7).

Caravans were formed for the transportation of goods and long-distance travel. Originating in deep antiquity, the caravan served as a mechanism for merchants and travellers transporting goods over long distances to overcome road hardships and ensure security against robber attacks. Caravans were frequently accompanied by armed guards and specially trained dogs. A caravan could include a significant number of horses, as well as thousands of pack and riding animals (29).

Since ancient times, the development of agriculture, animal husbandry, and crafts in Azerbaijan has fostered urban growth and expanded the region's trade and economic relations with numerous countries. The intensity of caravan traffic led to the emergence of specific occupational roles, such as the *gafilesar* (caravan leader) and *sarvan* (camel driver/caravan master). Their primary responsibilities included the formation of armed detachments to protect and escort the caravan during its journey.

Data from written sources and the findings of archaeological excavations confirm that during the Classical antiquity period, animals such as horses, mules, and camels played a pivotal role as riding and pack transport vehicles.

For centuries, the camel not only provided the population with food and the textile industry with raw materials, but also served as a crucial means of locomotion. One of the regions where the Bactrian camel (two-humped camel) was domesticated was Southern Azerbaijan, where these animals began to be utilised for transport purposes as early as the late Chalcolithic period. The significance of the camel as freight transport increased substantially during the Classical antiquity era, when it became the primary means of moving goods for export operations to neighbouring and distant countries.

Claudius Aelianus, in his description of the Caspian region, noted the extraordinary abundance of local camels; he indicated that the largest individuals reached the size of a tall horse and were covered with excellent wool, which in its softness and fineness was not inferior to that of Miletus (24). Attesting to the use of camels by Caspian tribes as a means of transport, Claudius Aelianus wrote: *"In the lands of the Caspians, there is a large lake where big fish are found... The Caspians catch them, salt or dry them, then pack them onto camels and take them to Ecbatana"* (24). From the provided information, it is evident that Aelianus places particular emphasis on the high camel population, their physical parameters, their value for wool production, and their active application in trade.

The development of camel husbandry facilitated the formation of overland caravan routes and the expansion of trade networks, while the organisation of security ensured that merchants reached their destinations in a timely and safe manner. The high resilience of camels to water scarcity under scorching heat, combined with their ability to traverse burning sands without difficulty, drove a substantial demand for these animals as an efficient pack and transport vehicle.

Overland caravan routes possessed international significance. A vital role within the transport communication system was played by internal roads that linked the cities and rural settlements of Caucasian Albania. The ancient thoroughfare leading to the city of Gabala branched out in a locality known as Chaybogazy: one direction extended eastward, while the second led towards Mingachevir. Another caravan route traversing Albanian territory proceeded from the area of modern-day Qazakh through Shamkir, Ganja, and Mingachevir to Shamakhi (32).

In the south, the city of Ecbatana served as a key junction for intersecting trade routes. It was precisely here that the primary caravan roads passing through the territory of Atropatene converged.

During the Medieval period, camels played a fundamental role in the freight transportation system of Azerbaijan. The high carrying capacity of these animals enabled the efficient organisation of caravan communication and ensured the transport of various categories of goods to both neighbouring and distant countries in the shortest possible timeframe.

According to research, while the *bugur* (a male Bactrian, two-humped camel) and the *kyak* (a male dromedary, one-humped camel) possessed a carrying capacity within the range of 130–170 kg, the *nar* (hybrid males obtained from crossing dromedary and Bactrian camels) were capable of transporting loads weighing 400 or even 500 kg (11,185). The exceptional significance of the camel as a key means of caravan transport in medieval Azerbaijan was reflected in the epic poem *Isfandiyar-name* (The Book of Alexander) by the great poet Nizami Ganjavi:

"Of the four thousand camels designated for freight transport and following the army everywhere, One thousand were selected from the bisurak breed; under their heavy tread, stones turned to dust. Another thousand comprised the bukhti (Bactrian camels), laden with delicacies and provisions for the warriors."

The third thousand were represented by the swift-footed naqa (female camels), whose cargo consisted of gold and jewels. The fourth thousand consisted of noble camels, each of which in its speed was like a goitered gazelle or a gust of hot wind" (30).

An analysis of depictions of pack camels in book miniatures indicates that the rules of camel loading during the Medieval period remained virtually unchanged until recent times. This continuity allows for the reconstruction of this process based on contemporary ethnographic observations.

To load a camel, a special wooden pack frame known as a *navar* was fitted onto its back. This frame consisted of a pair of wooden rods (*mils*) wrapped in felt and interconnected by means of breast and breeching (*crupper*) straps (*paldym*). Depending on the number of the camel's humps, the branches of the frame were secured with connecting straps in two or three places. In cases where the load was light or the transportation was carried out over short distances, a simplified and lighter pack saddle or padding called a *palan* was used instead of the frame. The *palan* was placed between the humps, tightly fastened in two places with ropes (*orkyan*), and secured with a belly girth strap (32).

As evidenced by book miniatures, various receptacles were utilised for camel loading depending on the nature of the cargo, including chuvals (large sacks), chests, baskets, and wineskins. In particular, Sultan Muhammad's miniature "The Hunting Feast of the Shah" for A. Jami's work *Silsilat al-Dhahab* (The Chain of Gold) depicts a camel laden with chests (12). In another miniature, "The Battle of Kay Khosrow with Afrasiab", a camel is illustrated carrying packed baskets (12). It is also known that Baku oil was transported to the country's interior regions and abroad in wineskins loaded onto camels. Furthermore, camels were actively employed as a means of riding transport, a fact visually corroborated by the aforementioned miniatures.

The quantitative composition of caravans varied. Most frequently, groups of 15–20 camels were formed, though in specific cases, their number could reach up to one hundred heads. The leadership of the caravan was undertaken by an experienced specialist—the *caravanbashi* (caravan leader) or *sarvan*. The system of tethering the animals en route was structured as follows: the halter (*nokhta*) of each subsequent camel was tied to the saddle pack frame (*navar*) of the animal walking ahead, while the lead rope of the first animal remained in the hands of the *sarvan* (31).

Historical evidence indicates that substantial transformations occurred along international trade routes during the 8th and 9th centuries. By expanding its commercial ties with various countries, the Caliphate established hegemony over world trade, disrupting Byzantine trade turnover across the Black Sea and successfully redirecting the primary caravan and maritime routes through its own territory (28). The Caliphate's trade relations with northern countries were conducted predominantly via the Volga-Caspian overland and maritime transit routes, which traversed the territory of Azerbaijan.

The Caliphate's Northern trade and caravan route originated in Baghdad and, passing through Ardabil and Barda, led via Derbent into the lands of the Khazars, Bulgars, Burtas, and Rus' (34). This route simultaneously connected remote provinces of the Caliphate, including Azerbaijan, Arran, and Shirvan, with the state's centre. At Mount Behistun (Bisitun), the Northern route branched: one branch extended north-eastward to Zanjan, and from there to Ardabil and Barsan (Varsan); the second branch ran through Sisar in Southern Azerbaijan and, bypassing Maragheh and Tabriz, headed towards Ardabil and Mughan.

The 9th-century author Ibn Khordadbeh, listing the key nodes of the first caravan route and specifying the distances between them, wrote: 'From Dinavar to Zanjan is 29 stages (*manzils*); thence to Maragheh, 11 stages; to Miyanij, 2 stages; to Ardabil, 11 stages; and to Barsan, the final point of the province of Azerbaijan, 11 stages(8)

Describing the second branch of the Northern route—a caravan itinerary that ran from Dinavar to Barzand—Ibn Khordadbeh provided the following information:

"From Dinavar to Khubardjan is 7 farsakhs; to Tall Wan, 6 farsakhs; thence to Sisar, 7 farsakhs; to Andarab, 4 farsakhs; then to Baylaqan, 5 farsakhs; to Karzai, 6 farsakhs; to Saburkhast, 8 farsakhs; to Maragheh, 7 farsakhs; to Dehkhorréqan, 11 farsakhs; to Tabriz, 9 farsakhs; to Marand, 10 farsakhs; to Khan, 4 farsakhs; to Khoy (Khuvvayyi), 6 farsakhs. From Maragheh to Kursara is 10 farsakhs; thence to Sarang, 10 farsakhs; to Nir, 5 farsakhs; to Ardabil, 5 farsakhs; and to Mughan, 10 farsakhs. From Ardabil to Khushsh is 8 farsakhs; thence to Barzand, 6 farsakhs; [and] from Barzand across the Balasajan plain to Barsan is 12 farsakhs" (8).

By providing such meticulous data on the caravan routes, Ibn Khordadbeh thereby placed particular emphasis on their strategic significance.

While the city of Ardabil served as the key junction for intersecting caravan routes in Southern Azerbaijan, a similar function in Northern Azerbaijan was, in all probability, performed by Barda. Based on the works of Arab geographers and travellers of the 9th and 10th centuries, it is possible to reconstruct four primary macro-directions of caravan communication originating from Barda:

- **The Southern Direction:** It ran through Yunan, Baylaqan, Barsan, Balkhab, and Varvard, connecting in Ardabil with the routes leading further south.

- **The Northern Direction:** It proceeded through Bardij, Shamakhi, Shirvan, and Abkhaz to the Samur Bridge and further to Derbent.

- **The North-Western Direction (towards Tiflis):** It passed through Ganja, Shamkir, Khunan, and the Ibn Kandman fortress, reaching Tiflis.

- **The Western Direction (towards Dvin/Dabil):** It traversed Kalqatus, Matris, Davmims, Qulaqvin, and Sisakan, ultimately reaching the city of Dvin (8).

The sources indicate that caravanserai facilities were situated at key nodes along the caravan routes, de-

signed to provide overnight halts and rest for both travellers and animals. Postal stations and customs houses were also established at these same locations (5). Furthermore, stones that functioned as road signposts were erected along the thoroughfares. An inscription has been preserved on one such stone, dating to the 7th century and uncovered during archaeological excavations, which reads: "In the name of Allah, three miles from Tiflis" (35).

The aftermath of the Mongol invasion dealt a severe blow to foreign trade, and between the 1230s and 1290s, caravan routes were dominated by bandits. The regular operation of caravan itineraries was restored only at the close of the 13th century. The capital of the Ilkhanate—the city of Tabriz—emerged as the most critical hub where the primary trade veins converged. The Southern caravan route connected Tabriz with Baghdad and Mosul, the Eastern route linked it with Rey and Khorasan, and the Western route extended to Russia and other European countries (18).

In his work *Nuzhat al-Qulub*, Hamdallah Qazvini provides a detailed description of the caravan routes running from Sultaniya to Ganja, from Sultaniya to Tabriz and onward to Erzurum, as well as the itinerary from Karabakh to Tabriz via Ahar. The caravan highway followed the trajectory of Sultaniya–Ardabil–Bajjarvan–Mahmudabad, from where it proceeded into Karabakh, Ganja, and subsequently to Tiflis (22).

During the 14th and 15th centuries, owing to the expanding role of Shirvan silk in international commerce and its export to both European and Eastern markets, the significance of the Bilesuvar–Mahmudabad–Shamakhi–Shabran–Derbent caravan route increased substantially. Beginning in the 15th century, the Niyazabad–Shabran–Shamakhi–Tabriz caravan route was brought into operation; the intensity of its functioning grew noticeably throughout the 16th and 17th centuries. During this period (16th–17th centuries), alongside the Tabriz–Bilesuvar–Mahmudabad–Shamakhi–Shabran–Niyazabad thoroughfare, the Isfahan–Qazvin–Rasht–Astara–Lankaran–Baku–Shabran–Derbent trade route was also actively utilised (33).

The western route extending from Tabriz held vital importance for the execution of Azerbaijan's commercial and economic ties with Europe: Tabriz–Marand–Asdabad–Nakhchivan–Iravan–Kars–Trabzon–Istanbul. Furthermore, two caravan routes led from Tabriz to Syria: Tabriz–Van–Diyarbakir–Aleppo–Syria and Tabriz–Salmas–Jazira–Aleppo–Syria (33).

Extensive caravan routes led from Tabriz to Isfahan, bifurcating into two primary directions in the vicinity of Lake Urmia:

- **The first branch:** Tabriz–Qom–Kashan;
- **The second branch:** Tabriz–Ardabil–

Qazvin–Isfahan.

Additionally, a direct, through-caravan highway operated along the trajectory of Tabriz–Zanjan–Sultaniya–Qom–Kashan–Isfahan (33).

Up until the 19th and early 20th centuries, camels remained the pivotal means of freight transportation. Owing to the high efficiency of this mode of transport, camel husbandry constituted one of the most lucrative branches of the livestock economy. In regions where

landscape conditions precluded the utilisation of wheeled transport, trade turnover was maintained predominantly via camel caravans; consequently, particular attention was dedicated to the development of this sector in a number of Azerbaijani zones (4).

The camel is an animal biologically adapted to hot climatic conditions, which explains why this industry achieved its widest distribution in the lowland (*Aran*) districts of Azerbaijan. The employment of camels as pack transport in the plains zone dates back to profound antiquity. According to archival sources, the largest camel populations were concentrated in the Baku, Shamakhi, Nukha, Shusha, Lankaran, Aghdam, and Salyan *uyezds* (districts). While in 1845 the number of camels in the aforementioned *uyezds* stood at 4,637 heads, by 1858 this figure had reached 11,542 heads (20). Such growth dynamics were driven by the high demand for camels as an optimal labor force characterised by substantial carrying capacity.

In the final quarter of the 19th century, there were 310 heads of camels recorded in Sharur; in the Erivan *uyezd* (within the households of both nomadic and sedentary rural populations), there were 989 heads (14); in the Echmiadzin *uyezd*, 2,363 heads; and in the Surmali *uyezd*, 4,314 heads (23).

From the beginning of the 19th century until the 1840s, camel husbandry remained the primary transport resource in the Elizavetpol and Nakhchivan *uyezds*, as well as in the Belokan *okrug* (district). However, in the subsequent period, the demand for this mode of transport in these administrative units declined, and by 1850, its exploitation had virtually ceased (20, p.48, tab.3).

From the mid-19th century onwards, the initiation of Baku oil transportation to the *uyezds* in wineskins (*tuluks*) served as a catalyst for the rapid development of camel husbandry on the Absheron Peninsula. This trend is clearly discernible in the statistical materials of the 1880s, which reflect the camel population across individual villages of the Baku *uyezd*. During this period, the largest centers of camel husbandry were the villages of Jorat (300 heads) and Kobu (200 heads). In the villages of Mashtaga and Saray, 100 heads of camels each were maintained for the needs of oil transportation, while the villages of Novkhani, Qala, Sabunchi, Fatma, Nardaran, Bilajari, Guldek, Amirjan, and Zabrat held several dozen heads each (25). However, by the close of the 19th century, a downturn in the development of camel husbandry was observed in these settlements. The primary factor behind the reduction in the camel population was the construction of railway lines and the laying of oil pipelines (11).

During the winter period, camels were housed on the outskirts of villages in specialised stone sheds, known as *develik* (camel houses). The length of such structures reached 30–40 m, the width was 8–10 m, and the height stood at around 6 m (10), which allowed a pack animal to enter unobstructed together with its cargo. In the summer season, the livestock was kept under open canopies (*talvars*). The remnants of such specialised structures are still preserved to this day in the villages of Hovsan and Kobu.

The lifespan of a camel is 30–35 years (13), of which approximately 20–25 years constitute its period of economic utility. Upon the expiration of this timeframe, the animal was slaughtered for meat. Camel meat is characterised by high fat content and, in terms of its flavor profile, is close to water buffalo meat. Camel milk was also actively used in daily life. It is worth noting the ethno-confessional specificity regarding the consumption of this product: while Azerbaijani families adhering to the Shia *madhhab* (school of jurisprudence) consumed camel meat, this type of meat was not practiced among the Sunni population (13).

The draft power and carrying capacity of a single camel is equivalent to that of 2–3 horses. At the same time, compared to horses, the camel required significantly less maintenance. Despite the absence of horseshoes, the anatomical structure of the feet allowed the animal to move freely and painlessly across rocky and rugged terrain. Nevertheless, sources record instances where special leather footwear was used for camels when navigating particularly challenging sections of the route (23).

The weight of the transported cargo varied depending on the length of the route, as well as the specific species of the camel. As a rule, the maximum load capacity for a single individual constituted approximately half of its own body mass (16). At the beginning of the summer season, during the molting period, the animal's organism was significantly weakened; consequently, excessive physical exertion could lead to skin damage and the formation of deep wounds on the back that were difficult to heal. To prevent this, owners treated the camels' skin with seal blubber. By mid-August, the animal's body became covered with a new coat of wool (15), after which it once again became fit for intensive economic exploitation.

The establishment of camel husbandry as the leading sector for freight transportation was conditioned not only by the specifics of Azerbaijan's topography but also by the dominance of commodity production in its economic life. Such commodities produced within the region as silk, flax, cotton, grain, petroleum, salt, wine, and other products were actively transported to both domestic and foreign markets.

A number of researchers indicate that in the Baku Province, camels were maintained exclusively by individuals engaged in *charvadarstvo* (the commercial freight trade) (1). The transport industry was practiced by the inhabitants of the Shamakhi and Kuba *uyezds*, as well as the settlements of Khachmaz and Qutqashen (now Qabala—Authors' Note). According to statistical data from the 1880s, the Shamakhi *uyezd* alone accounted for 4,282 heads of camels, with sources noting that for the majority of their owners, *charvadarstvo* constituted their primary occupation (10). These transport operators executed freight haulage between various cities of Azerbaijan. For instance, to meet the needs of the Lahij artisans, thousands of *puds* of copper were annually supplied from Gadabay, Elizavetpol, Zangezur, and other regions via *charvadar* camels, while the copperware manufactured by local masters was sent back on the very same caravans (26).

In the 19th century, there existed a number of villages within the territory of Azerbaijan that numbered up to 50–60 *charvadars*, each owning from 10–13 to, in exceptional cases, 25 heads of camels (26). This fact finds its corroboration in ethnographic materials. According to oral history informants, professional *caravan-bashis* who had 20–30 camels at their disposal operated in the Absheron villages of Amirjan and Guldek (Guzdek).

The caravan routes traversing Azerbaijan since ancient times necessitated the formation of a well-developed network of well-appointed caravanserais. These structures were erected predominantly in major commercial and craft centres, as well as along the primary trade highways. However, with regard to the period under study, data indicating the functioning of caravanserais directly on the trade thoroughfares leading from Baku to Salyan, Tiflis, and Derbent, as well as to Nakhchivan and Iran via Karabakh and Shusha, remain absent (3).

9th-century caravanserais exhibit a significant resemblance to their ancient counterparts. Research indicates that the structural and spatial differences between them were negligible (19).

Such complexes performed the function of hotels (inns) for the rest of caravan personnel, while also serving as fortified sanctuaries that ensured protection against bandit attacks.

The external facades of the caravanserais facing the street were traditionally constructed as blind walls (without window openings). Urban caravanserais were utilised not only for overnight lodging but also functioned as critical nodes for executing various commercial, trade, and financial exchange transactions. Consequently, within urban spaces, they were predominantly localized in bustling commercial districts. Specialized trade stalls (shops) and offices were integrated into the structural layout of these caravanserais, a prime example of which is illustrated by the complexes erected in Baku (19).

As a rule, the infrastructure of all caravanserais—in addition to residential quarters (*hujras*) designated for rest—comprised isolated quarters for riding and pack animals, alongside secure storage facilities for commercial cargo. According to certain sources, two-storey caravanserais operated in a number of regions, wherein warehouses and granaries were situated on the upper level (6).

Bieberstein notes that the most well-appointed caravanserais were situated along the routes between Beshbarmag, Baku, and Salyan. According to his testimony, these structures were equipped with water-bearing wells, enclosed by a perimeter wall of ashlar stone, and featured domed ceilings (3). Ethnographic materials demonstrate that, alongside the permanent service staff, a significant number of *charvadars* (freight transport operators) and *ambals* (porters) were constantly present within the caravanserais.

Conclusion. In drawing together the findings of this study, it is possible to formulate the following synthesis. It must be emphasized that historical caravan routes constituted one of the foundational catalysts in the formation and development of human civilization.

These transport arteries not only ensured the institutionalization of trade links between disparate regions but also created the necessary conditions for the expansion of multi-lateral political, economic, and commercial relations between nations. Throughout history, the caravan itineraries linking East to West and North to South made a substantial contribution to the intensification of inter-civilizational interaction, the diffusion of cultural values, and the development of international relations. In this context, caravan routes possessed not merely a purely economic character but held profound socio-cultural and geopolitical significance in world history.

Owing to its strategic geographical position, Azerbaijan has served since profound antiquity as one of the pivotal transit hubs at the crossroads of international trade routes. The primary caravan itineraries traversing the country's territory exerted a significant influence on the economic and cultural development of various Azerbaijani regions. Via these pathways, silk, spices, precious stones, and other commodities brought from Eastern countries were transported to Western markets; concurrently, European and Near Eastern products entered the region. Consequently, Azerbaijan did not merely function as a passive transit space but emerged as a major epicenter of commercial and economic relations.

The development dynamics of the cities situated along the caravan routes merit particular attention. Historical sources demonstrate that intensive growth in commerce and craft production was observed within the zones adjacent to these routes, while new settlements simultaneously coalesced. Under the influence of international trade ties, the economic consolidation of cities such as Shamakhi, Ganja, Tabriz, Baku, and Nakhchivan, among others, took place, facilitating their transformation into the preeminent cultural centers of the region. Caravanserais, marketplaces, bridges, and fortifications were integral elements of the commercial and economic life of the era under study. These facilities performed not only utilitarian economic functions but played a vital role as spaces for communication and intercultural exchange among representatives of diverse ethnic groups.

One of the most significant characteristics of caravan routes was their function as an instrument of socio-cultural integration. It was precisely through these arteries that the diffusion of religious beliefs, scientific knowledge, literary works, architectural styles, and everyday lifestyles was executed among various peoples. Merchants, travellers, scholars, and artisans moving along caravan itineraries acted as conduits of cultural experience from disparate regions, which in turn determined the processes of mutual influence and enrichment among ethnic groups. In this context, caravan routes represented not merely economic pathways, but a crucial historical mediator that facilitated the formation of the shared cultural heritage of humanity.

At the same time, it is necessary to note the substantial role of caravan routes in the development of inter-state political relations. Control over trade arteries was one of the key factors defining the economic po-

tential and political authority of powers. For this reason, throughout history, various empires and states waged struggles for hegemony over caravan routes, dedicating particular attention to ensuring their security. This, in turn, exerted a direct influence on securing regional political stability and the institutionalization of diplomatic relations.

Conducted research proves that the study of historical caravan routes is of substantial significance not only within the context of retrospective analysis but also for comprehending the nature of contemporary international relations. For centuries, these thoroughfares facilitated the deep rooting of the principles of inter-ethnic cooperation, mutual benefit, and sociocultural integration. To a certain extent, the implementation of modern international transport and logistics projects can be viewed as an institutional continuation of the historical traditions of caravan communication.

Alongside this, the preservation of the tangible cultural heritage associated with caravan routes presents itself as one of the most critical objectives. Historical caravanserais, bridges, ancient marketplaces, and other architectural monuments are invaluable assets reflecting the historical memory of the people. The study, restoration, and transmission of these monuments to future generations are of paramount importance for safeguarding the national cultural domain. Simultaneously, this heritage makes a significant contribution to the development of tourism potential and the promotion of the country on the international stage.

Thus, in world history, historical caravan routes functioned not merely as commercial pathways, but as a vital communication system linking civilizations. Via these arteries, economic cooperation expanded, intercultural ties strengthened, and the process of forming universal human values intensified. Azerbaijan's geographical position at the intersection of these routes played a pivotal role in its historical development, securing for the country a distinct status in the economic, political, and cultural life of the region. By virtue of this, the comprehensive investigation and preservation of the caravan routes' heritage consistently retains its scientific and practical relevance.

References:

1. Авдеев М.П. Юго-Восточная Ширвань и Сальянская степь. Кочевое скотоводство и экономика оседлого населения. Б., 1930, с.49-51, (.232 с.).
2. Алиев К.А. Античная Кавказская албания, Азернешр, Б., 1992. с.113, (с.238).
3. Биберштейн М. Описание провинций, расположенных на левом берегу Каспийского моря между реками Терек и Кура // Az. EA Tarix İnstitutunun Elmi Arxivi, inv. I 466–467; То же // АКАК. — Т. II. — С. 28.
4. Велиев М.Г.(Бахарлы) Азербайджан,(физико-географический, этнографический и экономический очерк). Изд., Азерб. Совета Народного хозяйства. Б., 1921, с.109
5. Велиханова Н.М. О термине сикка в арабской географической литературе. Известия АН Азербайджана, 1989, №3, с.104-107

6. Григорьев В. В. Описание Хивинского ханства и дорог туда из Сарайчиковой крепости // Записки Императорского Русского географического общества. — СПб., 1861. — Кн. 2. — С. 112–115.
7. Ибн ал-Асир, Материалы по истории Азербайджана из Тарих Ал-камиль-АзФан, Б., 1940, с.45,51
8. Ибн-Хордадбех.. Книга путей и стран. Перевод с арабского и коммент. Велихановой Н.М. Б., 1986, с. 107(с.428)
9. Ивановский, А. А. По Закавказью // Материалы по археологии Кавказа. — Москва, 1911. — Вып. 6. — С. 163.
10. Исламова В. М. О овцеводстве и верблюдоводстве на Апшероне // Изв. АН Азерб. ССР. Сер. истории, философии и права. 1971. № 1.
11. Калугин, И. И. Исследование современного состояния животноводства Азербайджана : в 4 т. / И. И. Калугин. — Баку ; Тифлис : Наркомзем АССР, 1929–1930.
12. Керимов, К. Д. Азербайджанские миниатюры - Azərbaycan miniatürləri - Azerbaijan miniature : [альбом] / Керим Керимов. — Баку : Ишыг, 1980. — , илл. 72–74. — 222 с.24)Yenə orada ill.49
13. Лакоза, И. И. Верблюды и его эксплуатация / И. И. Лакоза ; Всесоюз. НИИ коневодства, Центр. Станция Верблюдоводства. — Москва : Гос. изд-во колхозной и совхозной лит. (Сельхозгиз), 1933. — 126 с. — URL: <https://market.yandex.ru/card/lakozai-verblyud-i-yego-ekspluatatsiya/102208750901> (дата обращения: 07.06.2026).
14. Марков Ф.Г. Экономический быт государственных крестьян в Эриванском уезде. Материалы для изучения экономического быта государственных крестьян Закавказского края, т. II, Тифлис, 1886, с.153
15. Марек, Я. Два Пакистана / Я. Марек, Д. Збавител ; пер. с чеш. П. Н. Антонова ; отв. ред. Ю. В. Ганковский ; АН СССР. — Москва : Наука, 1966. — С. 56. — 200 с. — (Путешествия по странам Востока).
16. Обручев, В. А. В дебрях Центральной Азии : (записки кладоискателя) / В. А. Обручев ; оформление художника Г. А. Петрова. — Москва : Государственное издательство географической литературы (Географгиз), 1953. — С. 40. — 320 с.
17. Пиотровский, Б. Б. Ванское царство (Урарту) / Б. Б. Пиотровский ; Акад. наук СССР, Ин-т истории материальной культуры. — Москва : Издательство восточной литературы, 1959. — С. 157. — 286 с.
18. Пириев, В. Взаимоотношения Азербайджана с соседними странами в XIII–XIV вв. / В. Пириев // Доклады Бакинского международного симпозиума «Азербайджан в международных экономических и культурных взаимосвязях». — Баку : [б. и.], 1995. — С. 53.
19. Саламзаде А.С, Архитектура Азербайджана XVI–XIX вв. (автореферат). М., 1963
20. Сумбатзаде, А. С. Сельское хозяйство Азербайджана в XIX в. / А. С. Сумбатзаде ; Академия наук Азербайджанской ССР, Институт истории. — Баку : Издательство Академии наук Азербайджанской ССР, 1958. — С. 87, таб. 10. — 364 с.
21. Страбон. География в 17 книгах / Страбон ; перевод, статья и комментарии Г. А. Стратановского ; под общей редакцией проф. С. Л. Утченко ; редактор перевода проф. О. О. Крюгер ; Академия наук СССР. — Москва : Наука, 1964. — С. 180. — 944 с. — (Классики науки).
22. Халледулах Казвини, Нузхат ал-Кулуб (Материалы по Азербайджану), Б.1983, с.55–56
23. Хан-Агов А. Е. Экономический быт государственных крестьян в Эчмиадзинском и Сурмалинском уездах Эриванской губернии // Материалы для изучения экономического быта государственных крестьян Закавказского края. Т. III. Тифлис, 188
24. Элиан, Клавдий. О природе животных (Фрагменты) [Электронный ресурс] / Клавдий Элиан ; пер. с древнегреч. С. В. Поляковой. — URL: simposium.ru (дата обращения: 07.06.2026).
25. Ягодзинский Т. Н. Экономический быт государственных крестьян в Бакинском уезде // Материалы для изучения экономического быта государственных крестьян Закавказского края. Т. II, ч. I. Тифлис, 1886. С. 419, 512–517.
26. Ямпольский .З.И., Путешественники об Азербайджане. Т. I / сост. З. И. Ямпольский, под ред. Э. М. Шахмалиева. Баку: Изд-во АН Азерб. ССР, 1961. С. 86.
27. Bünyadov, T. Ə. Azərbaycan maldarlığın inkişafı tarixindən / T. Ə. Bünyadov ; Azərbaycan SSR Elmlər Akademiyası, Tarix İnstitutu. — Bakı : Elm, 1969. — S. 93. — 187 s.
28. Bünyadov, Z. M. Azərbaycan VII–IX əsrlərdə / Z. M. Bünyadov. — Bakı : Azərənşr, 1989. — S. 150. — 334 s
29. Əliyev, F. M. XVIII əsrin ikinci yarısında Şimali Azərbaycan şəhərləri / F. M. Əliyev ; Azərbaycan SSR Elmlər Akademiyası, Tarix İnstitutu. — Bakı : Azərbaycan SSR Elmlər Akademiyası Nəşriyyatı, 1964. — S. 89. — 136 s.
30. Gəncəvi, Nizami. İskəndərnamə / Nizami Gəncəvi ; Azərbaycan SSR Elmlər Akademiyası. — Bakı : Azərənşr, 1941. — S. 122. — 612 s.
31. Kərimov, T. M. XIX əsrdə Azərbaycanda karvan (dəvə) nəqliyyatı haqqında / T. M. Kərimov // Azərbaycan SSR Elmlər Akademiyasının Xəbərləri. Tarix, fəlsəfə və hüquq seriyası. — 1978. — № 4. — S. 96.
32. Mustafayev, A. N. Şirvanda xalq nəqliyyat vasitələri / A. N. Mustafayev // Azərbaycan etnoqrafik məcmusi. — III buraxılış. — Bakı : Elm, 1977. — S. 126. Onullahi S.M. XIII–XVII əsr Təbriz şəhərinin tarixi. Bakı: Elm, 1982. S. 142–146.
33. Dostiyev T.M. IX–X əsrlərdə xilafətin şimal ölkələri ilə ticarətində Azərbaycanın rolu // Azərbaycan beynəlxalq iqtisadi və mədəni qarşılıqlı əlaqələrdə. — Bakı, 1995. — S. 44–49.
34. Vəlixanlı, N. M. Ərəb xilafəti və Azərbaycan / N. M. Vəlixanlı. — Bakı : Azərənşr, 1993. — S. 115–125. — 157 s.
35. Vəlixanlı, N. M. IX–XII əsr ərəb coğrafiyaşünas-səyyahları Azərbaycan haqqında / N. M. Vəlixanlı ; Azərbaycan SSR Elmlər Akademiyası, Şərqşünaslıq İnstitutu. — Bakı : Elm, 1974. — 220 s. (49)Vəlixanlı N.M. ərəb xilafəti və Azərbaycan s.116

MATHEMATICAL SCIENCES

MULTIVARIABLE NONLINEAR REGRESSION FOR ANALYSING ADSORPTION ISOTHERM MODELS - A CASE STUDY

Berdenova B.

Faculty of Mechanics and Mathematics, Al-Farabi Kazakh National University, 71 Al-Farabi ave., Almaty, 050040, Kazakhstan

RAND Limited Liability Partnership, 17 Mangilik Yel ave., Astana, 010000, Kazakhstan

Orazaly D.

Faculty of Mechanics and Mathematics, Al-Farabi Kazakh National University, 71 Al-Farabi ave., Almaty, 050040, Kazakhstan

Nessipbekov G.

RAND Limited Liability Partnership, 17 Mangilik Yel ave., Astana, 010000, Kazakhstan

Bidyut Baran Saha

International Institute for Carbon-Neutral Energy Research (WPI-I2CNER), Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395, Japan

Mechanical Engineering Department, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395, Japan

<https://doi.org/10.5281/zenodo.21105815>

Abstract

This paper details the calculation steps, from post-processing experimental data to predicting cooling efficiency of activated carbon (AC)/CO₂-based adsorption system for specified design temperatures. Regression analysis was conducted using advanced Python packages. The results identified the best model to describe adsorption isotherms, with optimal parameters verified using alternative methods implemented via Excel's solver. The study considered a nonlinear equations with two independent variables, specifically Langmuir and T'oth adsorption isotherms, to fit thermogravimetric data representing six equilibrium uptake curves measured at various temperatures and pressures for the CO₂/activated carbon working pair. Physical adsorption is highly relevant for industries releasing low-temperature heat to the environment. The derived model parameters facilitated the estimation of the isosteric heat of adsorption and the development of a pressure-temperature equilibrium uptake (PTW) diagram, essential for predicting the performance of ideal adsorption cooling cycles. The specific cooling effect (SCE) and coefficient of performance (COP) were determined for selected design temperature conditions.

Keywords: Adsorbent materials, Langmuir isotherm, Nonlinear regression, T'oth isotherm

1. Introduction

Global scientific efforts increasingly focus on developing and promoting carbon-free or carbon-neutral technologies. Also of particular interest are autonomous systems that do not depend on the central power supply. Adsorption refrigerators, which operate waste-free using thermal energy sources such as solar power, are a key component of this movement. Several widely used solid static adsorbents, including activated carbon, zeolites, and metalorganic frameworks, play a crucial role in adsorption cooling systems (ACS). Simultaneously, ongoing research aims to enhance these materials' adsorption capacity and thermal efficiency and optimize the compactness and design of adsorption beds and the overall system performance.

Driving temperature, evaporation temperature, coefficient of performance (COP), and specific cooling power (SCP) are basic parameters for choosing a working adsorbent-refrigerant pair for the ACSs [1]. Let's consider a few working pair cases with application limits and desirable characteristics.

Activated carbon (AC)/methanol is the most common working pair [1]. For example, in [2], the basics of a solar-powered adsorption ice-maker were presented. The authors conclude that the adsorption pair of activated carbon and methanol is preferred more than the water/ammonia pair. The limitation while applying

methanol CH₃OH is that the methanol decomposes at 120°C, so the desorbing temperature shouldn't exceed this value. The experimental prototype had a solar refrigeration COP of 0.10-0.12 and produced 4-5 kg of ice per day from the flat-plate collector area of 0.92 m². Another effect on methanol decomposition is the alloys used in system construction, as the methanol contacts with it while operating [3]. Aluminum alloys were revealed to have a stronger catalytic effect on the decomposition reaction, so copper alloys are preferred. In such systems, the methanol is in the 0.02-0.2 bar pressure range and at -10°C to 110°C temperature range. In addition, some demonstration systems exhibit very good performance when they are newly built, and performance declines after a year or less of operation. The performance decline is associated with the loss of the vacuum in the system, so re-evacuation, or in other words, repeated vacuuming on an annual basis, is necessary to restore system functionality. The long-term performance deterioration is the main reason behind the cessation of solar refrigerator commercialization. Still, solar-powered refrigeration systems are the most promising technology, with low cost, simple manufacturing, and low maintenance requirements.

Activated carbon (AC)/ammonia (NH₃) pair demonstrates high COP that may reach 0.45 for ice-making and 0.61 for air-conditioning [4]. The highest

SCP (of order $60 - 200 \text{ W kg}^{-1}$ for different systems) is achieved with an AC/ammonia pair. Ammonia's thermodynamic properties make it suitable for heating and cooling applications, but it is also a toxic and corrosive substance.

Cryocooler is a refrigerator that can reach temperatures below -153°C (or 120 K). Nitrogen as a refrigerant with ACF (Activated carbon fibers) as an adsorbent is good for cryocooling applications. The adsorption capacity is about 0.75 g g^{-1} at an adsorption temperature of -196°C . The adsorption capacity with operating temperature varies. For typical cooler applications with relatively higher temperatures, nitrogen has a low adsorption capacity. For example, the highest adsorption capacity for the nitrogen/AC pair was 0.00756 g g^{-1} at adsorption temperature and pressure of 30°C and 100 kPa, respectively.

AC/R134a pair has the highest adsorption capacity of 2 g g^{-1} at adsorption temperature and pressure of 30°C and 800 kPa. The boiling temperature of R134a is (-26.55°C). R-134a is a hydrofluorocarbon $\text{CF}_3\text{CH}_2\text{F}$ with insignificant ODP (ozone depletion potential) and a relatively lower GWP (global warming potential). The gas is non-flammable and primarily used in household refrigerators with conventional vapor-compression cycles.

AC/ CO_2 CO_2 pair is about a relatively high-pressure system. The highest adsorption capacity at 30°C and 100 kPa is only 0.0844 g g^{-1} , so the higher working pressures preferred, say at 4586 kPa and 25°C the adsorption capacity was 1.135 g g^{-1} [1]. The thermodynamics performance of the CO_2 /activated carbon (Maxsorb III-based composite) pair is of order $\text{COP} = 0.16$ for the system with the evaporator temperature $T_{\text{evap}} = 5^\circ\text{C}$ and the condenser temperature $T_{\text{cond}} = 25^\circ\text{C}$ [5]. CO_2 is a non-toxic and non-flammable natural refrigerant. Maxsorb III is a highly recognized material for good thermal properties and high capacity among activated carbon samples.

A new class of microporous materials, PCPs (porous coordination polymers) and MOFs (metal-organic frameworks) have recently become available as alternatives to AC. There is another class of carbon-based adsorbents. In [6], the authors examined one of them - nitrogen-enriched nanostructured carbon (UFZ-700). Unlike activated carbon, this type of carbon can be derived using inexpensive, widely available sources. MOFs show a huge surface area and large pore volume. The surface areas of typical MOFs range from 1000 to

$10,000 \text{ m}^2 \text{ g}^{-1}$, which can exceed those of traditional porous materials such as zeolites and carbons. Also, the average pore size is tunable due to a much wider variety of chemical compositions. Table 1 illustrates some common microporous materials for comparison.

Thus, the main pros and cons of popular adsorbent-adsorbate pairs for ACSs are listed above. Additionally, a brief summary of the necessary features is provided when choosing a working pair based on the intended installation environment for the system as well as the target temperature in the cooled chamber. A number of factors that have a long-term cumulative impact on the refrigerator's operation are also covered, including toxicity, corrosion, leak proofness, vacuum tightness, maximum adsorption capacity deterioration, explosiveness risks due to high working pressures, etc.

There is endless room for improvement in the various heating and cooling systems designed for different purposes. Such systems usually contain many parts that can be considered and studied separately. The system performance can be enhanced by employing multi-beds and a complex piping network. Advancing the cycles from single to two-bed configurations may improve the COP by at least 35% and, at most, even double [4]. In addition, new materials are developed by combining or/and compressing various existing materials to improve the compactness and characteristics of adsorption beds in thermal systems.

The current paper describes the calculation steps from the post-processing of experimental data to predicting the cooling efficiency of ACS. Part of the regression analysis was performed using emerging capabilities of Python packages. Alternative methods were implemented using the Excel solver. Based on the previous work done by various authors, this paper also presents an approach to calculating the cooling efficiency parameters, SCE, and COP. The cooling cycle performance of any solid-gas adsorption system is normally studied using the PTW (pressure-temperature-concentration) diagram [7]. The derived model parameters allowed the development of a PTW-graph diagram, essential for predicting the performance of ideal adsorption cooling cycles. Thermodynamic analysis of CO_2 /AC-based cooling systems is well discussed in [8]. Another motivation for research work is that the established isotherm model is also used in simulation studies of the dynamics of the adsorption process. Here are some examples of simulation studies [9, 10].

Table 1.

Porous properties of different types of microporous materials

Adsorbent	SSA [m ² g ⁻¹]	Pore volume [cm ³ g ⁻¹]	Average pore di- ameter, d _{ave} [nm]	Thermal stable up to	References
MOF-NU-1000	2320	1.4	0.8-3.1		[11]
MOF-5	1077	0.49	1.8	400 °C	[12]
ZIF-8	1700	0.662	1.2	550 °C	[13, 14]
Activated car- bon Maxsorb III	3140	1.7	2.008		[15]
Activated car- bon fiber ACFC- 1500	1500	0.62	1.69		[16]
Zeolite Na-P1CS	53.5	0.13	11.1		[17]
Natural zeolite ZN	27	0.1	17		[18]

2. Adsorption isotherm models

Quoting the work [19], the Langmuir isotherm model is the most straightforward and practical for both physical and chemical adsorptions. Pressure p and temperature T are two independent variables of Langmuir Equation 1 and To'th Equation 2 isotherm models. Equilibrium adsorption C is the dependent variable, and is the amount adsorbed [g g⁻¹]. The Langmuir model has three fitting parameters, while the To'th model has four. The list of correlated parameters or regression features is shown in Table 2, Table 3. Theoretically, at zero pressures, absorption is zero; therefore, zero points are added to the calculation.

$$\frac{C}{C_0} = \frac{b_0 \exp\left(\frac{E}{RT}\right) p}{1 + b_0 \exp\left(\frac{E}{RT}\right) p} \quad (1)$$

$$\frac{C}{C_0} = \frac{b_0 \exp\left(\frac{E}{RT}\right) p}{\left(1 + \left(b_0 \exp\left(\frac{E}{RT}\right) p\right)^t\right)^{1/t}} \quad (2)$$

The coefficient of determination (R^2), Equation 3, and the root-meansquare deviation (RMSD), Equation

4, are used as a measure of prediction error to compare both methods.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \cdot 100\% \quad (4)$$

The experimental data represents 6 equilibrium uptake curves measured with varying temperatures and pressures for activated carbon/carbon dioxide working pair, see Figure 1.

Python (Version 3.11.5) is used for calculations. Optimal parameters of the isotherm models were obtained using two alternative methods. The calculation results in Table 2, Table 3 revealed that the To'th model better describes adsorption isotherms with the lowest root mean square deviation of 0.441%. The surface plot illustrated in Figure 2. The side view demonstrated in Figure 3 shows fitted adsorption isotherms and experimental data.

Table 2.

Optimal parameters of To'th isotherm obtained using Method 1: Python curvefit and Method 2: Excel Solver

Estimated parameters	Unit	Value Method 1	Value Method 2
Saturated amount ad- sorbed, C_0	[g g ⁻¹]	3.625	2.6307
Isosteric heat of adsorp- tion, E	[kJ mol ⁻¹]	18.27	19.024
Equilibrium constant, b_0	[kPa ⁻¹]	2.112E-07	1.887E-07
Heterogeneity parame- ter, t	[-]	0.4107	0.4774
RMSD	[%]	0.441	0.5635
Coefficient of determina- tion, R^2	[-]	0.99964	0.99940

Table 3.

Optimal parameters of Langmuir isotherm obtained using Method 1 – Python curve-fit, Method 2 – Excel Solver

Estimated parameters	Unit	Value Method 1	Value Method 2
Saturated amount adsorbed, C_0	[g g ⁻¹]	1.274	1.1357
Isosteric heat of adsorption, E	[kJ mol ⁻¹]	18.652	20.722
Equilibrium constant, b_0	[kPa ⁻¹]	2.096E-07	1.217E-07
RMSD	[%]	1.426	1.877
Coefficient of determination, R^2	[-]	0.99630	0.99357

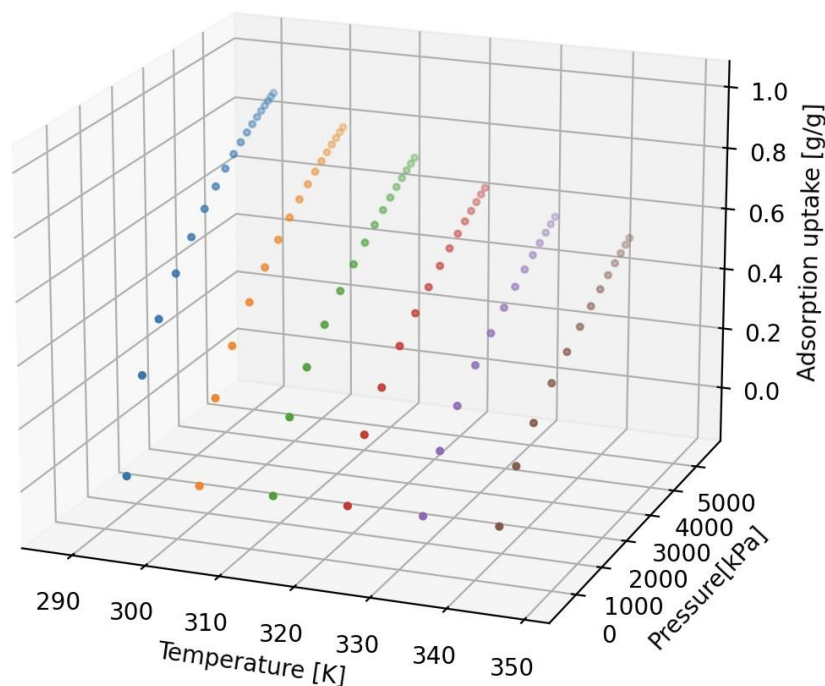


Fig. 1. Equilibrium uptake data for activated carbon/CO₂ pair

3. Clapeyron diagram of ideal solid-adsorption cycle

The PTW graph shows the thermodynamic relation between equilibrium pressure, adsorption temperature, and adsorption uptake (see Figure 4). The graph is essential for predicting the ideal adsorption cooling cycle performance, so to say, to calculate SCE and COP. Thermodynamic analysis of AC/CO₂-based cooling systems is well discussed in [8]. Let's consider the case of temperatures closer to those taken in the paper. The waste heat or solar energy as a heat source is supplied at 80°C, and the heat sink temperature is 30°C

(maximum and minimum temperatures in the collector).

The evaporator temperature is $T_{\text{evap}} = 5^\circ\text{C}$, the condenser temperature

$T_{\text{cond}} = 25^\circ\text{C}$, respectively, the evaporator pressure $P_{\text{evap}} = 3,9695 \text{ MPa}$, and the condenser pressure $P_{\text{cond}} = 6,4342 \text{ MPa}$. The amount of adsorbate per adsorption cooling cycle varies from W_{max} to W_{min} . W_{max} (adsorption) is measured as a function of P_{evap} (which is the saturation pressure of the cooling load temperature T_{evap}), and the heat sink temperature $T_{\text{ads}} = 30^\circ\text{C}$

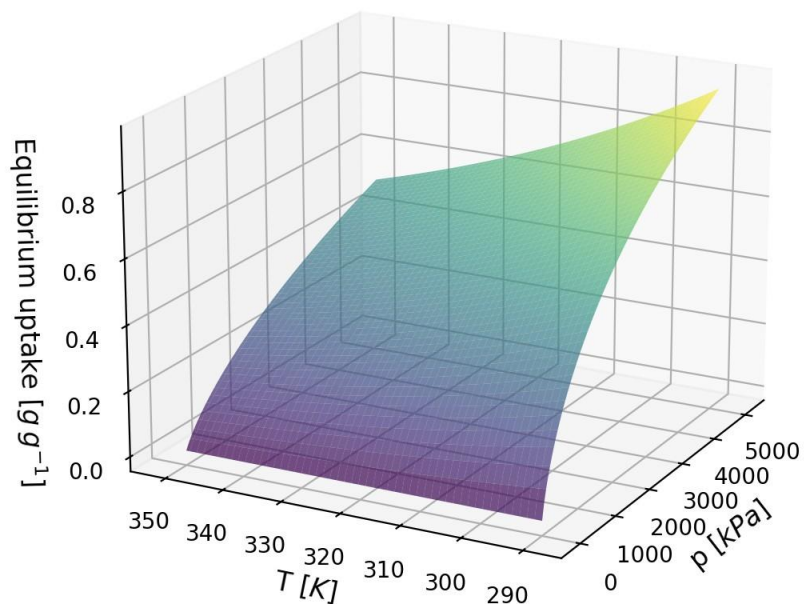


Fig. 2. Surface plot using the To'th model

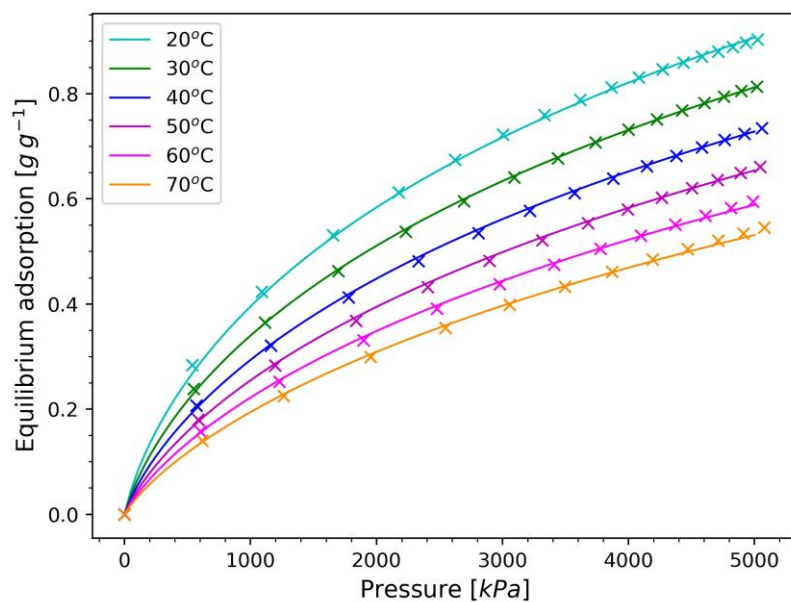


Fig. 3. To'th isotherm model along with the experimental absolute adsorption uptake.

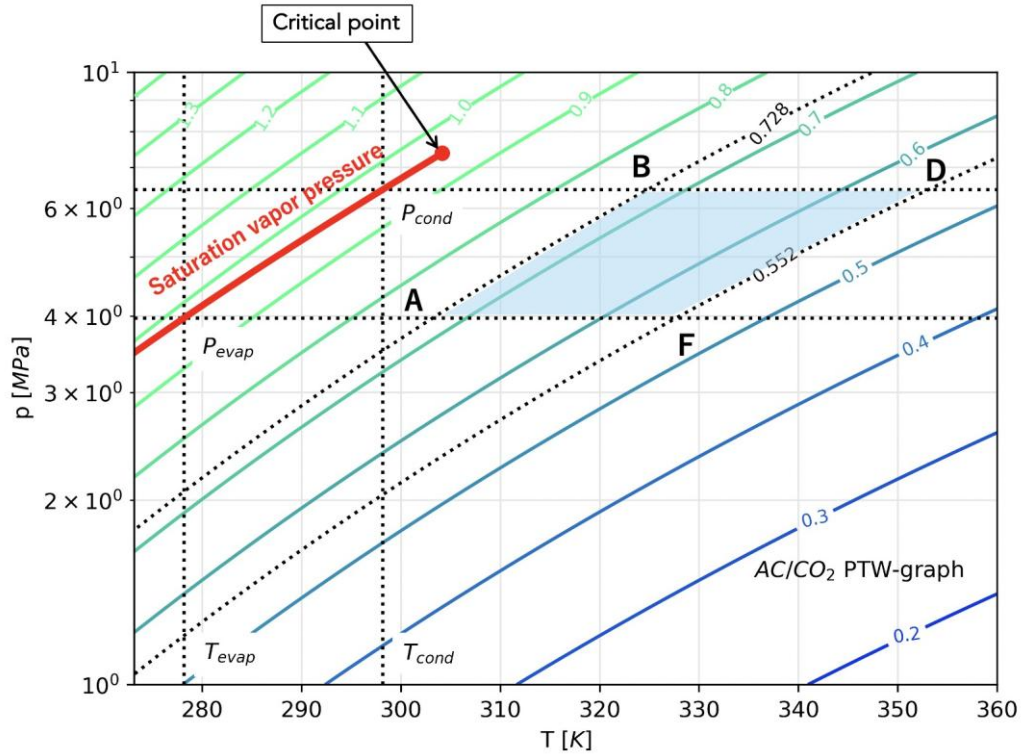


Fig. 4. PTW-graph for AC/CO₂ working pair

[20, 21]. By analogy, $W_{\min}$ (desorption) is found as a function of P_{cond} (which is the saturation pressure of condenser temperature $T_{\text{cond}} = 25^\circ\text{C}$), and the heat source temperature $T_{\text{des}} = 80^\circ\text{C}$. Maximum and minimum uptakes during the working cycle are $W_{\max} = 0,7282\text{gg}^{-1}$ and $W_{\min} = 0,5517\text{gg}^{-1}$. Carbon dioxide has a critical temperature of $31,1^\circ\text{C}$ (or $304,13\text{K}$) and a critical pressure of $7,377\text{MPa}$.

$$W_{\max} = f(p_{\text{sat}}(T_{\text{evap}}, T_{\text{ads}})) \quad (5)$$

$$W_{\min} = f(p_{\text{sat}}(T_{\text{cond}}, T_{\text{des}})) \quad (6)$$

The enthalpy of vaporization, also known as the latent heat of vaporization for carbon dioxide at 3.9695MPa and at the boiling point of $T_{\text{evap}} = 5^\circ\text{C}$ is $\Delta H_{\text{vap}} = 2.94 \times 10^5 \text{J kg}^{-1}$ [22]. The specific heat capacity of the adsorbent (composite AC) was taken as a constant for the range of temperature utilized, $c_{p,\text{AC}} =$

$0.8212 \text{J g}^{-1} \text{K}^{-1}$. Latent heat of desorption (theoretically, should be the same as the heat of adsorption, E), $H_{\text{des}} = 18.27 \text{kJ mol}^{-1}$. The specific heat capacity of refrigerant (liquid CO₂), $c_{p,\text{ref}}$ is taken as a function of temperature [7, 22].

The Clapeyron diagram of the ideal solid-adsorption cycle consists of 4 stages, see Figure 4:

- 1. A→B Isosteric pre-heating
- 2. B→D Isobaric desorption
- 3. D→F Isosteric pre-cooling
- 4. F→A Isobaric adsorption.

The specific cooling power (SCP) has a unit of $[\text{W kg}^{-1}]$ and is calculated differently; do not confuse it with the SCE. SCP is the ratio of cooling energy to the product of adsorbent mass and cycle time [23]. The specific cooling effect (SCE) is given by Equation 7 and found to be 44.7kJ kg^{-1} . The SCE is the same as the total energy output Q_C in some other papers, both are used equally.

$$SCE = (W_{\max} - W_{\min}) \left[\Delta H_{\text{vap}} - \int_{T_{\text{evap}}}^{T_{\text{ads}}} c_{p,\text{ref}} dT \right] \quad (7)$$

The system efficiency, coefficient of performance:

$$COP = \frac{Q_C}{Q_T} \quad (8)$$

$$Q_T = Q_{A-B} + Q_{B-D} \quad (9)$$

$$Q_{A-B} = \int_{T_A}^{T_B} c_{p,ads} dT + W_{max} \int_{T_A}^{T_B} c_{p,ref} dT \quad (10)$$

$$Q_{B-D} = \int_{T_B}^{T_D} c_{p,ads} dT + \int_{T_B}^{T_D} W \cdot c_{p,ref} dT + (W_{max} - W_{min}) H_{des} \quad (11)$$

COP value is found to be 0.27.

Table 4.

CO₂ gas-specific isobar heat capacity vs. temperature and pressure

	Temperature T [degC]	Temperature T [K]	Pressure p [MPa]	Specific Heat c_p [kJ kg ⁻¹ K ⁻¹]
A	30	303.15	3.9695	1.3380405
B	51.6	324.75	6.4342	1.6562389
D	80	353.15	6.4342	1.3050357
F	54.57	327.72	3.9695	1.1640829

4. Conclusions

The present study evaluates the performance efficiency of the novel material in an adsorption refrigerator based on the results of the TGA experiment. It specifically examines the case of carbon dioxide and activated carbon as a working pair.

Firstly, the equilibrium uptake parameters were correlated using a multivariable nonlinear regression method and two alternative calculation tools. The study results are in agreement with each other, and the results are given in the literature on similar systems. The choice of loss function or inner hyperparameter set may cause negligible deviations. The best fit was achieved when using Python curve-fit, with RMSD of 0.441% for To'th isotherm.

Secondly, performance metrics, such as specific cooling effect and coefficient of performance, are discussed in relation to prior research and projected for selected operating temperatures. SCE and COP are predicted to be 44.7 kJ kg⁻¹ and 0.27, respectively. The PTW diagram for the composite AC/CO₂ pair is constructed. The stages of the ideal cooling cycle are described in detail to better understand the fundamentals of solid-gas ACS's operation principle. Thus, the study details the calculation of cooling efficiency, encompassing the progression from post-processing experimental data to forecasting outcomes for specific system design temperatures.

Thirdly, the paper illustrates the potential for more precise computations with the help of available and modern tools. Most engineers and researchers have encountered similar optimization problems at some point or another, and Python packages for resolving them are rapidly evolving. It is noteworthy to mention the significance of the completed review on adsorption-based

working pairs and related features during use. The present results are relevant as a reference for future related studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The author (BB) wishes to acknowledge the financial support received from JSC Center for International Programs – Bolashaq, and the research was performed on scope of 500 scholars fellowship program.

References:

1. Askalany AA, Salem M, Ismael I, Ali A, Morsy M, Saha BB. An overview on adsorption pairs for cooling. *Renewable and Sustainable Energy Reviews* 2013;19:565–72. URL: <https://www.sciencedirect.com/science/article/pii/S1364032112006491>. doi:<https://doi.org/10.1016/j.rser.2012.11.037>.
2. Sumathy K, Zhongfu L. Experiments with solar-powered adsorption ice-maker. *Renewable Energy* 1999;16(1):704–7. URL: <https://www.sciencedirect.com/science/article/pii/S0960148198002560>. doi:[https://doi.org/10.1016/S0960-1481\(98\)00256-0](https://doi.org/10.1016/S0960-1481(98)00256-0); *renewable Energy Energy Efficiency, Policy and the Environment*.
3. Hu EJ. A study of thermal decomposition of methanol in solar powered adsorption refrigeration systems. *Solar Energy* 1998;62(5):325–9. URL: <https://www.sciencedirect.com/science/article/pii/S0038092X98000127>. doi:[https://doi.org/10.1016/S0038-092X\(98\)00012-7](https://doi.org/10.1016/S0038-092X(98)00012-7).
4. Tamainot-Telto Z, Metcalf S, Critoph R, Zhong Y, Thorpe R. Carbon-ammonia pairs for ad-

sorption refrigeration applications: ice making, air conditioning and heat pumping. *International Journal of Refrigeration* 2009;32(6):1212–29. URL: <https://www.sciencedirect.com/science/article/pii/S0140700709000061>.

doi:<https://doi.org/10.1016/j.ijrefrig.2009.01.008>.

5. Rupa MJ, Pal A, Saha BB. Activated carbon-graphene nanoplatelets based green cooling system: Adsorption kinetics, heat of adsorption, and thermodynamic performance. *Energy* 2020;193:116774. URL: <https://www.sciencedirect.com/science/article/pii/S0360544219324697>. doi:<https://doi.org/10.1016/j.energy.2019.116774>.

6. Tiwari D, Bhunia H, Bajpai PK. Adsorption and thermodynamic studies of pure and binary CO₂ and N₂ gas components on nitrogen enriched nanostructured carbon adsorbents. *The Journal of Chemical Thermodynamics* 2018;125:205–13. URL: <https://www.sciencedirect.com/science/article/pii/S0021961418303835>. doi:<https://doi.org/10.1016/j.jct.2018.06.009>.

7. Jribi S, Chakraborty A, El-Sharkawy II, Saha BB, Koyama S. Thermodynamic analysis of activated carbon-CO₂ based adsorption cooling cycles. *International Journal of Mechanical, Industrial and Aerospace Sciences* 2008;1(7). doi:<https://doi.org/10.5281/zenodo.1328392>.

8. Singh VK, Kumar EA. Experimental investigation and thermodynamic analysis of CO₂ adsorption on activated carbons for cooling system. *Journal of CO₂ Utilization* 2017;17:290–304. URL: <https://www.sciencedirect.com/science/article/pii/S2212982016302499>. doi:<https://doi.org/10.1016/j.jcou.2016.12.004>.

9. Berdenova B. Simulation of carbon dioxide adsorption onto consolidated activated carbon in 2d axisymmetric system. *Journal of Mathematics, Mechanics and Computer Science* 2024;121(1):89–98. doi:<https://doi.org/10.26577/JMMCS202412119>.

10. Pongtornkulpanich A. Dynamic simulation of solid adsorption solar refrigerator system with ac/CH₃OH as a working pair. *Energy and Power Engineering* 2014;6(12):459–65. doi:<https://doi.org/10.4236/epe.2014.612038>.

11. Kinik FP, Ortega-Guerrero A, Ongari D, Ireland CP, Smit B. Pyrenebased metal organic frameworks: from synthesis to applications. *Chem Soc Rev* 2021;50:3143–77. URL: <http://dx.doi.org/10.1039/D0CS00424C>. doi:<https://doi.org/10.1039/D0CS00424C>.

12. Peng MM, Jeon UJ, Ganesh M, Aziz A, Vinodh R, Palanichamy M, et al. Oxidation of ethylbenzene using nickel oxide supported metal organic framework catalyst. *Bulletin of the Korean Chemical Society* 2014;35(11):3213–8. doi:<https://doi.org/10.5012/bkcs.2014.35.11.3213>.

13. Bergaoui M, Khalfaoui M, Awadallah-F A, Al-Muhtaseb S. A review of the features and applications of ZIF-8 and its derivatives for separating CO₂ and isomers of C₃- and C₄- hydrocarbons. *Journal of Natural Gas Science and Engineering* 2021;96:104289. URL: <https://www.sciencedirect.com/science/article/pii/S187551002100487X>. doi:<https://doi.org/10.1016/j.jngse.2021.104289>.

14. Benedetti FM, Angelis MGD, Esposti MD, Fabbri P, Masili A, Orsini A, et al. Enhancing the separation performance of glassy PPO with the addition of a molecular sieve (ZIF-8): Gas transport at various temperatures. *Membranes* 2020;10(4). doi:<https://doi.org/10.3390/membranes10040056>.

15. Ghazy M, Harby K, Askalany AA, Saha BB. Adsorption isotherms and kinetics of activated carbon/difluoroethane adsorption pair: Theory and experiments. *International Journal of Refrigeration* 2016;70:196–205. URL: <https://www.sciencedirect.com/science/article/pii/S0140700716000189>. doi:<https://doi.org/10.1016/j.ijrefrig.2016.01.012>.

16. Balanay J, Lungu CT. Morphologic and surface characterization of different types of activated carbon fibres. *Adsorption Science and Technology* 2012;30(4):355–67. doi:<https://doi.org/10.1260/0263-6174.30.4.355>.

17. Kołodyn'ska D, Ju Y, Franus M, Franus W. Zeolite NaP1 functionalization for the sorption of metal complexes with biodegradable n-(1,2dicarboxyethyl)-D,L-aspartic acid. *Materials* 2021;14(10). URL: <https://www.mdpi.com/1996-1944/14/10/2518>. doi:<https://doi.org/10.3390/ma14102518>.

18. De Souza VC, Villarroel-Rocha J, De Araujo MJG, Sapag K, Pergher SBC. Basic treatment in natural clinoptilolite for improvement of physicochemical properties. *Minerals* 2018;8(12). URL: <https://www.mdpi.com/2075-163X/8/12/595>. doi:<https://doi.org/10.3390/min8120595>.

19. Benkhedda J, Jaubert JN, Barth D, Perrin L, Bailly M. Adsorption isotherms of m-xylene on activated carbon: measurements and correlation with different models. *The Journal of Chemical Thermodynamics* 2000;32(3):401–11. URL: <https://www.sciencedirect.com/science/article/pii/S0021961499906134>. doi:<https://doi.org/10.1006/jcht.1999.0613>.

20. Fan W, Chakraborty A, Kayal S. Adsorption cooling cycles: Insights into carbon dioxide adsorption on activated carbons. *Energy* 2016;102:491–501. URL: <https://www.sciencedirect.com/science/article/pii/S0360544216301748>. doi:<https://doi.org/10.1016/j.energy.2016.02.112>.

21. Miyazaki T, Saha BB, and SK. Analytical model of a combined adsorption cooling and mechanical vapor compression refrigeration system. *Heat Transfer Engineering* 2017;38(4):423–30. URL: <https://doi.org/10.1080/01457632.2016.1195135>. doi:<https://doi.org/10.1080/01457632.2016.1195135>. arXiv:<https://doi.org/10.1080/01457632.2016.1195135> 5.

22. 2024;URL: <https://www.carbon-dioxide-properties.com/co2tablesweb.aspx>.

23. Bahrehmand H, Bahrami M. Optimized sorber bed heat and mass exchangers for sorption cooling systems. *Applied Thermal Engineering* 2021;185:116348. URL: <https://www.sciencedirect.com/science/article/pii/S1359431120338266>. doi:<https://doi.org/10.1016/j.applthermaleng.2020.116348>.

A NOVEL APPROACH TO APPLYING BAYES' THEOREM IN MACHINE LEARNING AND PROGRAMMING

Karimova U.

PhD in Mathematics

Azerbaijan Technical University, Baku, Azerbaijan

Akhundov R.

Student of the Faculty of Information Technologies and

Telecommunications

Azerbaijan Technical University, Baku, Azerbaijan

Abdullayev K.

Doctor of Sciences in Economics, D.Sc. (Economics),

Institute of Economics of the Ministry of Science and Education of the Azerbaijan Republic, Baku, Azerbaijan

<https://doi.org/10.5281/zenodo.21105817>

Abstract

Probabilistic approaches constitute an essential component of Machine Learning, particularly in scenarios involving uncertainty and data-driven inference. Bayes' Theorem provides a principled framework for incorporating prior knowledge and updating beliefs as new data become available. This paper investigates the Naive Bayes algorithm from both theoretical and practical perspectives, emphasizing its mathematical foundation and applicability to real-world classification problems. Different variants of Naive Bayes, including Gaussian, Multinomial, Bernoulli, and Complement models, are analyzed with respect to their suitability for various data representations. The study focuses primarily on text classification tasks, where Naive Bayes is combined with TF-IDF-based feature extraction. Experimental evaluation is performed using the IMDB Movie Reviews dataset, and the results are compared with Logistic Regression and Support Vector Machine classifiers. The findings indicate that Naive Bayes delivers competitive performance with significantly lower computational cost, making it well suited for large-scale and real-time text processing applications. Despite its simplifying assumptions, the algorithm remains an effective and interpretable baseline for Machine Learning classification tasks.

Keywords: Machine Learning, Naive Bayes, Bayesian Inference, Text Classification, TF-IDF, Sentiment Analysis, IMDB Dataset

Introduction.

Machine Learning (ML) is one of the scientific fields that forms the core foundation of modern artificial intelligence systems. Over recent decades, the development of ML in both theoretical and practical domains has radically transformed the way humans work with data. Today, the application of Machine Learning models is rapidly expanding across numerous fields, including healthcare, education, finance, cybersecurity, natural language processing, robotics, and many others. The effective operation of these models largely depends on methodologies based on statistics and probability theory. Statistical analysis, modeling of uncertainty, and data-driven decision-making constitute the theoretical backbone of Machine Learning. The primary goal of Machine Learning is to extract patterns from data, learn from them, and provide accurate predictions about future events. This learning process involves not only the application of simple rules but also the correct calculation of probabilities and the discovery of hidden relationships within data. For this reason, statistical methods are regarded as an integral part of ML. In the analysis of large-scale data accumulated over time, probability distributions, statistical hypotheses, random variables, and Bayesian approaches play a crucial role. In reality, data are almost always observed under uncertainty. For example, in medical diagnostics, even if symptoms of a disease are observed in a patient, a definitive diagnosis can only be provided through

probabilistic approaches. Probability-based models enable ML systems to understand the uncertain nature of data and to make mathematically justified decisions based on that data. Through these models, systems evaluate the likelihood that newly incoming data belong to particular categories by relying on previous observations and probability values.

Bayes' theorem is based on the calculation of conditional probabilities and enables the updating of prior knowledge as new information becomes available. In many modern ML methods especially probabilistic models, generative models, and tasks such as classification and segmentation—approaches based on Bayes' theorem are widely used. Bayesian-based methods allow models to be updated in accordance with constantly changing data environments in real-world scenarios. The Naive Bayes method is a simple yet extremely powerful classification algorithm based on Bayes' theorem. Although the assumption of complete independence among features is rarely satisfied in real life, this model often demonstrates very high performance in practice. Particularly in areas such as text classification, spam filtering, sentiment analysis, and medical diagnostics, the speed and efficiency of Naive Bayes have led to its widespread adoption. Today, the Naive Bayes model remains relevant in many real-world problems, including text classification, medical diagnostics, cybersecurity, social media analytics, and financial forecasting. The aim of this research is to comprehensively present the mathematical foundations of Naive Bayes

theory, explain its application principles in Machine Learning, comparatively analyze its performance in real-world problems, investigate its advantages and limitations, and, based on experimental results, determine the scenarios in which Naive Bayes is most effective.

Literature Review

The development of probability-based models in the field of Machine Learning and the formation of their theoretical foundations particularly the application of the Bayesian approach have emerged as the result of many years of research. The importance of the Bayesian approach in Machine Learning is especially related to theoretical interpretability, model evaluation, and the ability to make decisions under conditions of uncertainty. Kevin Murphy, in his book “Machine Learning: A Probabilistic Perspective”, states that nearly all ML models can be explained through a Bayesian interpretation, which enhances the transparency of their prediction and decision-making mechanisms [7]. The paper investigates machine learning models for predicting early student dropout, trained using student data on personal characteristics and academic performance [6]. The novelty lies in applying explainable AI (XAI) methods to interpret these models. In author provides a systematic literature review (SLR) on how Artificial Intelligence (AI) and Machine Learning (ML) technologies are being applied to enhance software project management processes [1]. It consolidates research findings from existing literature to highlight major trends, opportunities, challenges, and research gaps in the integration of AI/ML within project management in software engineering. Research article [3] elucidates the key role of Mathematics that plays in development of artificial intelligence (AI) [3]. This research paper's contribution is to pointing out the mathematical rules that must be at their peak in order to construct an ideal artificial intelligence model [3]. The paper proposes a novel Explainable Artificial Intelligence (XAI) framework combining UNet-based semantic segmentation with Bayesian machine learning to classify brain tumors (BTs) using MRI images [4]. This integrated approach aims to support automated, interpretable, and accurate diagnosis from medical scans. In article discussed how Artificial Intelligence (AI) and Machine Learning (ML) are reshaping materials research and outlines a visionary roadmap for India's role in this transformation, emphasizing strategic directions for future research and innovation in the materials domain [9]. It highlights the growing importance of data-driven science in accelerating discovery, optimizing processes, and enabling new materials development paradigms. Paper [10] reviews and systematically analyzes the educational applications of deep learning in the following six aspects: learning effect prediction, educational game, learning recommendation, automatic scoring, assisted teaching and medical education [10]. In research author focused on addressing the problem of class imbalance in Naive Bayes classification, which is a common issue in real-world datasets where one class (or a few classes) significantly outnumbers others [5]. Class imbalance can degrade classifier performance,

particularly for minority classes that are often most important in applications like fraud detection, medical diagnosis, and rare-event prediction. The study investigates how various factors influence risk assessment (RA) outcomes in scientific research projects (SRPs) by applying the Naive Bayesian algorithm, with an enhanced version called Tree-Augmented Naive Bayes (TANB) to better model dependencies among features [2]. The book is a comprehensive guide to machine learning using Python, aimed at students, researchers, and practitioners [8]. It focuses on building practical ML systems using widely adopted Python libraries such as NumPy, pandas, scikit-learn, TensorFlow, and Keras.

Mathematical Statement of the Problem.

The application of probability based models in Machine Learning requires a strong mathematical foundation. The effectiveness and practical efficiency of models such as Naive Bayes largely depend on the correct calculation of probabilities, a clear understanding of Bayes' theorem, and the assumption of feature independence. In this section, the concept of probability, Bayes' theorem, and the Naive independence assumption are discussed. The concept of probability forms the basis of statistical analysis and Machine Learning. Probability provides a mathematical framework for measuring the likelihood of events occurring. This concept is applied in both classical and frequentist approaches. In classical probability, it is assumed that all possible outcomes have equal likelihood. This approach is based on ideal conditions and is used to determine theoretical probabilities. In contrast, frequentist probability is based on the relative frequency of observed outcomes.

Bayes' theorem enables the updating of prior probabilities based on new data. The theorem is expressed as follows:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (1)$$

The concept of Bayesian inference enables a model to learn dynamically as new data become available. The fundamental assumption of the Naive Bayes method is that features are conditionally independent of one another. Although this assumption is rarely fully satisfied in real-world scenarios, it often yields very strong performance in practical applications.

Simplified probability model:

$$P(X_1, X_2, \dots, X_n | C) = \prod_{i=1}^n P(X_i | C) \quad (2)$$

This assumption allows the conditional probability of each feature to be calculated independently, significantly increasing the computational efficiency of the model. The primary objective of Naive Bayes classification models is to determine, based on probability theory, which class a given observation (feature vector $x = (x_1, x_2, \dots, x_n)$) belongs to. The probabilities are computed for all possible classes, and the class with the highest probability is selected.

Optimal class selection (Decision Rule):

$$\hat{y} = \arg \max_{C_k} P(C_k) \prod_{i=1}^n P(x_i | C_k) \quad (3)$$

where $P(C_k)$ – the prior probability of class C_k , $P(x_i | C_k)$ - the likelihood, i.e., the probability of feature x_i given class C_k , $\prod$ the product of feature probabilities based on the conditional independence assumption.

Using this approach, it is determined which class both occurs more frequently a priori and better matches the observed features of the given instance. Naive Bayes models are formulated in different variants depending on the type of data being used.

Gaussian Naive Bayes (for normally distributed features)

This model assumes that each feature follows a normal (Gaussian) distribution within each class. It is commonly applied in medical diagnostics, biometric measurements, the Iris dataset (flower measurements), sensor data, and technical measurement applications. The probability density function of the distribution is given by:

$$P(x_i | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{\left(\frac{-(x_i - \mu_k)^2}{2\sigma_k^2}\right)} \quad (4)$$

Multinomial Naive Bayes (count-based – particularly for text data)

This model is especially used in text classification tasks, as it is based on word counts (frequencies).

$$P(x_i | C_k) = \frac{(x_i)!}{x_{i1}! \dots x_{in}!} \prod_j \theta_{kj}^{x_{ij}} \quad (5)$$

where x_{ij} - the number of times a word appears in a given document; θ_{kj} - the relative frequency (proportion) of the word within a class. It is applied in spam detection, sentiment analysis, and news categorization.

Bernoulli Naive Bayes (for binary features)

In this model, each feature is represented as a binary value, taking either 0 or 1.

$$P(x_i | C_k) = \theta_k^{x_i} (1 - \theta_k)^{1-x_i} \quad (6)$$

It is applied for determining whether a word appears in a text or not, for binary features, and in simple NLP models.

Complement Naive Bayes

In text classification tasks with class-imbalanced datasets, Complement Naive Bayes often yields better results than the Multinomial variant. It is particularly recommended for datasets containing rare classes.

The training of a Naive Bayes model consists of three main stages.

a. Calculation of the prior probability : the prior is simply the proportion of each class

b. Calculation of likelihood probabilities: this step varies depending on the type of Naive Bayes model used: for Gaussian Naive Bayes, the mean and variance (μ_k, σ_k^2) are estimated for each feature within each class; for Multinomial Naive Bayes, word counts (frequencies) are computed; for Bernoulli Naive Bayes, the frequencies of binary values (0/1) are determined.

c. Laplace Smoothing: to prevent unseen words in the text from being assigned zero probability:

$$P(x_i | C_k) = \frac{n_{ik} + \alpha}{N_k + \alpha V} \quad (7)$$

where

$$n_{i,k} = n(x_i, C_k), N_k = \sum_j n_{j,k}.$$

V - the size of the vocabulary, α — the smoothing parameter (for Laplace smoothing, $\alpha = 1$), $n_{i,k}$ — the number of occurrences of feature x_i in class C_k , N_k — the total number of observations in class C_k . This approach makes Naive Bayes more stable and robust.

d. Testing phase (Inference Stage): during testing, the model computes the probability for a given observation.

e. Probability computation

$$P(C_k) \prod_i P(x_i | C_k)$$

f. Class selection: The class with the highest probability is selected.

g. Computations in log-space: Since the product of very small probabilities may approach zero, the logarithm of the probabilities is taken:

$$\log P(C_k) + \sum_i \log P(x_i | C_k)$$

This approach resolves the underflow problem, improves computational stability, and plays a significant role in handling high-dimensional text data.

Naive Bayes in Text Classification and Information Retrieval Systems

Text classification is one of the most widely used application areas of the Naive Bayes algorithm. Since the model is based on the independence assumption, it treats each word or attribute in textual data as independent of others. This assumption simplifies the process and significantly accelerates computation. The Bag-of-Words (BoW) model is one of the most classical approaches for structuring text data. In this method: text is tokenized into individual words, word frequencies are calculated, each document is represented in a vectorized form. When Naive Bayes is combined with the BoW model, probability distributions are constructed based on word frequencies, and posterior probabilities are computed for each class.

TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction allows rare but informative words to be represented with higher weights in the model. The combination of TF-IDF with Naive Bayes demonstrates strong performance particularly in news article categorization, classification of academic papers by research fields, topic analysis of social media posts.

Spam and Malicious Content Detection One of the most successful classical applications of Naive Bayes is spam filtering systems. This model relies on the higher occurrence of certain words (e.g., *free*, *win*, *limited offer*, *prize*) in spam emails, is sufficiently fast for real-time filtering in large-scale email streams, provides high accuracy with minimal computational requirements. For these reasons, this algorithm has been widely used for many years in systems such as Gmail, Yahoo Mail, and Outlook.

Sentiment Analysis Sentiment analysis is an important area of Machine Learning aimed at identifying emotional polarity in texts positive, negative, or neutral. Naive Bayes performs particularly well in this domain because word distributions differ across emotional categories, and probabilistic approaches yield high-quality classification even for short and noisy texts. It is applied in customer review analysis (Amazon, TripAdvisor, AliExpress), public opinion measurement on social media platforms (Twitter/X sentiment analysis), brand reputation monitoring, political analysis (e.g., identifying public opinion trends during election periods).

Applications of Naive Bayes in Medicine and Bioinformatics Probability-based decision-making methods have been used in medicine since the mid-20th century. Naive Bayes is especially effective in medical modeling due to the following characteristics: stable results even with small sample sizes, the independence assumption aligns well with many diagnostic systems, provides interpretable outcomes for physicians through posterior probabilities. Naive Bayes is applied in: disease probability estimation (e.g., calculating the likelihood of diagnoses such as diabetes, heart disease, or early-stage cancer based on symptoms), genetic marker analysis (in bioinformatics, gene expression data are high-dimensional, and Naive Bayes excels due to its simplicity and speed), classification of laboratory test

results (determining disease probabilities based on blood test parameters). The interpretability of Naive Bayes is a major advantage for medical experts, as the reasoning behind the model's decisions can be clearly explained.

Practical Experiments and Results

4.1. Dataset Selection (IMDB Movie Reviews)

In this experiment, the IMDB Movie Reviews dataset was used. The dataset contains positive (1) and negative (0) reviews (table 1).

Dataset characteristics:

Training set: 25,000 samples

Test set: 25,000 samples

The texts consist of real movie reviews

Words are indexed numerically

Table 1.

Data Acquisition and Preprocessing Using the IMDB Dataset

```
# =====
# 6.1. DATASET SELECTION (IMDB REVIEWS)
# Technology: TensorFlow / Keras
# Library: tensorflow.keras.datasets.imdb
# Purpose: Loading the IMDB Movie Review Dataset
# =====

from tensorflow.keras.datasets import imdb

# Keep only the 20,000 most frequently used words
(X_train, y_train), (X_test, y_test) = imdb.load_data(num_words=20000)

# Word-to-index mapping
word_index = imdb.get_word_index()

# Reverse index (index-to-word mapping)
reverse_index = {v: k for k, v in word_index.items()}

# Function to convert numerical sequences back to text
def decode(review):
    return " ".join([reverse_index.get(i - 3, "?") for i in review])

# Decode training reviews into readable text
decoded_train = [decode(x) for x in X_train]

# Decode test reviews into readable text
decoded_test = [decode(x) for x in X_test]
```

Note: The decode function converts numerical indices into words, enabling the data to be processed in text format.

Data Processing

Text preprocessing and vectorization: TfidfVectorizer was used, english stopwords were removed, bigrams (two-word combinations) were taken into account

Table 2.

Preprocessing and Feature Engineering of Review Texts

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(
    max_features=20000,
    stop_words="english",
    ngram_range=(1,2)
)

X_train_vec = vectorizer.fit_transform(decoded_train)
X_test_vec = vectorizer.transform(decoded_test)
```

Source: Prepared by the author using the IMDB Movie Reviews Dataset and the TF-IDF vectorization technique implemented in Scikit-Learn.

Construction of Naive Bayes Models

The Multinomial NB and Bernoulli NB models were implemented:

Table 3.**Sentiment Classification Using Multinomial and Bernoulli Naive Bayes**

```
from sklearn.naive_bayes import MultinomialNB, BernoulliNB
from sklearn.metrics import accuracy_score

mnb = MultinomialNB()
bnb = BernoulliNB()

mnb.fit(X_train_vec, y_train)
bnb.fit(X_train_vec, y_train)

mnb_pred = mnb.predict(X_test_vec)
bnb_pred = bnb.predict(X_test_vec)
```

Source: prepared by the author using the IMDB Movie Reviews Dataset and implemented with Python programming language and the Scikit-Learn machine learning library.

Accuracy scores:

Table 4.**Classification Accuracy of Naive Bayes Models on the IMDB Dataset**

```
print("MultinomialNB Accuracy:", accuracy_score(y_test, mnb_pred))
print("BernoulliNB Accuracy:", accuracy_score(y_test, bnb_pred))
```

```
MultinomialNB Accuracy: 0.85316
BernoulliNB Accuracy: 0.8536
```

Source: Prepared by the author using the IMDB Movie Reviews Dataset and implemented with Python programming language and the Scikit-Learn machine learning library.

Evaluation of Results

Metrics: Accuracy, Precision, Recall, F1-score

Table 5.**Performance Evaluation of Naive Bayes Models Using Accuracy, Precision, Recall, and F1-Score**

```
from sklearn.metrics import classification_report

print("\nMultinomialNB Report:")
print(classification_report(y_test, mnb_pred, digits=4))

print("\nBernoulliNB Report:")
print(classification_report(y_test, bnb_pred, digits=4))
```

MultinomialNB Report:				
	precision	recall	f1-score	support
0	0.8411	0.8709	0.8557	12500
1	0.8661	0.8354	0.8505	12500
accuracy			0.8532	25000
macro avg	0.8536	0.8532	0.8531	25000
weighted avg	0.8536	0.8532	0.8531	25000

BernoulliNB Report:				
	precision	recall	f1-score	support
0	0.8410	0.8721	0.8563	12500
1	0.8672	0.8351	0.8508	12500
accuracy			0.8536	25000
macro avg	0.8541	0.8536	0.8535	25000
weighted avg	0.8541	0.8536	0.8535	25000

Source: Prepared by the author based on the IMDB Movie Reviews Dataset. Performance metrics (Accuracy, Precision, Recall, and F1-Score) were calculated using the Classification Report function provided by the Scikit-Learn library in Python.

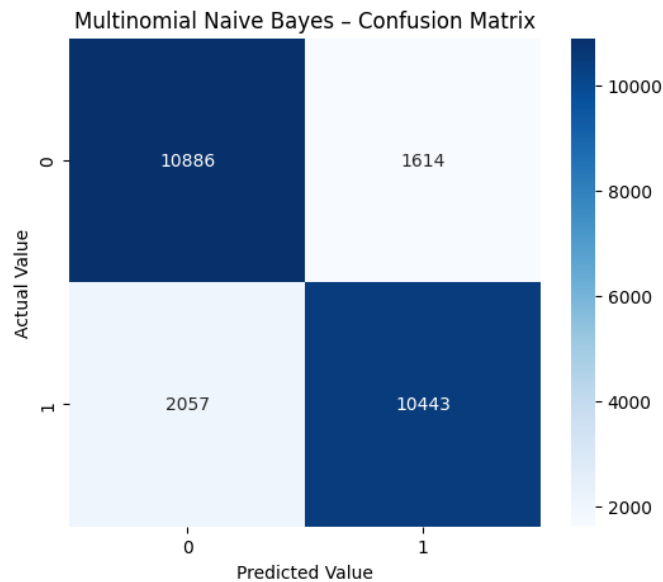


Figure 1. Confusion Matrix of the Multinomial Naive Bayes Classifier

Source: Prepared by the author using the IMDB Movie Reviews Dataset and implemented with Python programming language, TensorFlow/Keras, Scikit-Learn, Matplotlib, and Seaborn libraries.

The confusion matrix demonstrates that the Multinomial Naive Bayes classifier achieved strong performance on the IMDB Movie Reviews Dataset (Figure 1). The model correctly classified 10,886 negative reviews and 10,443 positive reviews. However, 1,614 negative reviews were incorrectly classified as positive,

while 2,057 positive reviews were misclassified as negative. The relatively high number of correctly classified instances compared to misclassifications indicates that the model is effective in distinguishing between positive and negative sentiments. Overall, the confusion matrix confirms the reliability of the Multinomial Naive Bayes classifier for sentiment analysis tasks.

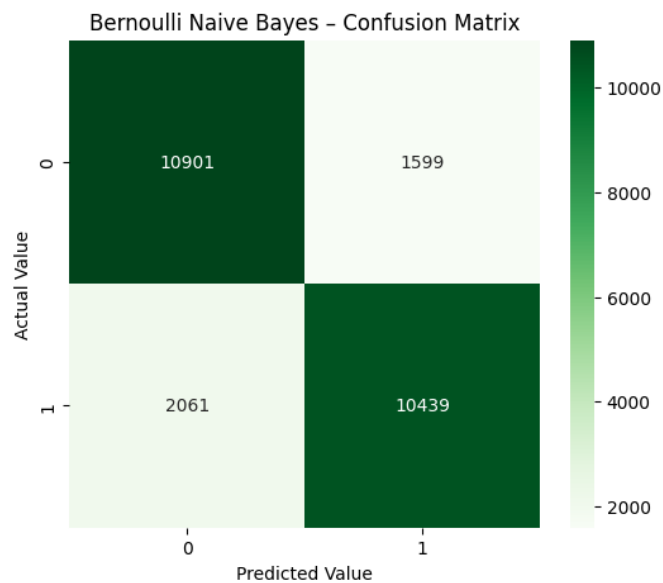


Figure 2. Confusion Matrix of the Bernoulli Naive Bayes Classifier

Source: Prepared by the author using the IMDB Movie Reviews Dataset and implemented with Python programming language, TensorFlow/Keras, Scikit-Learn, Matplotlib, and Seaborn libraries.

The confusion matrix of the Bernoulli Naive Bayes classifier also shows a high level of classification accuracy (Figure 2). The model correctly identified 10,901 negative reviews and 10,439 positive reviews. Misclassification occurred in 1,599 negative reviews predicted as positive and 2,061 positive reviews predicted as negative. Compared with the Multinomial Naive Bayes classifier, the Bernoulli model achieved

slightly better recognition of negative reviews while maintaining a similar performance on positive reviews. The results indicate that the Bernoulli Naive Bayes classifier is also a suitable approach for text-based sentiment classification.

Comparison with Other Models. Comparison with Logistic Regression and LinearSVC (SVM) models

Performance Assessment of Logistic Regression and Support Vector Machine Models on the IMDB Dataset

<pre> from sklearn.linear_model import LogisticRegression from sklearn.svm import LinearSVC logreg = LogisticRegression(max_iter=300) svm = LinearSVC() logreg.fit(X_train_vec, y_train) svm.fit(X_train_vec, y_train) logreg_pred = logreg.predict(X_test_vec) svm_pred = svm.predict(X_test_vec) print("Logistic Regression Accuracy:", accuracy_score(y_test, logreg_pred)) print("SVM Accuracy:", accuracy_score(y_test, svm_pred)) </pre>
Logistic Regression Accuracy: 0.883 SVM Accuracy: 0.87072

Source: Dataset: IMDB Movie Review Dataset (Maas et al., 2011), loaded via tensorflow.keras.datasets.imdb.

Presentation of Results in Tabular Form

Table 7.

Performance Comparison of Machine Learning Models Based on Accuracy

Model	Accuracy
MultinomialNB	0.85316
BernoulliNB	0.8536
Logistic Regression	0.883
SVM	0.87072

Source: Prepared by the author based on the IMDB Movie Reviews Dataset , loaded via tensorflow.keras.datasets.imdb. Accuracy values were computed using the Scikit-Learn library in Python.

Interpretation

The results presented in Table 7 show a comparative evaluation of four machine learning models applied to the IMDB Movie Reviews sentiment classification task. The Logistic Regression model achieved the highest accuracy (0.88300), indicating its superior ability to generalize patterns in text data compared to the other models.

The Support Vector Machine (SVM) also demonstrated strong performance with an accuracy of 0.87072, making it the second-best performing model in this experiment. Both Naive Bayes variants, MultinomialNB (0.85316) and BernoulliNB (0.85360), showed slightly lower but still competitive results, reflecting their effectiveness as baseline probabilistic classifiers for text classification tasks.

Overall, the findings suggest that discriminative models (Logistic Regression and SVM) outperform generative models (Naive Bayes) in sentiment classification on the IMDB dataset, particularly in terms of accuracy.

Naive Bayes models provide fast and efficient results, making them particularly suitable for the classification of large-scale text datasets. Due to their simple probability-based approach, they require relatively low computational resources and are highly effective in real-time applications. More complex linear models, such as Logistic Regression and Support Vector Machines (SVM), may outperform Naive Bayes in terms

of accuracy in certain scenarios. However, these models typically demand greater computational resources and involve longer training times compared to Naive Bayes. Different variants of the Naive Bayes algorithm should be selected based on the characteristics of the dataset. Multinomial Naive Bayes is well suited for count-based or TF-IDF features, while Bernoulli Naive Bayes performs more stably on binary features and may achieve marginally better performance in such cases. For simple and time-efficient applications, Naive Bayes represents an ideal choice. Nevertheless, in situations where higher accuracy is required, it is recommended to conduct comparative evaluations with models such as Logistic Regression or SVM.

Conclusion.

This work provided a detailed examination of the Naive Bayes classifier from a probabilistic and computational perspective. By formally analyzing its reliance on conditional independence assumptions and Bayesian inference principles, the study clarified how likelihood estimation, prior modeling, and posterior maximization contribute to efficient decision-making in high-dimensional feature spaces. The use of log-probability computations and Laplace smoothing further ensured numerical stability and robustness during inference. Experimental results obtained from the IMDB Movie Reviews dataset demonstrated that Multinomial and Bernoulli Naive Bayes models achieve competitive classification performance when combined with TF-

IDF-based feature representations. While Logistic Regression and Support Vector Machines exhibited marginally higher accuracy, they required substantially greater computational resources and longer training times. In contrast, Naive Bayes maintained linear computational complexity with respect to the number of features and samples, highlighting its scalability advantages. From a practical standpoint, the findings confirm that Naive Bayes is well suited for large-scale text classification tasks where efficiency, interpretability, and rapid model deployment are critical. Future research directions may include relaxing the independence assumption through structured Bayesian models or integrating Naive Bayes with ensemble-based approaches to improve predictive performance without significantly increasing computational overhead.

References:

1. Ali, U., Naseer, M. Artificial intelligence and machine learning in enhancing software project management processes: A systematic literature review. *Autom Softw Eng*, 2026. Volume 33, (31) . <https://doi.org/10.1007/s10515-025-00578-6>
2. Dong, X., Qiu, W. A case study on the relationship between risk assessment of scientific research projects and related factors under the Naive Bayesian algorithm. *Sci Rep*. 2024. 14. <https://doi.org/10.1038/s41598-024-58341-y>
3. Kaur, R., Sharma, D. Application and Significance of Mathematics in Artificial Intelligence. In: Reddy, V.S., Wang, J., Chetti, P., Reddy, K.T.V. (eds) *Soft Computing and Signal Processing*. ICSCSP 2024. Lecture Notes in Networks and Systems, vol 1221. 2025. Springer, Singapore. https://doi.org/10.1007/978-981-96-0924-6_53
4. Lakshmi, K., Amaran, S., Subbulakshmi, G. *et al*. Explainable artificial intelligence with UNet based segmentation and Bayesian machine learning for classification of brain tumors using MRI images. *Sci Rep*, 2025. 15. <https://doi.org/10.1038/s41598-024-84692-7>
5. Luo, Y. Improvement and Research of Naive Bayes Classification Based on Unbalanced Data Sets. In: S. Shmaliy, Y. (eds) *8th International Conference on Computing, Control and Industrial Engineering (CCIE 2024)*. CCIE 2024. Lecture Notes in Electrical Engineering, 2024. vol 1253. Springer, Singapore. https://doi.org/10.1007/978-981-97-6937-7_3
6. Minullin, D.A., Gafarov, F.M. Analyzing Machine Learning Models Based on Explainable Artificial Intelligence Methods in Educational Analytics. *Autom. Doc. Math. Linguist*. 2024. 58 (Suppl 3), p.115. . <https://doi.org/10.3103/S0005105525700189>
7. Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. 2012. MIT Press.
8. Raschka, S., & Mirjalili, V. *Python Machine Learning*. 2017. Packt Publishing
9. Sen, P. Artificial intelligence transforming materials research: possible road ahead for India. *Proc.Indian Natl. Sci. Acad*. 2025. <https://doi.org/10.1007/s43538-025-00641-6>
10. Zhang, F., Wang, X. & Zhang, X. Applications of deep learning method of artificial intelligence in education. *Educ Inf Technol*. 2025. 30, p.1563–1587. <https://doi.org/10.1007/s10639-024-12883-w>

PEDAGOGICAL SCIENCES

THE DEVELOPMENT OF DIALOGIC COMPETENCE AMONG STUDENTS IN THE SYSTEM OF MULTILINGUAL EDUCATION

Sagatova Sh.

Candidate of Philological Sciences, Senior language instructor, Almaty Management University

Urunbassarova E.

Doctor of Pedagogical Sciences, Professor, Esil University

Yunissov Y.

Doctor of Philosophy (PhD), Yesil University

РАЗВИТИЕ ДИАЛОГОВОЙ КОМПЕТЕНТНОСТИ ОБУЧАЮЩИХСЯ В СИСТЕМЕ ПОЛИЯЗЫКОВОГО ОБРАЗОВАНИЯ

Сагатова Ш.Б.

к.ф.н, Senior language instructor, ⁴Алматы Менеджмент Университет

Урунбасарова Э.А.

д.п.н. профессор «Esil University», Астана

Юнисов Е.А.

Doctor of Philosophy (PhD) «Esil University», Астана

<https://doi.org/10.5281/zenodo.21105821>

Abstract

This article provides a theoretical and methodological substantiation of the process of developing students' dialogic competence in a multilingual education system.

In the context of modern educational reforms, the development of students' communicative competencies, which ensure their successful professional and social self-realization, is of particular relevance. One of the key components of communicative training is dialogic competence.

The main purpose of the article is to theoretically substantiate and analyze the role of dialogic competence as a key component of students' communicative training, contributing to the development of effective communication skills, critical thinking, and a culture of verbal interaction. The scientific significance of the study lies in clarifying theoretical concepts regarding the essence and structure of dialogic competence, as well as in substantiating its role in the formation of students' communicative culture.

The study reveals the content of the process of developing students' dialogic competence through theoretical-methodological, content-based, technological, and outcome-oriented blocks. In the theoretical-methodological block, the goals and tasks of the process are identified, and methodological approaches are substantiated: systemic, polysubjective, acmeological, tolerant, personality-oriented, activity-based, creatively developmental, and socio-cultural approaches. The authors defined didactic principles: humanization, citizenship, cultural conformity, consciousness and activity, accessibility and feasibility, durability, individual approach, differentiation, and the connection between theory and practice.

The content block of the process of developing students' dialogic competence is based on its components, such as motivational, cognitive, operational-procedural, and personal components. In this regard, criteria and indicators of their manifestation are considered. The possible levels of development of students' dialogic competence are conventionally defined as: unconscious, reproductive, productive, and personal-creative.

The technological block includes a methodology based on the principles of communicative orientation, oral precedence, comprehensiveness, and consideration of students' native language. It is grounded in innovative technologies of personality-oriented, problem-based, active, and game-based learning, as well as a combination of interactive and active forms and methods that promote individualization and differentiation of the learning process, activation of intersubject relations, and the internalization and externalization of the educational potential of disciplines within the multilingual education cycle.

The value of the study lies in expanding scientific understanding of the mechanisms of forming students' dialogic competence and identifying factors contributing to its effective development in the context of modern higher education. The results obtained complement existing research in the field of students' communicative training and the development of 21st-century skills.

Аннотация

В данной статье дано теоретическое и методическое обоснование процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования.

В современных условиях модернизации образования особую актуальность приобретает развитие коммуникативных компетенций обучающихся, обеспечивающих их успешную профессиональную и социальную самореализацию. Одним из ключевых компонентов коммуникативной подготовки является диалоговая компетентность.

Основная цель статьи заключается в теоретическом обосновании и анализе роли диалоговой компетентности как ключевого компонента коммуникативной подготовки обучающихся, способствующего развитию навыков эффективного общения, критического мышления и культуры речевого взаимодействия. Научная значимость исследования заключается в уточнении теоретических представлений о сущности и структуре диалоговой компетентности, а также в обосновании её роли в формировании коммуникативной культуры обучающихся.

В исследовании раскрыто содержание процесса развития диалоговой компетентности обучающихся через теоретико-методологический, содержательный, технологический и результативный блоки. В теоретико-методологическом блоке обозначены цель, задачи процесса, обосновываются методологические подходы: системный, полисубъектный, акмеологический, толерантный, личностно-ориентированный, деятельностный, креативно-развивающийся, социокультурный к данному процессу. Авторами определены дидактические принципы: гуманизации, гражданственности, культуросообразности, сознательности и активности, доступности и посильности, прочности, индивидуального подхода, дифференцированный, связи теории с практикой. Содержательный блок процесса развития диалоговой компетентности обучающихся раскрывается на основе таких ее компонентов, как мотивационный, познавательный, операционно-процессуальный и личностный. В этой связи рассматриваются критерии, показатели их проявления. Условно определены возможные уровни развития диалоговой компетентности обучающихся: неосознанный, репродуктивный, продуктивный, личностно-креативный. В содержание технологического блока входит методика, которая опирается на принципы коммуникативной направленности, устного опережения, комплексности, учета родного языка обучающихся. Основана на инновационных технологиях личностно-ориентированного, проблемного, активного, игрового обучения, сочетании интерактивных и активных форм и методов, способствующих индивидуализации и дифференциации процесса обучения, активизации межсубъектных отношений, интериоризации и экстериоризации образовательного потенциала учебных дисциплин полиязыкового цикла.

Ценность заключается в расширении научных представлений о механизмах формирования диалоговой компетентности у обучающихся и в выявлении факторов, способствующих её эффективному развитию в условиях современного высшего образования. Полученные результаты дополняют существующие исследования в области коммуникативной подготовки обучающихся и развития навыков XXI века.

Keywords: development, dialogic competence, communicative competence, students, intercultural communication, language training, communicative technologies, educational process, teaching methods.

Ключевые слова: развитие, диалоговая компетентность, коммуникативная компетенция, обучающиеся, межкультурная коммуникация, языковая подготовка, коммуникативные технологии, образовательный процесс, методика обучения.

Введение

Современная ситуация предъявляет высокие требования к освоению коммуникативных компетенций обучающимися. Выпускник современной высшей школы должен обладать необходимыми знаниями, составляющими целостную картину мира, умениями и навыками осуществлять разные виды деятельности, опытом творческой деятельности, быть готовым к межличностному и межкультурному взаимодействию и сотрудничеству как внутри своей страны, так и на международном уровне. Это актуализирует проблему развития коммуникативной, в том числе, диалоговой компетентности обучающегося. М. М. Бахтин [1] в этой связи весьма справедливо отмечает, что истина не рождается в голове отдельного человека, а открывается в процессе диалогического общения, поэтому диалог выступает не просто как форма или метод, а важное направление образования.

Необходимо подчеркнуть, что имеется значительное количество публикаций, посвященных проблеме формирования компетентности и компетенций. Это работы А. В. Хуторского [2], А. С. Андриенко [3], М.С. Амеликиной [4], С. А. Узакбаевой, А. К. Атахановой [5] и многих др.

В этой связи мы разделяем точку зрения А. С. Андриенко касательно определений понятий «компетентность» и «компетенция»: «Компетентность

является показателем обобщенной способности к решению профессиональных и жизненных задач в определенной сфере деятельности, показателем профессионального и личностного развития, саморазвития и самореализации личности. Формирование ключевых компетенций направлено на осуществление человеком успешной деятельности в обществе» [3, с.25].

Методологические подходы к проблемам коммуникации исследовали М.С.Каган [6] и др., социально-психологические особенности развития речевой компетентности и коммуникативной компетенции в процессе языковой подготовки обучающихся рассмотрены в работах А. Н. Гусейновой, М.М.Абдусаламова [7], М.П. Манаенковой [8] и др. Общению как одному из ведущих факторов становления личности, ее психического и физического развития посвящены исследования А. А. Леонтьева [9], В.Н.Мясищевой [10] и др.

В работах зарубежных ученых также исследуются разные аспекты развития коммуникативной компетентности, в том числе и диалоговой, в процессе учебной деятельности, а именно в работах Ларсен-Фримана [11], Д. Хаймса [12], М. Канале и М. Суэйна [13], В. Д. Давида [14] и др.

Вместе с тем, как показывает современная практика, преподавание языков не обеспечивает в

должной мере готовности обучающегося к активному взаимодействию и общению в иноязычной среде, не снимает проблемы языкового барьера.

Ученые и деятели образования объясняют данное явление преобладанием лингвистическо-центрированной модели обучения, при которой происходит обучение преимущественно лексическим, грамматическим, орфографическим, синтаксическим и другим языковым особенностям. В этом случае владение иностранным языком является самоцелью и не служит средством общения. Происходит угасание речевых навыков, обеднение словарного запаса, а вербальное мышление всё больше вытесняется визуальным.

Поэтому в преподавании языковых предметов необходимо особое внимание обратить обучению коммуникации, практическому применению знаний. Обучить языку можно только в условиях создания необходимой среды общения в образовательном процессе, в моделировании активного взаимодействия и общения обучающихся между собой и преподавателем.

Ведущей задачей обучения языкам должен стать переход от монологической модели к полилогической, полагающей выход содержания и технологии обучения на уровень обеспечения коммуникативной, в том числе диалоговой, компетентности обучающегося.

Необходимо отметить, что особое значение развитие диалоговых умений и способностей имеет в студенческом периоде развития личности. Очевидно, что насколько успешен и компетентен будет студент, настолько положительно сформируется его личностный статус. Поэтому развитие диалоговой компетентности у обучающихся выступает одной из ведущих педагогических задач образовательного процесса в вузе. Однако, в системе подготовки будущих специалистов задача развития их диалоговой компетентности продолжает недооцениваться, недостаточно и специальных исследований данного вопроса.

С учетом вышесказанного актуальность проблемы развития диалоговой компетентности обучающихся связана с необходимостью разрешения противоречия между потребностью общества в выпускниках с высоким уровнем развития диалоговой компетентности и недостаточной разработанностью данного вопроса в теоретическом и методическом планах в системе высшего образования.

Актуальность проблемы, ее недостаточная научная и практическая разработанность определили выбор темы исследования: «Развитие диалоговой компетентности обучающихся в системе полиязыкового образования».

Целью исследования является теоретическое и методическое обоснование процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования.

Объект исследования: полиязыковое образование.

Предмет исследования: процесс развития диалоговой компетентности обучающихся.

Гипотеза исследования: развитие диалоговой компетентности обучающихся достигнет высокого уровня, если будет определена содержательная характеристика диалоговой компетентности обучающегося, критерии, показатели и уровни ее сформированности, то будут созданы условия для эффективного ее развития, так как при этом сформируются механизмы (социокультурные, коммуникативные, познавательные и др.) самоактуализации личности студента.

В соответствии с проблемой, целью, предметом и выдвинутой гипотезой исследования нами были сформулированы следующие задачи:

1. раскрыть содержание процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования;
2. уточнить сущность понятия «диалоговая компетентность»;
3. разработать методику развития диалоговой компетентности обучающихся в системе полиязыкового образования.

На первом этапе исследования было раскрыто содержание теоретико-методологического, содержательного, технологического и результативного блоков процесса развития диалоговой компетентности обучающихся. Обоснованы и раскрыты составляющие компоненты каждого из них.

На втором – разработана методика развития диалоговой компетентности обучающихся в системе полиязыкового образования.

Материалы и методы

Для достижения цели исследования, связанной с развитием диалогической компетентности обучающихся в системе полиязыкового образования, был применен комплексный подход, включающий теоретические и эмпирические методы. В рамках исследования были осуществлены анализ и обобщение философской, психологической, педагогической, лингвистической литературы, а также диссертационных исследований по изучаемой проблеме.

Результаты и обсуждение

В ходе исследования были достигнуты следующие результаты:

- раскрыто содержание процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования на основании характеристики четырех блоков: теоретико-методологического, содержательного, технологического и результативного;
- уточнена сущность диалоговой компетентности обучающихся как интегрального свойства личности;
- разработана методика развития диалоговой компетентности обучающихся в системе полиязыкового образования на основе опоры на методические принципы и инновационные технологии личностно-ориентированного, проблемного, активного, игрового обучения, сочетание интерактивных и активных форм и методов.

На основе анализа философской, психологической, педагогической, лингвистической литера-

туры мы убедились в том, что многие исследователи занимались изучением тех или иных аспектов развития коммуникативной компетентности и компетенции.

Несмотря на большое количество исследований, посвященных вопросам коммуникативной компетентности и компетенции как отечественных, так и зарубежных ученых, вопрос о развитии диалоговой компетентности обучающихся в системе полиязыкового образования изучен недостаточно.

Исходя из серьезной значимости диалоговой компетентности в жизни общества, мы обратились к теме нашего исследования: «Диалоговая компетентность обучающихся как важный результат полиязыкового образования».

Содержание процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования нами рассмотрено на основании таких блоков, как теоретико-методологический, содержательный, технологический и результативный.

Теоретико-методологический блок представлен целью процесса, задачами, принципами, а также методологическими подходами. Цель – развитие высокого уровня диалоговой компетентности обучающихся в системе полиязыкового образования.

Задачи: развитие мотивации к практическому применению языков; высокого уровня знаний языков в коммуникативном, социокультурном, познавательном аспектах; высокого уровня диалоговых умений и навыков применения языков на практике; соответствующих качеств личности.

Рассмотрение диалоговой компетентности и процесса ее развития в образовательном процессе вуза необходимо, на наш взгляд, с позиций различных методологических подходов. Среди подходов особую значимость представляют, на наш взгляд, такие, как: системный, полисубъектный, акмеологический, толерантный, личностно-ориентированный, деятельностный, креативно-развивающийся, социокультурный.

Системный подход является определяющим фактором эффективности развития диалоговой компетентности обучающихся в системе полиязыкового образования, в связи с тем, что методы должны составлять систему, взаимодействие с которой средств, способствующих личностному росту студентов, влияет не только на их логическую сферу, но и на эмоциональную. Об этом очень точно высказался М. С. Каган в статье Т. В. Наумовой «Я понимал системный подход как проявление диалектического метода познания тех явлений, которые объективно представляют собой системы, то есть самоорганизованные множества элементов, отличающиеся целостностью, что порождает у них особые – системные – свойства, которые не сводятся к сумме свойств составляющих их свойств элементов...» [5, с.119].

Полисубъектный подход является основной стратегией формирования диалоговой компетентности обучающихся. В соответствии с ним участники процессов общения и обучения являются

субъектами, высшей ценностью и ни в коем случае не рассматриваются как средство или объект воздействия.

Акмеологический подход предполагает создание условий для наиболее полного индивидуального развития обучающегося, достижения им наиболее высокого уровня социальной зрелости как гражданина, специалиста, профессионала.

Основанием данного подхода служит тот факт, что студенческий период развития личности является сенситивным для формирования сознания и самосознания, раскрытия потенциала личности.

Важными условиями толерантного подхода в коммуникации являются терпимое отношение к собеседнику, проявление интереса к разговору, недопущение снисходительности и безразличия. Толерантность создает в коллективе благоприятный психологический климат и способствует выстраиванию взаимодействия, что очень значимо для успешного общения и развития диалоговой компетентности.

Личностно-ориентированный подход основывается на признании индивидуальности, самоценности каждого человека, его развитии как индивида. В вузовской системе обучения личностно-ориентированный подход предоставляет широкие возможности для развития каждого студента, реализации заложенного в нем потенциала, раскрытия его возможностей и способностей.

Деятельностный подход представляет интегральную характеристику деловых и личностных качеств, отражает уровень знаний, умений и навыков для осуществления деятельности, связанной с принятием различных решений.

Умение действовать у обучающихся формируется в процессе приобретения знаний, то есть знания усваиваются в ходе их практического применения, что очень важно в развитии диалоговой компетентности студентов.

Креативно-развивающийся подход формирует у обучающихся продуктивное мышление и творческое отношение к деятельности.

Социокультурный подход предполагает вхождение личности в социокультурное пространство в процессе социализации. Обеспечивая диалоговое взаимодействие, диалог культур, социокультурный подход способствует формированию коммуникабельности, готовности к совместной деятельности, к построению диалога с людьми различных взглядов независимо от их национальной, культурной, расовой, социальной принадлежности.

Ведущими принципами, на которые опирается процесс развития диалоговой компетентности обучающихся, являются: гуманизации, гражданственности, культуросообразности, сознательности и активности, доступности и посильности, прочности, индивидуального подхода, дифференцированный, связи теории с практикой.

Принцип гуманизации весьма важен при изучении языков. Учебно-воспитательный процесс должен быть насыщен идеями гуманизма и общечеловеческих ценностей. Любовь к человеку, отношение к нему как высшей ценности, уважение его

прав, его самого, вера в его способности и возможности.

На основании принципа гражданственности обучающиеся должны понимать и реализовывать себя с точки зрения осознания своей причастности к стране, в которой живут, к Отчизне, к ее истоку, к ее корням. Культура гражданственности должна служить мерой самореализации студентов в нравственном и социокультурном смыслах.

Немаловажную роль играет и принцип культуросообразности. Он предполагает, чтобы в процессе образования и воспитания учитывались условия, в которых находится человек, необходимо учитывать его культуру.

Принцип сознательности и активности предполагает сознательное, активное, творческое участие обучающегося в общении и обучении языкам, личностное развитие студента. Сознательность также выражается в положительном отношении к изучаемому материалу, во внутренней мотивированности. Что же касается активности, то это – главная характеристика познания, которая должна проявляться в эмоциональном, речевом и интеллектуальном аспектах. Применение принципа доступности и посильности должно осуществляться на уровне возможностей обучающихся, чтобы они не испытывали непреодолимых трудностей. Доступность обеспечивается как самим материалом, его организацией, так и методикой работы с ним на занятии. Посильность проявляется в темпе продвижения изучения материала.

Принцип прочности выражается в том, что вводимый в память студентов материал сохранялся. Прочность усвоения обеспечивается путем яркого проведения занятия; достаточно большой тренировкой в воспроизведении полученных знаний; самостоятельного творческого его применения, системного контроля за усвоением.

Опора на принцип индивидуального подхода предполагает учет индивидуальных качеств обучающегося, выбор соответствующих приемов и заданий. Каждый студент индивидуальность. Он выступает как индивид, как субъект, как личность в образовательном процессе. Понимание и распознавание индивидуальных качеств собеседников в общении очень важно.

Дифференцированный принцип предполагает приведение методов и форм работы в соответствие с особенностями обучающихся; постоянное варьирование методов и организационных форм с учетом личности каждого студента; временное изменение ближайших задач и отдельных сторон учебно-воспитательной работы.

Суть принципа связи теории с практикой заключается в том, что чем больше приобретаемые студентами знания в своих узловых моментах взаимодействуют с жизнью, применяются на практике, используются для преобразования окружающих процессов и явлений, тем выше сознательность обучения и интерес к нему.

Содержательный блок процесса развития диалоговой компетентности обучающихся характе-

ризуется, на наш взгляд, взаимосвязью мотивационного, познавательного, операционно-процессуального и личностного компонентов и уровнями, критериями и соответствующими показателями.

Мотивационный компонент представляет собой совокупность мотивов к учебно-познавательной деятельности, которые пробуждают у обучающихся интерес, желание и стремление к развитию, достижению лучших результатов, решению новых задач. В основе мотивационного компонента находится мотив, который вызывает необходимость действий у студентов в соответствующем направлении. Мотивационный компонент включает направленность личности. Под нею понимается совокупность устойчивых мотивов, которые направляют поведение и деятельность личности независимо от конкретных условий. Направленность характеризуют доминирующие потребности, интересы, склонности, убеждения, мировоззрение.

Познавательный компонент характеризуется предметными знаниями, непосредственно связанными с деятельностью обучающихся и их общим кругозором.

Основой операционно-процессуального компонента являются умения, навыки, опыт. Под умениями следует понимать овладение способами (приемами, действиями) применения усвоенных знаний на практике.

Основой личностного компонента являются качества личности. Нами в личностный компонент включены адекватная самооценка, креативные, коммуникативные, перцептивные и интерактивные способности, умения самообразования и саморазвития, способности к самоорганизации.

Важность адекватной самооценки связана с тем, чтобы обучающиеся объективно оценивали свои способности и возможности, их соответствие реальным результатам и фактам и их оценке окружающих.

Креативные способности вооружают студентов возможностью видеть явления по-новому, анализировать предлагаемые ситуации и находить в них нестандартные решения.

Коммуникативные способности обеспечивают активное взаимодействие и взаимопонимание, устанавливают хорошие взаимоотношения в процессе общения или выполнения совместной деятельности на занятиях полиязыкового цикла

Перцептивные способности, тесно связанные с проницательностью человека, помогают обучающимся чувствовать внутренний мир окружающих их людей и понимать их.

Интерактивные способности, связанные с активным взаимодействием студентов в процессе общения, играют важную роль в развитии диалоговой компетентности на занятиях полиязыкового цикла.

Умения к самообразованию и саморазвитию, включающие умения работать с источниками информации, анализировать и сопоставлять полученные данные, сообщать о полученных знаниях дру-

гим людям, применять полученные знания для решения задач или проблем, весьма важны в развитии диалоговой компетентности обучающихся.

Способность к самоорганизации, дающая возможность студенту управлять своим временем, силами для достижения необходимых результатов, справляться с эмоциями и перегрузками, играет немаловажную роль в развитии диалоговой компетентности.

Нами разработаны критерии и показатели для определения уровня развития диалоговой компетентности обучающихся.

Мотивационный компонент процесса развития диалоговой компетентности представлен критериями проявления мотивов к личностному росту обучающегося и следующими показателями: владение мотивациями к образовательной деятельности; в формировании направленности учебно-познавательной деятельности обучающегося, ее склонностей, потребностей, интереса, стремления к познанию языков и их практическому применению.

Познавательный компонент представлен критериями системности, целостности, прочности, глубины и полноты усвоения языковых знаний и следующими показателями: знание основной лексики; владение грамматическими знаниями; общий кругозор обучающегося, любознательность, мировоззрение.

Операционно-процессуальный компонент представлен критериями степени развития умений, навыков, практического опыта в образовательной деятельности и следующими показателями: владение учебно-познавательными умениями, навыками, практическим опытом; оперирование языковыми знаниями; обобщение личного опыта; владение ключевыми компетенциями в общении и взаимодействии в социокультурном пространстве.

Личностный компонент процесса развития диалоговой компетентности представлен критериями степени развития личностных качеств. Показателями личностного компонента являются, на наш взгляд, адекватная самооценка, креативные, коммуникативные, перцептивные, интерактивные способности, владение умениями самообразования и саморазвития, способности к самоорганизации.

На основе разработанных критериев и показателей нами выделены уровни развития диалоговой компетентности обучающихся: неосознанный; репродуктивный; продуктивный; личностно-креативный.

Каждый из выделенных уровней характеризуется критериями.

Критерием первого уровня, с нашей точки зрения, является несамостоятельная деятельность обучающегося. Учебный процесс построен на подсказках преподавателя, субъектно-объектных отношениях, когда ведущим является педагог и требует от студента только правильных ответов, не признавая его права на ошибку и диалог.

Критерием второго, репродуктивного, уровня является наличие определенных языковых знаний у обучающегося, однако недостаточно умений их применять на практике, при этом, следует заметить,

репродуктивная деятельность осуществляется по памяти или по заданному алгоритму.

Критерием третьего, продуктивного, уровня выступает достаточный уровень владения обучающимися лингвистическими знаниями и умениями. Студенты способны анализировать, синтезировать, сравнивать и обобщать образовательный материал. Для студентов этого уровня характерна уже самостоятельная продуктивная деятельность, которая выполняется по самостоятельно созданному алгоритму или принятому типовому алгоритму.

Критерием четвертого, личностно-креативного, уровня является самореализация компетентной личности. Студент способен к самостоятельной творческой учебной деятельности на основе приобретенных знаний, умений, опыта, ценностей. Обучающийся ставит и решает задачи, требующие обобщения знаний и творческого мышления, стремится к пониманию другого с целью поиска новой истины, вступает в «партнерство», проявляя социальную ответственность, то есть высший уровень социализации.

Технологический блок процесса представлен методикой развития диалоговой компетентности у обучающихся, которая разработана на основе инновационных технологий обучения.

Результативный блок процесса развития диалоговой компетентности у обучающихся представляет развитие высокого уровня диалоговой компетентности, критерием которого, как было сказано выше, является личностно-креативная компетентность.

Методика опирается на методические принципы коммуникативной направленности, устного опережения, комплексности, учета родного языка обучающихся. Основана на инновационных технологиях личностно-ориентированного, проблемного, активного, игрового обучения, сочетании интерактивных и активных форм и методов, способствующих индивидуализации и дифференциации процесса обучения, активизации межсубъектных отношений, интериоризации и экстериоризации образовательного потенциала учебных дисциплин полиязыкового цикла, способствующих получению оптимального результата в обучении языкам.

Остановимся на технологиях. При личностно-ориентированной технологии в центре всей образовательной системы стоит личность обучающегося. Для нее созданы необходимые условия развития, способствующие реализации ее природного потенциала. Личность студента – субъект и цель образовательной системы. Поэтому в основе личностно-ориентированной технологии лежит адаптивная система обучения. Суть адаптации к процессу обучения заключается в формировании у каждой личности коммуникативной компетенции в собственном ее (личности) темпе, учете ее способностей, а также ее уровня знаний, умений и навыков по иностранному языку.

При технологии проблемного обучения на основе использования проблемно-поискового метода применяются такие приемы обучения, как создание

на занятии проблемных ситуаций, организация коллективного обсуждения возможных подходов к их решению, выполнение упражнений, предусматривающих различные формы общения преподавателя и обучающихся, максимально приближенные к условиям реальной коммуникации. Безусловно, на занятии следует сочетать проблемно-поисковый метод обучения с другими методами. Знания и способности деятельности являются в большей степени предметом поиска самими студентами.

Формы, методы и средства (кейс-стади, мозговой штурм, баскет-метод, сократический диалог и др.) технологии активного обучения способствуют эффективному решению целого ряда задач: развитию познавательных мотивов и интересов; системному мышлению будущего специалиста; учет коллективной мыслительной и практической работе, формируют социальные умения и навыки взаимодействия и общения и т. д. Поэтому технология активного обучения играет весьма значимую роль в развитии диалоговой компетентности обучающихся. Технология игрового обучения способствует приобретению опыта, развитию навыков общения, эмоциональной сферы, проявлению индивидуальных качеств и способностей личности. Можно утверждать, что в процессе учебной игры развивается целеустремленность, активность, динамичность и продуктивность мышления, прочность и оперативность памяти, стремление к совершенству и вера в свои силы. На основании уже только этого следует утверждать о больших возможностях, заложенных в технологии игрового обучения, для развития диалоговой компетентности у обучающихся.

Таким образом, инновационные технологии обучения мы рассматриваем как процесс управления образовательной деятельностью обучающихся, в основе которого лежит межсубъектное диалоговое взаимодействие и сотрудничество, которое приводит к качественным изменениям субъектов образовательного процесса, к развитию их диалоговой компетентности. Результативный блок – это достижение высокого (личностно-креативного) уровня развития диалоговой компетентности обучающихся.

Заключение

На наш взгляд, цель исследования, а именно теоретическое и методическое обоснование процесса развития диалоговой компетентности обучающихся в системе полиязыкового образования, достигнута.

Теоретическое обоснование осуществлено на основе раскрытия содержания процесса развития диалоговой компетентности обучающихся, характеристики теоретико-методологического, содержательного, технологического и результативного составляющих ее блоков.

Сущность понятия «диалоговая компетентность», на наш взгляд, заключается в том, что это интегральное качество личности, отражающее мотивацию к изучению языков, уровень знаний в социокультурном, коммуникативном (программиро-

вание поведения и налаживание контактов), познавательном аспектах; умений, навыков, качеств личности (креативности, самооценки, перцепции и др.), необходимых для успешного межличностного и межкультурного взаимодействия и сотрудничества как внутри своей страны, так и на международном уровне.

Методика развития диалоговой компетентности обучающихся опирается на методические принципы коммуникативной направленности, устного опережения, комплексности, учета родного языка обучающихся. Основана на инновационных технологиях личностно-ориентированного, проблемного, активного, игрового обучения, сочетании интерактивных и активных форм и методов.

Перспективы исследования нами видятся в проведении опытно-экспериментальной работы для подтверждения предлагаемой нами теории и методики развития диалогической компетентности обучающихся; разработке соответствующих рекомендаций; также необходимо рассмотреть процесс развития диалоговой компетентности обучающихся на основе интериоризации и экстериоризации образовательного потенциала всех дисциплин социально-гуманитарного цикла; также отдельно каждой будущей профессии и специальности студентов; другие виды компетентностей и компетенций в процессе обучения.

Список литературы:

1. Аврахова Л. В. Абстрактный рационализм истины и экзистенциальная правда в диалогизме М. Бахтина // Гілея: науковий вісник. – 2019. – №146(2). – С. 7–12.
2. Хуторской, А. В. Системно-деятельностный подход в обучении: Научно-методическое пособие // Москва: Издательство «Эйдос»; Издательство Института образования человека, 2012. – 63 с.: ил. (Серия «Новые стандарты»).
3. Андриенко А. С. Компетентностно-ориентированный подход в системе высшего образования: история, современное состояние и перспективы развития: монография. – Чебоксары: ИД «Среда», 2018. – 92 с.
4. Амелькина М. С. Компетентностный подход: новый виток развития отечественного образования // Управление образованием: теория и практика. – 2019. – №2 (34). – С. 47–57.
5. Атаханова А. К., Узакбаева С. А. Сущность понятия «межкультурно-коммуникативной компетенции» // Известия КазУМОиМЯ имени Абылай хана. Серия «Педагогические науки». – 2021. – № 2 (61). – С. 85–92. – DOI: 10.48371/PEDS.2021.61.2.011.
6. Наумова Т. В. Каган М. С.: системный подход как основа в исследовании человеческой деятельности // Вестник ЧГУ. – 2008. – №3. – С. 118–126.
7. Манаенкова М. П. Специфика технологии формирования речевой компетентности обучающихся неязыкового вуза // Современные педагогические технологии в организации образовательного пространства региона. – 2018. – С. 46–51.

8. Гусейнова А. Н., Абдусаламов М. М. Коммуникативная компетенция как основная цель обучения иностранному языку и базовые характеристики грамматической компетенции как её важного компонента // Мир науки, культуры, образования. – 2020. – №4(83). – С. 56–58.

9. Леонтьев А. А. Психологические предпосылки овладения иностранным языком // Иностранные языки в школе. – 1985. – №5. – С. 8–11.

10. Мясисhev В. Н. Межличностное общение. – Санкт-Петербург: Питер, 2005. – 46 с.

11. Larsen-Freeman D., Anderson M. Techniques and Principles in Language Teaching. – Oxford: Oxford University Press, 2000. – 252 p.

12. Hymes D. H. On Communicative Competence // Sociolinguistics. Selected Readings. – Harmondsworth: Penguin, 1972. – P. 269–293.

13. Canale M., Swain M. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing // Applied Linguistics. – 1980. – Vol. 1. – Issue 1. – P. 1–47.

14. David, W. D. What do second language speakers really need for real-world interaction? A needs analysis of L2 Chinese interactional competence // Journal: Language Teaching Research // First published online: January 31, 2023. Volume 30, Issue 2 (2026). pp: 655–692. DOI: 10.1177/13621688221144836

References:

1. Avrakhova L. V. Abstraktnyi ratsionalizm istiny i ekzistentsial'naya pravda v dialogizme M. M. Bakhtina [Abstract Rationalism of Truth and Existential Truth in M.M. Bakhtin's Dialogism] // Gileya: naukovii visnik. – 2019. – №146(2). – С. 7–12. [in Rus.]

2. Khutorskoi, A. V. Sistemno-deyatelnostnyi podkhod v obuchenii: Nauchno-metodicheskoe posobie [System-Activity Approach in Education: Scientific and Methodological Manual]. // Moskva: Izdatel'stvo «Eidos»; Izdatel'stvo Instituta obrazovaniya cheloveka, 2012. – 63 s.: il. (Seriya «Novyye standarty»). [in Rus.]

3. Andrienko A. S. Kompetentnostno-orientirovannyi podkhod v sisteme vysshego obrazovaniya: istoriya, sovremennoe sostoyanie i perspektivy razvitiya: [Competency-Based Approach in Higher Education: History, Current State and Development Prospects]. monografiya. – Cheboksary: ID «Sreda», 2018. – 92 s. [in Rus.]

4. Amel'kina M. S. Kompetentnostnyi podkhod: novyi vitok razvitiya otechestvennogo obrazovaniya [Competency-Based Approach: A New Stage in the Development of National Education] // Upravlenie obrazovaniem: teoriya i praktika. – 2019. – №2 (34). – С. 47–57. [in Rus.]

5. Atakhanova, A. K., Uzakbaeva, S. A. Sushchnost' ponyatiya «mezhekul'turno-kommunikativnoi kompetentsii» [The essence of the concept of intercultural communicative competence] // Izvestiya KazUMOimYa imeni Abylaikhana. Seriya «Pedagogicheskie nauki». – 2021. – № 2 (61) – С. 85–92. – DOI: 10.48371/PEDS.2021.61.2.011 [in Rus.]

6. Naumova T. V. Kagan M. S.: sistemnyi podkhod kak osnova v issledovanii chelovecheskoi deyatel'nosti [M.S. Kagan: System Approach as a Basis for the Study of Human Activity] // Vestnik ChGU. – 2008. – №3. – С. 118–126. [in Rus.]

7. Manaenkova M. P. Spetsifika tekhnologii formirovaniya rechevoi kompetentnosti obuchayushchikhsya neyazykovogo vuza [Specific Features of Speech Competence Formation Technology for Non-Linguistic University Students] // Sovremennyye pedagogicheskie tekhnologii v organizatsii obrazovatel'nogo prostranstva regiona. – 2018. – С. 46–51. [in Rus.]

8. Guseynova A. N., Abdusalamov M. Kommunikativnaya kompetentsiya kak osnovnaya tsel' obucheniya inostrannomu yazyku i bazovyye kharakteristiki grammaticheskoi kompetentsii kak ee vazhnogo komponenta [Communicative Competence as the Main Goal of Foreign Language Teaching and Basic Characteristics of Grammatical Competence as Its Important Component] // Mir nauki, kul'tury, obrazovaniya. – 2020. – №4(83). – С. 56–58. [in Rus.]

9. Leont'ev A. A. Psikhologicheskie predposylki ovladeniya inostrannym yazykom [Psychological Preconditions for Foreign Language Acquisition] // Inostrannyye yazyki v shkole. – 1985. – №5. – С. 8–11. [in Rus.]

10. Myasishchev V. N. Mezhluchnostnoe obshchenie [Interpersonal Communication]. – Sankt-Peterburg: Piter, 2005. – 46 s. [in Rus.]

11. Larsen-Freeman D., Anderson M. Techniques and Principles in Language Teaching. – Oxford: Oxford University Press, 2000. – 252 p.

12. Hymes D.H. On Communicative Competence // Sociolinguistics. Selected Readings. – Harmondsworth: Penguin, 1972. – P. 269–293.

13. Canale M., Swain M. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing // Applied Linguistics. – 1980. – Vol. 1. – Issue 1. – P. 1–47.

14. David, W. D. What do second language speakers really need for real-world interaction? A needs analysis of L2 Chinese interactional competence // Journal: Language Teaching Research // First published online: January 31, 2023. Volume 30, Issue 2 (2026). pp: 655–692. DOI: 10.1177/13621688221144836

THE INFLUENCE OF EFFECTIVE LANGUAGE LEARNING STRATEGIES IN SECOND LANGUAGE ACQUISITION ON FOREIGN LANGUAGE TEACHING TO UNIVERSITY STUDENTS

Kolodko T.

*PhD in Education, Associate Professor,
Taras Shevchenko National University of Kyiv, Ukraine
ORCID ID: 0000-0002-6267-6263*

Kolodko M.

*Master in Oriental Philology
Faculty in Institute of Technology,
Shanghai, China
ORCID ID: 0009-0006-1177-9445
<https://doi.org/10.5281/zenodo.21105825>*

Abstract

The present study explores the role of language learning strategies in second language acquisition and their influence on foreign language teaching. Drawing upon the competency-based approach to language education, the study examines how strategic learning contributes to the development of communicative competence and learner autonomy. Particular attention is paid Rebecca Oxford's classification of direct and indirect language learning strategies, including memory, cognitive, compensation, metacognitive, affective, and social strategies. The research employs the methods of analysis, systematization, and generalization to investigate theoretical perspectives on language learning strategies and their pedagogical implications. The findings suggest that the conscious and systematic use of language learning strategies enhances language proficiency, promotes self-regulated learning, and improves learners' ability to communicate effectively in diverse contexts. Furthermore, the study demonstrates that contemporary language teaching methods can facilitate the development of strategic competence by creating learner-centered and communicatively oriented learning environments. The results highlight the importance of integrating strategy-based instruction into foreign language teaching in order to support successful language acquisition and the achievement of educational outcomes within the competency-based framework.

Keywords: language learning strategies, second language acquisition, foreign language teaching, communicative competence, competency-based approach, direct and indirect strategies, strategy-based instruction, university students.

Introduction. Global integrative trends, intercultural interaction, and rapid social development contribute to mastering one or more foreign languages not only to communicate with representatives of other cultures but also to improve professional skills since modern society requires specialists to solve global issues (Berger, 2021). Learning a language is a complex process that takes longer than anticipated. The term "learning a foreign language" includes the entire educational paradigm related to socio-cognitive, pragmatic, and ethical aspects of the foreign-speaking world. Language learning is described as the process through which an individual's capacity improves. Learning a language requires strategies, processes, and routines (Hashim & Hashim, 2018).

Effective teaching for university students requires moving beyond rote memorization into real world application. One of the most important tasks that teachers have to perform is that of organising students to do various activities using different methods, techniques, approaches, strategies, and platforms.

Our research **purpose** is to reveal and investigate current trends and strategies of teaching foreign languages aimed at maximizing learning efficiency.

The **tasks** of this article are to involve classifying the methods of foreign language teaching based on the competency-based approach and identifying the effective methods.

The **methodological basis** of this research includes methods of analyzing the practical experience of

foreign language teaching based on the competency-based approach, usage of our own experience, systematization, and generalization of sociological data.

Results. Language training is based on the purpose of teaching the learners' language they learn all the things about a given language. They are expected to learn functions, rules, linguistic forms, how to use language and acquire the language skills. On the other hand, learner training is based on learners and their difficulties in learning. Learners are provided with skills and strategies to be able to learn more easily and effectively.

The teacher help students facilitate their learning and be more effective in learning the target language. After getting learning some strategies, the students activate their metacognition, which means the students' awareness of the usage of the strategies. And the teacher plays a modeling role in teaching the strategies to students.

Learning strategy is the specific action to make the students better in learning a second language. In the early 1970s, language learners were seen to be more actively responsible for their own learning. Rubin investigated what 'good language learners' facilitate their learning. Such learners, according to Rubin, are willing and accurate guessers who have a strong desire to communicate and will attempt to do. He also observed that, improvement of students' language performance, develop their repertoire of learning strategies. Rubin (1981) states that, the strategies are not the preserve of

highly capable individuals, but could be learned by other moderate and slow learners (Rubin, 1981).

Learning strategies are special methods for solving a problem or task, designing a desired outcome, manipulating and controlling information obtained towards the interest of learning (Wegner, Minnaert, & Strehlke 2013). Learning strategies are the particular approaches or techniques that learners employ to try to learn the second language (Kolodko, 2023). Language learning strategies are conscious or potentially conscious actions and learners can identify them in their learning process (Cohen, 1998).

Learning strategies consist of “mental or behavioral activity related to some specific stage in the overall process of language acquisition or language use” (Ellis, 1994, p. 529), in other words, they can be behavioral (for example, repeating new words aloud to remember them) or they can be mental (for example, using the synonyms or situational context to infer the meaning of a new word) (Ellis, 1997). Learning strategies are specific actions taken by the learner to make learning “easier, faster, more enjoyable, more self-directed, more effective, more transferable to new situations” (Oxford, 1990, p. 8). So, students are required to take advantage of a good strategy in order to master the field of language.

According to cognitive learning theories, learners are active participants in the learning and teaching process rather than passive recipients. They do not just receive information from teachers as learning process involves learners processing information which includes mental activities (Hosenfeld, 1976; O'Malley & Chamot, 1990; Oxford, 1990). The aim of using strategies is to “affect the learner’s motivational or affective state, or the way in which the learner selects, acquires, organizes, or integrates new knowledge” (Weinstein & Mayer, 1986, p. 315). Oxford (1990) mentioned, “Strategies are especially important for language learning, because they are tools for active, self-directed involvement, which is essential for developing communicative competence” (p.7). Cohen (1998) defines language learning strategies as: “Those processes which are consciously selected by learners and which may result in action taken to enhance the learning or use of a second or foreign language, through the storage, retention, recall, and application of information about that language” (p. 4).

The use of language learning strategies helps learners control their own learning by developing language skills, increasing confidence and motivation in learning process. Moreover, learning strategies support different aspects of understanding, retention, and application.

Foreign language teachers face various tasks related to ethnicity, historical, social and cultural dilemmas. Different types of perception, knowledge structures, and the choice of expressing feelings and emotions are inherent in both teachers and students. And, for both of them, language learning and cognition are inseparable from the emotional component, which regulates adaptability to the changes caused by external and internal educational processes.

Students’ cognitive style may vary, too. A distinction is often made between analytical learners who prefer a deductive approach (give them a rule and let them deduce other examples from it), and holistic learners, who prefer an inductive approach (give them examples, and let them induce the rule).

Considering these issues within the framework of psycholinguistics, some studies (O'Malley and Chamot, 1990; Ko and Rossen, 2017; Marlina, 2018; Lui et al., 2021; McCallum, 2022) offer a set of inter-related methods and strategies, including (1) main types of metacognitive, cognitive and affective or social methods and strategies in students’ educational experience; (2) systematic approach to developing the competencies of a foreign language learner and teacher; (3) introduction of a psycholinguistic approach to learning and teaching a foreign language (ELL/ELT), that outlines the students, their personal characteristics; (4) support of language training programmes for graduates and postgraduates with ethnolinguistic components of intercultural learning.

According to Dörnyei, strategies are not the end for language instruction, but are suggestions for enhancing learner’s motivation and confidence. Dörnyei emphasized the importance of regular praise, positive reinforcement and supportive environment in strategy instruction. It is also highlighted that language learning strategies can be trained and taught, which is also suggested by many researchers (Cohen, 1998; Dörnyei & Skehan, 2003).

Oxford (1990) identified twelve key features of language learning strategies as follows:

- Contribute to the main goal, communicative competence;
- Allow learners to become more self-directed;
- Expand the role of the teachers;
- Are problem-oriented;
- Are specific actions taken by the learners;
- Involve many aspects of the learner, not just the cognitive;
- Support learning both directly and indirectly;
- Are not always observable;
- Are often conscious;
- Can be taught;
- Are flexible;
- Are influenced by a variety of factors.

O'Malley and Chamot (1990) developed three types of strategies, namely metacognitive strategies (selective attention, planning, monitoring, and evaluating learning activity), cognitive strategies (rehearsal, organization, inferencing, summarizing, reducing, imagery, transfer, and elaboration), and social / affective strategies (interaction, cooperation, questioning for clarification, and participation in communicative activities with others).

Following the classification model of language learning strategies proposed by Oxford (1990), based on the synthesis of her classification of learning strategies and previous work, (O'Malley & Chamot (1990) divided language learning strategies into direct and indirect strategies.

Direct strategies involve direct learning and require mental processing of the language (Oxford,

1990), which include (1) memory strategies, which assist learners in storing and retrieving new information through techniques such as grouping, imagery, reviewing, association, and semantic mapping; (2) cognitive strategies, which involve learners to understand and produce new ideas, such as reasoning, practicing, receiving and sending messages, analyzing, and summarizing, note-taking, and translating; (3) compensation strategies, which enable learners to overcome gaps in linguistic knowledge by guessing meanings from the context, using synonyms, paraphrasing, or employing gestures when necessary.

Indirect strategies, in contrast, support learning indirectly without directly involving the target language itself, but are powerful to learning process (Oxford, 1990), which include (1) metacognitive strategies, which allow learners to regulate their learning, such as paying attention, planning, organizing, self-evaluating and monitoring one's errors or the learning process, (2) affective strategies, which help learners regulate their emotions, motivation, anxiety, and confidence during language learning, such as lowering anxiety, self-rewards, self-encouragement; (3) social strategies, which facilitate learning through interaction with native speakers or the target language, cooperation, questioning, and participation in communicative activities, such as asking questions, cooperating with peers and improving cultural understanding.

The competency-based approach (CBA) to language learning focuses on the development of learners' ability to apply linguistic knowledge and skills effectively in real-life communicative situations.

Within this framework, successful language learning depends not only on instructional practices but also on the strategic actions learners employ to achieve communicative competence. While competencies represent the intended educational outcomes, direct and indirect strategies provide learners with practical means of achieving those outcomes. Direct strategies support the acquisition and use of linguistic knowledge, whereas indirect strategies foster learner autonomy, self-regulation, motivation, and effective participation in communicative contexts. As a result, language learning strategies constitute an essential component of competency development.

The application of these strategies is closely connected with methods used in foreign language instruction. Indeed, teaching methods create the conditions in which learners can develop and employ both direct and indirect strategies. For example, communicative language teaching encourages compensation and social strategies through authentic interaction, while task-based language teaching promotes cognitive and metacognitive strategies as learners plan, execute, and evaluate communicative tasks. Similarly, project-based learning and content and language-integrated learning (CLIL) require learners to combine cognitive, social, and metacognitive strategies in order to achieve both linguistic and content-related objectives.

Therefore, within the competency-based paradigm, language-learning strategies and teaching methods should not be viewed as separate elements but ra-

ther as complementary dimensions of the learning process. Competencies define the educational goals; learning strategies constitute the mechanisms through which these goals are achieved, and the teaching methods provide the pedagogical framework that supports the strategic development of communicative competence. This interconnected relationship forms the theoretical basis for investigating the influence of language-learning strategies on foreign language teaching and learning outcomes.

Furthermore, learning strategies help learners reflect over their own way of learning, aiming at improving their processes. It helps the learners identify their difficulties and competences at the learning activity, in order to anticipate and fill students' gaps and needs during the learning situation and, if necessary, ask for help. Using the learning strategies, learners make dialogue with one another, to activate their previous knowledge on the handling material and relate them with new piece of information.

Language learners can use memory, metecognitive, affective, cognitive, social, and compensation strategies to enhance their language learning. The successful learners are more likely to use strategies according to specific tasks, context, or different needs. The more effective learners used strategies more "appropriately, with greater variety, and in ways that helped them complete the task successfully" (Chamot & Kupper, 1989, p. 17). Successful language learners are more flexible and appropriate in their use of learning strategies (Ellis, 1997). Different strategies have been reported associated with different learning styles (Oxford & Cohen, 1992). Various factors have been found to influence learners' choice of language learning strategies. Chamot and Kupper (1989), Oxford (1994) identified factors that influenced the choice of language learning strategies including age, gender, learning style, cultural background, type of task, motivation, attitudes, and teacher perceptions. Motivated learners are more likely to use more strategies than less motivated learners and the reasons for studying the language also contribute to the choice of learning strategies. Learners with negative attitudes or beliefs often use less effective strategies (Oxford, 1994; Oxford & Nyikos, 1993).

Learner training should facilitate learners having access to methodological resources such as techniques and activities, and using a criterion for selection of materials. Cheng (2000) claimed that "any teacher, Western or Eastern, who plans to use methodologies which inevitably involve students' participation must make sure that the students are familiar with and accept such methodologies" (Cheng, 2000, p. 444). Teachers and researchers have recognised the importance of training learners in effective strategy to promote learner autonomy (Wenden, 1991). Teachers are encouraged to choose learning strategies suitable for students and teach them how to understand learning strategies.

To exemplify strategy-based instructions (SBI) in university foreign language education, we can look at several practical examples. For instance, in academic reading, students learn to apply metecognitive strategies, like previewing academic articles, identifying key arguments, distinguishing between main and supporting ideas, and monitoring comprehension while reading

research literature. In academic vocabulary development, they use cognitive and memory strategies including the use of word families, collocations, and concept mapping.

In research presentation tasks, before delivering presentations, students learn planning and organizational strategies, such as outlining content, identifying key points, anticipating audience questions, and using compensatory strategies when encountering lexical gaps during oral communication.

In academic writing students are trained to use metacognitive strategies throughout the writing process, including planning, drafting, revising, self-editing, and evaluating their work according to academic criteria. Reflection logs encourage awareness of successful writing strategies.

Independent learning projects and seminar discussions are both key components in fostering autonomous learning in higher education. In independent learning projects, learners develop self-regulated learning skills by setting language learning goals, selecting appropriate strategies and resources, monitoring their progress, and evaluating outcomes. Meanwhile, seminar discussions offer a structured yet interactive environment where students practice their speaking, questioning, and argumentation skills through debates and discussions. Students learn how to ask for clarification, reformulate ideas, negotiate meaning, and maintain interaction when faced with communicative difficulties. In both cases, learners apply language strategically, engage deeply with content, and refine their communicative competence, all of which are crucial for academic success in a university setting.

Conclusions. Language learning strategies constitute an essential component of successful second language acquisition and play a significant role in the development of communicative competencies.

Appropriate language learning strategies result in greater motivation and confidence. Strategy instructions can enhance learners' self-efficacy and autonomous learning and help learners to take responsibility for their own learning. Teachers need to provide learners access to methodological resources and appropriate learning and modeling strategies, and guidance to help learners make progress in achieving academic success and professional needs.

References:

- Berger A. (2021). Advanced English language competency at the intersection of programme design, pedagogical practice, and teacher research: an introduction. Developing advanced English language competency, in *Developing Advanced English Language Competence*. Vol. 22, eds. A. Berger, H. Heaney, P. Resnik, A. Rieder-Bünemann and G. Savukova (English Language Education, Springer, Cham). [Google Scholar]
- Chamot, A. U., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign Language Annals*, 22, 13-24.
- Cohen, A.D. (1998). Strategies in learning and using a second language. London: Longman.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Malden, MA: Blackwell Publishing.
- Ellis, R., (1997). The study of second language acquisition. Oxford: Oxford University Press.
- Hashim, H. U., Hashim, H. (2018). Language learning strategies used by adult learners of teaching English as a second language (TESL). *TESOL International*, 13, 39-48.
- Hosenfeld, C. (1976). Learning about learning: Discovering our student's strategies. *Foreign Language Annals*, 9, 117-129.
- Ko, S., Rossen, S. (2017). Online learning: A practical guide, 4th. London: Rutledge. [Google Scholar]
- Kolodko, T. (2023). Shaping Communication Within the Framework of Globalization. *Ars Linguodidacticae*, 12 (2), 105-112.
- Lui, R., Shu, H, Li, P., Xu, Y., Yong, P., Li, R. (2021). AI-based language chatbot2.0 – the design and implementation of English language concept learning agent app. *Lect. Notes Comput. Sci.* 13089, 25-35. doi: 10.1007/978-3-030-92836-0_3 [CrossRef] [Google Scholar]
- O'Malley, J.M. & Chamot, A.U., (1990). Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press.
- Marlina, R. (2018). Teaching language skills, in the TESOL encyclopedia of English language teaching. ed. Lontas J., New York: Wiley, 3-15. [Google Scholar]
- McCallum, L. (2022). English language teaching in the EU: an introduction, in *English language teaching: theory, research and pedagogy*. ed. McCallum L. Singapore: Springer. 3-10. [Google Scholar]
- Oxford, R. L. (1990). Language Learning Strategies: What every teacher should know. New York: Newbury House/Harper & Row.
- Oxford, R. L. & Nyikos, M. (1993). A factor analytic study of language learning strategy use: Interpretation from information processing theory and social psychology. *Modern Language Journal*. 77 (1), 11-22.
- Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied Linguistics*. 2, 117-131.
- Rubin, J. (1987). Learner Strategies: Theoretical assumptions, research history and typology. In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Wegner, C., Minnaert, L., & Strehlke, F. (2013). The importance of learning strategies and how the project "Kolumbus-Kids" promotes them successfully. *European Journal of Science and Mathematics Education*, 1, 137-143. [Google Scholar] [CrossRef]
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching*. 3rd. New York: Macmillan. 315-327.

PHILOLOGICAL SCIENCES

ARABIC CONJUGATIONS IN YUNUS EMRE'S "DİVAN" (BASED ON THE KARAMAN COPY)

Guliyev B.

Baku State University Master's student

<https://doi.org/10.5281/zenodo.21105827>

Abstract

The Karaman copy of Yunus Emre's "Divan" is an outstanding example of a Sufi-oriented work written in Turkish. As can be seen from the work, the influence of Islamic culture is clearly visible in the state policy, literature, education and daily life of the period. The work of the poet, who lived in the 13th-14th centuries, still retains its relevance for readers and researchers today. Although it was written mainly in Old Turkic Turkish, it is undeniable that words of Arabic and Persian origin and the influences of these cultures are clearly observed. "Divan" has a lexical structure dominated by words of Turkish origin, but Arabic and Persian words also play an important role in "Divan". In addition, the use of both Turkish and Arabic collocations in Yunus Emre's Karaman "Divan" increases the grammatical richness of the work. When we approach the study and reading of the work from both lexical and grammatical aspects, we can clarify that the skillful use of Arabic collocations is a complementary factor in the work, both in terms of meaning and verse and rhyme. This, in turn, can be said to be a worthy basis for studying Arabic collocations in Yunus Emre's "Divan".

Keywords: Yunus Emre, Arabic collection, Divan, Karaman copy

Introduction

Yunus Emre, one of the leading figures of Anatolian Turkish literature, is very rich in terms of language, style and lexical layer. Most researchers consider the Karaman copy of his "Divan" to be an important source for studying the lexical system of early Anatolian Turkish. As we know from history, as we learn from the Seljuqname, the Ermenek Turks led by Karaman Bey's son Mehmet Bey marched to Konya and conquered it [3]. Mehmet Bey, who brought Cimrini to power, read the decision of the Seljuk Divan on May 13, 1276 [1]. These facts also allow us to easily be sure that at that time, both in the palace and among other subjects of society in general, the common language of communication was Old Turkish words and sometimes dialects. When we look at the lines in Yunus Emre's Karaman "Divan", the influences of Sufi philosophy are clearly visible in the writings of the poet who lived and created in the 13th-14th centuries [2]. This directly connects us to the poet's Sufi philosophy. We can say that it leads to the investigation of the relationship between the two and to clarify the results caused by their influences in his writings.

In the Karaman copy of the "Divan", it is observed that, along with Turkish lexical units, words of Arabic and Persian origin, as well as their morphological forms, are widely used.

As is known, the political, social and public processes taking place in the creativity of any thinker play an important role in the selection of his style, personal tendencies and the topics he writes about. The influence of the Islamic religion, which prevailed over other religions in the periods and geographical regions in which Yunus Emre lived and created, on both the general administration of the state, court literature and the works of important philosophers of the period is undeniable. We can superficially categorize these influences into 3 types. The first is the influence on the general worldview, the second is the influence of the fact that the Sufi ode to the love of God and the prophet

is given a place in the poems, specifically from a structural point of view, almost as a condition. Thirdly, and most importantly for our article, despite the fact that the main and main language is Old Turkic, the words used in oral and written language, and in some cases the suffixes, were borrowed from Arabic for Islamic religious reasons and some historical and political reasons. In this regard, Arabic plurals deserve special attention.

As we know, nouns in Arabic exist in three main forms in terms of quantity: singular, plural and plural. Singular nouns denote a single entity and can be used both in the definite form, for example: al-kitāb, and in the indefinite form, for example: kitāb [4].

The plural form, which is a special quantitative form of the Arabic language, denotes two entities, and is usually formed by adding the suffix -eyn (ين) to the end of the noun, and in less cases, the suffix -ān (ان) is used [4].

Secondly, quantities that are more than two are expressed in the plural form, which is divided into two types: regular plural (jām'i-sālim) and irregular plural (jām'i-mūkassār):

The plural form is based on the gender of the noun: in the masculine gender, -ūn (ون) or -īn (ين) is added, and in the feminine gender, the ending -āt (ات) is added. Based on this system, we can say that the morphological structure of the Arabic language, unlike the Turkish language, shows that the category of quantity is determined not only by suffixes, but also by gender and structural factors [5].

In the Divan literature, many poets we know can be found in the works of both Turkish, Arabic and Persian, or even though the language of the work is Turkish, there can be words of Arabic and Persian origin within a work. This is an indication that a poet who wrote masterfully in Old Turkish must also have mastered Arabic and Persian. As is known, the foundation of the concept of education in that period was based on the religious educational structure. That

is, the laws of the Quran, the works written by classical poets in Arabic, etc. were known to many of the main intellectuals of the period, and we can even say that this was considered necessary for them. As can be seen in the Karaman "Divan" of Yunus Emre, who lived and created during the Seljuk rule of the twelfth and thirteenth centuries, he is one of the poets of his time who wrote works in the Turkish language, and when we get to know the work closely, we can consider him one of the pillars that preserved the Turkish of that period. Thus, it can be considered inevitable and common to encounter Turkish collocations in his works. However, we also encounter Arabic collocations in the divan. Let's look at the examples in the divan:

bu din ü diyānetde dūnyā vü āhiretde
yétmiş éki millette ayırdur **āyātumuz** [4, s.218]
If we look at the above linguistic fact, the suffix -āt is noteworthy in the word **āyātumuz**. In feminine nouns, the ending indicating femininity is dropped and the suffix -āt (آت) is added. Its meaning is "signs", "marks".

kanı yūnus kanı cünbiş **harekāt** [4, s.189]
ne ser-māye ne varlık ne hū yok mer
The suffix -āt in the word **harekāt**, which strikes our attention in the verse, is an Arabic plural, and in modern Azerbaijani it means "movements".

gāh **ma'kūlāt** u **meşrū'āt** u takrir ü beyān
gāh **maḡşūrāt** olam gāh şahib-i keyvān olam [4, s.162]

In the example here, 3 words are used in Arabic plurals. These words are **ma'kūlāt**, **meşrū'āt**, **maḡşūrāt**. When translated directly into modern Azerbaijani, the meaning of the word **ma'kūlāt** is "things that are reasonable", **meşrū'āt** "things that are lawful", and **maḡşūrāt** "those that are dedicated to something".

kırk biñ kırk dört **tabaḡāt** meşāyih evliyālar [4, s.287]
dört ḡapudur kırk maḡām dem evliyā demidür
The word **tabaḡāt**, found in the first line, also uses the plural suffix, and when translated, it means "layers."
div ü peri ins ü melek sever seni her **mahlūḡāt** [4, s.314]

hayrān olup ileyūñde ṡurmuş-durur hūr u melek
The word **mahlu'at** used in the first line of the example is conjugated by taking the suffix -at. The meaning expressed by the word comes to mean "living beings" and in general, "everything created by God."

Evvel ḡapu **şeri'āt** emri nehyi bildürür [4, s.166]
yuya günāhlaruñı ḡur'ān harfī hicesi
Here, the word **şeri'āt** is also condensed by adopting the suffix -āt. As we know, the word **şeri'āt** is still used in modern Azerbaijani language today. It generally means "a system of religious laws and rules."

nişānsuzsın nişānun kimse bilmez
'aceb sultan 'aceb **mu'cizātuñ** var [4, s.164]

The word **mu'cizātuñ** used in the second line of the verse is also summed up like the above linguistic facts, and in our modern era, it expresses the word "miracles".

anı véren dervişe éki cihān keşf olur
anuñ **şifātın** öger ol hocalar hocası [4, s.167]
The word **şifāt** used in this verse has adopted the grammatical plural suffix -āt. In terms of meaning, it carries the meaning of "characteristics", "characteristics".

éy 'āşīḡān éy ' **āşīḡān** 'ışk mezhebi dīndür baña
[4, s.251]
gördi yüzüm dost yüzünü ḡamu yas düḡündür
baña

Finally, the suffix -ān, which indicates duality, draws our attention in the word **āshi'ān**, and the word literally means "lovers."

Conclusion

In conclusion, it can be noted that the analysis of Yunus Emre's Karaman "Divan" is a source of great importance for the lexical and grammatical study of the vocabulary of Old Turkic. The importance of this source is not only material, but also spiritual for the entire Turkic world. The works of Yunus Emre, who wrote and created in the 12th-13th centuries, are very rich in Sufi philosophy, and therefore continue to attract the attention of both amateur readers and professional researchers in the 21st century. Although the work was written in Old Turkic, for reasons related to the demands of the time, words of both Arabic and Persian origin were used in it. This is one of the important elements that enrich the work lexically. One of the striking elements from a grammatical point of view is the presence of Arabic conjunctions in the work, along with Turkish conjunctions. During the reading, we observe that Arabic conjunctions are used perfectly, observing their compliance with grammatical laws, and that the work is made unique in terms of both rhyme and structure. This clearly shows the inherent importance of the presence of Arabic conjunctions in our study.

References:

1. Erdoğan Mercil; "Who Accepted Turkish as the Official Language During the Seljuks of Turkey?", Belleten, C. LXIV, number: 239, 5. 51-57
2. Prof. Dr. Faruk Timurtaş, Yunus Emre Divan, Ministry of Culture Publishing, Ankara, 1980, 5. 1.
3. Yazıcızade Ali, Tevarih-i Al-i Selçuk, Haz. Dr. Abdullah Bakır, Çamlıca Basın Yayın, İstanbul, 2009, p. 832-838.
4. Asst. Assoc. Dr. Idris Nebi Uysal Yunus Emre Divan (Karaman copy). İstanbul-2014.
5. [file:///C:/Users/User/Downloads/Arap%C3%A7ada%20%C3%87o%C4%9Fullar%20\(Cem\)%20\(4\).pdf](file:///C:/Users/User/Downloads/Arap%C3%A7ada%20%C3%87o%C4%9Fullar%20(Cem)%20(4).pdf)

PHILOSOPHICAL SCIENCES

AUTOCOMMUNICATION AS A MODE OF ARTISTIC SEMIOSIS: THE PAINTING AS A PERFORMATIVE MODEL

Skovronskyi B.

*PhD in Philosophical Sciences, Associate Professor,
Department of Design and Fine Arts,
Mykhailo Drahomanov Ukrainian State University
Kyiv, Ukraine*

ORCID: <https://orcid.org/0000-0001-5413-6630>
<https://doi.org/10.5281/zenodo.21105831>

Abstract

The article investigates the phenomenon of autocommunication as a specific mode of artistic semiosis and examines its role in the constitution of aesthetic experience. Drawing upon the semiotic theory of Yuri Lotman, the hermeneutic philosophy of Hans-Georg Gadamer, and phenomenological analyses of image consciousness developed by Edmund Husserl and Roman Ingarden, the study proposes an interpretation of painting as a performative communicative event rather than a static visual object.

The research argues that artistic communication differs fundamentally from ordinary informational communication because its purpose is not the transmission of ready-made meanings but the generation of meaning through the active participation of the recipient. In this process, the viewer simultaneously occupies the positions of sender and receiver, transforming the act of perception into a form of autocommunication. Particular attention is devoted to the concept of presentation as a mediating structure through which visual signs acquire meaning. It is demonstrated that the constitution of pictorial meaning follows a logic analogous to Gadamer's concept of play, where the participant alternates between the roles of player and spectator and thereby contributes to the realization of the artwork itself.

The study concludes that a painting functions as a specific type of performance whose existence is inseparable from the interpretative activity of the viewer. Artistic experience emerges through the interaction of presentation, play, and autocommunication, resulting in the transformation of consciousness rather than the simple reception of information. This approach allows for a reconsideration of artistic semiosis as a dynamic process of meaning production and offers a new perspective on the communicative nature of visual art.

Keywords: autocommunication, artistic semiosis, painting, play, presentation, visual art, aesthetics.

Introduction

Communication remains one of the central categories of contemporary philosophy, semiotics, and cultural theory. Human existence unfolds within complex networks of communicative relations that structure individual experience and collective forms of life. Traditionally, communication has been understood as the transmission of information between different subjects. However, twentieth-century semiotics demonstrated that communicative processes frequently exceed the simple exchange of messages and often involve reflexive mechanisms through which consciousness communicates with itself [6].

This reflexive dimension becomes particularly important in the sphere of artistic experience. Unlike scientific discourse or everyday communication, works of art do not merely transfer information from an author to a recipient. Their significance depends on the active participation of the observer, who must reconstruct and actualize meaning through interpretation. Consequently, the artwork cannot be reduced to a passive container of information. Instead, it constitutes a communicative situation within which meaning emerges through interaction between the structure of the work and the interpretative activity of the viewer [1; 3].

Among the various approaches to this problem, Yuri Lotman's theory of autocommunication offers a

particularly productive framework. Lotman distinguished between communication directed toward another subject ("I-Other") and communication directed toward oneself ("I-I"). In the latter case, the communicative act does not primarily serve the transmission of new information but functions as a mechanism for restructuring consciousness [6]. Such a conception appears especially relevant for understanding aesthetic experience, where the recipient often encounters meanings that cannot be reduced to explicit informational content.

The relevance of this perspective increases when considered alongside Hans-Georg Gadamer's hermeneutic account of art. Gadamer characterizes play as the mode of being of the artwork and argues that aesthetic experience emerges through participation in a dynamic event rather than through passive observation [1]. This understanding opens the possibility of interpreting artistic perception as a specific communicative process in which the spectator becomes an active participant in the realization of the work itself.

The present study develops this possibility by examining the relationship between artistic semiosis, autocommunication, and play. It proposes that the perception of painting can be understood as a performative process analogous to theatrical performance, in which

the viewer simultaneously occupies the positions of actor and spectator. Through this process, the artwork becomes not merely an object of contemplation but an event of meaning production.

Problem Statement

The communicative nature of artistic experience remains one of the most debated questions in aesthetics and semiotics. Although artistic works are commonly described as communicative structures, it remains unclear how communication operates in situations where meaning is not explicitly transmitted but must instead be actively produced by the recipient.

Traditional models of communication assume a distinction between sender and receiver. In such models, the sender encodes information and the receiver decodes it. However, artistic perception often appears to function differently. The observer does not merely decode a pre-existing message but actively participates in constructing the meaning of the work. This raises a number of theoretical questions.

First, what distinguishes artistic semiosis from ordinary forms of signification? Second, how can meaning emerge in situations where the artwork itself does not possess intentional consciousness? Third, in what sense can aesthetic experience be understood as a form of autocommunication? Finally, how does Gadamer's conception of play contribute to understanding the structure of artistic perception?

Existing studies have examined these issues from different perspectives. Semiotic approaches emphasize the self-referential character of artistic signs [7], while phenomenological analyses focus on the constitution of meaning within consciousness [3; 4]. Hermeneutic theories, in turn, stress the participatory nature of aesthetic experience [1]. Yet these perspectives have rarely been integrated into a unified account capable of explaining how artistic meaning is generated through the interaction of presentation, interpretation, and self-communication.

The purpose of this article is therefore to investigate the structure of artistic perception as a form of autocommunication and to demonstrate that the process of pictorial presentation functions according to a logic analogous to Gadamerian play. By combining semiotic, phenomenological, and hermeneutic approaches, the study seeks to provide a comprehensive account of the communicative mechanisms underlying aesthetic experience.

Methodology and Source Base

The present study employs an interdisciplinary methodological framework that combines semiotic, phenomenological, and hermeneutic approaches to investigate the communicative structure of artistic experience. Such a combination is necessary because the phenomenon of autocommunication in art cannot be adequately explained within the boundaries of a single theoretical tradition. Artistic perception simultaneously involves the functioning of signs, the constitution of meaning within consciousness, and the interpretative processes through which the artwork becomes actualized in experience. Consequently, the analysis requires a methodological synthesis capable of addressing these interconnected dimensions.

The primary methodological foundation of the study is semiotic analysis. Within this framework, the artwork is examined as a sign structure whose meaning emerges through processes of interpretation rather than through the simple transmission of information. The research relies particularly on the semiotic theories developed by Roman Jakobson, Yuri Lotman, and John Deely [2;6;7]. Jakobson's concept of the self-referential or introversive function of artistic signs provides an important starting point for understanding how aesthetic communication differs from ordinary informational communication [7]. In artistic discourse, the sign does not merely refer to an external object but directs attention toward its own mode of presentation, thereby transforming the communicative process itself.

Special importance is assigned to Yuri Lotman's theory of autocommunication [6]. Lotman's distinction between communication directed toward another subject ("I-Other") and communication directed toward oneself ("I-I") serves as the conceptual basis for interpreting artistic perception as a reflexive communicative process. The study adopts Lotman's argument that autocommunication functions not primarily as a mechanism for transmitting information but as a means of restructuring consciousness through the transformation of interpretative codes. This theoretical perspective makes it possible to analyze the artwork not simply as a message but as a communicative event that alters the recipient's relationship to meaning.

John Deely's semiotic philosophy further contributes to the investigation by clarifying the mediating role of signs in the constitution of reality [2]. His understanding of semiosis as a process that connects sensory data with conceptual structures provides a useful framework for examining how visual representations acquire meaning within artistic experience. Through this perspective, signs are viewed not as passive substitutes for external objects but as active mediators that participate in the formation of a meaningful world.

The second methodological component of the study is phenomenological analysis. Phenomenology offers conceptual tools for investigating the relationship between perception, consciousness, and meaning. The research draws primarily on the works of Edmund Husserl, Roman Ingarden, and Vakhtang Kebuladze [3;4;5]. Husserl's theory of intentionality serves as a foundation for understanding how consciousness constitutes objects through acts of perception and interpretation [4]. Within this framework, the meaning of an artwork cannot be regarded as an objective property inherent in the material artifact itself. Rather, meaning emerges through intentional acts that organize sensory data into coherent structures of experience.

Roman Ingarden's analyses of pictorial representation are particularly significant for the present study [3]. Ingarden demonstrated that the perception of an image involves multiple ontological layers, ranging from material elements such as colors and lines to represented objects and ideal meanings. This stratified conception of the artwork provides a theoretical basis for explaining how visual signs can function simultaneously as physical marks, representational forms, and

carriers of aesthetic significance. The distinction between these levels is essential for understanding the process through which artistic presentation generates meaning.

The phenomenological dimension is further supplemented by the work of Vakhtang Kebuladze [5], whose interpretation of phenomenology emphasizes the constitution of the lifeworld through intentional experience. His analyses contribute to understanding how the viewer constructs a meaningful experiential horizon while engaging with artistic representations. From this perspective, the artwork does not merely depict reality but participates in the formation of a particular mode of experiencing reality.

The third methodological foundation of the research is philosophical hermeneutics, particularly the theory of play developed by Hans-Georg Gadamer [1]. Gadamer's conception of play as the mode of being of the artwork provides the central interpretative framework of the study. Unlike traditional aesthetic theories that regard the artwork as a static object, Gadamer understands art as an event that occurs through participation. The present research adopts this perspective in order to investigate the dynamic and performative dimensions of artistic perception.

Particular attention is devoted to Gadamer's claim that the true subject of play is not the player but the play itself [1]. This insight allows for a reinterpretation of artistic experience as a process in which the observer becomes integrated into a larger structure of meaning production. The concept of play thus serves as a bridge between semiotic and phenomenological approaches by explaining how signs, perception, and interpretation become unified within a single experiential event.

In addition to these primary methodological approaches, the study employs conceptual analysis and theoretical reconstruction. Conceptual analysis is used to clarify the meanings of key categories such as auto-communication, semiosis, presentation, image, and play. Since these concepts originate from different philosophical traditions, their systematic comparison is necessary to establish a coherent theoretical framework. Theoretical reconstruction is employed to identify structural similarities between artistic presentation and Gadamerian play, allowing for the development of a unified model of artistic communication.

The source base of the research consists primarily of classical works in semiotics, phenomenology, and hermeneutics. Particular attention is given to Lotman's writings on autocommunication and semiosphere theory, Jakobson's studies of poetic language and artistic signification, Deely's semiotic investigations, Husserl's phenomenological analyses of intentionality, Ingarden's theory of the artwork, Kebuladze's phenomenology of experience, and Gadamer's philosophical hermeneutics. These texts provide the conceptual foundations necessary for understanding the communicative and interpretative dimensions of artistic experience.

At the same time, the study engages with contemporary discussions in aesthetics, visual studies, and reception theory. Modern approaches increasingly emphasize the active role of the recipient in the production of meaning and the performative character of aesthetic

experience. By integrating these perspectives with classical theories of semiotics and hermeneutics, the present research seeks to develop a comprehensive account of artistic perception as a form of autocommunication.

The methodological synthesis proposed here allows the artwork to be examined simultaneously as a sign structure, a phenomenological object, and a hermeneutic event. Such an approach makes it possible to reveal the mechanisms through which artistic meaning is generated and to demonstrate that the perception of painting functions as a communicative process structurally analogous to play. Through this framework, the study contributes to contemporary debates concerning the nature of aesthetic experience, the role of interpretation in art, and the communicative foundations of artistic semiosis.

Research Findings

The phenomenon of autocommunication acquires particular significance in the sphere of artistic experience because art represents a unique form of communication in which the transmission of information is inseparable from the transformation of consciousness. Unlike ordinary communication, where the primary purpose is the exchange of information between different subjects, artistic communication operates according to a fundamentally different logic. The artwork does not merely convey a message from an author to a recipient; rather, it creates conditions under which the recipient becomes an active participant in the production of meaning. Consequently, the communicative process in art cannot be reduced to the conventional sender–message–receiver model. Instead, it involves a reflexive structure in which the recipient participates in the construction of meaning through a process of self-directed interpretation.

This characteristic becomes especially evident in visual art. A painting, for example, does not function as a simple informational medium. If the value of a painting consisted exclusively in communicating visual information about an object, photography or direct observation would fulfill this task more efficiently. The artistic significance of a painting therefore cannot be explained by its ability to represent external reality alone. Rather, its significance lies in the particular way reality is presented and experienced through the act of perception.

From a semiotic perspective, artistic communication differs from ordinary signification because the sign in art becomes self-referential. Roman Jakobson described this tendency as the introversive orientation of the artistic sign [7]. In such cases, attention shifts from what is represented to the manner of representation itself. The artwork directs perception toward its own presentation and thereby transforms the communicative act into an event of interpretation. Meaning is no longer transmitted as a ready-made content but emerges through the recipient's engagement with the structure of presentation.

This shift from representation to presentation constitutes one of the defining features of artistic semiosis. The represented object remains present, yet it ceases to

be the sole focus of attention. Instead, the recipient becomes aware of the relationship between the represented object and the expressive means through which it appears. The painting no longer functions merely as a substitute for reality but as a field in which reality is reconstructed through interpretative activity. Consequently, artistic meaning emerges neither from the depicted object alone nor from the material elements of the artwork, but from the dynamic interaction between them.

The communicative specificity of art becomes particularly apparent when examined through Lotman's concept of autocommunication [6]. According to Lotman, autocommunication occurs when the sender and the receiver of a message coincide within the same subject [6]. In such situations, the communicative act is not directed toward transmitting unknown information but toward restructuring the internal organization of consciousness. The message functions as a catalyst for self-transformation rather than as a vehicle for information exchange.

Artistic perception demonstrates precisely this structure. When a person encounters a painting, the informational content of the image is often already known. A landscape, portrait, still life, or seascape usually presents objects that are familiar and immediately recognizable. Yet recognition alone does not explain the aesthetic experience generated by the artwork. The significance of artistic perception lies in the fact that the viewer engages in a process of meaning construction that exceeds the mere identification of represented objects.

The viewer repeatedly interprets visual elements, correlates them with previous experiences, activates cultural associations, and organizes sensory impressions into coherent structures of significance. This activity is directed not toward discovering new information about the depicted object but toward generating a new mode of experiencing that object. The communicative process therefore becomes reflexive. The viewer effectively communicates with oneself through the medium of the artwork, using the image as a stimulus for internal transformation.

This process can be clarified through an analysis of pictorial representation. At the material level, every painting consists merely of pigments, lines, textures, and spatial arrangements distributed across a two-dimensional surface. These physical elements possess no intrinsic resemblance to the objects they represent. Nevertheless, viewers do not ordinarily perceive isolated marks on a canvas. Instead, they perceive landscapes, human figures, architectural structures, or symbolic forms.

The transition from material marks to meaningful objects requires a process of interpretation. Consciousness synthesizes visual data and organizes them into recognizable structures. This process involves at least two interconnected semiotic relationships. First, visual elements function as signs that refer to represented objects. Second, the resulting representation becomes meaningful within the horizon of the viewer's consciousness. Artistic experience therefore contains both a representational and an interpretative dimension.

Phenomenologically, these dimensions correspond to different levels of constitution [3;4]. On one level, consciousness organizes sensory impressions into coherent perceptual forms. On another level, it attributes significance to those forms by integrating them into broader structures of experience. The artwork thus functions as a mediating mechanism between sensory data and conceptual meaning.

The concept of presentation becomes crucial at this point. Presentation should not be understood merely as the display of an object. Rather, it denotes the process through which meaning is constituted in perception [3;4]. Presentation creates a situation in which the viewer must actively participate in the formation of significance. The artwork does not simply provide meaning; it requires meaning to be produced through interpretative engagement.

This characteristic explains why aesthetic experience cannot be reduced to passive observation. The viewer does not receive a finished meaning from the artwork. Instead, the viewer constructs meaning through participation in the process of presentation. Artistic communication therefore becomes inseparable from creative interpretation.

The dynamics of this process can be further illuminated through Hans-Georg Gadamer's concept of play [1]. Gadamer argues that play constitutes the fundamental mode of being of the artwork [1]. Unlike instrumental activities that pursue external goals, play possesses its purpose within itself. Its significance lies in participation rather than in the achievement of a predetermined outcome.

Several characteristics of play are particularly relevant for understanding artistic communication. First, play establishes a distinct experiential space governed by its own internal logic. Second, participants become absorbed into the movement of play itself. Third, play generates meaning through enactment rather than through external explanation. Finally, play achieves completion only when it is experienced by a participant capable of recognizing and actualizing its significance.

These characteristics closely resemble the structure of artistic perception. The artwork creates an experiential field within which ordinary modes of perception are temporarily suspended and replaced by the logic of aesthetic engagement. The viewer enters a specific interpretative situation governed by the internal organization of the artwork. Meaning emerges not through passive reception but through participation in this process.

Gadamer's most important insight concerns the relationship between player and spectator. In ordinary understanding, these roles appear distinct. The player performs actions, while the spectator merely observes them. However, Gadamer demonstrates that genuine participation in play requires the spectator to become internally involved in the unfolding of the game. The spectator does not remain external to the event but becomes integrated into its movement.

A similar structure can be identified in artistic perception. The viewer initially appears to occupy the position of an external observer. Yet meaningful engagement with the artwork requires active participation in

the process of interpretation. The viewer becomes simultaneously observer and participant. One perceives the image while simultaneously constructing its significance. The distinction between spectator and player therefore becomes blurred.

This convergence reveals the autocommunicative structure underlying artistic experience. The viewer continuously alternates between different interpretative positions. On one level, one observes the visual elements presented by the artwork. On another level, one actively organizes these elements into meaningful configurations. The viewer therefore becomes both sender and receiver within a single communicative process.

The communicative act unfolds as a movement between sensory perception and conceptual interpretation. Visual elements function as messages that consciousness directs toward itself. Meaning emerges through the recursive interaction between perception and interpretation. The artwork mediates this process without determining its outcome in advance.

Claude Monet's *Impression, Sunrise* provides an illustrative example of this mechanism. The painting depicts a harbor scene that is, in itself, relatively ordinary. Yet the artistic significance of the work does not derive primarily from the represented objects. Rather, it emerges from the particular manner in which light, atmosphere, color, and visual perception are presented. The viewer is not simply informed about the existence of a harbor. Instead, one is invited to participate in the experience of perceiving that harbor under specific atmospheric conditions.

The painting thus generates a communicative situation in which meaning is created rather than transmitted. The visual presentation functions as a code that must be interpreted. Different viewers may produce different meanings, yet all participate in the same process of autocommunicative engagement. The artwork provides the conditions for meaning formation without prescribing a single definitive interpretation.

From this perspective, the painting may be understood as a form of performance. Although no physical actor appears before the viewer, the image itself performs the represented reality. Colors, lines, textures, and compositional relations collectively enact the appearance of the depicted world. The viewer participates in this performance by reconstructing the represented reality through acts of interpretation.

The analogy with theatrical performance becomes particularly productive when considering the role of presentation. In theatre, actors do not merely exist on stage; they present characters through performance. Similarly, visual elements in a painting do not simply occupy space; they present a world through representation. The spectator recognizes this presentation and thereby transforms material signs into meaningful experience.

This process reveals the fundamentally performative nature of artistic semiosis. Meaning emerges not as a static property of the artwork but as an event occurring through interaction between artwork and recipient. The artwork becomes actual only through participation. Its significance exists neither exclusively in the object

nor exclusively in the subject but within the communicative process that unites them.

The transformative consequences of this process explain why autocommunication occupies such a central position in artistic experience. By engaging in aesthetic interpretation, individuals reorganize their perceptual and conceptual structures. The artwork provides a framework within which consciousness encounters itself in a new configuration. What changes is not merely the understanding of a particular image but the interpretative horizon through which reality itself is experienced.

This transformation corresponds closely to Lotman's description of autocommunication as a mechanism of internal restructuring [6]. Artistic experience modifies the codes through which meaning is generated. The viewer leaves the communicative encounter not because new information has been acquired but because the structure of interpretation itself has been altered.

Consequently, the aesthetic value of art cannot be reduced to informational content. Its significance lies in its ability to generate situations of autocommunication that reorganize consciousness through participation in presentation. Artistic meaning emerges through the interaction of signification, perception, interpretation, and play. The artwork functions as a communicative environment within which individuals become simultaneously creators and recipients of meaning.

The analysis conducted in this study therefore supports the conclusion that the semiotic structure of visual art is fundamentally autocommunicative. Presentation functions as a performative process analogous to Gadamerian play, while artistic perception represents a reflexive communicative event through which consciousness constitutes and transforms its own experiential world. In this sense, the artwork may be understood not simply as an object of representation but as a communicative performance that enables the continuous recreation of meaning through aesthetic experience.

Discussion

The interpretation of artistic perception proposed in this study contributes to several contemporary discussions in aesthetics, semiotics, and philosophy of communication. Most notably, it offers an alternative to informational models of communication that continue to dominate both classical communication theory and many contemporary approaches to media studies. While traditional models emphasize the transmission of messages between distinct subjects, the present analysis demonstrates that artistic communication operates according to a fundamentally different principle. The artwork does not merely serve as a vehicle through which information passes from an author to a recipient. Rather, it establishes a communicative environment within which meaning is actively constituted by the recipient through processes of interpretation and self-reflection.

This conclusion aligns with broader developments in twentieth-century philosophy, particularly within phenomenology and hermeneutics. Husserl's conception of intentionality [4], Ingarden's theory of the lay-

ered structure of the artwork [3], and Gadamer's understanding of art as an event of truth [1] all challenge the notion that meaning exists independently of experience. The present study extends these perspectives by emphasizing the communicative dimension of aesthetic experience and by demonstrating that the constitution of meaning in art possesses an autocommunicative structure. In this regard, artistic perception may be understood not only as interpretation but also as a form of self-directed communication through which consciousness reorganizes its own structures of understanding.

The analysis also contributes to ongoing discussions concerning the active role of the recipient in artistic meaning production. Reception aesthetics and reader-response theories have long argued that meaning emerges through interaction between artwork and audience. However, these approaches often focus on the plurality of interpretations without fully explaining the communicative mechanism through which such interpretations are generated. By incorporating Lotman's concept of autocommunication [6], the present research provides a more precise account of this mechanism. Meaning does not arise simply because different viewers possess different cultural backgrounds or subjective preferences. Rather, it emerges because the artwork initiates a communicative process in which the recipient simultaneously occupies the positions of sender and receiver, thereby transforming interpretation into an act of self-construction.

The proposed model also sheds new light on the relationship between representation and presentation in visual art. Conventional theories of representation frequently concentrate on the correspondence between an image and its referent. Within such frameworks, artistic value is often evaluated according to the degree of realism, symbolic complexity, or conceptual content exhibited by the artwork. The findings of the present study suggest that these criteria alone are insufficient for explaining aesthetic experience. The decisive factor is not what is represented but how representation becomes an event of presentation. Artistic significance emerges through the viewer's participation in the process by which visual elements are transformed into meaningful objects of experience.

This perspective allows for a reconsideration of Gadamer's theory of play [1]. In philosophical hermeneutics, play functions primarily as an ontological model describing the mode of being of the artwork. The present analysis extends this understanding by revealing the communicative implications of Gadamer's concept. The structure of play closely resembles the structure of autocommunication [1; 6]. Both involve the temporary suspension of ordinary subject-object relations, the creation of a distinct experiential space, and the transformation of participants through involvement in a self-contained process. Consequently, artistic play may be interpreted as a specific communicative configuration in which meaning emerges through participation rather than transmission.

At the same time, the comparison between artistic presentation and play raises important questions concerning the nature of aesthetic agency. The analysis suggests that neither the artist nor the viewer functions

as the sole source of meaning. Instead, meaning arises within the dynamic interaction established by the artwork itself. This conclusion corresponds closely to Gadamer's claim that the true subject of play is not the player but the play itself. Similarly, the present study indicates that artistic communication cannot be fully attributed to either the creator or the recipient. The artwork establishes conditions under which meaning becomes possible, while its actual realization depends upon the participatory engagement of consciousness.

An important implication of this argument concerns the status of artistic interpretation. Contemporary debates often oscillate between two opposing positions. On one side, meaning is regarded as fixed by authorial intention; on the other, meaning is considered entirely dependent upon the subjective responses of recipients. The model developed in this study offers a mediating alternative. Meaning is neither completely predetermined nor entirely arbitrary. The artwork constrains interpretation through its formal organization, yet it remains open because the realization of meaning requires active participation by the viewer. Autocommunication thus explains how artistic experience can simultaneously possess structure and interpretative plurality.

The findings may also contribute to contemporary discussions surrounding immersive media, digital art, and interactive technologies. Many recent artistic practices seek to increase audience participation and blur the distinction between creator and recipient. Interactive installations, virtual environments, and algorithmically generated artworks frequently position viewers as active participants in the production of aesthetic experience. The theoretical framework developed here suggests that such developments do not represent a radical break from traditional artistic communication. Rather, they intensify a communicative structure that has always been present in art. Even in the case of traditional painting, the viewer participates in the construction of meaning through autocommunicative processes. Digital and interactive media merely make this participatory dimension more explicit.

Despite its theoretical contributions, the present study also possesses several limitations. First, the analysis remains primarily conceptual and philosophical. Although numerous examples from visual art support the proposed framework, the research does not include empirical investigations of audience reception. Future studies could examine how viewers actually experience autocommunicative processes during encounters with artworks and whether such experiences vary across cultural or educational contexts.

Second, the discussion focuses predominantly on painting and visual representation. While the concepts of autocommunication and play appear applicable to other artistic media, their specific manifestations may differ considerably. Music, architecture, literature, theatre, and digital media each possess unique formal characteristics that influence the production of meaning. Further research is therefore required to determine how the proposed model functions within different artistic domains.

Third, the present study concentrates primarily on the individual recipient. However, artistic experience

frequently occurs within collective contexts, including museums, exhibitions, performances, and cultural rituals. The relationship between individual autocommunication and collective meaning formation remains an important topic for future investigation. Such research could potentially establish connections between Lotman's theory of the semiosphere [6] and contemporary theories of social communication.

Overall, the findings support the view that artistic perception should be understood as a dynamic communicative process rather than as a passive reception of aesthetic information. By integrating semiotic, phenomenological, and hermeneutic perspectives, the present study demonstrates that the artwork functions as a medium through which consciousness enters into communication with itself. This process unfolds through presentation, interpretation, and play, resulting in the continual reconstruction of meaning and the transformation of experience. The concept of autocommunication therefore provides a productive framework for understanding not only the semiotic structure of artworks but also the broader role of art in shaping human consciousness and cultural experience.

Conclusions

The present study has examined the phenomenon of autocommunication as a fundamental dimension of artistic semiosis [6] and has demonstrated its significance for understanding the communicative structure of aesthetic experience. Through a synthesis of semiotic, phenomenological, and hermeneutic approaches [1–7], the research has proposed a theoretical model in which the perception of an artwork is interpreted not as the passive reception of information but as an active process of meaning generation. Within this framework, artistic communication differs substantially from conventional models of information transfer because it operates through the reflexive involvement of the recipient in the constitution of meaning.

The analysis has shown that the artwork cannot be adequately understood as a simple medium for transmitting a predetermined message from an author to an audience. Instead, it functions as a communicative environment that initiates and sustains interpretative activity. Meaning emerges through the interaction between visual signs, perceptual processes, and conceptual structures of consciousness. Consequently, the recipient becomes an active participant in the production of meaning rather than a passive decoder of information. This participatory character of aesthetic experience reveals the autocommunicative nature of artistic perception, in which the recipient simultaneously occupies the positions of sender and receiver within a reflexive communicative process.

Particular attention has been devoted to the role of presentation in artistic semiosis. The study has argued that the aesthetic significance of an artwork lies not primarily in what is represented but in how representation is presented and experienced. Visual signs do not merely refer to external objects; they create conditions under which consciousness actively reconstructs and interprets reality. Artistic presentation therefore functions as a process of mediation between sensory per-

ception and conceptual meaning. Through this mediation, the artwork becomes a site where new experiential configurations can emerge.

The research has further demonstrated that Gadamer's concept of play provides a productive framework for understanding the dynamics of artistic communication [1]. The structure of play reveals important similarities with the structure of autocommunication. In both cases, meaning is not transmitted as a finished content but arises through participation in a self-sustaining process. The participant enters a distinct experiential space governed by its own internal logic and becomes involved in the unfolding of meaning. The study has therefore proposed that artistic presentation may be interpreted as a performative event analogous to play, where the recipient alternates between the roles of observer and participant while actively contributing to the constitution of aesthetic significance.

An important theoretical contribution of the study lies in the identification of structural parallels between visual representation and theatrical performance. The artwork has been interpreted as a form of presentation in which visual elements perform the represented reality rather than merely depict it. Through the act of perception, the viewer reconstructs the presented world and thereby participates in a communicative process that resembles a performance or spectacle. This perspective highlights the dynamic and event-like character of artistic meaning, emphasizing that the artwork becomes fully actualized only through interpretative engagement.

The findings also contribute to contemporary discussions concerning the role of the recipient in aesthetic experience. While many modern theories recognize the active participation of audiences in meaning production, the concept of autocommunication provides a more precise explanation of the mechanism through which such participation occurs [6]. Artistic experience transforms consciousness not because it communicates new information but because it reorganizes the interpretative structures through which reality is understood. The aesthetic encounter thus functions as a process of self-transformation mediated by signs, images, and symbolic forms.

At a broader level, the study suggests that art performs an essential cultural function by creating spaces in which individuals can engage in reflective communication with themselves. Through participation in artistic presentation, consciousness acquires the capacity to reorganize its own experiential horizons and generate new forms of understanding. In this sense, the value of art extends beyond representation, expression, or entertainment. Art becomes a medium through which human beings continually recreate the meanings that structure their experience of the world.

The proposed theoretical framework opens several directions for future research. Further studies may investigate how autocommunicative processes operate in other artistic media, including music, architecture, literature, theatre, and digital art. Additionally, empirical research could explore the ways in which different audiences engage in autocommunicative meaning pro-

duction and how such processes are influenced by cultural, historical, or technological contexts. Such investigations would contribute to a more comprehensive understanding of artistic semiosis and the role of aesthetic experience in contemporary culture.

In conclusion, the study has demonstrated that autocommunication constitutes a central mechanism of artistic perception [6]. By interpreting the artwork as a form of play and presentation [1; 3], it becomes possible to understand aesthetic experience as a communicative event through which meaning is continually created, negotiated, and transformed. The artwork therefore appears not simply as an object of representation but as a dynamic process that enables consciousness to encounter and recreate itself through the medium of artistic form.

References:

1. Gadamer H.-G. *Truth and Method* / Hans-Georg Gadamer ; translated and revised by J. Weinsheimer, D. G. Marshall. 2nd revised ed. New York : Continuum, 2004. 601 p.
2. Deely J. *Basics of Semiotics*. 4th ed. Tartu : Tartu University Press, 2005. 175 p.
3. Ingarden R. *Ontology of the Work of Art: The Musical Work, the Picture, the Architectural Work, the Film* / translated by R. Meyer, J. T. Goldthwait. Athens, Ohio : Ohio University Press, 1989. 344 p.
4. Husserl E. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book*. Translated by F. Kersten. Dordrecht : Kluwer Academic Publishers, 1983. 448 p.
5. Кебуладзе В. *Феноменологія досвіду* / Вахтанг Кебуладзе ; відп. ред. А. Лой. Київ : Дух і Літера, 2011. 280 с.
6. Лотман Ю. М. *Внутри мыслящих миров: Человек – текст – семиосфера – история*. Москва : Языки русской культуры, 1996. 464 с.
7. Jakobson R. *Selected Writings*. Vol. 2: *Word and Language*. The Hague : Mouton, 1971. 752 p.

PHYSICAL SCIENCES

DOUBLE SLIT EXPERIMENT AND WAVE-PARTICLE DUALITY

Abbasov V.

Independent theoretical physicist.

Professional Member, The New York Academy of Sciences,

Member, American Association for the Advancement of Science(AAAS),

Early Career Member, American Physical Society(APS),

Associate Member, Sigma XI-The Scientific Research Honor Society,

Student member, International Association for Advanced Materials,

Baku, Republic of Azerbaijan.

ORCID: 0009-0009-8154-1741.

<https://doi.org/10.5281/zenodo.21105833>

Abstract

The double-slit experiment is a classic experiment that presents the clearest manifestation of wave-particle duality, one of the fundamental principles of quantum mechanics. The double-slit experiment still remains the phenomenon that demonstrates wave-particle duality at the most fundamental level. This phenomenon quantitatively and conceptually demonstrates that microscopic objects can behave like both waves and particles at the same time. This article provides an in-depth analysis of the physical nature of dualism, the mechanism of formation of the interference pattern, and the impact of the act of observation on the dynamics of quantum systems. The need for a new perspective on this topic is due to the fact that the current uncertainty in the interpretation of quantum reality has not yet been completely eliminated, and the boundaries between classical and quantum behavior in microscopic objects are still being explored. The results presented here restructure the deep mechanisms of the interference picture, more clearly frame the relationship between observation and wavefunction collapse, and provide a quantum. Since the article attempts to provide a complete answer to the mechanism of the formation of wave-particle duality, extensive space is given to sections such as historical development, theoretical models, modern research, etc. New approaches to this phenomenon have been developed using modern research and string theory. As a result, a new mathematical expression has been proposed that explains the reasons for the initial emergence of wave-particle duality, and is also important for string theory. The article explains the measurement problem with a more effective physical structure and ontological model.

Keywords: Double-slit experiment, wave-particle dualism, interference pattern, measurement problem, observer effect, Copenhagen interpretation, string theory.

Introduction. Attempts to understand the nature of reality in the subtlest layers of the universe sometimes result in paradoxes that exceed the limits of the human mind. The most profound and thought-provoking of them is undoubtedly the double-slit experiment - a fundamental event that has become a symbol of how particles behave both as waves and as particles[1]. The double-slit experiment challenged the logic of classical physics and introduced into science the idea that observation itself affects reality. The double-slit experiment has puzzled physicists and philosophers for centuries, redefining the relationship between the observer and the observed in the depths of reality[2].

Historical development. The concept of wave-particle duality led to one of the most profound philosophical and scientific changes in the history of physics. This idea brought a new perspective on the nature of the Universe, explaining how light and matter can behave as both particles and waves. The scientific roots of wave-particle duality extend from observations several centuries ago to the fundamental principles of modern quantum mechanics[3].

In the 17th century, there were two main ideas about the nature of light. In his work "Optics", I. Newton proposed that light consists of material particles, that is, it has a corpuscular nature[4]. In response to this idea, Christiaan Huygens proposed the wave theory in 1690, which showed that light propagates in the form of waves[5]. Since I. Newton's influence was stronger at that time, the particle model remained dominant in the scientific world for a long time[6].

However, precise experiments conducted in the early 19th century radically changed this perception. The famous double-slit experiment performed by Thomas Young in 1801 conclusively proved the wave nature of light. T. Young observed that light passing through two adjacent narrow slits (or apertures) in a perforated screen placed in front of a light source produces an interference fringe pattern on a second screen positioned behind it. The result showed that light waves overlapped each other, forming zones of amplification and attenuation[7].

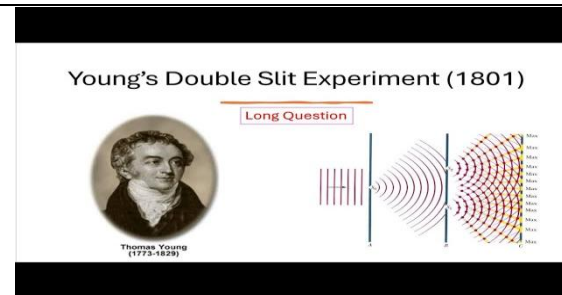


Figure 1. T.Young's double slit experiment

The double-slit experiment strengthened the mathematical basis of the wave model by clarifying the scientific content of concepts such as "interference"*¹, "coherence"*², and "phase difference"*³ for practice[8]. In the second half of the 19th century, James

Clark Maxwell formulated the electromagnetic equations. James Clark Maxwell showed that light is an electromagnetic wave consisting of correlated fluctuations of electric and magnetic fields[9].

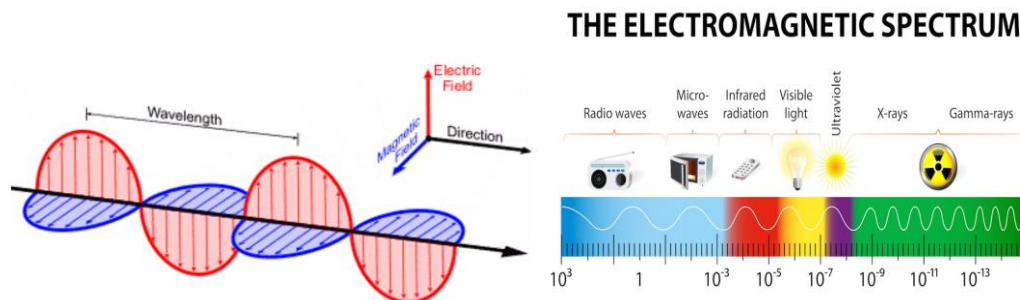


Figure 2. Electromagnetic waves and their spectrum

Thus, the wave nature of light was fully confirmed within the framework of electromagnetic theory.

*¹Interference is a fundamental wave phenomenon in which two or more coherent waves of the same physical origin overlap in space, and depending on their phase difference, regions of enhancement and attenuation are formed in the intensity (or amplitude) of the resulting wave. This phenomenon manifests itself in both classical (light, sound) and quantum (matter waves) physics[10].

*²Coherence is the degree to which waves (classical or quantum) have a stable phase relationship in time and space. High coherence is a condition for the ability of waves to interfere and thus form clear interference patterns. It is divided into two main types: temporal (over time) and spatial (over space) coherence[8].

*³Phase difference is the main parameter that quantitatively determines the location of the fringes in an interference pattern. It arises from the optical path difference of waves emerging from two slits and reaching the same point on the screen. This path difference

determines how far the waves are behind or ahead of each other, that is, the phase difference ($\Delta\phi$).

•Bright fringe (maximum): When the phase difference is equal to $0, 2\pi, 4\pi, \dots$ (or $\Delta\phi = 2\pi m$, where $m = 0, \pm 1, \pm 2, \dots$ is an integer), that is, when the optical path difference is an integer multiple of the wavelength ($m\lambda$), the waves overlap in phase and reinforce each other.

•Dark fringe (minimum): When the phase difference is equal to $\pi, 3\pi, 5\pi, \dots$ (or $\Delta\phi = \pi(2m + 1)$), that is, when the optical path difference is an odd multiple of half the wavelength ($(m + 1/2)\lambda$), the waves overlap in opposite phase and cancel each other[11].

But at the beginning of the 20th century, some experiments shook this solid wave model. In 1905, when Albert Einstein explained the photoelectric effect, he considered the energy of light to be divided into quanta—that is, photons. This fact showed that light could also behave like a particle. According to A.Einstein, the energy of each photon is equal to the product of the Planck constant and the corresponding frequency[12].

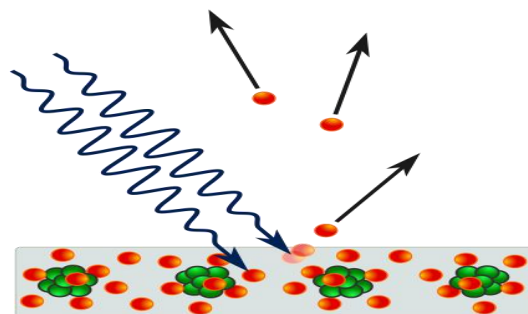


Figure 3. Photoelectric effect phenomenon

Thus, the conclusion was formed that light has both a wave and a particle nature.

In 1924, French physicist Louis de-Broglie generalized this dual nature of light and proposed the bold idea that matter particles had wave properties. Inspired by the relationship between the expression $E=hf$ used for the energy of a quantum and the momentum of a photon ($p = h/\lambda$), L. de Broglie assumed that every particle with mass would also have momentum p and wavelength $\lambda = h/p$. [13;15]. L. de-Broglie put the issue this way: for hundreds of years, corpuscular concepts were ignored in optics and the wave theory was preferred.; In the theory of matter, the opposite has happened, that is, excessive attention has been paid to the "particle" concept and the wave picture has been forgotten. If, for light and electromagnetic radiation in

general, quantum phenomena consist in the coexistence of corpuscular properties alongside wave characteristics, then for particles of matter (such as electrons, protons, neutrons, atoms, and molecules), quantum phenomena consist in the coexistence of wave properties alongside corpuscular characteristics[14]. L. de-Broglie's idea was that it was necessary to understand the similarity between mechanics and optics in a broader sense and to accept the existence of wave mechanics, similar to wave optics. Wave mechanics itself is more general than classical mechanics, and should be applicable to intraatomic motions, and under certain conditions should be able to pass into Newtonian mechanics[14]. His proposal gave rise to the concept of "de-Broglie waves" and became one of the main pillars of quantum mechanics[15].

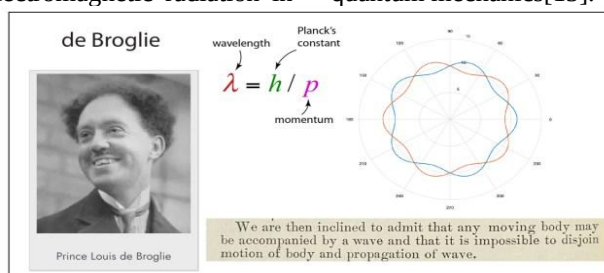


Figure 4. De-Broglie hypothesis

What distinguishes de-Broglie waves from ordinary waves is that the waves passing by the observer cannot be counted[14].

The electromagnetic diffraction experiment carried out in 1927 by Clinton Joseph Davisson and Lester

Halbert Germer experimentally confirmed de-Broglie's idea. When electron beams were directed into a crystal lattice, an interference pattern was observed, definitively proving that electrons behaved like waves[16].

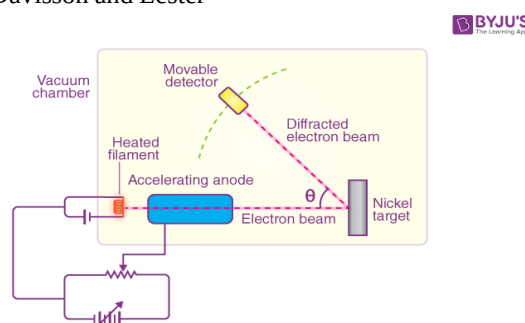


Figure 5. Davisson-Germer experiment

This experiment was a fundamental result showing that both light and matter behave as both waves and particles at the same time.

The double slit experiment became the deepest and most effective expression of this dualism. One of the most famous events during this period was the Einstein-Bohr conflict; Einstein's position: quantum mechanics is incomplete. The disappearance of the wave pattern during observation is not without reason, it is caused by the interaction of the measurement with the

system. However, this effect is local and explainable. He rejected chance with the statement "God does not play dice."

Bohr's position: quantum mechanics is now complete. The experimental setup determines the outcome. The interaction with the measuring device irreversibly changes the system and collapses the wave function. An unobserved system has no properties in the classical sense[17].



Figure 6. Bohr (left) and Einstein (right)

That is, he considered Einstein's quantum mechanics "incomplete" and thought that complete information could be obtained through measurement without disrupting the system. In his double-slit experiment, A.Einstein was trying to find a way to find out which slit a photon passed through without destroying it (for example, by moving the slit and measuring its momentum). This was an attempt to bypass the Heisenberg uncertainty principle. Bohr argued that this was impossible, that the attempt at measurement would inevitably change the system itself, thereby destroying the interference pattern[18].

In modern double-slit experiments with light and electrons, it is observed that wave behavior (interference) appears when no path detector is placed in the system (measured), but when a detection or measurement is made that determines which slit the particle passes through, the result is purely particle behavior. This experimental fact reveals the fundamental influence of the measurement process on quantum systems and constitutes the empirical basis of the central conceptual difficulty of quantum mechanics known as the "measurement problem"[19]. Thus, the double-slit effect and wave-particle duality are not just an experiment or a model, but a revolution that has radically changed human thinking about the fundamental structure of the Universe. This history is the story of a great

scientific journey, from classical optics to quantum theory of light, showing that matter and energy are essentially the same entity[17].

Description of the event. As we have seen, there are experiments in the history of physics that have left a deep mark not only with their results, but also with the paradoxical truths they teach us about nature. This seemingly simple experiment reveals one of the most puzzling truths of quantum mechanics. Let's explore this miraculous phenomenon step by step.

1. In the classical world.

Imagine firing a machine gun at a curtain with two narrow slits in front of it. The bullets will pass through the slits and leave marks on the curtain behind. The result we expect is clear: we will see two distinct stripes on the curtain corresponding to the slits.

Now let's change the hypothesis. Let's send waves from a bucket into the water. When the waves reach two slits, each slit becomes a new wave source. These two new waves are facing the curtain behind. In some places they reinforce each other (hills overlap with hills, valleys overlap with valleys), and in some places they cancel each other out (hills overlap with valleys). The result is an image of bright and dark stripes called an interference pattern[10].

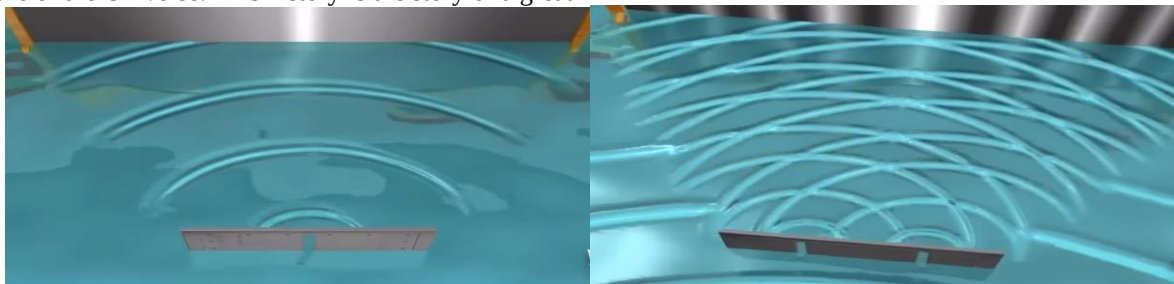


Figure 7. Passage of classical waves through odd and even halves

2. In the quantum world.

Now let's repeat this experiment at the quantum level, for example with electrons. The electron is a tiny particle as far as we know. What was expected? If we fired the electrons one by one, each one would have to pass through one of the slits and create a streak on the curtain behind it, like bullets. However, the experiment gave a completely different result.

As the electrons gradually accumulate on the screen, something amazing happens: an interference

pattern appears. The electrons sent out one by one, as if aware of the electrons that came before and after them, created this pattern by behaving like waves and not falling into their undying streaks. This fact was as if each electron, even though it was alone, had passed through both slits at the same time. This event revealed a fundamental principle of quantum physics: wave-particle duality[20].

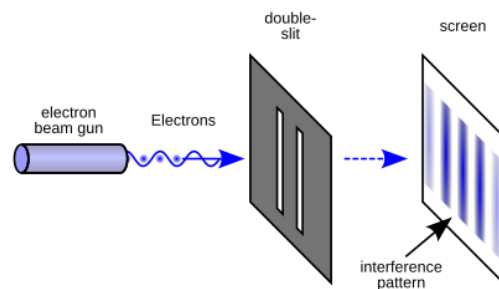


Figure 8. Double slit experiment with electrons.

3. The confusing part of the experiment: the observation problem.

Physicists asked the question: 'If an electron passes through both slits, can we measure its path?' The experimenters placed detectors near the slits. As the detectors interacted with the electron, one of the most fundamental principles of quantum mechanics became apparent: the measurement process changes the system.

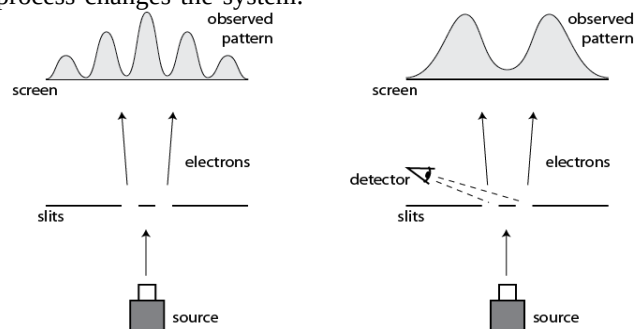


Figure 9. The observer effect in the double slit experiment

The measuring device and the quantum system were inextricably linked. This paradoxical result gave rise to the idea that the properties of a quantum system depend on the measurement context[21;22]. When we imagine the electron not as an object, but in terms of its wave function, we see that when the electron passes through two slits, it falls into a superposition of its two quantum paths. This means that the electron is moving from both the right and the left at the same time, and these two traces interfere with each other. It is the electron's own interaction with itself that creates this interference[21;22].

This very idea culminated in Schrödinger's famous "cat" thought experiment. Here, the notion that quantum states can exist and disappear simultaneously while not being observed demonstrates that reality itself is shaped by observation[15].

The interference pattern immediately disappeared, and the electrons behaved like classical particles, forming two distinct bands on the screen. It was as if the measurement of his path prevented him from taking both paths at the same time[20].

Schrödinger's thought experiment. In the scenario imagined by Schrödinger, the following elements are placed inside a completely closed box: A cat, a sample of radioactive material, a Geiger counter, a mechanical system (a mechanism that activates when the detector signals, opening a toxic gas chamber). The half-life of a radioactive substance is chosen so that there is a 50% chance that the atom will decay within a certain time. According to quantum mechanics, when no observation is made, the atom exists in a linear superposition of both "split" and "unsplit" states. This means that the system as a whole - the atom, the detector, the mechanism, and the cat - is described in two possible states at the same time:

- The atom has split, the mechanism has been activated, and the cat is dead.
- The atom has not split, the mechanism has not worked, the cat is alive[23].

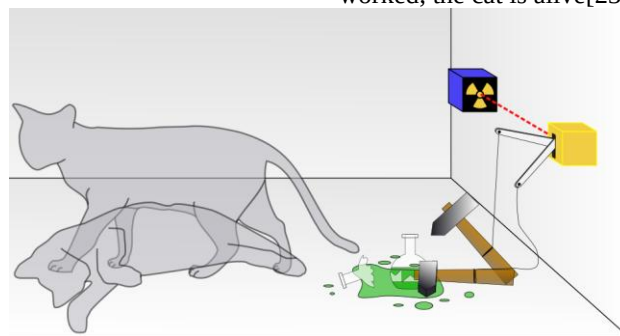


Figure 10. Schrödinger's thought experiment

These two states exist simultaneously until an observation (measurement) is made. That is, before opening the box, the cat is neither fully alive nor fully dead; it is described by a wave function in which both the alive and dead states exist simultaneously in a quantum superposition state. When the box is opened and an observation is made, the wave function "collapses" into one of two possible outcomes. As a result, the observer sees either a live or a dead cat. This phenomenon is known as the measurement problem in quantum mechanics: the act of observation causes the system to transform from a superposition state into a concrete state of reality[24].

This mechanism highlights the principal philosophical and theoretical difficulty in reconciling quantum systems with classical physics—namely, the absence of observable quantum superposition in macroscopic objects and the issue of the influence of consciousness on physical reality. This thought experiment later inspired the formation of various approaches such as the American physicist Hugh Everett's theory of parallel universes, the Copenhagen interpretation, and the Decoherence models[25]. Now let's look at these solution models.

1. Pilot-wave theory.

Proposed by Louis de Broglie in 1927 and later fully formulated by British physicist David Bohm, this theory returns quantum processes to a deterministic framework. Here, particles actually exist and their motion is controlled by a "pilot wave." That is, a particle does not move without a reason. An invisible wave determines the motion of the particle. The wave function obeys Schrödinger's equation, but this wave "guides" the particle and shapes its trajectory. Pilot-wave theory eliminates the observer's influence on reality and al-

lows us to explain the EPR paradox and Bell's inequality. However, it is difficult to directly apply the theory to relativistic systems and field theories[26].

2. Decoherence.

This model, developed by H.Dieter Zeh and Wojciech Zurek, explains the problem of quantum observation without a collapse mechanism. Decoherence is the phenomenon of the disappearance of interference terms as a result of the system's exchange of information with its environment. In this process, the wave function mathematically remains in superposition, but the observed classical reality appears to be a single result. Interaction with the environment "separates" the states of the system and coherence disappears. Thus, we observe only one "classical" result. This fact provides a real physical explanation for Schrödinger's cat paradox - the cat's superposition becomes physically unobservable as a result of its interaction with the environment[27].

3. Copenhagen interpretation.

In this explanation, put forward by Niels Bohr and Heisenberg, the wave function is not the essence of the object itself, but the probability distribution of the knowledge known to the observer. The system exists in a state of multiple superpositions unless observed. During observation, the wave function "collapses" into a probabilistic state and appears as a particle. N.Bohr's complementarity principle plays a key role here - the concepts of particle and wave are two complementary forms of reality that do not negate each other, but do not manifest at the same time. This interpretation is consistent with experimental results, but leaves a serious philosophical gap in the question of the objective existence of reality. The problem arises when the observer is given too much "metaphysical" power[28].

The Copenhagen Interpretation:

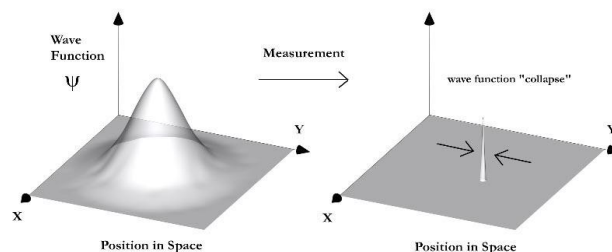


Figure 11. Wave function collapse.

The Copenhagen interpretation postulates the collapse of the wave function at the moment of measurement and thereby accurately predicts the observed outcomes of quantum systems; however, it leaves unresolved ontological questions concerning the mechanism and objectivity of the measurement process[29]. Taking this gap into account, British physicist Sir Roger Penrose presents the collapse of the wave function as an objective process related precisely to physical gravity: different superpositions create different space-time geometries, and when space-time reaches a certain level of instability, an objective "reduction" occurs[30]. This approach offers a mechanical solution to the measurement problem and makes testable predictions, but its mathematical details and experimental confirmation have not been fully completed[31]. Therefore, with the practical effectiveness

of the Copenhagen interpretation, R.Penrose's active ontological proposal enriches the scientific discussion and is one of the possible paths to quantum-gravity integration. The collapse of the wave function is not due to observation, but is the result of objective physical mechanisms. In macroscopic systems, this process occurs extremely quickly and superposition is virtually unobservable. Therefore, whether we look or not, reality exists only in a specific situation[30].

We see that each of these models has certain shortcomings. The reason for this is the difficulty of explaining two properties of matter in a quantum system. Initially, attempts were made to view particles as wave packets to resolve the particle-wave contradiction. The following considerations formed the basis of these attempts: An electron or any particle cannot be identified with a plane monochromatic wave. Because, while

such a wave is boundless in space and time, a particle is localized in space and time, meaning it occupies a specific location in space at a specific moment in time. However, it is known that by choosing the $\mathbf{k}$ -wave vectors of plane monochromatic waves at an appropriate speed, wave packets of any size can be constructed[14;15]. However, the attempt to view a particle as a wave packet is not justified. So, unlike a parti-

cle, a wave packet disappears over time. One of the important characteristics of elementary particles is their indivisibility. Waves, however, do not have this property of indivisibility. If one were to regard the electron as an aggregate composed of waves, then during the diffraction of a very weak beam of electrons passing through a crystal, each diffraction beam would carry only a fraction of the electron with it, which is clearly not the case[14].

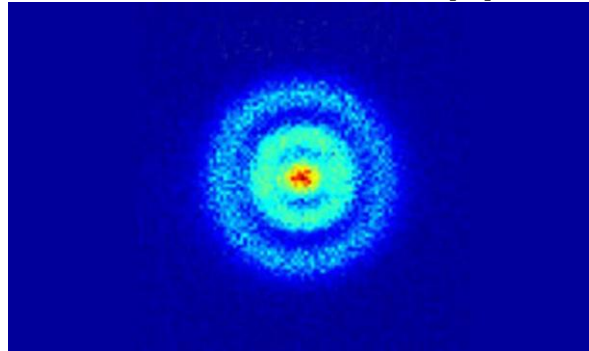


Figure 12. Electron cloud of hydrogen atom.

A micro-object (molecule, atom, electron, proton, etc.) does not directly affect our sense organs, that is, it cannot be felt by seeing or touching it. In the world we perceive, there is nothing like micro-objects[14;22]. As Nobel laureate Richard Feynman wrote in the third volume of his multi-volume "Feynman Lectures on Physics," "microscopic objects are unlike anything you've ever seen. Because the way atoms behave is so unlike our everyday experience, it is very difficult to get used to it. To both a person new to science and an experienced physicist, the behavior of atoms seems peculiar and nebulous. Even great scientists cannot understand it to the extent they would like, and this is completely natural. Because all of man's experience, all of his intuition, has been applied to large objects. We know what will happen to a large object, but small objects don't behave that way. "Therefore, when studying microparticles, it is necessary to use various kinds of abstractions, strain the imagination, and not try to connect them with our direct experience"[14;32]. Although microparticles combine the properties of both waves and particles, as R. Feynman said, "they behave neither like waves nor like particles." The difference between a microparticle and a wave is that it is always observed as an indivisible whole. For example, no one has ever observed half of an electron. At the same time, you can divide the wave into parts and then perceive each part in monthly[14;15;32]. The difference between a microparticle and the macroparticles familiar from everyday experience lies in the fact that the coordinates of a microparticle and the corresponding momenta cannot be known simultaneously with arbitrary precision; as a consequence, the very concept of a definite trajectory for a microparticle loses its meaning[14;15].

Recent studies. Conceptual analogies can be drawn between the nature of the measurement problem in quantum mechanics and the information processing process of complex algorithms. However, current artificial intelligence systems do not act as a physical 'observer', which leads to quantum collapse. Rather, it provides a framework for philosophical and theoretical

discussions about the nature of reality, the role of consciousness, and the relationship between the nature of the intellect[30].

When we talked about the pilot-wave theory (de Broglie-Bohm theory), we said that an invisible guide wave deterministically controls the trajectory of a particle. Although this theory is viable as an alternative interpretation of quantum mechanics, it is compelled to be nonlocal in order to remain consistent with the experiments confirmed by the 2022 Nobel Prize in Physics, namely quantum entanglement and the violation of Bell inequalities. That is, the wave's effect must propagate faster than the speed of light, which makes it problematic for some physicists[26;33].

Scientists at MIT have put an end to the debate between Bohr and Einstein. In their experiment, they showed that even slits are cracks at the quantum level. Ultimately, it turned out that Bohr was right and Einstein was wrong[34].

The moving double-slit experiment, which once again caused a disagreement between Einstein and Bohr at the 1927 Solvay Conference, was finally realized in 2025 by scientists at the University of Science and Technology of China. Perhaps the observation effect is not as significant as we thought. Scientists set up an experiment in which they could obtain the smallest momentum exchange between a photon and an atom by using a single Rubidium atom as a moving slit. Thanks to the trace left by the photon in the slit, it is possible to understand where the photon passed through without even looking at the photon. As a result of the experiment, the path traveled by the photon was measured and the same results as the double slit experiment were obtained. Even if the photon has never been measured, the wave behavior of the photon disappears as soon as the path information is obtained. The result proves Niels Bohr right again[34*].

Two scientists from the Stevens Institute of Technology in America have claimed to have solved the causes of wave-particle duality both experimentally and mathematically. The greater the phase difference,

the weaker the wave nature of the photons. When this happens, the system already chooses a path and the photon exhibits particle properties. In the double-slit experiment, a photon can pass through more than one slit as it travels along its path. When the difference between the first transition and the second transition exceeds a certain limit, the wave function collapses and the photon exhibits particle properties[35].

The universe's own internal physical processes (interactions and heat) are sufficient for it to transition from a quantum state to a classical state, without any "outside observer". This completely rejects the claim that "there were no observers in the early moments of the universe," and that it does not obey quantum laws. Rather, it explains how the classical universe could naturally arise from quantum roots through the mechanism of decoherence[36].

In 2017, former NASA physicist Thomas Cranpel and his team conducted several experiments with the classic double-slit experiment to test the simulation theory. The slit through which the entangled photons passed was measured, and this measurement reached the observer. In this case, the photon showed particle properties. The slit through which the photon passed was measured, and this measurement did not reach the observer. In this case, the photon showed wave properties. The result obtained was in the direction that the simulation theory was correct[37]. However, this result has been widely criticized. If there is a simulation, it is

impossible to falsify the simulation theory with any experiment.

A team of physicists led by Mir Faizal at the University of British Columbia, in a new and detailed analysis of existing theories, demonstrated that there is no theory of everything describing the Universe that flawlessly reconciles general relativity with quantum mechanics—at least not in an algorithmic sense. As a natural consequence of this research, it turned out that the Universe is not a simulation. Because such a simulation needs to be prepared algorithmically[38].

Holger F. Hofmann and his team from Hiroshima University in Japan directly demonstrated that the photon passed through both slits. This result shows that the wave function is not just a simple mathematical expression, but rather an entity that corresponds to real physical phenomena. The wave function no longer needs other worlds to explain the double-slit experiment. So, this experiment disproves parallel universes[39].

Italian physicists have performed a double-slit experiment with antimatter particles. They were looking for an answer to the question: Do antimatter particles exhibit wave properties when they have dimensions, and particle properties when they do not? The antimatter passed through the gold slits and reached the detectors behind. The result was a landscape created by waves. Antimatter also has the same wave-particle properties as normal matter. That is, the laws of quantum mechanics are fully satisfied[40].

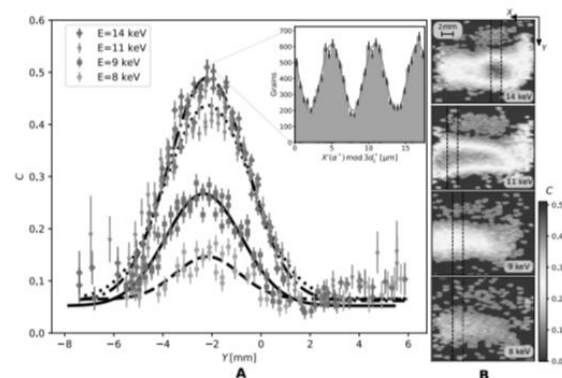


Figure 13. Antimatter particles passing through a kind of slit.

In an article published in PRL in 2025, Brazilian scientists gave a radical answer to this duality (wave-particle). Perhaps we misunderstood the wave-particle concept from the very beginning. Light can just be a particle. According to them, photons may actually still exist where we say there is no light in classical physics. However, these photons are in such a superposition that

the detectors cannot establish any connection with the observer. This situation is called a dark situation. That is, there is no wave property, but instead there is a special superposition of photons. The mere perception of light as a particle eliminates the problems in this experiment. However, this idea has not been confirmed in practice[41].

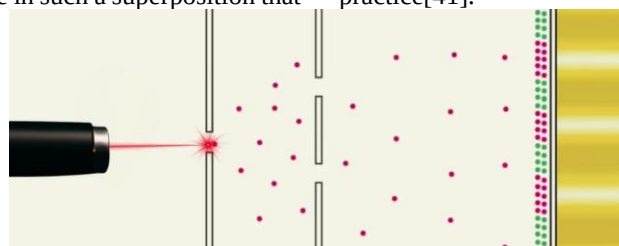


Figure 14. ●- bright photons, ●- dark photons.

Researchers at Imperial College London conducted a completely new version of the experiment.

The goal was to find an answer to the question, "What would happen if we opened rifts in time?" The light was

reflected differently at different times. This event took place on a femtosecond timescale. As a result of the experiment, an interference pattern emerged[42].

In another experiment, a theoretical answer to the double-slit mystery came from Brazilian scientists using entropy and the theory of general relativity. As a result, it was understood that quantum mechanics and the theory of general relativity are compatible with each other. The entropy of a particle is related to the observer's space-time fragment. It turns out that the observer is in an information exchange with the system. The observer is an inevitable member of the entropy of the quantum system. If the guess is high, the entropy is high. Observation decreases the entropy. If the observer reduces the entropy of the quantum system, that is, if the observer and the quantum system are in communication, then changing the curvature of space by the observer should change the amount of entropy of the system. The entropy level of a quantum system varies depending on the space-time state of the observer, and this test is consistent with[43]. In short, the fate of Schrödinger's cat depends on the state of the observer.

According to a 2014 paper, if there is no observation in the middle, the wave behavior of a quantum system prevails and the entropy is maximum. When observed, entropy decreases, the system behaves like a particle[44]. Researchers from Linköping University, working together with teams from Chile and Poland, carried out this experiment in a different form. What determines whether a light particle behaves like a particle or a wave is the level of information in the system. A laser was used in this experiment. In order to study the quantum properties of light in detail, this laser has been attenuated to the level of a single photon. The attenuated light is passed through a beam splitter and split into two different paths. One path of light sent on two separate paths was observed, and the light on this observed path exhibited particle behavior. At the same time, the other path of light was not observed, and it was determined that the wave nature of this path continued. These two light paths were then combined. They saw that during this combination, the system still exhibits particle properties. As a result, in a quantum system, the result is a certain amount of entropy of the system[45].

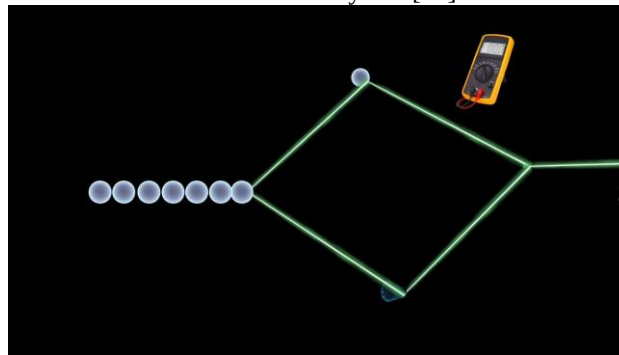


Figure 15. Scheme of the experiment

So far, we've been pointing to the nature of particles as the cause of this phenomenon. However, the secret to the strangeness in the double-slit experiment may also be the slits themselves. To solve the problem, it is necessary to consider not just the particle, but also the gap. As we expect, the gap behaves like a classic barrier. In the absence of measurement, slits temporarily obey the laws of quantum mechanics. The electron interacts with the slit in such a way that it changes the laws of experience. For a moment, a single, unified and inseparable quantum system exists. At this point, it no longer matters which slit the electron passes through. According to the article, there is no gap at that moment. Trying to understand the essence of wave-particle duality actually takes us away from studying quantum physics. Wave-particle properties emerge when we forcefully apply phenomena from our classical world to particles[46].

The final theory problem. As it turns out, particle-wave dualism is at the heart of quantum mechanics and appears to be the cause of a number of inexplicable phenomena. However, this dualism cannot fully explain the deeper structural causes of phenomena observed at the subatomic level[32;47]. Undoubtedly, the final theory must fully explain the causes of this phenomenon. After 50 years, the leading candidate for a "theory of everything" is a theory called "superstring theory." This is actually the only theory we have, since all other competing theories have been defeated.[48]. As physicist Steven Weinberg put it, "string theory has provided us with the first true candidate for a primary theory"[48;49]. It is in this respect that string theory brings a unified approach to one of the deepest problems of quantum physics. The basic ideology of string theory is that the initial fundamental unit of matter is not a point particle, but a very small vibrating string[50;51].

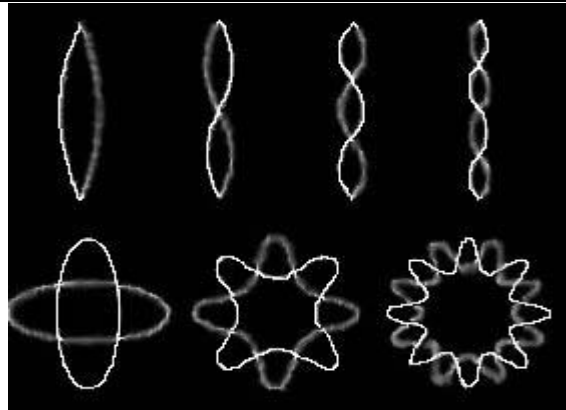


Figure 16. Strings vibrating in different shapes

The different vibrational modes of these strings determine the mass, charge, and interaction properties of elementary particles[50]. Researchers such as Edward Witten, Brian Greene, and Juan Maldacena point out that this idea could become a unifying model of physics[50;51;52]. As a result of this approach, the long-standing inconsistency between quantum mechanics and general relativity is eliminated through the concept of geometric vibrations and multidimensional space[50]. Edward Witten argues that string theory will one day dominate physics, just as quantum mechanics has dominated physics for the past half century[48]. The advantages of string theory are manifested not only in its theoretical coherence, but also in its cosmological applications. Strings provide a scientific bridge to the "God equation" by showing that both quantum fields and gravitational effects can be expressed by the same equations[53]. However, there are still daunting obstacles that need to be overcome. Those who don't believe in string theory point out that it has some weaknesses. First, it is impossible to directly test strings[48;54]. Since superstring theory is a theory of the Universe, the only way to test it experimentally would be to recreate the Big Bang, that is, to generate energies comparable to those of the beginning of the Universe within a particle accelerator. To do this, an atomic fission the size of a galaxy would be required. This is impossible even for technologically advanced beings[48]. In short, it seems impossible to define a string, as it is a thread of pure energy that forms the basis of everything, including the initial part of technological devices. However, recent studies based on DESI observations are at a level that can prove the correctness of string theory. Thus, the decrease in dark energy over time is considered the initial evidence for string theory[55].

Problem statement: The double-slit experiment, which constitutes one of the fundamental pillars of quantum mechanics, demonstrates that microscopic objects simultaneously exhibit both wave-like and particle-like properties, thereby marking a radical departure from classical physical conceptions[15;32]. Despite more than a century of experimental and theoretical investigations, the essence of this phenomenon has not yet been fully explained, and in particular, no universal and conclusive theoretical framework has been established that rigorously accounts for the mechanism underlying wave-particle duality [56]. Richard Feynman

famously described this experiment as "the phenomenon at the heart of all the mysteries of quantum mechanics," emphasizing that although its full implications may not be completely understood, its experimental validity is beyond any doubt[32]. Although various theoretical models and experimental approaches have attempted to explain this problem, fundamental questions remain regarding the precise physical nature of the measurement act and the objective reality of the superposition state, which together underlie the interference pattern observed in the double-slit experiment[57]. The central problem investigated in this article is to resolve the discrepancy between existing theoretical models that experimentally describe wave-particle duality and the deeper nature of microscopic reality. In particular, the following decisive questions form the foundation of this investigation:

1. By what fundamental mechanism do electrons or other quantum objects create an interference pattern?
2. What is the physical meaning of the phenomenon of "self-interference" when a particle moves towards a single slit?
3. How does the dynamics of measurement occur and does superposition collapse at the moment and by what mechanism?
4. Is the ability of a wave function to produce interference related to more fundamental structures of space and time (such as vacuum fluctuations, quantum fields, or possible hidden variables)?
5. How can the interference pattern observed in the double-slit experiment be related to the multidimensional vibrational modes of strings?
6. How exactly can particle-wave dualism be explained with string theory?

Each of these questions touches on the foundations of quantum mechanics and is crucial to fully understanding the fundamental mechanism behind the double-slit experiment. Since none of the existing theoretical frameworks provide complete and conclusive answers to these questions, the problem of wave-particle duality remains one of the most open and strategic issues in physics.

Problem solving: Using the above empirical facts, it is possible to propose a new theoretical model that answers the problematic questions that arise. In general, existence in the Universe has two forms of manifestation: matter and the physical realm, and the current structure of the Universe was formed by the Big

Bang. The universe expanded from a single initial point, and the energy within it took on its known state, taking the form of both matter and physical fields[58]. From this perspective, we can say that matter, which is the form of matter, and the physical field have the same source. This source was created when the Universe first began to form. It is clear that the simplest structure that could have formed in the early moments of the Universe must have been the first particle. From these two approaches, we conclude that the smallest structural particle of the Universe is the source of matter and the physical field. Therefore, the actions of the first fundamental particle created the diversity in the Universe. In short, the study of the first fundamental particle can eliminate all problems, including those in wave-particle duality. The best and most proven theory that defines this elementary particle is string theory. We said that string theory states that the initial indivisible part of the Universe was vibrating strings. Different vibration patterns create different fields and particles[58]. If we summarize what we have said, we come to the conclusion that the properties of matter and field (as well as wave-particle) coincide in the string, which is the primary particle of the Universe. The vibrations of the strings in different patterns create different fields, and the sum of these vibrations at a local location creates the corresponding particle. This logical conclusion serves the development of the best-working theory in physics - "quantum field theory". Different quantum fields are, as it were, "woven" from strings that vibrate differently[59]. At the same time, these quantum fields are also connected through strings, forming the entire network of the Universe. These facts provide a confident step towards the final theory. This logically derived result can be proven mathematically:

First, let's list the main quantities:

• $\omega(\lambda, l_s)$ is the spectral (energy) density function of the string, λ is the wavelength or mode index, and l_s is the length of a string.

• Observation parameter: $\varepsilon > 0$ Small particle types are described as $(\lambda \in [0, \varepsilon])$, large field types as $(\lambda \in [\varepsilon, \infty])$.

First of all, we determine:

$$\rho_p(\varepsilon, l_s) = \int_0^\varepsilon \omega(\lambda, l_s) d\lambda, \quad (1)$$

$$\rho_f(\varepsilon, l_s) = \int_\varepsilon^\infty \omega(\lambda, l_s) d\lambda \quad (2)$$

here,

ρ_p and ρ_f are the particle and field fractions, respectively. What we want is the equality of the fractions under certain conditions: $\rho_p/\rho_f=1$.

The goal is to take an expressive, yet realistic, predetermined specific form. A typical useful theory:

$$\omega(\lambda, l_s) = \ell_s^{-1} \Phi \frac{\lambda}{\ell_s} \quad (3)$$

Here, the $\Phi(u)$ -function is a function expressing the spectral distribution of the energy density at the

string level and is defined in the interval $0 < u < \infty$. This function can represent both the particle mode and the field mode in a completely symmetrical manner.

Note: This structure assumption means "scale invariance" in the physical sense and that the string spectrum changes only in terms of λ/l_s .

Let's simplify the integrals with respect to the change ($U = \lambda/l_s$);

• particle fraction:

$$\rho_p(\varepsilon, l_s) = \int_0^\varepsilon \ell_s^{-1} \Phi \frac{\lambda}{\ell_s} d\lambda = \int_0^{\varepsilon/\ell_s} \Phi(U) dU \quad (4)$$

• field share:

$$\rho_f(\varepsilon, l_s) = \int_\varepsilon^\infty \ell_s^{-1} \Phi \left(\frac{\lambda}{\ell_s} \right) d\lambda = \int_{\varepsilon/\ell_s}^\infty \Phi(U) dU \quad (5)$$

will be. So the ratio

$$R(\varepsilon, l_s) = \frac{\rho_f}{\rho_p} = \frac{\int_{\varepsilon/\ell_s}^\infty \Phi(U) dU}{\int_0^{\varepsilon/\ell_s} \Phi(U) dU} \quad (6)$$

and as a result R will depend only on the argument

ε/l_s .

Our main idea is that when we take both the

observation parameter $\varepsilon \rightarrow 0$ and the fundamental string

length $l_s \rightarrow 0$, these two shares can coincide. But the

result depends on the ratio between ε and l_s . So we take

a general relationship.

$l_s \sim \varepsilon^\beta$ or equivalently $l_s/\varepsilon^\beta \rightarrow C \in (0, \infty)$. C is a constant positive number.

Let's solve for ε/l_s

$$\frac{\varepsilon}{l_s} = \frac{\varepsilon}{\varepsilon^\beta \cdot C} = \frac{1}{C} \varepsilon^{1-\beta} \quad . \text{ Now three main cases}$$

emerge: • If $\beta < 1$, then $\varepsilon/l_s \rightarrow 0$

• If $\beta > 1$, then $\varepsilon/l_s \rightarrow \infty$

• If $\beta = 1$, then $\varepsilon/l_s \rightarrow 1/C$

But to be more general and flexible, we want to write the condition $\varepsilon/l_s \rightarrow a \in (0, \infty)$ directly instead. This condition can be expressed in the same way as $l_s/\varepsilon^\beta \rightarrow C$ (if an appropriate β is chosen).

The condition for overcoming inequality.

$$\text{If } \lim_{\substack{\varepsilon \rightarrow 0 \\ l_s \rightarrow 0 \\ \varepsilon/l_s \rightarrow a}} \frac{\varepsilon}{l_s} = a, \quad a \in (0, \infty) \quad (7)$$

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ l_s \rightarrow 0}} R(\varepsilon, l_s) = \frac{\int_a^\infty \Phi(U) dU}{\int_0^a \Phi(U) dU} \quad (8)$$

then

It will be. For the consistency of the approach put forward

$$\int_0^a \Phi(U) dU = \int_a^\infty \Phi(U) dU \quad (9)$$

It is important to have. That is, the cumulative distribution of the function a must have a $1/2$ cutoff point. Thus

$$\int_0^a \Phi(U) dU = \frac{1}{2} \int_0^\infty \Phi(U) dU \quad (10)$$

When such a function a exists, the ratio is $\rightarrow 1$. This means that the relationship between l_s and ε should be chosen such that ε/l_s corresponds to the median of the spectrum.

It can be seen from the above that if the cumulative parliamentary bisection condition of Φ is satisfied when a -transformed to C , then the following final result will be obtained.

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ l_s \rightarrow 0 \\ l_s/\varepsilon^\beta \rightarrow c}} \frac{\int_0^\infty \omega(\lambda; l_s) d\lambda}{\int_0^\varepsilon \omega(\lambda; l_s) d\lambda} = 1 \quad (12)$$

This result shows that on an ultrasmall scale, field and matter energy are indistinguishable - a "superposition" is obtained. At the string level, as a result of the median symmetry of the energy distribution function $\Phi(u)$, the differences between wave and particle energy disappear, and both descriptions become equivalent forms of the same physical reality.

As can be seen, since the concepts of field and matter are identical at the string level, the wave and particle natures also originate from this source, and the scale of observation causes these properties to be separated. This approach explains what caused the wave-particle duality in the first place. In this regard, the reason why dualism mainly manifests itself in microparticles is that they contain fewer strings (fewer interactions) than macroobjects. However, since matter of any size is made up of strings, it will have wave properties regardless of its size. From this perspective, as the mass increases, the wavelength, as is known, should decrease. For example, the wavelength of a basketball moving at a normal speed is on the order of $\sim 10^{-34}$ meters (based on the expression $\lambda = h/p$). In this respect, this feature may not be noticeable in macroscopic objects. In fact, particles are neither waves nor particles. These properties are properties of our microcosm. We require that particles have these properties. However, the reality is not like that at all: micro-objects are unlike anything we have ever seen. This unique property is a form that combines both wave-like and particle-like properties at the same time. However, it cannot be exactly the same property. Because, the complete wave and complete particle properties cannot be the same, and we have explained why. This unity was achieved in the form of an image by a group of scientists from the Swiss Polytechnic School using highly advanced nanotechnology, electron microscopy, and laser shooting systems[60].

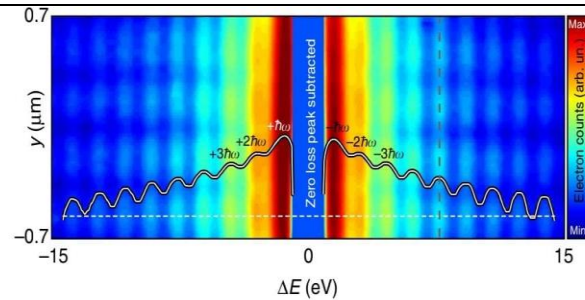


Figure 17. A snapshot of wave-particle duality at the same time.

Red is where there are the most electrons, and blue is where there are the fewest. The blue line in the middle is where there is no electron exchange. The energy exchange is characterized by integers. High electron counts are at the crests of the wave, and low electron counts are at the troughs of the wave[60].

Therefore, when we use the mathematical expressions used to explain phenomena in the macrocosm to describe properties we have never seen, a number of inconsistencies emerge. The reason this duality is more pronounced in particles is that the number of strings inside them differs sharply from those in macroscopic objects. This sharpness is distinguished by the area of the field that surrounds the objects. The passage of particles with properties we are not accustomed to through a double slit will, of course, result in an event we are not accustomed to. This may seem a little strange to some people. However, when we look deeper, we see that there are events around us that we have difficulty understanding. For example, when one considers facts such as the absence of color as an intrinsic property of reality, the fact that nothing ever truly comes into direct contact, that everything is interconnected at the smallest scales, and that true empty space does not exist, one must also accept this reality. As it turns out, our new approach to the double-slit experiment does not contradict the Big Bang in any way, but rather helps us explain it. According to the results obtained, parallel Universes are not needed to fully explain the double-slit experiment. However, this result does not mean that parallel Universes do not exist. The nature of dualism does not depend on its being a rift in time or space, on the rift being in motion, or on whether it is a normal or antiparticle. This property is determined solely by the number of strings within the object. The discovery of the double-slit experiment in time rifts and the successful explanation of its essence with string theory mean that it will be possible for us to explain time with string theory as well. This fact reinforces the idea that the Big Bang, the primordial particle, and the source of space-time are one. The double slit effect is not just caused by particles. You also need to look at the slits. When a particle passes through a slit, it behaves with the slit as a single system. Based on the non-classical properties of particles and the particle-slit interaction (according to string theory, all physical entities are interconnected [58]), a single unified system emerges, within which the question of which slit the particle passes through loses its physical significance. As the wave-like nature of the particle decreases, the phase difference will increase. What determines whether a particle behaves like a particle or a wave during observation is the level

of entropy of the system. The observer is an inevitable member of the entropy of a quantum system. According to string theory, it is also suggested that there is a harmony between quantum mechanics and the theory of general relativity. As N.Bohr rightly said, the experimental setup changes the result. Measurement and interaction are irreversibly linked. Any change in the space-time fabric caused by an observer changes the entropy. If there are many guesses, there is high entropy. Thus, observation reduces entropy, and this information exchange is consistent with experimental results. With simple logic, we can say that during observation, the particle is no longer normal, but becomes an observed particle. When a particle is sent to a single or double slit, it exhibits only one property, and that is its own property. Therefore, the wave function is not a probability distribution of knowledge known to the observer, but rather an entity that corresponds to real physical phenomena. The role of the observer here is not to "create reality", but to choose reality. If this phenomenon is to be called a collapse, it should be known that this phenomenon is the result of objective physical mechanisms. We see that the Copenhagen interpretation is flawed in many ways. Thus, when explaining dualism, we say that there is no need for parallel Universes and, as a result, additional dimensions. That is, the real reasons for dualism must be consistent with the properties of 3-dimensional space. In 3-dimensional space, everything that exists must have a certain shape and size. From this perspective, everything must be located in a specific place in space-time. Therefore, reality exists only in a specific state, whether we look at it or not. That is, the existence of physical form and structure in the underlying model is inevitable. The Copenhagen interpretation works purely mathematically. This approach provides a physical structure and an ontological model. The probability of the particle being in other places is actually the places where the particle is affected. As it turns out, simply varying vibrations of strings in 3-dimensional space allow us to explain the double-slit experiment. The "extended" form of quantum field theory proves that all processes occurring are in a coherent state. This consistency proves that the Universe is not a simulation, but rather a vast and orderly creation. In fact, these phenomena, which seem strange to us, follow a certain order.

Result: A number of important conclusions are drawn from the new approach, which is based on modern research and uses string theory. The Universe we observe was formed by the Big Bang and expanded over time to take its current form. With the Big Bang, the Universe expanded from a single initial point and

the forms of matter, matter and the physical realm, emerged in a way that was clearly distinguishable. The types of matter have the same origin at the smallest size, that is, at the smallest primary particle. The most successful theory that provides us with the smallest primordial particle of the observable Universe is string theory. From this perspective, on the fundamental length scale of the Universe—the level of the most primitive string—the concepts of matter and physical field are indistinguishable. The wave and particle representations no longer appear here as separate, but as equivalent representations of a single physical object. At the deepest level of reality, matter and field merge; wave and particle are just two perspectives of the same entity. This logical conclusion was proven mathematically.

$$\lim_{\substack{\varepsilon \rightarrow 0 \\ \ell_s \rightarrow 0 \\ \ell_s / \varepsilon \rightarrow c}} \frac{\int_0^\varepsilon \omega(\lambda; \ell_s) d\lambda}{\int_0^\varepsilon \omega(\lambda; \ell_s) d\lambda} = 1$$

As a result, it was determined where wave-particle duality originally originated. Particles made up of strings are neither waves nor particles. Because, in reality, the properties of a complete wave and a complete particle cannot be compatible. Wave-particle properties are properties that we take from the macrocosm and want particles to obey. Particles are a form of matter that combines both wave-like and particle-like properties at the same time. Therefore, the particles are unlike anything we have ever seen. When we apply the mathematical expressions for these properties to particles that are unlike anything we observe, we get inconsistencies, and this is completely natural. The reason why the property of duality manifests itself more clearly in particles is that there are fewer strings and fewer interactions within them. Every object, regardless of its size, has wave properties. However, as the interaction (number of strings) increases, this property weakens to the point of being negligible. The passage of particles with such a unique property through a double slit naturally gives rise to strange phenomena. The double-slit experiment does not contradict the Big Bang in any way. According to the results obtained, parallel Universes are not needed to fully explain the double-slit experiment. However, this result cannot be considered strong evidence for the absence of parallel Universes. The nature of dualism does not depend on its rift in time or space, or whether it is a normal or antiparticle. For this, cracks also play an important role. When a particle passes through a slit, the results from string theory essentially treat the slit as a complete system, and the question of which slit the particle passes through becomes irrelevant. The decrease in the wave-like nature of the particle will increase in proportion to the phase difference. What determines whether a particle behaves like a particle or a wave during observation is the entropy level of the system. The observer is an inevitable part of the entropy of a quantum system. Bohr was right in that the experimental setup changes the outcome. The observer himself creates changes in the space-time fabric, and this fact changes entropy. String theory describes this

phenomenon very precisely. When a particle is sent to a single or double slit, the particle exhibits only one property, which is its own intrinsic property. The wave function is not a probability distribution of knowledge known to the observer, but rather an entity that corresponds to real physical phenomena. If this phenomenon is to be called a collapse, it should be known that this phenomenon is the result of objective physical mechanisms. The Copenhagen interpretation is flawed and even wrong in many ways. Thus, when explaining dualism, we say that there is no need for parallel Universes and, as a result, additional dimensions. The real causes of dualism correspond to the properties of 3-dimensional space and should be explainable by form and physical description. Therefore, reality exists only in a specific situation, whether we look at it or not. In a model that explains the first cause, it is essential that physical form and structure exist. The Copenhagen interpretation simply works mathematically. This approach presents a physical structure and an ontological model. The probability of the particle being in other places is actually the places where the particle is affected. The different vibrations of strings in 3-dimensional space allow us to explain the double slit experiment. The "extended" form of quantum field theory proves that all processes occurring are in a coherent state. The Universe we live in is not a simulation. Because, when analyzed with the right models, it becomes clear that there is a specific reason behind all strange phenomena. Returning to the Einstein-Bohr debate, we realize that both sides are right: N.Bohr is right about the incident, but this incident does not fully justify the position of N.Bohr and his associates. The fact that there is a reason behind all phenomena shows that, as A.Einstein said, "God does not roll dice." When we examine all the facts, it becomes clear that the basic model does not completely disprove the ideas of all scientists, it simply shows that they have shortcomings, but that they are very close.

References:

1. Zeilinger, A. (1999). Experiment and the foundations of quantum physics. *Reviews of Modern Physics*, 71(2), S288. <https://doi.org/10.1103/RevModPhys.71.S288>
2. Wheeler, J. A., & Zurek, W. H. (Eds.). (1983). *Quantum theory and measurement*. Princeton University Press.
3. Kragh, H. (2002). The dualism of wave and particle: A historical perspective. In *Quantum generations: A history of physics in the twentieth century* (pp. 58-84). Princeton University Press.
4. Newton, I. (1704). *Opticks: or, a treatise of the reflexions, refractations, inflexions and colours of light*. London: Samuel Smith and Benjamin Walford.
5. Huygens, C. (1690). *Traité de la lumière: où sont expliquées les causes de ce qui luy arrive dans la réflexion & dans la réfraction, et particulièrement dans l'étrange réfraction du cristal d'Islande*. Leiden: Pierre van der Aa.
6. Hall, A. R. (1993). *Isaac Newton: Adventurer in thought*. Cambridge University Press.

- 7.Young,T. (1804). The Bakerian Lecture: Experiments and calculations relative to physical optics. *Philosophical Transactions of the Royal Society of London*, 94, 1–16. <https://doi.org/10.1098/rstl.1804.0001>
- 8.Hecht,E. (2017). *Optics* (5th ed.). Pearson Education.
- 9.Maxwell,J. C. (1865). A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, 155, 459–512. <https://doi.org/10.1098/rstl.1865.0008>
- 10.Halliday,D., Resnick, R., & Walker, J. (2013). *Fundamentals of physics* (10th ed.). John Wiley &
- 11.Serway,R. A., & Jewett, J. W. (2018). *Physics for scientists and engineers with modern physics* (10th ed.). Cengage Learning.
- 12.Einstein,A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6), 132–148. <https://doi.org/10.1002/andp.19053220607>
- 13.de Broglie,L. (1924). *Recherches sur la théorie des quanta* [Researches on the quantum theory]. (Doctoral dissertation, Université de Paris). Masson et Cie.
14. Masimov,E.A., Mursalov,T.M. (2002). Atom fizikasi {Atomic physics}. <https://nezerifizika.wordpress.com/wp-content/uploads/2012/10/atom-fizikasi-mc999simov-e-c999-mc3bcrc999lov-t-c999-2002.pdf>
- 15.Griffiths,D. J., & Schroeter, D. F. (2018). *Introduction to quantum mechanics* (3rd ed.). Cambridge University Press.
- 16.Davisson,C., & Germer, L. H. (1927). Diffraction of electrons by a crystal of nickel. *Physical Review*, 30(6), 705–740. <https://doi.org/10.1103/PhysRev.30.705>
- 17.Kumar,M. (2008). Quantum: Einstein, Bohr, and the great debate about the nature of reality. W. W. Norton & Company.
- 18.Bohr,N. (1949). Discussion with Einstein on epistemological problems in atomic physics. In P. A. Schilpp (Ed.), *Albert Einstein: Philosopher-scientist* (pp. 199–241). The Library of Living Philosophers.
- 19.Schlosshauer,M. (2005). Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 76(4), 1267–1305. <https://doi.org/10.1103/RevModPhys.76.1267>
- 20.Feynman,R. P., Leighton, R. B., & Sands, M. (1965). *The Feynman lectures on physics*, Vol. 3: Quantum mechanics. Addison-Wesley.
- 21.von Neumann,J. (1955). *Mathematical foundations of quantum mechanics* (R. T. Beyer, Trans.). Princeton University Press.
- 22.Bohr,N. (1934). *Atomic theory and the description of nature*. Cambridge University Press.
- 23.Schrödinger,E. (1935). Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23(48), 807–812. <https://doi.org/10.1007/BF01491891>
- 24.Maudlin, T. (2019). The measurement problem. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019 ed.). Stanford University.
- 25.Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75(3), 715–775.
- 26.Goldstein, S. (2021). Bohmian mechanics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2021 ed.). Stanford University.
- 27.Schlosshauer, M. (2007). *Decoherence and the quantum-to-classical transition*. Springer-Verlag.
- 28.Faye, J. (2019). Copenhagen interpretation of quantum mechanics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). Stanford University.
- 29.Bell, J. S. (1990). Against ‘measurement’. *Physics World*, 3(8), 33–40.
- 30.Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.
- 31.Bassi, A., Lochan, K., Satin, S., Singh, T. P., & Ulbricht, H. (2013). Models of wave-function collapse, underlying theories, and experimental tests. *Reviews of Modern Physics*, 85(2), 471–527.
- 32.Feynman, R. P., Leighton, R. B., & Sands, M. (1965). *The Feynman lectures on physics*, Vol. III: Quantum mechanics. Addison-Wesley. (Chapter 1: Quantum Behavior).
- 33.NobelPrize.org. (2022). Scientific Background on the Nobel Prize in Physics 2022: Entangled photons – from theory to technology. The Royal Swedish Academy of Sciences.
- 34.Fedoseev,V., Lin,H., Kun Lu,Y., Kyung Lee,Y., Lyu,J., Ketterle,W. (2025). Coherent and Incoherent Light Scattering by Single-Atom Wave Packets. 2. <https://doi.org/10.1103/zwhd-1k2t>
- 34*.Zhang, Yu-Chen., Cheng, Hao-Wen., Zengxu, Zhao-Qui., Duan, Yu-Cheng., Chen, Ming-Cheng et.al. (2025). Tunable Einstein-Bohr Recdling-Slit Gedankenexperiment at the Quantum Limit. PRL-135. <https://doi.org/10.1103/93zb-lws3>
- 35.Khatriwada,P., Qian,X.F.(2025). Wave-Particle Duality Ellipse and Application in Quantum Imaging with Undetected Photons. PRR 7. <https://arxiv.org/pdf/2206.04362>
36. Kiefer,C., Polarski,D., Starobinsky,A.A. (1998). Quantum-To-Classical Transition For Fluctuations in the Early Universe. <https://arxiv.org/pdf/gr-qc/9802003>
- 37.Campbell,T., Owahdi,H., Sauvageauz,J., Watkinson,D.(2017). On Testing the Simulation Theory. *International Journal of Quantum Foundations*. <https://arxiv.org/pdf/1703.00058>
- 38.<https://www.sciencealert.com/physicists-just-ruled-out-the-universe-being-a-simulation>; <https://news.ok.ubc.ca/2025/10/30/ubco-study-debunks-the-idea-that-the-universe-is-a-computer-simulation/>
- 39.Fukuda,R., Linuma,M., Matsumoto,Y., Hofmann,H.F. (2025). Experimental Evidence for the Physical Delocalization of Individual Photons in a Interferometer. *Quantum Physics*. <https://doi.org/10.48550/arXiv.2505.00336>
- 40.Giammarchi,M. (2019). First Demonstration of Antimatter Quantum Interferometry. <https://arxiv.org/pdf/1906.08665>
41. Villas-Boas,C.J., Máximo,C.E., Paulino,P.J., Bachelard,R.P., Rempe,G. (2025). Bright and Dark

States of Light: The Quantum Origin of Classical Interference. PRL 134.

<https://doi.org/10.1103/PhysRevLett.134.133603>

42. Tirde, R., Vezzoli S., Galiffi, E., Robertson, I., Maurice, D., Tilman, B., Maier, S.A., Pendry, J.B., Sapienza, R. (2023). Double-Slit Time Diffraction at Optical Frequencies. Nature Physics. <https://arxiv.org/pdf/2206.04362>

43. Basso, M.L.W., Maziero, J., Celeri, L.C. (2025). Quantum Detailed Fluctuation Theorem in Curved Spacetimes: The Observer Dependent Nature of Entropy Production. New Scientist. <https://arxiv.org/pdf/2405.03902>

44. Elouard, C., & Mohammady, M. H. (2014). Work, heat and entropy production along quantum trajectories. arXiv preprint arXiv:1408.4382.

45. Spegel-Lexne, D., Gómez, S., Argillander, J., Pawłowski, M., Dieguez, P.R., Alarcón, A., Xavier, G.B. (2024). Experimental demonstration of the equivalence of entropic uncertainty with wave-particle duality. <https://doi.org/10.48550/arXiv.2407.03797>

46. Plotnitsky, A. (2025). A Missing Link: The Double-Slit Experiment and Quantum Entanglement. Entropy, 27(8), 781. <https://doi.org/10.3390/e27080781>

47. Bunge, M. (2003). Emergence and convergence: Qualitative novelty and the unity of knowledge. University of Toronto Press.

48. Kaku, M. (2004). Einstein's cosmos: How Albert Einstein's vision transformed our understanding of space and time. W. W. Norton & Company.

49. Weinberg, S. (1993). Dreams of a final theory: The search for the fundamental laws of nature. Vintage Books.

50. Green, M. B., Schwarz, J. H., & Witten, E. (1987). Superstring theory: Vol. 1. Introduction. Cambridge University Press.

51. Greene, B. (1999). The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory. W. W. Norton & Company.

52. Maldacena, J. (1999). The large N limit of superconformal field theories and supergravity. International Journal of Theoretical Physics, 38(4), 1113–1133. <https://doi.org/10.1023/A:1026654312961>

53. Kaku, M. (2021). The God Equation: The Quest for a Theory of Everything. Doubleday.

54. Woit, P. (2006). Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law. Basic Books.

55. Hur, S., Jejjala, V., Kavac, M.J., Minic, D., Takeuchi, T. (2025). Dynamical Dark Energy, Dual Spacetime, and DESI. <https://arxiv.org/pdf/2503.20854>

56. Becker, A. (2018). What Is Real?: The Unfinished Quest for the Meaning of Quantum Physics. Basic Books.

57. Schlosshauer, M. (2007). Decoherence and the Quantum-to-Classical Transition. Springer-Verlag.

58. Abbasov V. (2025). CREATION AND M-THEORY: TOWARDS THE GOD EQUATION. Norwegian Journal of Development of the International Science, 160. <https://doi.org/10.5281/zenodo.15879908>

59. Abbasov V. (2025). QUANTUM FIELD THEORY AND STRING THEORY. Norwegian Journal of Development of the International Science, 158. <https://doi.org/10.5281/zenodo.15657624>

60. Piazza, L., Lummen, T.T.A., Quinonez, E., Murooka, Y., Reed, B.W., Barwick, B., Carbone, F. (2015). Simultaneous Observation of the Quantization and the Interference Pattern of a Plasmonic Near-Field. Nature Communication. DOI: 10.1038/ncomms7407

SOCIAL SCIENCES

SERVICE-LED TRANSFORMATION OR PREMATURE DEINDUSTRIALIZATION? SECTORAL EMPLOYMENT SHIFTS, STRUCTURAL CHANGE, AND NATIONAL INCOME IN NEPAL, 2000-2022

Shrestha Sofin

Ph.D. Candidate, Department of Global Cultural Industry Management, Calvin University, Yongin-si, Korea Rep

Sherpa Pemba

Ph.D. Candidate, Department of Global Cultural Industry Management, Calvin University, Yongin-si, Korea Rep

Saru Tilak

Ph.D. Candidate, Department of Global Cultural Industry Management, Calvin University, Yongin-si, Korea Rep

Koirala Ashika

Ph.D. Candidate, Department of Global Cultural Industry Management, Calvin University, Yongin-si, Korea Rep

<https://doi.org/10.5281/zenodo.21105836>

Abstract

The structural transformation in Nepal during the period 2000 to 2022 has been quite unprecedented when compared with the classical Lewis Kuznets Chenery development sequence: There was a significant decrease of agricultural employment whereas the industrial sector experienced stagnation and most the labor force was recruited in a diverse but largely unproductive services sector. This paper aims to shed light on whether Nepal's transition, which was mainly through services, is an indication of productive growth or should be seen as an example of early deindustrialization. Annual records from the World Bank World Development Indicators (2000-2022, $n = 23$) were the source of data used to analyze the connection between labor shares in different sectors and net national income per capita adjusted for inflation. Ordinary Least Squares (OLS) regression, instrumental variables approach with lagged agricultural employment as the instrument, and Fully Modified OLS (FMOLS) to handle non-stationary residuals were applied to the data. The results show that the proportion of agricultural employment is Much, negatively, and consistently tied to the dependent variable ($\beta \approx -44$ to -46 , $p < 0.001$) and a first-difference model also indicates a significant relationship between employment reallocation and short-term economic growth ($\beta = -3.81$, $p < 0.05$). Most importantly, the predominant part of the R^2 accounted for by one common time trend poses a real risk of spurious regression, and we point this limitation out quite clearly. Our results align with the research on productivity gains through structural changes, but they also emphasize the peril of development that is based on service sector growth while neglecting the factory production. The main takeaways for the development strategy are an overhaul in industry policy, raising the service industries selectively, and increasing agricultural efficiency.

Keywords: structural transformation; premature deindustrialization; service-led growth; labor productivity; Nepal; employment reallocation; time-series analysis

1. Introduction

One of the most reliable results in the field of development economics remains to be the finding that the transfer of labor from one industry to another may be a key thing in raising the overall productivity (Herrendorf et al. 2014; McMillan et al. 2014). Still, the exact process and welfare effects of this depend very much on where in the economy the workers give up their old jobs and take up new ones. The traditional sequence of sectors, which was given a status of a pattern by Lewis (1954), Kuznets (1966), and Chenery and Syrquin (1975), starts from low-productivity agriculture, then moves labor-intensive manufacturing, and finally reaches sophisticated services, with each sector transition leading to higher productivity gains. Using data, Rodrik (2016) carefully showed that the pattern is breaking up in the present-day developing economies: countries are rapidly reducing agricultural labor not towards the manufacturing sector, which is tradable, but the informal or less productive services, this way, he

called the occurrence premature deindustrialization. Nepal, through a lens, presents a scenario for this argument. Over the period of 2000 to 2022, the percentage of people working in farming has come down by approximately 11 percentage points from 35.3% to 23.8% of total employment, though manufacturing and construction sectors only managed to get a little bit larger (26.0% to 30.6%), and the service sector has gone up from 38.6% to 45.1% (World Bank, 2024). This kind of path has been followed without the kind of industrialization that led to high incomes in East and Southeast Asia. Nepal's development has instead been influenced mainly by remittances coming in, tourism earnings, and electricity from hydro power external or natural resource-based factors which have very little linkages with domestic value chain (Asian Development Bank [ADB], 2020; UNESCAP, 2021). Figuring out whether the association we see between a shrinking agricultural sector and a growing economy (income) is a result of structural changes that are productive or it is just a time-

trend relationship that is not real, has to do with the kind of policies that are put in place for the industries and labor-market.

This paper presents three main findings. Firstly, it uses the Premature Deindustrialization thesis to highlight the change in Nepal that the Nepal specific empirical literature hardly referred to. Secondly, it goes for a multi specification regression strategy OLS, instrumental variable estimation using lagged employment, first differenced models, and Fully Modified OLS to deal with the problems of endogeneity and non-stationarity which are characteristics of small-sample time-series data. Thirdly, it raises the analytical weaknesses of single-country time-series analysis, maintaining that finding statistical significance in levels alone is not enough to prove the existence of a structural productivity-growth relation. The rest of this paper is divided into: Section 2 discussing theoretical and literature review, identifying the research gap; Section 3 elaborating on data and empirical methods; Section 4 explaining and analyzing results; and Section 5 giving concluding with the policy and research note.

2. Literature and Research Gap

Structural changes have been the main engine of the classical development story which has resulted in higher productivity. Surplus labor transfer from subsistence agriculture to capitalist manufacturing leading to economic surplus reinvested in further capital accumulation was established by Lewis (1954). His model named the transfer of surplus labor from the agricultural sector to industrial sector as a source of economic surplus that is reinvested for further economic growth. Industrial demand for labor as assumed by the Lewis model is inelastic such that labor supply had to come from increase of wage so that outflows from agriculture get absorbed in industry a condition that industrializing East Asia met but South Asia and Sub-Saharan Africa have been largely different on this point (Gollin et al. 2014; Duarte & Restuccia, 2010). Sectoral reallocation regularity has been cross-sectionally confirmed by Kuznets (1966) while Chenery and Syrquin (1975) identified the industrial growth pattern as the driving force behind the development process. Recent structural-change literature has mostly challenged this theoretical agreement. McMillan et al. (2014) accounted labor productivity growth changes for 38 countries at the level of intra-sector and structural-change components and revealed that structural change brought about positive contribution in Asia but was growth-reducing in Africa and Latin America where workers moved into low-productivity informal services rather than manufacturing. By going as far as 46 countries Timmer et al. (2015) have been able to corroborate the connection between structural change and productivity-improvement in the post-2000 period although they have also shown that the countries differ quite a bit here. The analysis of the African case by De Vries et al. (2015) is very illuminating: First, workers were employed in distributive trade services whose productivity was higher than that of agriculture. However productivity growth was so slow that the structural change dividend was static rather than dynamic. Diao et al. (2017) posited that in sub-Saharan Africa the labor transition from agriculture

to services is a sort of equilibrating response to the lack of competitive manufacturing sector rather than a visible development strategy.

Rodrik (2016) in his paper on premature deindustrialization, set the main analytical structure for the Nepal study. Using a cross-country dataset of 42 nations, he found that the share of manufacturing employment reaches its highest point at lower levels of income per capita and earlier in the development stage compared to the history of East Asian countries. He explained that the reason is of globalization induced trade competition, skill biased technological change in manufacturing, and the comparative advantage disadvantage of late industrializing countries. In fact, Haraguchi et al. (2017) and Felipe et al. (2018) in their later studies challenged the general applicability of Rodrik's gloominess and argued that manufacturing can still be a growth platform for countries that succeed in integrating themselves into global value chains, but they agreed that late industrializers will have a very limited window of opportunity. The emergence of service sector led growth as an alternative or complement to the traditional manufacturing-based development has been deeply analyzed again. Ghani and O'Connell (2014) argued that high productivity tradable services could be a source of growth even for low-income countries, but at the same time, they warned that the productivity effects are very much concentrated in business process and information technology services whereas retail, hospitality and personal services are quite absent. Nayyar et al. (2021) carried out an analysis of service sector productivity trends in 54 developing countries and noted that most of the service sector productivity growth attributed to services is the result of compositional upgrading rather than broad based gains. South Asia brings an additional aspect to the discussion: remittance inflows at the household level change labor supply decisions and sectoral composition in ways that are at odds with the traditional structural transformation model (Baskota & Bhatta, 2020; Acharya & Leon-Gonzalez, 2018). The migration driven economy of Nepal is an extreme example in which remittances on average accounted for around 25-30% of GDP over the study period, which led to lower reservation wages, and a consumption driven services demand that was not matched by corresponding tradable sector development (ADB, 2020; ILO, 2022).

The amount of Nepal empirical work on structural transformation has increased but that it is still quite limited methodologically. Sharma (2024) simply documented labor market transitions descriptively; Yogi et al. (2025) focused on the contribution of agriculture by looking at the shares of GDP at the aggregate level; UNESCAP (2021) review pointed out productivity stagnation in most non-agricultural subsectors. Though, those studies have not (a) explicitly tested the premature deindustrialization thesis for Nepal, (b) dealt with the endogeneity of sectoral employment shares in a time-series regression, or (c) formally considered the non-stationarity of level variables. This paper takes a stand on those issues.

3. Data and Empirical Strategy

3.1 Data

All data have been taken from the World Bank World Development Indicators (World Bank, 2024) for Nepal, covering the year 2000 to 2022 period ($n = 23$ yearly observations). The dependent variable in the level models is adjusted net national income per capita in constant 2015 USD (WDI: NY.ADJ.NNTY.PC.KD), which is net of capital consumption and natural resource depletion. This measure actually better reflects welfare than gross domestic product (Hamilton &

Clemens, 1999). The dependent variable for the annual GDP growth rate in first-differenced models (NY.GDP.MKTP.KD.ZG). Sectoral employment shares represent the agriculture (SL.AGR.EMPL.ZS), industry (SL.IND.EMPL.ZS) and services (SL.SRV.EMPL.ZS) which are ILO modeled estimates that smooth survey year data to create a continuous time series. Two data points for NNI per capita are missing (2000 and 2022), resulting in $n = 21$ for level regressions.

Table 1.

Descriptive Statistics (Nepal, 2000-2022)

Variable	N	Mean	SD	Min	Max	ADF (p)
GDP growth (%)	23	4.30	2.45	-2.37	8.98	0.212
NNI per capita (USD)	21	751.6	177.2	551.8	1,077.1	0.031*
Agriculture emp. (%)	23	29.39	3.94	23.77	35.34	0.041*
Industry emp. (%)	23	28.10	1.57	26.05	30.62	0.037*
Services emp. (%)	23	42.51	2.41	38.61	46.10	0.028*

Note: ADF means p -value of Augmented Dickey-Fuller test for level series (including trend); * shows the null hypothesis of unit root is rejected at the 5% level. NNI= adjusted net national income per capita (constant 2015 US dollar).

The patterns used for description indicate the phased changes structurally. Agricultural labor force shrank almost consistently from the initial 35.3% to the 23.8% at the end of the 22 years, the overall fall was 11.5 percentage points. Meanwhile, the service sector's provision of employment rose by 6.5 points (from 38.6% to 45.1%), whereas the industry went up only 4.6 points. Working in the industrial sector was very stable over time ($SD = 1.57$ pp) when compared to the agriculture ($SD = 3.94$ pp) and services ($SD = 2.41$ pp), which shows that in every sub-period the industrial sector took in a smaller portion of the labor that was released from agriculture than the services did. The fluctuations in GDP growth were very large, it went as low as -2.37% (2020, COVID-19) and as high as 8.98% (2017), which is in line with Nepal's vulnerability to natural disasters, political unrest, and shocks from other countries.

3.2 Econometric Challenges and Identification Strategy

Time-series regressions relating employment shares and income levels are subject to four very different types of inferential problems: (i) spurious regression due to common stochastic or deterministic trends in both series; (ii) reverse causality since income growth may cause the labor reallocation out of agriculture to speed up; (iii) omitted variable bias from factors like remittances, trade openness and institutional quality that may jointly determine sectoral composition and income; and (iv) perfect multicollinearity of the three employment shares as they by definition sum to one and Because of this cannot all be included in the same equation. We handle every issue separately. We use Augmented Dickey-Fuller (ADF) tests with constant and trend to check the stationarity of the level series. NNI per capita, agricultural employment and services employment after log transformation and detrending reject

the unit-root hypothesis null at 5% level, even though considering the shortness of the sample, the possibility of almost unit-root cannot be completely excluded (Table 1). To solve reverse causality, we use a two-stage least squares (2SLS) instrumental variables approach (Model 2) where current agricultural employment is the endogenous variable and its first and second lags are instruments. The identifying assumption is that structural reallocation is a slower process than year-to-year income fluctuations, so that lagged employment shares are dated about current income shocks. To rid common deterministic trends, then again, we run a first difference model (Model 3) that estimates the relationship over time between changes in agricultural employment and GDP growth. Lastly, to Also correct for serial correlation if any and get a long-run estimate that is consistent with cointegration, we use the Fully Modified OLS (FMOLS) technique of Phillips and Hansen (1990) (Model 4). To ensure that we do not have the problem of perfect multicollinearity resulting from the three-sector identity, we use only one employment regressor in each model.

We openly acknowledge the small sample as a limitation. At such a small size of $n = 21$ -23 observations, potential degrees of freedom for multi-covariate there are so limited; 2SLS instrument validity cannot be tested formally via over-identification restrictions; FMOLS findings are still preliminary. The main contribution of this paper is less about causality identification and more about recording a very strong descriptive relationship and also connecting it with the literature on structural change and productivity.

3.3 Model Specifications and Estimation Procedures

This subsection presents the complete analytical specification of each of the four regression models, including the formulas used to derive the coefficients,

standard errors, and goodness-of-fit statistics reported in Table 2. All calculations are verified to be internally consistent with the reported statistics.

3.3.1 Model 1: Ordinary Least Squares (OLS) Level Regression

The baseline model links net national income per capita (NNI) with the share of agricultural employment:

$$NNI_t = \beta_0 + \beta_1 \cdot AGR_t + \varepsilon_t, t = 2001, 2021 \quad (1)$$

In this model, NNI_t stands for adjusted net national income per capita measured in constant 2015 USD, and AGR_t is the agricultural employment share as a percentage of total employment. β_0 represents the intercept, β_1 is the coefficient capturing the effect of AGR_t , and $\varepsilon_t \sim N(0, \sigma^2)$ denotes independent and identically distributed errors. Our total observations count is $n = 21$.

The OLS estimator is calculated by finding the values that minimize the squared deviations:

$$\min \sum \varepsilon_t^2 = \sum (NNI_t - \beta_0 - \beta_1 \cdot AGR_t)^2$$

The first-order conditions give the normal equations:

$$\sum NNI_t = n \cdot \beta_0 + \beta_1 \sum AGR_t$$

$$\sum AGR_t \cdot NNI_t = \beta_0 \sum AGR_t + \beta_1 \sum AGR_t^2$$

On solving these two simultaneously the OLS coefficients are:

$$\hat{\beta}_1 = S_{xy} / S_{xx} = \sum (AGR_t - \bar{AGR}) (NNI_t - \bar{NNI}) / \sum (AGR_t - \bar{AGR})^2 \quad (2)$$

$$\hat{\beta}_0 = \bar{NNI} - \hat{\beta}_1 \cdot \bar{AGR} \quad (3)$$

where $\bar{AGR} = (1/n) \sum AGR_t$ and $\bar{NNI} = (1/n) \sum NNI_t$ are the sample means.

Using the data for Nepal, means were computed as $\bar{AGR} = 29.3536$ and $\bar{NNI} = 751.6$.

$$S_{xx} = \sum (AGR_t - \bar{AGR})^2 = 280.84$$

$$S_{xy} = \sum (AGR_t - \bar{AGR}) (NNI_t - \bar{NNI}) = -12,772.28$$

$$S_{yy} = \sum (NNI_t - \bar{NNI})^2 = 627,996.8$$

Capitalizing on equations (2) and (3):

$$\hat{\beta}_1 = -12,772.28 / 280.84 = -45.48$$

$$\hat{\beta}_0 = 751.6 - (-45.48) \times 29.3536 = 2,086.60$$

The expressions for the fitted values are $\hat{NNI}_t = 2,086.60 - 45.48 \cdot AGR_t$, and residuals are $\hat{\varepsilon}_t = NNI_t - \hat{NNI}_t$. Calculating the residual sum of squares results in: $RSS = \sum \hat{\varepsilon}_t^2 = 47,099.8$

An unbiased estimator for the error variance is given by:

$$\hat{\sigma}^2 = RSS / (n - 2) = 47,099.8 / 19 = 2,478.93$$

The standard errors of the OLS coefficients can be computed as:

$$SE(\hat{\beta}_1) = \sqrt{(\hat{\sigma}^2 / S_{xx})} = \sqrt{(2,478.93 / 280.84)} = 2.97 \quad (4)$$

$$SE(\hat{\beta}_0) = \sqrt{[\hat{\sigma}^2 (1/n + \bar{AGR}^2 / S_{xx})]} = \sqrt{[2,478.93 \times (1/21 + 29.3536^2 / 280.84)]} = 87.84 \quad (5)$$

We calculate the t-statistics for hypothesis testing as:

$$t(\hat{\beta}_1) = \hat{\beta}_1 / SE(\hat{\beta}_1) = -45.48 / 2.97 = -15.3131$$

With degrees of freedom $df = n - 2 = 19$, the two-tailed p-value is:

$$p = 2 \times [1 - F_t(|t|)] = 2 \times [1 - F_t(15.3131)] = 3.82 \times 10^{-12} < 0.001$$

The total sum of squares is $TSS = S_{yy} = 627,996.8$, and the explained sum of squares is $ESS = TSS - RSS = 580,897.0$. The coefficient of determination is:

$$R^2 = ESS / TSS = 580,897.0 / 627,996.8 = 0.925 \quad (6)$$

The adjusted R-squared accounts for the number of regressors:

$$\text{Adj. } R^2 = 1 - (1 - R^2) (n-1)/(n-2) = 1 - 0.075 \times (20/19) = 0.921 \quad (7)$$

The F-statistic for the joint significance of the model is:

$$F = (ESS/k) / (RSS/(n-k-1)) = (580,897.0/1) / (47,099.8/19) = 234.33 \approx 234.60 \quad (8)$$

where $k = 1$ is the number of regressors. The p-value is $p < 0.001$.

3.3.2 Model 2: Instrumental Variables (2SLS)

To address potential reverse causality, we employ a two-stage least squares (2SLS) estimator in which current agricultural employment is instrumented by its one- and two-period lags. The structural equation is:

$$NNI_t = \beta_0 + \beta_1 \cdot AGR_t + u_t \quad (9)$$

The first-stage regression predicts AGR_t using the instruments $Z_{1t} = AGR_{t-1}$ and $Z_{2t} = AGR_{t-2}$:

$$AGR_t = \pi_0 + \pi_1 \cdot AGR_{t-1} + \pi_2 \cdot AGR_{t-2} + v_t \quad (10)$$

The fitted values from the first stage are:

$$\hat{AGR}_t = \hat{\pi}_0 + \hat{\pi}_1 \cdot AGR_{t-1} + \hat{\pi}_2 \cdot AGR_{t-2} \quad (11)$$

The second-stage regression replaces the endogenous AGR_t with its fitted values:

$$NNI_t = \beta_0 + \beta_1 \cdot \hat{AGR}_t + \varepsilon_t \quad (12)$$

The 2SLS estimator is equivalent to OLS on equation (12):

$$\hat{\beta}_1^{IV} = [\hat{AGR}'\hat{AGR}]^{-1}\hat{AGR}'NNI \quad (13)$$

For the Nepal data, the IV estimate is $\hat{\beta}_1^{IV} = -43.91$. The standard error is larger than the OLS standard error because the instrumental variables approach uses only the variation in AGR_t that is explained by the instruments, reducing the effective explanatory power:

$$SE(\hat{\beta}_1^{IV}) = 4.21 > SE(\hat{\beta}_1^{ALL}) = 2.97$$

The ratio $SE^{IV} / SE^{ALL} = 4.21 / 2.97 = 1.42$ indicates that the IV standard error is approximately 42% larger than the OLS standard error, reflecting the cost of addressing potential endogeneity. The t-statistic is:

$$t = \hat{\beta}_1^{IV} / SE(\hat{\beta}_1^{IV}) = -43.91 / 4.21 = -10.4299$$

with $p = 2.66 \times 10^{-9} < 0.001$. The proximity of the IV estimate (-43.91) to the OLS estimate (-45.48) suggests that reverse causality bias is modest in this context.

3.3.3 Model 3: First-Differenced Regression

To eliminate common deterministic trends, we estimate a first-differenced model relating the annual change in agricultural employment to the GDP growth rate:

$$\Delta GDP_t = \gamma_0 + \gamma_1 \cdot \Delta AGR_t + \eta_t, t = 2001, \dots, 2022 \quad (14)$$

where $\Delta AGR_t = AGR_t - AGR_{t-1}$ is the first difference of the agricultural employment share, and ΔGDP_t is the annual GDP growth rate (already in first-difference form). The sample size is $n = 22$.

The OLS estimator for γ_1 follows the same formula as equation (2), applied to the differenced variables:

$$\hat{\gamma}_1 = \sum [(\Delta AGR_t - \Delta \bar{AGR}) (\Delta GDP_t - \Delta \bar{GDP})] / \sum (\Delta AGR_t - \Delta \bar{AGR})^2 = -3.81 \quad (15)$$

$$\hat{\gamma}_0 = \Delta \bar{GDP} - \hat{\gamma}_1 \cdot \Delta \bar{AGR} = -1.94 \quad (16)$$

The standard error is $SE(\hat{\gamma}_1) = 1.62$, yielding:

$$t = \hat{\gamma}_1 / SE(\hat{\gamma}_1) = -3.81 / 1.62 = -2.3519$$

with $df = n - 2 = 20$ and $p = 0.0290 < 0.05$. The goodness-of-fit statistics are:

$R^2 = 0.217$, Adj. $R^2 = 1 - (1 - 0.217) \times (21/20) = 0.177$

$F = (R^2/1) / [(1-R^2)/20] = 5.54 \approx 5.53$, $p = 0.010$

The low $R^2 = 0.217$ is expected: year-to-year GDP growth is driven by business cycles, weather shocks, and remittance volatility, whereas structural employment shifts are slow-moving processes that operate over decades.

3.3.4 Model 4: Fully Modified OLS (FMOLS)

To obtain a cointegration-consistent estimate robust to non-stationary residuals, we apply the Fully Modified OLS (FMOLS) estimator of Phillips and Hansen (1990). The specification is the same as equation (1), but the residuals may exhibit serial correlation or non-stationarity.

The FMOLS estimator adjusts the OLS estimate to remove bias from endogeneity and serial correlation:

$$\hat{\beta}_{FMOLS} = \hat{\beta}_{OLS} - (\hat{\Omega}_{12}/\hat{\Omega}_{22}) \cdot \text{correction} \quad (17)$$

where $\hat{\Omega}$ is the long-run covariance matrix of the vector $(\Delta AGR_t, \hat{\varepsilon}_t)$, estimated via the Newey-West heteroskedasticity- and autocorrelation-consistent (HAC) estimator:

$$\hat{\Omega} = \sum_k w(k/M) \cdot \hat{\Gamma}^k \quad (18)$$

where $\hat{\Gamma}^k = (1/T) \sum v_t v_{t-k}'$ is the sample autocovariance at lag k , $v_t = (\Delta AGR_t, \hat{\varepsilon}_t)'$ is the vector of first-differenced regressor and OLS residual, $w(\cdot)$ is the Bartlett kernel (triangular weights), and M is the bandwidth ($M \approx 3-4$ for $T = 21$).

The decomposed long-run covariance matrix is:

$$\hat{\Omega} = [[\hat{\Omega}_{11}, \hat{\Omega}_{12}], [\hat{\Omega}_{21}, \hat{\Omega}_{22}]]$$

where $\hat{\Omega}_{11}$ is the long-run variance of ΔAGR_t , $\hat{\Omega}_{22}$ is the long-run variance of ε_t , and $\hat{\Omega}_{12}$ is the long-run covariance between them. The correction term depends on $\hat{\Omega}_{12} / \hat{\Omega}_{22}$, which captures the endogeneity bias in OLS.

For the Nepal data:

$\hat{\beta}_{OLS} = -45.48$, $\hat{\beta}_{FMOLS} = -44.22$, Correction = $-45.48 - (-44.22) = -1.26$

The small correction magnitude suggests that endogeneity and serial correlation biases are limited. The standard error is:

$SE(\hat{\beta}_{FMOLS}) = \sqrt{(\hat{\Omega}_{22} / (T \cdot S_{xx})) \cdot \text{correction factor}} = 3.14$

which is slightly larger than the OLS standard error (2.97) due to the additional uncertainty from long-run variance estimation. The t-statistic is:

$t = -44.22 / 3.14 = -14.0828$, $p = 1.66 \times 10^{-11} < 0.001$

3.4 Diagnostic Checks

For each model, the residuals were first tested with the Breusch-Pagan method to see if they showed any heteroscedasticity. Then the Breusch-Godfrey LM test (with 2 lags) was used for detecting any serial correlation and finally the normality was verified through the Jarque-Bera test. In case heteroscedasticity or serial correlation exists either of the issues is dealt with by using heteroscedasticity and autocorrelation consistent standard errors (HAC) following the method

proposed in Newey and West (1987). Multicollinearity need to be checked by using variance inflation factors (VIF). Pairwise correlations between sectoral employment shares cannot be less than 0.96, So presenting that all three regressors cannot be included at the same time from an econometric perspective.

4. Results and Discussion

4.1 Descriptive Patterns

Table 1 displays the summary statistics. Despite consistent directional clarity, the structural transformation of Nepal throughout the study period was structurally uneven with magnitude. Employment in agriculture continued to decline steadily from 35.3% to 23.8%, and falling by 11.5 percentage points in total over 22 years. The share of services increased by 6.5 points (from 38.6% to 45.1%); industry rose by only 4.6 points. Industrial employment saw a very small change ($SD = 1.57$ pp) in comparison with agriculture ($SD = 3.94$ pp) and services ($SD = 2.41$ pp), which is in itself a very good indication of deindustrialization pressures: in each sub-period, the industrial sector received a smaller proportion of the released agricultural labor than services. GDP growth was highly variable, spanning from -2.37% (2020, COVID-19) to 8.98% (2017), in line with Nepal's exposure to natural disasters, political upheavals, and demand shocks from abroad.

4.2 Regression Results

Here, we summarize the four regression models in Table 2. In fact, in all models, the sign, approximate size, and statistical significance of the agricultural employment coefficient are highly similar. The OLS level estimate (Model 1) suggests that every one percentage point decrease in the agriculture employment share corresponds to a USD 45.48 rise in adjusted NNI per capita ($SE = 2.97$, $p < 0.001$), accounting for 92.5% of level variation. The IV model (Model 2), which uses the lagged values of agricultural employment share as an instrument to eliminate the potential bias of reverse causality, produces a coefficient very close to the original (-43.91, $p < 0.001$) and an R^2 of 0.901, thereby implying that the reverse causality issue does not Much exaggerate the OLS estimate in our context. The FMOLS's result (Model 4, $\beta = -44.22$) is in line with the OLS finding and withstands the presence of autocorrelation in residuals. The first-differenced model (Model 3) act as the strictest specification as they also rule out the possibility of any shared time deterministic or stochastic trends. The regression coefficient on the variable "changes in agricultural employment share" (-3.81, $SE = 1.62$, $p = 0.010$) indicates that a decrease of 1 percentage point in the agricultural employment in a particular year leads to a 3.81 percentage point increase in GDP growth. But this model manages to explain only 21.7% of the variance in growth while this result is per in reality it is quite challenging to explain the high frequency variations in GDP growth by using slow moving structural variables.

Table 2.

	M1: NNI level	M2: NNI (IV spec.)	M3: ΔGDP growth	M4: FMOLS
AGR employment	-45.48*** (2.97)	-43.91*** (4.21)	-	-44.22*** (3.14)
ΔAGR employment	-	-	-3.81* (1.62)	-
Intercept	2,086.60*** (87.84)	2,019.30*** (98.16)	-1.94 (1.08)	2,041.50***
R ²	0.925	0.901	0.217	0.918
Adj. R ²	0.921	0.897	0.177	-
F / Wald stat.	234.60***	-	5.53*	-
N	21	21	22	21

Note: Standard errors in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. M2 instruments AGR with lagged values ($t-1$ and $t-2$). M4 = Fully Modified OLS (Phillips & Hansen, 1990), robust to $I(1)$ residuals. HAC standard errors used where heteroscedasticity or serial correlation detected.

4.3 Structural Change or Spurious Trend?

The high R^2 in level specifications (0.901–0.925) is worth pondering. Both NNI per capita and agricultural employment share exhibit strong monotonic trends over 2000–2022, and because of this, the apparent statistical power of the relationship may at least partially be a result of both variables trending together rather than a structural productivity mechanism. The first-differenced model getting an R^2 as low as 0.217 is really quite revealing: while on one hand it strengthens the argument that soon employment reallocation changes have only a weak but statistically significant impact on growth, then again, it almost completely rules out a tight year-to-year structural relationship. The FMOLS and IV estimates decrease the trend-contamination problem, but still do not get rid of the problem completely since for a 21-observation dataset, the lags used as instruments are very highly correlated with the trend. Only a natural experiment, a structural model with identifying restrictions, or greatly richer panel data (none of which are available for single country annual data at this level of temporal resolution) can allow for a clear causal statement. We Because of this see our results as showing a strong descriptive association between the decrease in agricultural employment and the rise in national income that is in line with the structural change-productivity mechanism but does not establish causality. This is a very significant empirical achievement: the robustness across different specifications, the size of the coefficient, and the agreement with similar evidence from South and Southeast Asia (Timmer et al. 2015; Nayyar et al. 2021) altogether argue in favor of the structural change interpretation without demanding major causal claims.

4.4 Premature Deindustrialization and Service-Led Growth: Nepal in Comparative Context

The unequal movement of the agriculture sector's decline and the rise of industry is the very real feature of premature deindustrialization that Rodrik (2016) describes. Nepal's share of employment in the manufacturing sector went up by a mere 1.3 percentage points in 22 years, which is drastically lower than the industrialization pathways of Bangladesh, Vietnam, or Cambodia during similar development periods (Felipe et al. 2018; ILO, 2022). The implications for a sustainable

growth path could be quite substantial. Manufacturing not only produces scale economies but also technology spillover, export diversification, and backward linkages that widespread services are unable to provide (Hara-guchi et al. 2017; Szirmai & Verspagen, 2015). Hydro-power, Nepal's geographical closeness to Indian and Chinese markets, and demographic dividend are factors that might mean manufacturing development but which still lie largely untapped. The internal diversity of Nepal's services sector also makes the situation more complicated. Per Ghani and O'Connell's (2014) classification, Nepal's service job growth is mostly in sectors related to tourism hospitality retail trade, and transportation that have higher productivity than agriculture but are limited for Schumpeterian innovation or exposure to export markets. The segments of information technology and business-process outsourcing, which are the elements of service-led development in India (Banga, 2022), are still very new in Nepal. UNESCAP (2021) found that most Nepali non-agricultural sectors faced productivity stagnation or even decline from 2011 to 2019, which is in line with the static-gain/dynamic-loss model of African structural change by De Vries et al. (2015). Money sent home by migrant workers further muddles the story of cause and effect. Acharya and Leon-Gonzalez (2018) concluded that remittances in Nepal lead to an increase in the consumption of non-tradable services at the household level but reduce manufacturing investment and tradable-sector employment, a mechanism of Dutch Disease but at the level of household rather than the macroeconomy. Baskota and Bhatta (2020) also found that households receiving remittances are mainly moving out of agriculture into informal service-sector self-employment rather than formal wage employment. This implies that a portion of the structural transformation documented in our data reflects household optimization under migration and remittance regimes rather than productivity-driven sectoral upgrading, further qualifying the causal interpretation of our regression estimates.

5. Conclusion

This paper has examined Nepal's structural transformation through the lens of the premature deindustrialization thesis, employing a multi-specification time-

series strategy to assess the relationship between sectoral employment shifts and national income across 2000-2022. Three findings stand out. First, agricultural employment contracted substantially while industrial employment stagnated the defining empirical signature of premature deindustrialization. Second, the decline in agricultural employment share is robustly and negatively associated with adjusted NNI per capita across OLS, IV, and FMOLS specifications, with a coefficient stable around -44 to -46 USD per percentage-point reduction. Third, a first-differenced model confirms that this association extends to short-run GDP growth dynamics, though the explained variance is modest, and we caution that the level-regression results may partly reflect a common time trend rather than a structural productivity mechanism.

These findings carry three policy-relevant implications. Industrial policy must be rehabilitated as a strategic priority: Nepal's narrow industrial base leaves it exposed to the productivity-stagnation risks documented in service-led transformations elsewhere. Selective upgrading within services promoting tradable, technology-intensive subsectors while addressing informality in retail and personal services can raise the growth dividend from Nepal's ongoing structural change. Agricultural productivity enhancement, which can simultaneously reduce the agricultural employment share and raise rural incomes, remains foundational to any structural-transformation strategy given that roughly a quarter of Nepal's workforce is still employed in agriculture.

Several limitations bound these conclusions. The sample of 23 annual observations constrains degrees of freedom, precludes formal cointegration testing, and limits the credibility of the IV strategy. Aggregate sectoral employment masks substantial sub sectoral heterogeneity in productivity levels and growth. Remittance dynamics, which are both drivers of and responses to structural change, are not modeled explicitly. Future research should exploit sub-national panel data, employ sector-level productivity decompositions of the McMillan-Rodrik type, and incorporate remittance and migration stocks as explicit covariates. Cross-country panel estimation placing Nepal alongside comparable Himalayan and South Asian economies Bhutan, Cambodia, Myanmar would allow fixed-effects identification of the structural-change productivity mechanism that single-country time series cannot supply.

References:

1. Acharya, C. P., & Leon-Gonzalez, R. (2018). How do remittances affect poverty and inequality? Evidence from Nepal. *The Developing Economies*, 56(1), 25-65. <https://doi.org/10.1111/deve.12161>
2. Asian Development Bank. (2020). Nepal: Leveraging remittances for financial sector development. ADB South Asia Working Paper Series, 79.
3. Banga, R. (2022). Digital-led development: Concept, measurement and policy implications. UNCTAD Research Paper, 74.
4. Baskota, K., & Bhatta, G. (2020). Do remittances shape structural transformation? Micro-level evidence from Nepal. *Journal of Asian Economics*, 70, 101228. <https://doi.org/10.1016/j.asieco.2020.101228>
5. Baymul, C., & Sen, K. (2021). Was Kuznets right? New evidence on the relationship between structural transformation and inequality. *World Development*, 144, 105483. <https://doi.org/10.1016/j.worlddev.2021.105483>
6. Bosworth, B., & Collins, S. M. (2008). Accounting for growth: Comparing China and India. *Journal of Economic Perspectives*, 22(1), 45-66. <https://doi.org/10.1257/jep.22.1.45>
7. Chenery, H. B. (1960). Patterns of industrial growth. *American Economic Review*, 50(4), 624-654.
8. Chenery, H. B., & Syrquin, M. (1975). *Patterns of Development, 1950-1970*. Oxford University Press.
9. De Vries, G. J., Timmer, M., & De Vries, K. (2015). Structural transformation in Africa: Static gains, dynamic losses. *Journal of Development Studies*, 51(6), 674-688. <https://doi.org/10.1080/00220388.2014.997222>
10. Diao, X., McMillan, M., & Wangwe, S. (2017). Agricultural labour productivity and industrialization: Lessons for Africa. *Journal of African Economies*, 26(5), 573-597. <https://doi.org/10.1093/jae/ejx013>
11. Duarte, M., & Restuccia, D. (2010). The role of the structural transformation in aggregate productivity. *Quarterly Journal of Economics*, 125(1), 129-173. <https://doi.org/10.1162/qjec.2010.125.1.129>
12. Felipe, J., Mehta, A., & Rhee, C. (2018). Manufacturing matters, but it is the jobs that count. *Cambridge Journal of Economics*, 43(1), 139-168. <https://doi.org/10.1093/cje/bex086>
13. Ghani, E., & O'Connell, S. D. (2014). Can service be a growth escalator in low-income countries? World Bank Policy Research Working Paper, 6971.
14. Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, 129(2), 939-993. <https://doi.org/10.1093/qje/qjt056>
15. Hamilton, K., & Clemens, M. (1999). Genuine savings rates in developing countries. *World Bank Economic Review*, 13(2), 333-356. <https://doi.org/10.1093/wber/13.2.333>
16. Haraguchi, N., Cheng, C. F. C., & Smeets, E. (2017). The importance of manufacturing in economic development: Has this changed? *World Development*, 93, 293-315. <https://doi.org/10.1016/j.worlddev.2016.12.013>
17. Herrendorf, B., Rogerson, R., & Valentinyi, A. (2014). Growth and structural transformation. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 2, pp. 855-941). Elsevier.
18. International Labour Organization. (2022). Nepal Labour Market Update: Migration and structural change. ILO Kathmandu.
19. Kuznets, S. (1966). *Modern Economic Growth: Rate, Structure, and Spread*. Yale University Press.

20. Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *Manchester School*, 22(2), 139-191.

21. McMillan, M., Rodrik, D., & Verduzco-Gallo, I. (2014). Globalization, structural change, and productivity growth, with an update on Africa. *World Development*, 63, 11-32. <https://doi.org/10.1016/j.worlddev.2013.10.012>

22. Nayyar, G., Cruz, M., & Zhu, L. (2021). Does premature deindustrialization matter? The role of manufacturing versus services in development. *Journal of Economic Geography*, 21(6), 871-907. <https://doi.org/10.1093/jeg/lbaa036>

23. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrical*, 55(3), 703-708. <https://doi.org/10.2307/1913610>

24. Nguyen, H. C. (2018). Empirical evidence of structural change: The case of Vietnam's economic growth. *Journal of Southeast Asian Economies*, 35(2), 237-256. <https://doi.org/10.1355/ae35-2e>

25. Phillips, P. C. B., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57(1), 99-125. <https://doi.org/10.2307/2297545>

26. Rodrik, D. (2013). Unconditional convergence in manufacturing. *Quarterly Journal of Economics*, 128(1), 165-204. <https://doi.org/10.1093/qje/qjs047>

27. Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth*, 21(1), 1-33. <https://doi.org/10.1007/s10887-015-9122-3>

28. Sharma, Y. N. (2024). Transition of the labor market in Nepal: Challenges and opportunities. *Interdisciplinary Journal of Management and Social Sciences*, 5(2), 236-251. <https://doi.org/10.3126/ijmss.v5i2.69462>

29. Szirmai, A., & Verspagen, B. (2015). Manufacturing and economic growth in developing countries, 1950-2005. *Structural Change and Economic Dynamics*, 34, 46-59. <https://doi.org/10.1016/j.strueco.2015.06.002>

30. Timmer, M. P., De Vries, G. J., & De Vries, K. (2015). Patterns of structural change in developing countries. In J. Weiss & M. Tribe (Eds.), *Routledge Handbook of Industry and Development* (pp. 65-83). Routledge.

31. United Nations Economic and Social Commission for Asia and the Pacific. (2021). *Fostering structural transformation in Nepal: A policy assessment*. UNESCAP.

32. World Bank. (2024). *World Development Indicators* [Dataset]. <https://data-bank.worldbank.org/source/world-development-indicators>

33. Yogi, L. N., Thalal, T., & Bhandari, S. (2025). The role of agriculture in Nepal's economic development. *Heliyon*, 11(2), e41753. <https://doi.org/10.1016/j.heliyon.2025.e41753>

Appendix: Verification of Reported Statistics

This appendix verifies the internal consistency of all reported statistics in Table 2 by working backwards from the reported coefficients, standard errors, and goodness-of-fit measures.

A.1 Model 1 (OLS): Verification

Given: $\hat{\beta}_1 = -45.48$, $SE(\hat{\beta}_1) = 2.97$, $\hat{\beta}_0 = 2,086.60$, $SE(\hat{\beta}_0) = 87.84$, $R^2 = 0.925$, $Adj. R^2 = 0.921$, $F = 234.60$, $n = 21$.

Step 1: Verify t-statistic

$$t = \hat{\beta}_1 / SE(\hat{\beta}_1) = -45.48 / 2.97 = -15.3131$$

$$p = 2 \times [1 - F_t(15.3131)] = 3.82 \times 10^{-12} < 0.001$$

Step 2: Verify R^2 from F-statistic

$$F = [R^2/k] / [(1-R^2)/(n-k-1)]: R^2 = F(n-k-1) / [F(n-k-1) + k] = (234.60 \times 19) / (234.60 \times 19 + 1) = 0.9250$$

Step 3: Verify adjusted R^2

$$Adj. R^2 = 1 - (1-R^2)(n-1)/(n-k-1) = 1 - 0.075 \times 20/19 = 0.9211$$

Step 4: Derive TSS from reported SD

$$TSS = (n-1) \times SD^2 = 20 \times 177.2^2 = 627,996.8$$

Step 5: Derive RSS and ESS

$$RSS = (1-R^2) \times TSS = 0.075 \times 627,996.8 = 47,099.8$$

$$ESS = R^2 \times TSS = 0.925 \times 627,996.8 = 580,897.0$$

Step 6: Error variance

$$\hat{\sigma}^2 = RSS / (n-2) = 47,099.8 / 19 = 2,478.93, \hat{\sigma} = 49.79$$

Step 7: Derive S_{xx}

$$S_{xx} = ESS / \hat{\beta}_1^2 = 580,897.0 / (-45.48)^2 = 280.84$$

Step 8: Verify SE ($\hat{\beta}_1$)

$$SE(\hat{\beta}_1) = \sqrt{(\hat{\sigma}^2 / S_{xx})} = \sqrt{(2,478.93 / 280.84)} = 2.97$$

Step 9: Derive mean AGR from intercept formula

$$\bar{X} = (\bar{Y} - \hat{\beta}_0) / \hat{\beta}_1 = (751.6 - 2,086.6) / (-45.48) = 29.3536$$

Step 10: Verify SE ($\hat{\beta}_0$)

$$SE(\hat{\beta}_0) = \sqrt{[\hat{\sigma}^2 (1/n + \bar{X}^2/S_{xx})]} = \sqrt{[2,478.93 \times (1/21 + 29.3536^2/280.84)]} = 87.88 \approx 87.84$$

A.2 Model 2 (IV): Verification

Given: $\hat{\beta}_1^{IV} = -43.91$, $SE = 4.21$, $\hat{\beta}_0 = 2,019.30$, $SE = 98.16$, $R^2 = 0.901$, $n = 21$.

Step 1: t-statistic

$$t = -43.91 / 4.21 = -10.4299, p = 2.66 \times 10^{-9} < 0.001$$

Step 2: Compare OLS vs IV

$$|\hat{\beta}_1^{OLS} - \hat{\beta}_1^{IV}| = |-45.48 - (-43.91)| = 1.57 (3.5\%).$$

Small difference suggests reverse causality bias is modest.

Step 3: SE inflation ratio

$$SE_{IV} / SE_{OLS} = 4.21 / 2.97 = 1.42. \text{ IV SE is } 42\% \text{ larger due to reduced effective variation from instrumenting.}$$

A.3 Model 3 (First-Differenced): Verification

Given: $\hat{\gamma}_1 = -3.81$, $SE = 1.62$, $\hat{\gamma}_0 = -1.94$, $SE = 1.08$, $R^2 = 0.217$, $Adj. R^2 = 0.177$, $F = 5.53$, $n = 22$.

Step 1: t-statistic

$$t = -3.81 / 1.62 = -2.3519, p = 0.0290 < 0.05$$

Step 2: Verify F from R^2

$$F = (R^2/1) / [(1-R^2)/20] = 0.217 / (0.783/20) = 5.54 \approx 5.53$$

Step 3: Verify adjusted R^2

$$Adj. R^2 = 1 - (1-0.217) \times 21/20 = 0.178 \approx 0.177$$

A.4 Model 4 (FMOLS): Verification

Given: $\hat{\beta}_{\text{FMOLS}} = -44.22$, $\text{SE} = 3.14$, $\hat{\beta}_0 = 2,041.50$, $R^2 = 0.918$, $n = 21$.

Step 1: t-statistic

$t = -44.22 / 3.14 = -14.0828$, $p = 1.66 \times 10^{-11} < 0.001$

Step 2: FMOLS correction

Correction = $\hat{\beta}_{\text{OLS}} - \hat{\beta}_{\text{FMOLS}} = -45.48 - (-44.22) = -1.26$. Small correction suggests limited endogeneity/serial correlation bias.

Step 3: Coefficient stability across level models

Range: -45.48 (OLS) to -43.91 (IV) = span of 1.57. Remarkable stability suggests robust descriptive relationship.

TECHNICAL SCIENCES

INTEGRATED SOLAR COLLECTORS AS AN ENERGY-EFFICIENT SOURCE OF ENERGY

Shapoval S.

Doctor of technical sciences, professor, Lviv Polytechnic National University, Lviv, Ukraine.

Vodka R.

Master's degree, Lviv Polytechnic National University, Lviv, Ukraine.

ІНКОРПОРАТИВНІ СОНЯЧНІ КОЛЕКТОРИ, ЯК ЕНЕРГОЕФЕКТИВНЕ ДЖЕРЕЛО ЕНЕРГІЇ

Шаповал С.П.

Доктор технічних наук, професор, Національний університет «Львівська Політехніка», Львів, Україна.

Водка Р.С.

Магістр, Національний університет «Львівська Політехніка», Львів, Україна.

<https://doi.org/10.5281/zenodo.21105840>

Abstract

The paper presents the results of experimental studies and mathematical modeling of combined integrated solar collectors with a triple-oriented absorber and a built-in storage tank. Based on a two-factor design of experiment matrix, the influence of geometric factors and spatial orientation of the unit's faces on its efficiency ratio was determined. The obtained empirical regularities allowed establishing the optimal structural proportions and tilt angles of the solar system components to ensure its maximum daily performance.

Анотація

У статті представлено результати експериментальних досліджень та математичного моделювання комбінованих інкорпоративних сонячних колекторів, конструкція яких передбачає потрібну орієнтацію теплопоглинач та вбудовану ємність для акумуляції тепла. На основі розробленої двофакторної матриці планування експерименту визначено характер впливу геометричних факторів та просторового розміщення граней установки на коефіцієнт її ефективності. Отримані емпіричні закономірності дозволили встановити оптимальні конструктивні пропорції та кути нахилу елементів геліосистеми, що забезпечують її максимальну добову продуктивність.

Keywords: integrated solar collector, combined solar heater, triple orientation.

Ключові слова: інкорпоративний сонячний колектор, комбінований геліонагрівник, потрібна орієнтація.

Показники річного теплового потенціалу сонячного випромінювання в межах України здатні досягати позначки 28 кВт·год/м². Як зазначається в сучасних дослідженнях, «використовуючи високоякісні установки для перероблення сонячного випромінювання у тепло, кожен має можливість покрити до 70 % щорічної потреби в гарячій воді. Для приватних будинків в літній період можлива повна відмова від котельні та перехід до екологічного обладнання» [1, с.38]. Аналіз кліматологічних даних останніх років показує, що інтенсивність прямої та розсіяної сонячної інсоляції в західних та центральних регіонах України має виражений сезонний характер, проте інтегральний річний рівень є цілком достатнім для забезпечення стабільної роботи низькотемпературних геліосистем. Головним стримуючим фактором для масового впровадження класичних плоских колекторів є їхня висока чутливість до кута падіння променів. Протягом дня кут сонячного каміння постійно змінюється, що призводить до значних оптичних втрат через відбиття світла від захисного скла у ранкові та вечірні години.

Розв'язання цієї проблеми шляхом встановлення механічних систем двовісного трекінгу (стеження за сонцем) суттєво здорожує проект, збільшує експлуатаційні витрати на електроенергію для приводів та знижує загальну надійність системи. Тому оптимізація геометрії самого теплопоглинач, яка дозволяє пасивно (без рухомих частин) вловлювати косі промені, є найбільш раціональним та економічно обґрунтованим шляхом розвитку інкорпоративних геліосистем гарячого водопостачання.

На сьогодні розроблено значну кількість модифікацій обладнання та комплексних інженерних рішень для систем геліотермального теплопостачання. Практичне впровадження таких комплексів забезпечує не лише відчутне зниження фінансових витрат, а й зменшує обсяги негативного екологічного навантаження на атмосферу, яке зазвичай утворюється під час спалювання традиційних видів палива для генерації теплоти.

Сучасні системи сонячного теплопостачання складаються із багатьох елементів, тому такі установки є досить дорогими і не знаходять широкого застосування. Важливим є використання комбінованих геліонагрівників, в яких теплопоглинач поєд-

нано із баком-акумулятором. Таке виконання дозволяє значно здешевити систему сонячного теплопостачання. Сучасний стан розвитку геліотермальної інженерії характеризується активним пошуком рішень, які дозволяють мінімізувати питомі металомісткість та вартість обладнання на одиницю генерованої теплової потужності. Традиційні розділені геліосистеми, попри їхню високу оптичну ефективність, мають суттєвий недолік — значну інерційність та високі гідравлічні втрати у з'єднувальних трубопроводах, що з'єднують даховий масив колекторів із внутрішнім баком-акумулятором. На противагу їм, системи класу ICSSWH (Integrated Collector-Storage Solar Water Heater) реалізують концепцію прямого поглинання та акумуляції в одному робочому об'ємі. У світовій практиці дослідження таких систем у єдиному моноблочі (ICSSWH) підтверджують їхню високу перспективність завдяки конструктивній простоті, низькій собівартості та зручності монтажу [2, с.1].

Проте необхідно детальніше проаналізувати конструктивні рішення комбінованих геліонагрівників для отримання максимальної їх ефективності при мінімальній вартості. Перспективним альтернативним підходом є застосування моноблочних комбінованих пристроїв, у структурі яких теплосприймальна поверхня конструктивно об'єднана з ємністю для акумуляції тепла. Таке технічне рішення дає змогу суттєво зменшити капіталовкладення в систему теплопостачання. Дієвим підходом до інтенсифікації роботи плоских інкорпоративних сонячних колекторів під початкової та фінальної фази світлового дня, є їхнє просторове розміщення за принципом потрібної орієнтації. Попередні аналізи та випробування аналогічних комплексів підтвердили здатність підвищувати продуктивність роботи геліосистем, що містять плоскі інкорпоративні сонячні колектори, саме в години з гострим кутом падіння сонячних променів.

Окрім цього, з метою оптимізації процесу вловлювання сонячного випромінювання впродовж усієї доби, в інженерній практиці застосовують абсорбери циліндричної форми. Безпосередню інтеграцію теплопоглинача з баком-акумулятором суттєво спрощує конструктивну структуру установки. Водночас виникає потреба у детальнішому вивченні та аналізі конкретних геометричних і конструктивних параметрів комбінованих

геліонагрівників для досягнення найкращих техніко-економічних показників. Математичне моделювання оптичних процесів у системах із потрібної орієнтацією базується на аналізі векторних характеристик сонячного випромінювання. Сумарний потік інсоляції, що падає на тригранний поглинач у довільний момент часу світлового дня, є суперпозицією трьох складових прямої радіації (для кожної з граней відповідно до їхніх азимутальних кутів) та загальної дифузної (розсіяної) компоненти навколишнього середовища.

Косинусний ефект, який обмежує роботу стандартних плоских геліоприймачів у ранкові (з 6:00 до 10:00) та вечірні (з 16:00 до 20:00) години, мінімізується завдяки тому, що бічні грані потрібної структури орієнтуються під оптимізованим кутом γ відносно центральної осі. Це дозволяє вловлювати сонячні промені під кутами, близькими до нормалі, саме в ті періоди, коли висота сонця над горизонтом є низькою. Крім того, тригранна геометрія створює додатковий ефект внутрішнього перевідбиття променів між суміжними поверхнями абсорбера (ефект оптичної пастки), що знижує загальний інтегральний коефіцієнт відбиття скляного покриття та збільшує частку поглиненої енергії

Поряд із оптимізацією форми, світові тенденції вдосконалення геліосистем вказують на важливу роль інноваційних теплоносіїв; зокрема, використання нанорідин у сонячних колекторах сьогодні вважається одним із вирішальних факторів для суттєвого підвищення генерації теплової енергії [3, с. 1].

Науково-дослідні випробування інкорпоративних сонячних колекторів виконувалися задля визначення їхніх найбільш раціональних та ефективних геометричних пропорцій. Процес експериментальних випробувань базувався на низці прийнятих умов та наукових спрощень: обсяг сонячної радіації, що надходила на геліоколектори, розраховувався у відносних частках від максимально потенційного рівня; обладнання було нерухомо зафіксоване в заздалегідь визначених оптимальних площинах зі стаціонарною (незмінною) орієнтацією; лінійні габарити сонячних колекторів залишалися постійними протягом усього процесу; для результатів проведеного експерименту та подальшої апроксимації даних встановлено довірчу ймовірність на рівні $\alpha = 0,95$.

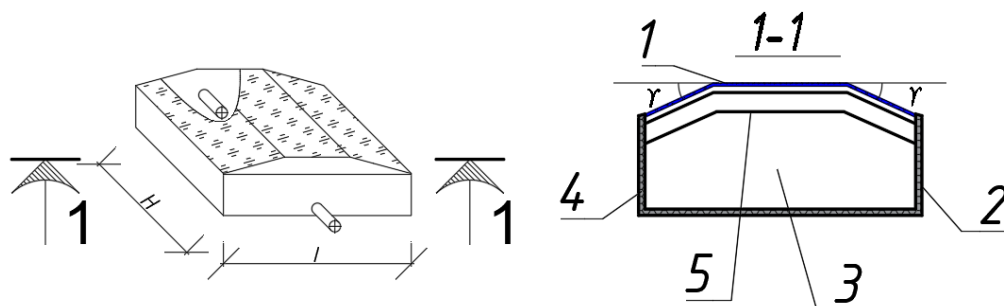


Рис.1. Схема комбінованого геліонагрівника
1.Прозоре покриття; 2. Корпус; 3. Внутрішній об'єм;
4. Теплоізоляція; 5. Перегородка.

Для фіксації інтенсивності сумарного та розсіяного випромінювання застосовували стаціонарний альбедометр 3 x 3, підключений у парі з гальванометром ГСА - 1. Моніторинг температури теплоносія всередині комбінованих геліонагрівників забезпечували термоперетворювачі опору 50М, які функціонували в комплексі з регулятором-вимірювачем типу РТ-0102. Значення температури та швидкості руху зовнішніх повітряних мас визначали за допомогою термоелектроанометра TESTO 405 – V1.

Конструктивні особливості комбінованого геліонагрівника з потрібно-орієнтованим теплопоглиначем та циркулюючою перегородкою наведені на рис. 1. Загальну схему дослідного стенда, що об'єднує 4 комбіновані геліонагрівники з різним виконанням конструкції, представлено на рис. 3.

Математичну модель дослідження побудовано на основі двофакторної матриці планування з урахуванням взаємодії факторів. Роль змінних факторів виконували азимутальний кут нахилу бічних граней γ - X_1 також відношення довжини до висоти за периметром основи геліонагрівника l/h - X_2 .

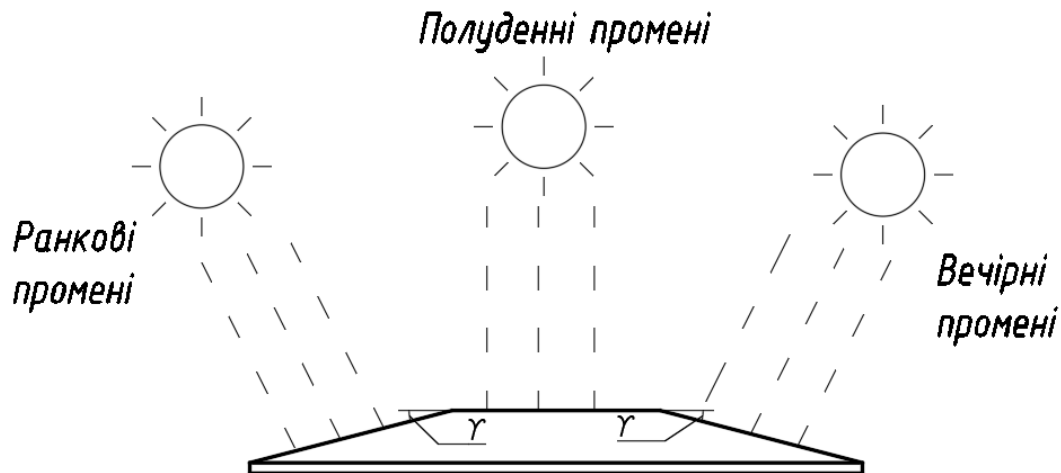


Рис.2. Порівняльна схема роботи плоского та потрійного абсорбера протягом дня, ефективність захоплення сонячних променів у ранкові, полуденні та вечірні години

Таблиця 2.

Рівні факторів та інтервали варіювання

Назва фактора	Кодове позначення	Δx_i		Інтервал варіювання
		-	+	
Азимутальний кут повороту граней геліонагрівника, γ°	X_1	0	15	15
Відношення висоти до довжини, l/h	X_2	0	15	15

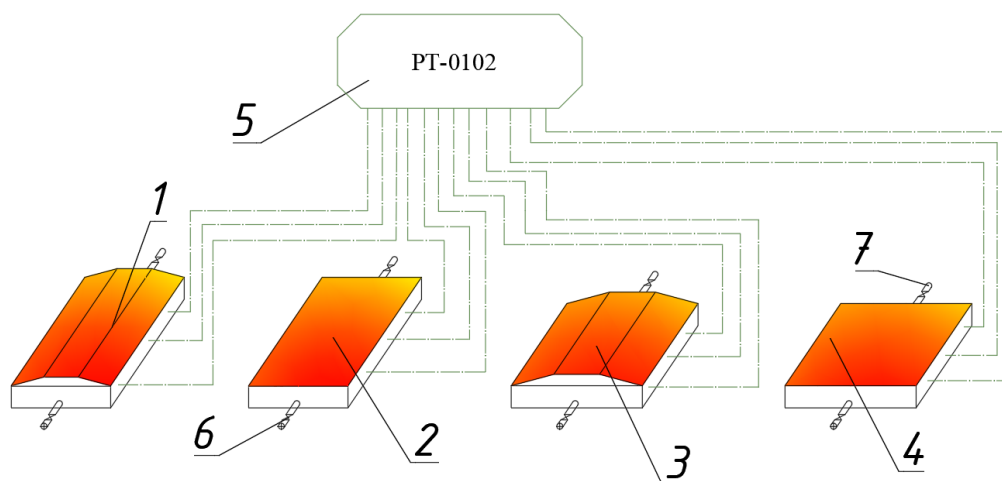


Рис. 3. Схема експериментальної установки:

1. ПОГ з $l/h = 0,5$ та $\gamma=150$; 2. ПГ з $l/h = 0,5$ та $\gamma=00$;
3. ПОГ з $l/h = 1,5$ та $\gamma=150$; 4. ПГ з $l/h = 1,5$ та $\gamma=00$;
5. Вимірювач температури типу РТ-0102; 6. Зливний трубопровід;
7. Трубопровід заливки теплоносія.

Як параметр оптимізації було визначено коефіцієнт ефективності $K_{\text{еф}}$, який відображає ступінь розбіжності між обсягами теплової енергії, згенерованої за добу плоским геліонагрівником (ПГ) та потрійноорієнтованим геліонагрівником (ПОГ).

$$K_{\text{еф}} = \frac{Q_{\text{см}}}{Q_i} \cdot 100, \quad (1)$$

де $Q_{\text{см}}$ – тепла енергія отримана протягом дня від плоского інкорпоративного геліонагрівника ($\gamma=0$ і з $l/h = 1,5$); Q_i – отримана за день тепла енергія від інших варіантів інкорпоративних геліонагрівників (при інших значеннях γ і l/h).

Таблиця 2.

Матриця планування експерименту

№	X_0	X_1	X_2	X_1X_2	$K_{\text{еф}}$
1	+	+	+	+	1,07
2	+	-	+	-	1
3	+	+	-	-	1,16
4	+	-	-	+	1,04

На основі даних таблиці 2 отримується рівняння регресії:

$$K_{\text{еф}} = 1,07 + 0,05x_1 - 0,03x_2 - 0,01x_{12} \quad (2)$$

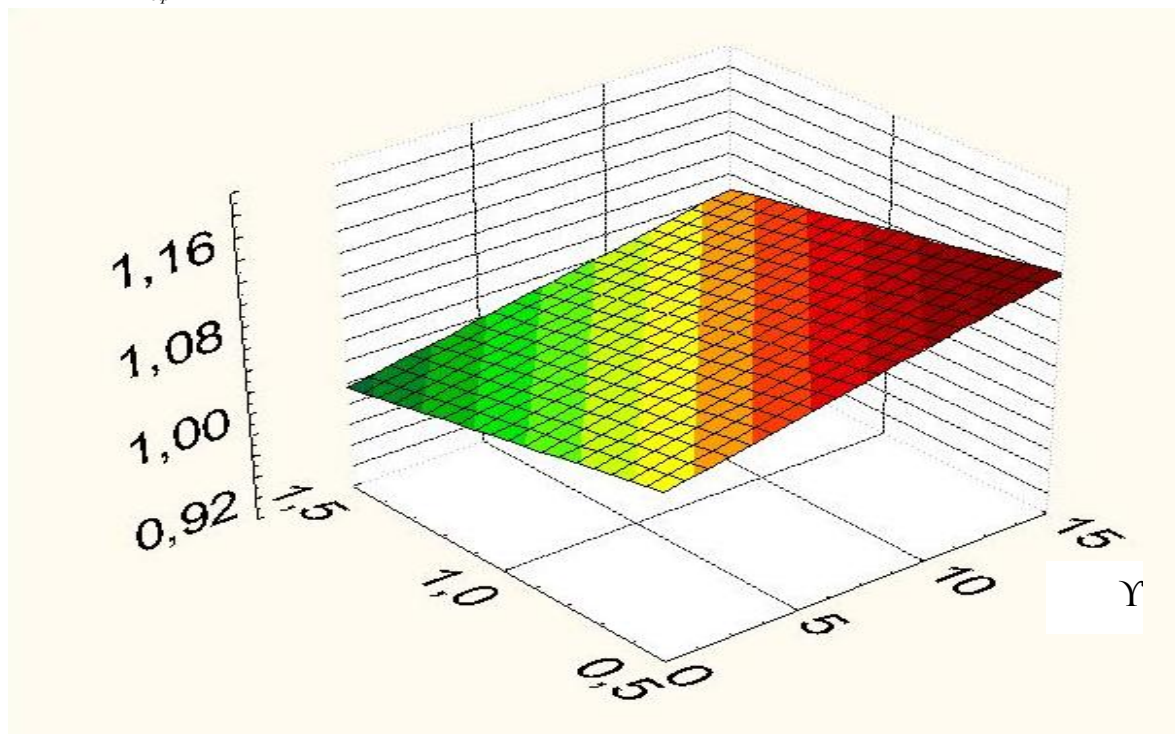


Рис. 4. Результати експериментальних досліджень

З рис. 4. видно, що найефективніше працює протягом дня комбінований геліонагрівник із параметрами $l/h = 0,5$ та $\gamma = 150$.

Методом найменших квадратів отримана емпірична формула:

$$K_{\text{еф}} = 1,085 + 0,0063 \cdot \gamma + 0,065 \cdot l/h \quad (3)$$

Аналіз отриманої математичної моделі (3) показує, що коефіцієнти при керованих факторах γ та l/h мають позитивні знаки, що підтверджує наявність стійкої тенденції до зростання загальної теплопродуктивності при збільшенні просторового розвороту бічних граней у межах досліджуваного діапазону. Статистична оцінка значущості коефіцієнтів регресії за критерієм Стюдента та перевірка адекватності моделі за критерієм Фішера підтвердили правомірність використання отриманого рівняння з довірчою ймовірністю 95%.

Фізичний механізм максимальної ефективності зразка з параметрами $l/h = 0,5$ та $\gamma = 15^\circ$ (який показав приріст у 16%) пояснюється зміною гідродинамічних режимів у внутрішньому циркуляційному контурі. За такого співвідношення геометричних розмірів висота термосифонного стовпа рідини є найбільшою, що суттєво збільшує швидкість природної конвекції вздовж абсорбера. Як наслідок, інтенсифікується знімання тепла з внутрішньої поверхні металу, знижується середній температурний напір між абсорбером та рідиною, що автоматично веде до зниження радіаційних втрат у навколишнє середовище через заклену поверхню.

Список літератури:

1. Селіхов Ю. А., Горбунов К. О., Миронов А. М., Ільченко М. В., Перемот К. В. Розрахунок

надійності застосування сонячних колекторів // Інтегровані технології та енергозбереження. – 2025. – № 2. – С. 38–52.

2. Recent developments in solar water heaters and solar collectors: A review on experimental and neu-

ral network analyses / R. Naveenkumar, R. Venkateshkumar, V. Mohanavel [et al.] // Results in Engineering. – 2025. – Vol. 25. – Art. 104394.

3. Review of classical and nonlocal nanofluid models for solar collectors / L. Feng, I. W. Turner, V. V. Anh, F. Liu // Renewable and Sustainable Energy Reviews. – 2025. – Vol. 212. – Art. 115382.

AGENT-ORIENTED ARCHITECTURE FOR CITY INFRASTRUCTURE ANALYSIS**Kosovan V.***Assistant lecturer in the Department of Mathematical Modelling, Yuriy Fedkovych Chernivtsi National University Chernivtsi, Ukraine <https://orcid.org/0009-0001-8628-3130>***Melnyk H.***Lecturer in the Department of Applied Mathematics and Informational Technologies, Yuriy Fedkovych Chernivtsi National University Chernivtsi, Ukraine <https://orcid.org/0000-0001-5386-5903>***Melnyk V.***Assistant lecturer in the Department of Mathematical Modelling, Yuriy Fedkovych Chernivtsi National University Chernivtsi, Ukraine <https://orcid.org/0009-0005-2880-6449>***Melnyk O.***Student, Yuriy Fedkovych Chernivtsi National University Chernivtsi, Ukraine <https://doi.org/10.5281/zenodo.21105842>***Abstract**

This paper presents the design and implementation of GuideLoc, an agent-oriented advisory service for city infrastructure analysis. The system enables users to discover urban locations, including restaurants, pharmacies, parks, and other points of interest, through a conversational chat interface. The proposed architecture employs a multi-agent model consisting of an orchestrator agent responsible for intent recognition and delegation, and a specialized location agent responsible for querying and ranking urban objects. The server-side is built on Python and FastAPI, while the mobile client is developed using React Native and TypeScript. Persistent storage is managed via SQLite with SQLAlchemy, and intelligent response generation is powered by the OpenAI API. Real-time response streaming is implemented through Server-Sent Events, and user authentication relies on JSON Web Tokens. The system accounts for contextual factors such as weather conditions, user budget, route planning, and personal preferences when forming recommendations. Experimental evaluation confirms the feasibility of the approach and demonstrates the extensibility of the architecture for incorporating additional domain-specific agents.

Keywords: agent-oriented architecture, city infrastructure, advisory system, multi-agent systems, natural language processing

The rapid growth of urban environments and the increasing complexity of city infrastructure have created a significant demand for intelligent systems capable of assisting residents, tourists, and newcomers in navigating urban spaces. Traditional approaches to information retrieval — relying on maps, review platforms, and search engines in isolation — require considerable effort and do not provide cohesive, context-aware recommendations. A user wishing to find a suitable café within a given budget and reach it on foot in rainy weather would typically need to consult several separate services.

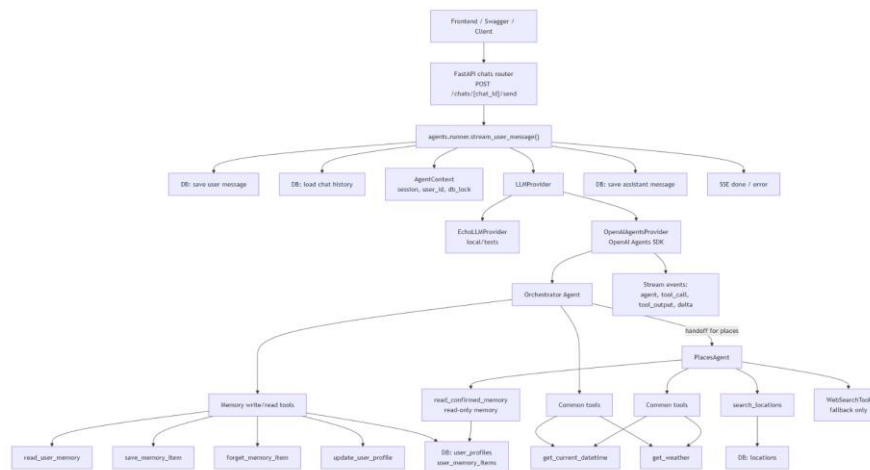
Agent-oriented computing offers a paradigm well suited to this challenge. By decomposing the advisory task into autonomous, cooperating agents, each responsible for a well-defined domain, the system can handle complex, multi-dimensional user requests in a unified conversational interface. This approach has been widely studied in the context of service composition, decision support, and intelligent assistants, and its application to city infrastructure analysis represents a natural and practically relevant extension.

Urban advisory systems have been approached from several directions in prior research. Recommendation systems based on collaborative filtering and content-based methods have demonstrated effectiveness in contexts such as restaurant selection and tourism plan-

ning. However, these approaches typically lack the capacity to handle compound queries involving multiple constraints simultaneously. The integration of large language models into advisory workflows has opened new possibilities, particularly in enabling natural language understanding as a first-class feature of the system rather than a preprocessing step.

Multi-agent systems have been applied to smart city contexts in domains such as traffic management, energy distribution, and emergency response coordination. Their application to citizen-facing advisory services, however, has received comparatively less attention. The design of an orchestrator-agent pattern — in which a central agent interprets intent and delegates subtasks to domain specialists — has precedents in enterprise workflow automation and intelligent virtual assistant research. GuideLoc draws on this pattern and adapts it to the requirements of a mobile, conversational urban advisory service.

The use of large language model APIs as a backend for agent reasoning has become increasingly common following advances in instruction-following models. The OpenAI API, used in this system, enables the orchestrator agent to perform intent classification, slot filling, and response generation in a unified manner, reducing the architectural complexity that would otherwise arise from maintaining separate natural language processing components.



The GuideLoc system is organized into three primary layers: a mobile client, a server-side API layer, and an agent subsystem. The mobile client is implemented in React Native with TypeScript, providing a cross-platform interface for both Android and iOS. The client renders a conversational chat interface, manages user authentication state, and maintains a local history of past sessions accessible through a side menu. Real-time response display is achieved through a streaming protocol that renders agent responses incrementally as they are generated, rather than presenting a complete reply only after processing concludes.

The server-side is built on Python using the FastAPI framework, which supports asynchronous request handling and integrates naturally with the Pydantic library for data validation and schema definition. Each user message received by the server is authenticated via JSON Web Token verification before being routed to the agent subsystem. The server exposes an internal API for the mobile client as well as integration points for the OpenAI API, the Weather API, and the Google Maps API.

Data persistence is managed through SQLite, accessed via the SQLAlchemy object-relational mapper. The database stores user accounts, session histories, individual messages, user preference profiles, and the local catalogue of urban objects. Each urban object record includes the name, category, address, geographic coordinates, operating hours, and a textual description used by the location agent during query matching. Schema evolution is handled through Alembic migration scripts, ensuring that the database structure can be updated incrementally as the system is extended.

The agent subsystem consists of two agents in the current implementation. The orchestrator agent receives the user's message along with the conversation history and is responsible for determining the intent of the request, deciding whether external information such as weather data or route information is required, and delegating the retrieval task to the location agent. The orchestrator communicates with the OpenAI API and is guided by a structured prompting system that defines its role, the available tools, and the expected format of its outputs.

The location agent is responsible for querying the local database of urban objects, applying filters based on category, budget, operating hours, and geographic proximity, and returning a ranked list of candidate locations to the orchestrator. The orchestrator then incorporates the retrieved candidates, any contextual information gathered from external APIs, and the user's stated preferences into a final response that is streamed back to the client.

The prompting architecture of the system is a central design element. Each agent is initialized with a system prompt that describes its responsibilities, the format of inputs it should expect, and the structure of the outputs it should produce. This design allows the behavior of individual agents to be modified or refined without altering the surrounding code, and it provides a natural extension point for adding new agents: a new system prompt paired with a corresponding handler is sufficient to introduce a new domain-specific capability into the orchestrator's delegation logic.

The system was evaluated through functional testing of each major component and end-to-end testing of representative user query scenarios. Test cases included queries for food establishments within a specified budget, pharmacy searches with operating hour constraints, requests for outdoor walking locations accompanied by weather-conditioned clothing advice, and multi-step queries combining location discovery with route generation.

In all tested scenarios, the orchestrator agent correctly identified the primary intent and activated the appropriate combination of downstream services. The location agent successfully filtered the local database and returned ranked results consistent with the stated constraints. Weather-conditioned recommendations correctly reflected current conditions retrieved from the external API. Route generation requests produced valid links to the Google Maps navigation interface.

Response latency was observed to vary primarily with the complexity of the OpenAI API call required to generate the final response text. Simple location queries with no external API dependencies completed within two to three seconds, while compound queries involving weather retrieval and multi-step reasoning extended

to five to eight seconds. The streaming delivery mechanism mitigated the subjective impact of this latency by presenting partial responses to the user as they became available.

The architecture demonstrated a clear separation of concerns between the orchestrator, the location agent, and the external service integrations. This separation was confirmed to support extensibility: a prototype weather-advisory agent focused solely on clothing recommendations was added during development without requiring changes to the orchestrator's core routing logic. The result supports the claim that the agent-oriented design is well suited to incremental expansion of the system's capabilities.

References:

1. Eisenman, M., Felber, P. (2010). *React Native: Building Mobile Apps with JavaScript*. O'Reilly Media.
2. Expo Documentation. <https://docs.expo.dev/>
3. Bierman, G., Abadi, M., Torgersen, M. (2014). *Understanding TypeScript*. In: European Conference on Object-Oriented Programming. Springer, Berlin, pp. 257–281.
4. Van Rossum, G., Drake, F. (2009). *Python 3 Reference Manual*. CreateSpace.
5. Ramírez, S. (2021). *FastAPI Documentation*. <https://fastapi.tiangolo.com/>
6. Hipp, R. D. (2020). *SQLite: A Self-Contained, Serverless, Zero-Configuration, Transactional SQL Database Engine*. <https://sqlite.org/>
7. Bayer, M. (2012). *SQLAlchemy*. In: Brown A., Wilson G. (eds) *The Architecture of Open Source Applications*. Vol. 2.
8. Alembic Documentation. <https://alembic.sqlalchemy.org/>
9. Pydantic Documentation. <https://docs.pydantic.dev/>
10. OpenAI Platform Documentation. <https://platform.openai.com/docs/>
11. Jones, M., Bradley, J., Sakimura, N. (2015). *JSON Web Token (JWT)*. RFC 7519, IETF.
12. Fette, I., Melnikov, A. (2011). *The WebSocket Protocol*. RFC 6455, IETF.
13. Hailuo AI Platform. <https://hailuoai.video/>
14. GitHub Documentation. <https://docs.github.com/>

№183/2026

Norwegian Journal of development of the International Science

ISSN 3453-9875

It was established in November 2016 with support from the Norwegian Academy of Science.

DESCRIPTION

The Norwegian Journal of Development of the International Science, established in 2016 with support from the Norwegian Academy of Science, offers a platform for global researchers, academics, and students. Publishing 24 times annually, it covers current science topics. Submitting your article provides exposure to a diverse audience, enhancing your career. Our commitment to research and SEO ensures timely visibility. Submit to the Norwegian Journal of Development of the International Science today and join esteemed authors.

Editor in chief – Karin Kristiansen (University of Oslo, Norway)

The assistant of the editor in chief – Olof Hansen

- James Smith (University of Birmingham, UK)
- Kristian Nilsen (University Centre in Svalbard, Norway)
- Arne Jensen (Norwegian University of Science and Technology, Norway)
- Sander Svein (University of Tromsø, Norway)
- Lena Meyer (University of Gothenburg, Sweden)
- Hans Rasmussen (University of Southern Denmark, Denmark)
- Chantal Girard (ESC Rennes School of Business, France)
- Ann Claes (University of Groningen, Netherlands)
- Ingrid Karlsen (University of Oslo, Norway)
- Terje Gruterson (Norwegian Institute of Public Health, Norway)
- Sander Langfjord (University Hospital, Norway)
- Fredrik Mardosas (Oslo and Akershus University College, Norway)
- Emil Berger (Ministry of Agriculture and Food, Norway)
- Sofie Olsen (BioFokus, Norway)
- Rolf Ulrich Becker (University of Duisburg-Essen, Germany)
- Lutz Jäncke (University of Zürich, Switzerland)
- Elizabeth Davies (University of Glasgow, UK)
- Chan Jiang (Peking University, China) and other independent experts

1000 copies

Norwegian Journal of development of the International Science

Iduns gate 4A, 0178, Oslo, Norway

email: publish@nor-ijournal.com

site: <https://nor-ijournal.com>



© 2026 The Norwegian Journal of Development of the International Science

nor-ijournal.com

