

ISSN 2518-1726 (Online),  
ISSN 1991-346X (Print)



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ

# Х А Б А Р Л А Р Ы

---

---

**ИЗВЕСТИЯ**

РОО «НАЦИОНАЛЬНОЙ  
АКАДЕМИИ НАУК РЕСПУБЛИКИ  
КАЗАХСТАН»

**N E W S**

OF THE NATIONAL ACADEMY  
OF SCIENCES OF THE REPUBLIC  
OF KAZAKHSTAN

**SERIES OF PHYSICS AND MATHEMATICS**

**1 (353)**

**JANUARY – MARCH 2025**

**PUBLISHED SINCE JANUARY 1963**

**PUBLISHED 4 TIMES A YEAR**

**ALMATY, NAS RK**

#### БАС РЕДАКТОР:

**МҮТАНОВ Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

#### РЕДАКЦИЯ АЛҚАСЫ:

**ҚАЛИМОЛДАЕВ Максат Нұрәділұлы**, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

**МАМЫРБАЕВ Өркен Жұмажанұлы** (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

**БАЙҒҮНЧЕКОВ Жұмаділ Жанабайұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

**ВОЙЧИК Вальдемар**, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

**СМОЛАРЖ Анджей**, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

**КЕЙЛАН Әлімхан**, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

**ХАЙРОВА Нина**, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

**ОТМАН Мохаммед**, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

**НЫСАНБАЕВА Сауле Еркебұланқызы**, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

**БИЯШЕВ Рустам Гакашевич**, техника ғылымдарының докторы, профессор, Информатика және басқару мәселелері институты директорының орынбасары, Ақпараттық қауіпсіздік зертханасының меңгерушісі (Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6603642864>, <https://www.webofscience.com/wos/author/record/3802016>

**КАПАЛОВА Нұрсұлу Алдаржарқызы**, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

**КОВАЛЕВ Александр Михайлович**, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

**МИХАЛЕВИЧ Александр Александрович**, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

**ТИГИНЯНУ Ион Михайлович**, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

«ҚР ҰҒА Хабарлары. Физика-математика сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **28.02.2025** ж. берген №**KZ20VPY00113741** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*

*<http://www.physico-mathematical.kz/index.php/en/>*

© «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ, 2025

## ГЛАВНЫЙ РЕДАКТОР:

**МУТАНОВ Галимканр Мутанович**, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

## Редакционная коллегия:

**КАЛИМОЛДАЕВ Максат Нуралилович**, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

**МАМЫРБАЕВ Оркен Жумажанович**, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

**БАЙГУНЧЕКОВ Жумадил Жанабаевич**, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Саппаева (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

**ВОЙЧИК Вальдемар**, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

**СМОЛАРЖ Анджей**, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

**КЕЙЛАН Алимхан**, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

**ХАЙРОВА Нина**, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

**ОТМАН Мохамед**, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

**НЫСАНБАЕВА Сауле Еркебулановна**, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

**БИЯШЕВ Рустам Гакашевич**, доктор технических наук, профессор, заместитель директора Института проблем информатики и управления, заведующий лабораторией информационной безопасности (Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6603642864>, <https://www.webofscience.com/wos/author/record/3802016>

**КАПАЛОВА Нурсулу Алдажаровна**, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

**КОВАЛЕВ Александр Михайлович**, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

**МИХАЛЕВИЧ Александр Александрович**, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

**ТИГИНЯНУ Ион Михайлович**, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

---

**«Известия НАН РК. Серия физико-математическая».**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № **KZ20VPU00113741**. Дата выдачи **28.02.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКРНВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*  
<http://www.physico-mathematical.kz/index.php/en/>

© РОО «Национальная академия наук Республики Казахстан», 2025

#### CHIEF EDITOR:

**MUTANOV Galimkair Mutanovich**, doctor of technical sciences, professor, academician of NAS RK, acting General Director of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

#### EDITORIAL BOARD:

**KALIMOLDAYEV Maksat Nuradilovich**, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

**Mamyrbayev Orken Zhumazhanovich**, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

**BAIGUNCHEKOV Zhumadil Zhanabaevich**, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

**WOJCIK Waldemar**, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

**SMOLARJ Andrej**, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

**KEILAN Alimkhan**, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

**KHAIROVA Nina**, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

**OTMAN Mohamed**, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

**NYSANBAYEVA Saule Yerkebulanovna**, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

**BIYASHEV Rustam Gakashevich**, doctor of technical sciences, professor, Deputy Director of the Institute for Informatics and Management Problems, Head of the Information Security Laboratory (Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6603642864>, <https://www.webofscience.com/wos/author/record/3802016>

**KAPALOVA Nursulu Aldazharovna**, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

**KOVALYOV Alexander Mikhailovich**, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

**MIKHALEVICH Alexander Alexandrovich**, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

**TIGHINEANU Ion Mihailovich**, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

---

#### News of the National Academy of Sciences of the Republic of Kazakhstan.

##### Series of Physics and Mathematics

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty).

Certificate No. **KZ20VPY00113741** on the re-registration of the periodical printed and online publication of the information agency, issued on **28.02.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan

Subject area: *information and communication technologies.*

Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*

Periodicity: *4 times a year.*

Editorial address: *28, Shevchenko str., of 219, Almaty, 050010, tel. 272-13-19*

<http://www.physico-mathematical.kz/index.php/en/>

<https://doi.org/10.32014/2025.2518-1726.337>

MPHTИ: 28.23.37

УДК: 004.89

**D. Oralbekova<sup>1\*</sup>, O. Mamyrbayev<sup>1</sup>, A. Akhmediyarova<sup>2</sup>, D. Kassymova<sup>3</sup>, 2025.**

<sup>1</sup> Institute of Information and Computational Technologies, Almaty, Kazakhstan;

<sup>2</sup> Satbayev University, Almaty, Kazakhstan;

<sup>3</sup> ALT University, Almaty, Kazakhstan.

E-mail: dinaoral@mail.ru

## **USING KAZAKH NER DATASETS FOR MULTICLASS CLASSIFICATION IN THE LEGAL DOMAIN: A COMPARATIVE STUDY OF BERT, GPT, AND LSTM MODELS**

**Oralbekova Dina** – PhD, senior researcher, associate professor, Institute of information and computational technologies, Almaty, Kazakhstan, E-mail: dinaoral@mail.ru, <https://orcid.org/0000-0003-4975-6493>;

**Mamyrbayev Orken** – PhD, professor, deputy general director, Institute of information and computational technologies, Almaty, Kazakhstan, E-mail: morkenj@mail.ru, <https://orcid.org/0000-0001-8318-3794>;

**Akhmediyarova Ainur** – PhD, professor, Satbayev University, Almaty, Kazakhstan, E-mail: a.akhmediyarova@satbayev.university, <https://orcid.org/0000-0003-4439-7313>;

**Kassymova Dinara** – PhD, assistant professor, ALT University, Almaty, Kazakhstan, E-mail: d.kassymova@alt.edu.kz, <https://orcid.org/0000-0001-6152-8317>.

**Abstract.** This study presents an in-depth comparative analysis of the performance of three key approaches in natural language processing (NLP) — transformers, recurrent neural networks, and traditional machine learning methods — in the task of multiclass text classification in the legal domain of the Kazakh language. A specialized dataset for named entity recognition, adapted to legal topics, was used for the analysis. The primary focus is on the classification of texts reflecting legal aspects and terminology. Evaluation metrics such as accuracy, recall, precision, and area under the curve (AUC) were applied, enabling an objective comparison of their effectiveness. Special attention was given to adapting methods for the Kazakh language, which is under-researched in computational linguistics. This underscores the importance of developing specialized algorithms that ensure effective processing of texts in this language, considering its agglutinative structure and complex legal terminology. The study not only highlights the strengths and weaknesses of modern NLP models in processing texts in low-resource languages but also reveals new opportunities for integrating these technologies into applied

fields. These include automating legal document management, analyzing court decisions, and developing intelligent decision-support systems for the legal sector. The findings also emphasize the critical importance of creating specialized datasets for training and testing models. This research expands the understanding of the capabilities of existing NLP models for processing texts in the Kazakh language and emphasizes the significance of further advancing technologies for automating legal services. The presented results can serve as a foundation for the development of accessible and scalable tools that enhance the efficient processing of legal information and contribute to the growth of digital jurisprudence.

**Keywords:** NLP, LSTM, BERT, GPT, text classification.

**Д. Оралбекова<sup>1\*</sup>, О. Мамырбаев<sup>1</sup>, А. Ахмедиярова<sup>2</sup>, Д. Қасымова<sup>3</sup>, 2025.**

<sup>1</sup> Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан;

<sup>2</sup> Satbayev университеті, Алматы, Қазақстан;

<sup>3</sup> М. Тынышпаев атындағы АЛТ университеті, Алматы, Қазақстан.

E-mail: dinaoral@mail.ru

## **ҚАЗАҚ ТІЛІНДЕГІ NER ДЕРЕКТЕР ЖИНАҒЫН ҚҰҚЫҚТЫҚ САЛАДА КӨПСАНАТТЫ ЖІКТЕУ ҮШІН ПАЙДАЛАНУ: BERT, GPT ЖӘНЕ LSTM МОДЕЛЬДЕРІНІҢ САЛЫСТЫРМАЛЫ ЗЕРТТЕУІ**

**Оралбекова Дина** – PhD, аға ғылыми қызметкер, Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан, E-mail: dinaoral@mail.ru, <https://orcid.org/0000-0003-4975-6493>;

**Мамырбаев Өркен** – PhD, профессор, бас директордың орынбасары, Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан, E-mail: morkenj@mail.ru, <https://orcid.org/0000-0001-8318-3794>;

**Ахмедиярова Айнұр** – PhD, профессор, Satbayev университеті, Алматы, Қазақстан, E-mail: a.akhmediyarova@satbayev.university, <https://orcid.org/0000-0003-4439-7313>;

**Қасымова Динара** – PhD, ассистент-профессор, М. Тынышпаев атындағы АЛТ университеті, Алматы, Қазақстан, E-mail: d.kassymova@alt.edu.kz, <https://orcid.org/0000-0001-6152-8317>.

**Аннотация.** Бұл зерттеуде табиғи тілдерді өңдеу (NLP) саласындағы үш негізгі тәсілдің – трансформерлердің, рекурренттік нейрондық желілердің және дәстүрлі машиналық оқыту әдістерінің – қазақ тіліндегі құқықтық саладағы мәтіндерді көпсанатты жіктеу міндеттерінде өнімділігіне терең салыстырмалы талдау ұсынылады. Талдау үшін заң тақырыбына бейімделген, атаулы объектілерді тануға арналған арнайы деректер жинағы пайдаланылды. Негізгі назар құқықтық аспектілер мен терминологияны қамтитын мәтіндерді жіктеуге аударылды. Модельдерді бағалау үшін дәлдік (accuracy), толықтық (recall), басқа дәлдік көрсеткіші (precision) және қисық астындағы аумақ (AUC) сияқты метрикалар қолданылды, бұл олардың тиімділігін объективті түрде салыстыруға мүмкіндік берді. Қазақ тіліне арналған әдістерді бейімдеуге ерекше көңіл бөлінді, өйткені бұл тіл компьютерлік лингвистикада жеткіліксіз зерттелген. Бұл қазақ тілінің агглютинативтік құрылымы мен

күрделі құқықтық терминологиясын ескере отырып, мәтіндерді өңдеуге арналған арнайы алгоритмдерді әзірлеу қажеттілігін көрсетеді. Зерттеу тек ресурстары шектеулі тілдердегі мәтіндерді өңдеуге арналған заманауи NLP модельдерінің күшті және әлсіз жақтарын көрсетіп қана қоймай, сонымен қатар осы технологияларды қолданбалы салаларға интеграциялаудың жаңа мүмкіндіктерін ашады. Оларға құқықтық құжат айналымын автоматтандыру, сот шешімдерін талдау, сондай-ақ құқықтық салаға арналған интеллектуалды шешім қабылдауды қолдау жүйелерін әзірлеу жатады. Зерттеу нәтижелері модельдерді оқыту және тестілеу үшін арнайы деректер жиынтықтарын әзірлеудің маңыздылығын растайды. Бұл зерттеу қолданыстағы NLP модельдерінің қазақ тіліндегі мәтіндерді өңдеудегі мүмкіндіктерін түсінуді кеңейтеді және құқықтық қызметтерді автоматтандыру технологияларын одан әрі дамытудың маңыздылығын көрсетеді. Ұсынылған нәтижелер құқықтық ақпаратты тиімді өңдеуді арттыратын және цифрлық құқықтанудың дамуына ықпал ететін қолжетімді әрі ауқымды құралдарды әзірлеуге негіз бола алады.

**Түйін сөздер:** NLP, LSTM, BERT, GPT, мәтінді классификациялау.

**Д. Оралбекова<sup>1\*</sup>, О. Мамырбаев<sup>1</sup>, А. Ахмедиярова<sup>2</sup>, Д. Касымова<sup>3</sup>, 2025.**

<sup>1</sup>Институт информационных и вычислительных технологий,  
Алматы, Казахстан;

<sup>2</sup>Satbayev University, Алматы, Казахстан;

<sup>3</sup>ALT университет имени М. Тынышпаева, Алматы, Казахстан.

E-mail: dinaoral@mail.ru

## **ИСПОЛЬЗОВАНИЕ НАБОРОВ ДАННЫХ NER НА КАЗАХСКОМ ЯЗЫКЕ ДЛЯ МУЛЬТИКЛАССИФИКАЦИИ В ПРАВОВОЙ СФЕРЕ: СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МОДЕЛЕЙ BERT, GPT И LSTM**

**Оралбекова Дина** – PhD, старший научный сотрудник, Институт информационных и вычислительных технологий, Алматы, Казахстан, E-mail: dinaoral@mail.ru, <https://orcid.org/0000-0003-4975-6493>;

**Мамырбаев Оркен** – PhD, профессор, заместитель генерального директора, Институт информационных и вычислительных технологий, Алматы, Казахстан, E-mail: morkenj@mail.ru, <https://orcid.org/0000-0001-8318-3794>;

**Ахмедиярова Айнур** – PhD, профессор, Satbayev Университет, Алматы, Казахстан, E-mail: a.akhmediyarova@satbayev.university, <https://orcid.org/0000-0003-4439-7313>;

**Касымова Динара** – PhD, ассистент-профессор, ALT университет имени М. Тынышпаева, Алматы, Казахстан, E-mail: d.kassymova@alt.edu.kz, <https://orcid.org/0000-0001-6152-8317>.

**Аннотация.** В данном исследовании представлен углубленный сравнительный анализ производительности трех ключевых подходов в области обработки естественного языка (NLP) — трансформеров, рекуррентных нейронных сетей и традиционных методов машинного обучения — в задачах мультиклассификации текстов в правовой сфере на казахском языке. Для анализа был использован специализированный набор данных для

распознавания именованных сущностей, адаптированный под юридическую тематику. Основное внимание уделено классификации текстов, отражающих юридические аспекты и терминологию. Для оценки моделей применены метрики точности (accuracy), полноты (recall), прецизионности (precision) и площади под кривой (AUC), что позволило провести объективное сравнение их эффективности. Особое внимание уделено адаптации методов для казахского языка, который является недостаточно исследованным в компьютерной лингвистике. Это подчеркивает важность разработки специализированных алгоритмов, обеспечивающих качественную обработку текстов на данном языке с учетом его агглютинативной структуры и сложной юридической терминологии. Исследование не только демонстрирует сильные и слабые стороны современных NLP-моделей в задачах обработки текстов на языках с ограниченными ресурсами, но и открывает новые перспективы для интеграции этих технологий в прикладные области. Среди них можно выделить автоматизацию юридического документооборота, анализ судебных решений, а также создание интеллектуальных систем поддержки принятия решений для правовой сферы. Результаты работы также подтверждают важность создания специализированных данных для обучения и тестирования моделей. Проведенное исследование расширяет понимание возможностей существующих моделей NLP для работы с текстами на казахском языке и подчеркивает значимость дальнейшего развития технологий автоматизации юридических услуг. Представленные результаты могут стать основой для создания доступных и масштабируемых инструментов, способствующих эффективной обработке правовой информации и развитию цифровой юриспруденции.

**Ключевые слова:** NLP, LSTM, BERT, GPT, классификация текстов.

*Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан (Грант BR24993166).*

**Введение.** Обработка естественного языка (NLP) охватывает широкий спектр вычислительных методов, предназначенных для интерпретации, генерации и анализа человеческого языка. Одной из ключевых задач в области NLP является классификация текста, которая находит применение в различных сферах, начиная от фильтрации спама до автоматизации работы служб поддержки (Narejo, et al., 2024). Однако сложность этой задачи многократно возрастает для языков с ограниченными ресурсами, таких как казахский язык, который не обладает достаточной вычислительной базой и объемом данных для проведения передовых исследований в области NLP.

Казахский язык, относящийся к тюркской языковой семье и являющийся официальным в Казахстане, привлекает все больше внимания в сообществе NLP, но его ресурсы для исследования и разработки остаются ограниченными по сравнению с более широко изучаемыми языками, такими как английский



(Mamyrbayev, et al., 2020; Oralbekova, et al., 2022). За последние годы был достигнут определенный прогресс, включая создание базовых словарей и морфологических анализаторов (Myrzakhmetov, et al., 2018; Abibullayeva, et al., 2022; Yeshpanov, et al., 2022). Однако дефицит аннотированных данных остается значительной проблемой, особенно в задачах классификации текста, которые требуют крупных и разнообразных наборов данных для эффективного обучения и тестирования моделей.

В настоящем исследовании рассматривается возможность использования существующего набора данных распознавания именованных объектов (NER) для решения задач мультиклассификации в контексте правовой сферы. Это исследование направлено на расширение использования данного набора данных за пределы его первоначальной области применения, адаптируя его для юридической классификации текстов. Юридические тексты на казахском языке обладают специфическими морфологическими и синтаксическими особенностями, которые делают задачу классификации нетривиальной. Стандартные модели NLP, разработанные для языков с другими структурами, не всегда способны адекватно обработать казахский текст, особенно в юридическом контексте.

Для решения этой задачи в нашем исследовании используются три различных модели: трансформеры, рекуррентные нейронные сети. Каждая из этих моделей имеет уникальные преимущества для задач классификации текстов. Рекуррентные нейронные сети (RNN), такие как LSTM, особенно полезны при работе с последовательными данными, что имеет важное значение для казахского языка, который является агглютинативным и имеет сложную структуру словоформ. Трансформеры, такие как BERT (Garrido-Merchan, et al., 2023; Jamshidi, et al., 2024), предоставляют глубокое контекстное понимание текста, что позволяет лучше учитывать юридические нюансы и специфические термины. Модели генеративного типа, такие как GPT, демонстрируют эффективность в задачах, связанных с языковой генерацией, и могут быть полезны для анализа и классификации правовых документов с неструктурированным текстом (Pandey, et al., 2024; Li, et al., 2024).

Мы проводим сравнительный анализ этих моделей для задач мультиклассификации казахских текстов, применяя метрики, такие как точность, полнота, F1-мера и AUC (Reusens, et al., 2024; Bhowmik, et al., 2024). Этот всесторонний подход позволяет не только выявить сильные и слабые стороны каждой модели, но и оценить их потенциал для применения в реальных задачах, таких как автоматизация юридической помощи, анализ судебных документов и классификация запросов в правовой сфере.

Таким образом, наше исследование подчеркивает важность разработки NLP-инструментов для казахского языка, способствуя прогрессу в области цифровой юриспруденции и обеспечивая новый уровень автоматизации юридических услуг на языках с ограниченными ресурсами.

**Материалы и методы исследования.** Ландшафт NLP для казахского языка постепенно обогащался благодаря различным целенаправленным исследованиям. Одной из основополагающих работ является исследование экспериментов по расширенному языковому моделированию казахского языка (Murzakhmetov, et al., 2018), которое заложило основу для понимания сложностей, связанных с созданием вычислительных ресурсов для недостаточно представленных языков. Это исследование подчеркивает первоначальные шаги по адаптации передовых методов NLP к казахскому языку, подчеркивая потенциал сложных языковых моделей для улучшения обработки и понимания текста.

Основываясь на этих основополагающих моделях, недавние достижения продемонстрировали возможность адаптации современных технологий, таких как BERT, к казахскому языку. В частности, исследование извлечения ключевых слов из наборов данных новостей на казахском языке с использованием BERT (Abibullayeva, et al., 2022) показало, как модели на основе Transformer могут эффективно использоваться для задач, требующих глубокого семантического понимания, таких как определение ключевых терминов и фраз, которые отражают суть новостных статей. Это приложение не только демонстрирует гибкость BERT, но и его потенциал для улучшения поиска информации и анализа контента на казахском языке.

Дальнейшее расширение сферы применения NLP на казахском языке, классификация научных документов с использованием глубоких нейронных сетей и слияние изображений и текста (Bogdanchikov, et al., 2022) иллюстрирует инновационные подходы, изучаемые в сообществе. В этой работе особое внимание уделяется интеграции мультимодальных данных для целей классификации, предлагая новые пути повышения производительности модели и ее применимости при обработке текстов на казахском языке.

Центральное место в текущем исследовании занимает набор данных KazNERD (Yeshpanov, et al., 2022), который является значительным вкладом в эту область и предоставляет богато аннотированный корпус для распознавания именованных объектов (NER) на казахском языке. Создание и использование этого набора данных знаменуют собой важнейший прогресс в исследованиях NLP для нашей страны, предлагая ценный ресурс для обучения и оценки моделей NLP для задач распознавания сущностей. Этот набор данных служит основой для нашей работы, позволяя исследовать проблемы мультиклассификации через призму NER.

Вместе эти исследования образуют связное повествование, которое подчеркивает развивающиеся возможности и применение технологий NLP для казахского языка. От базового языкового моделирования до практического применения моделей на основе Transformer и не только — в этой области наблюдаются значительные успехи в адаптации и применении передовых вычислительных методов для решения уникальных задач, связанных с казахским языком. Наша работа направлена на то, чтобы внести свой вклад

в этот растущий объем знаний, уделяя особое внимание практической реализации моделей LSTM и BERT для задач мультиклассификации в рамках разработанной структуры NER.

#### *Метрики оценки*

Прежде чем мы начнем практически анализировать модели, важно понять фундаментальные концепции и архитектуры, лежащие в основе современного ландшафта моделей NLP. Задачи NLP, начиная от классификации текста и заканчивая анализом настроений и ответами на вопросы, требуют сложных моделей, способных понимать и генерировать человеческий язык. Для достижения этой цели были разработаны различные модели глубокого обучения, каждая из которых имеет свой уникальный подход к обработке и интерпретации текста.

Эволюция моделей NLP началась с более простых нейронных сетей с прямой связью и продолжилась через рекуррентные нейронные сети (RNN), включая сети долгосрочной краткосрочной памяти (LSTM), чтобы удовлетворить потребность в фиксации временных зависимостей в текстовых последовательностях. Однако эти модели часто имели проблемы с долгосрочной зависимостью и вычислительной эффективностью.

Внедрение архитектуры Transformer ознаменовало значительный сдвиг, приведший к разработке таких моделей, как BERT (Bidirectional Encoder Representations from Transformers) и GPT (Generative Pre-trained Transformer). Эти модели используют механизмы внимания для обработки текста таким образом, чтобы более эффективно и результативно улавливать контекст, чем их предшественники.

Понимание принципов построения, сильных сторон и ограничений этих моделей имеет решающее значение для их эффективного применения в задачах NLP. Используя свои преимущества, исследователи и практики могут решать широкий спектр задач NLP с беспрецедентной точностью и эффективностью.

#### *LSTM (Long Short-Term Memory)*

LSTM — это тип рекуррентной нейронной сети (RNN), способной изучать долгосрочные зависимости. Они созданы для того, чтобы избежать проблемы долгосрочной зависимости за счет включения ряда шлюзов, контролирующих поток информации (Minaee, et al., 2022). Эти модели оказались особенно эффективными для решения задач прогнозирования последовательностей и показали успех в таких задачах, как генерация текста, распознавание речи и языковой перевод. Несмотря на свою эффективность в фиксации временных зависимостей, LSTM могут бороться с очень длинными последовательностями и могут уступать моделям на основе Transformer, таким как BERT, для задач, требующих понимания более широкого контекста (Ezen-Can, et al., 2020).

#### *BERT (Bidirectional Encoder Representations from Transformers)*

BERT отличается своей глубокой двунаправленной природой, обрабатывая текст как слева направо, так и справа налево, что позволяет ему лучше понимать контекст. Это делает BERT особенно эффективным для задач, требующих

глубокого понимания языкового контекста, таких как анализ настроений, ответы на вопросы и распознавание именованных объектов. Его архитектура преобразователя позволяет ему лучше справляться с долгосрочными зависимостями по сравнению с традиционными моделями, такими как LSTM. Производительность BERT заметно повышается при точной настройке для конкретной задачи даже с относительно небольшими наборами данных (Tikaayat Ray, et al., 2023).

#### *GPT (Generative Pre-trained Transformer)*

Модели GPT, благодаря своей способности генерировать связный и контекстуально релевантный текст, превосходно справляются с задачами, требующими языковой генерации, такими как завершение текста, творческое письмо и диалоговые системы. Авторегрессионный характер GPT, обрабатывающий текст слева направо, позволяет ему прогнозировать вероятность последовательности слов и генерировать текст на основе вероятностей последующих слов. Архитектура преобразователя GPT позволяет эффективно изучать зависимости на больших расстояниях в тексте. Однако, в отличие от BERT, односторонний характер GPT может ограничивать понимание контекста по сравнению с двусторонним подходом BERT (Narejo, et al., 2024).

Таким образом, выбор между BERT, LSTM и GPT для задач классификации NLP зависит от конкретных требований поставленной задачи, включая характер набора данных, тип задачи NLP и доступные вычислительные ресурсы. BERT очень эффективен для задач, требующих глубокого понимания языкового контекста, LSTM превосходит в изучении долгосрочных зависимостей для прогнозирования последовательности, а GPT лучше всего подходит для задач, требующих последовательной и контекстно-зависимой генерации языка.

Оценка производительности модели является важнейшим компонентом любой задачи машинного обучения. Именно с помощью различных показателей мы оцениваем эффективность наших моделей, каждая из которых предлагает свой взгляд на возможности модели. Метрики — это призма, через которую мы рассматриваем и понимаем результаты, и выбор правильных из них имеет решающее значение для справедливой и информативной оценки. В этом исследовании мы используем набор хорошо зарекомендовавших себя показателей, каждый из которых выбран из-за его способности отражать определенный аспект производительности модели.

В нашем сравнительном исследовании мы используем три модели NLP — LSTM, BERT и GPT, каждая из которых предлагает различные возможности обработки и классификации текста. Чтобы точно оценить их эффективность, мы используем ряд показателей, каждый из которых касается конкретных аспектов поведения модели.

Точность — это отправная точка для оценки, которая дает нам широкую оценку производительности каждой модели. Однако точность не дает полной картины, особенно когда речь идет о дисбалансе классов или различных

издержках, связанных с различными типами ошибок. Чтобы копнуть глубже, мы используем точность и отзыв. Точность имеет решающее значение, когда последствия ложных срабатываний значительны: мы должны быть уверены, что положительные результаты, выявленные нашими моделями, являются истинными положительными.

Кривая AUC-ROC делает еще один шаг вперед, позволяя понять, насколько хорошо наши модели различают классы на всех возможных пороговых уровнях. Это особенно полезно для точной настройки моделей таким образом, чтобы они соответствовали конкретным оперативным целям или процессам принятия решений.

При сравнительном анализе моделей важно учитывать совокупное значение этих показателей, чтобы получить целостную картину производительности. Использование набора метрик в тандеме обеспечивает комплексное представление о производительности модели, особенно в контексте классификации по нескольким меткам. Каждая метрика предлагает уникальный взгляд на конкретный аспект прогнозных возможностей модели, и вместе они рисуют полную картину ее сильных и слабых сторон.

Когда эти показатели используются вместе, они предлагают надежную структуру оценки, которая учитывает как производительность на макроуровне (точность, AUC-ROC), так и динамику на микроуровне (точность и полнота для отдельных классов). Этот мультиметрический подход жизненно важен для задач мультиклассификации, поскольку:

- Это снижает риск чрезмерной зависимости от одной метрики, которая может не отражать все аспекты производительности.
- Он выделяет конкретные области для улучшения модели, например, какие классы нуждаются в большей точности или полноте, что дает информацию для уточнения модели и разработки функций.
- Это позволяет сделать сбалансированный выбор модели, не отдавая предпочтение какому-либо конкретному классу или типу ошибок, что может иметь решающее значение в приложениях с различными и значительными последствиями для разных типов ошибок.

Тщательно анализируя все эти показатели, исследователи и практики могут принимать более обоснованные решения о развертывании моделей, понимать связанные с этим компромиссы и направлять разработку моделей, которые лучше подходят для сложных реальных приложений NLP.

#### *Подготовка данных*

Для этапа подготовки данных в нашем проекте мы предприняли комплексный процесс по преобразованию набора данных распознавания именованных объектов (NER) на казахском языке в формат, подходящий для нашей схемы бинарной классификации, что в конечном итоге упростило задачу мультиклассификации в контексте правовой сферы и цифровой юриспруденции. Этот процесс включал несколько ключевых шагов, каждый из которых был критически важен для обеспечения целостности и полезности нашего набора

данных для обучения моделей, связанных с юридическими услугами и правовой помощью. В данном разделе мы подробно опишем создание четырех бинарных меток классов и последующий процесс мультиклассификации, акцентируя внимание на его значимости и методологии.

Учитывая подробные описания именованных объектов (NE) в правовой сфере, мы разработали схему бинарной классификации, которая отражает аспекты данных и разделяет их на две значимые категории. С учетом природы NE и контекста их использования, логическая бинарная классификация могла бы различать предложения, которые включают:

1. *Классификация правовых документов:*

Включает именованные объекты, такие как 'LAW', 'DOCUMENT', 'CASE', которые часто встречаются в юридических документах, нормативных актах и судебных разбирательствах.

*Пример использования:* Юридические фирмы или платформы для анализа правовых документов могут использовать данную классификацию для автоматической сортировки и анализа контрактов, судебных решений и других юридических текстов.

2. *Классификация юридической помощи:*

Содержит NE, такие как 'LAWYER', 'LEGAL\_SERVICE', 'CONTACT', указывающие на взаимодействие между адвокатами, клиентами и юридическими учреждениями.

*Пример использования:* Системы юридической поддержки могут автоматически определять сообщения, касающиеся запросов на юридическую помощь, и перенаправлять их соответствующим специалистам.

3. *Классификация судебных разбирательств:*

Включает предложения с NE, такими как 'COURT', 'JUDGE', 'CASE\_NUMBER', которые указывают на судебные процессы или упоминания дел, находящихся в стадии рассмотрения.

*Пример использования:* Платформы мониторинга судебных процессов могут классифицировать и отслеживать информацию по активным делам, обеспечивая своевременное уведомление клиентов или адвокатов.

4. *Классификация коммерческих правовых отношений:*

Содержит NE, такие как 'CONTRACT', 'PARTNERSHIP', 'MONEY', 'ORGANISATION', что указывает на правовые аспекты коммерческих отношений, включая договоры, сделки и финансовые обязательства.

*Пример использования:* Юридические отделы компаний могут использовать данную классификацию для отслеживания договоров и других коммерческих документов, которые требуют юридического рассмотрения.

В рамках данного исследования категории были выбраны вручную для демонстрации функциональности метода. Однако в дальнейшем планируется автоматизация процесса создания классификационных меток с использованием более сложных алгоритмов, учитывающих больший спектр именованных объектов (NE) и их контекст в предложении. Это усовершенствование

позволит не только упростить процесс классификации, но и повысить точность анализа правовых текстов на казахском языке, что будет особенно полезно для цифровой юриспруденции и предоставления юридических услуг.

Важным шагом в подготовке нашего набора данных была токенизация, при которой преобразовывали необработанные текстовые последовательности в структурированный формат, который могли обрабатывать наши модели. Используя BertTokenizer и GPT2Tokenizer, мы гарантировали совместимость нашего набора данных с современными моделями Transformer. Для модели LSTM был использован специальный процесс токенизации, включающий создание уникального сопоставления Token2Index. Этот процесс сыграл важную роль в преобразовании текстов на казахском языке в последовательности целых чисел, каждое из которых представляет собой токен или слово в нашем словаре, который превышает 72 000 уникальных слов. Этот всеобъемлющий словарный запас имел решающее значение, поскольку он отражал лингвистическое разнообразие и сложность казахского языка, закладывая прочную основу для точного обучения и оценки моделей.

**Результаты и обсуждение.** В рамках данного исследования был проведен эмпирический анализ с использованием трех широко известных архитектур для обработки естественного языка: LSTM, BERT и GPT. Целью исследования было оценить их эффективность в задачах мультиклассификации текста на казахском языке, с особым акцентом на юридическую сферу. Анализ был направлен на проверку возможностей моделей в классификации текстов, связанных с правовыми документами, судебными разбирательствами и юридической поддержкой. Наш набор данных, наполненный многомерными атрибутами, стал испытательной площадкой для этих моделей, предоставив богатый набор лингвистических особенностей для оценки.

#### *Функция потерь (Loss) и точность (Accuracy) моделей*

Мы начали нашу оценку с отказа от недоумения — меры, традиционно связанной с языковым моделированием — и вместо этого сосредоточили наш анализ на потерях и точности. Эти метрики являются наиболее важными и дают прямое представление об эффективности модели для задач классификации. Потери, будь то двоичная или категориальная кросс-энтропия, разграничивают расхождение прогнозируемых вероятностей от фактических меток, проливая свет на достоверность модели — или ее отсутствие. Точность в своей простой элегантности измеряет долю правильно классифицированных экземпляров, что отражает практическую эффективность модели.

Демонстрируя классические признаки RNN, путь LSTM начался с резкого снижения потерь в обучении, которые в конечном итоге вышли на плато — признак стабильности обучения (рис. 1). Тем не менее, его эффективность проверки колебалась, что является свидетельством борьбы с обобщениями и индикатором потенциального переобучения. Эта версия была дополнительно подтверждена оценками точности и полноты, которые, хотя и похвальны, но предполагали возможности для улучшения обобщения модели.

Благодаря своей базовой архитектуре Transformer путь BERT ознаменовался резким и неуклонным снижением потерь как на этапах обучения, так и на этапах проверки. Он продемонстрировал замечательную эффективность обучения, что отражено в высоких показателях точности, воспроизводимости и AUC. Мастерство BERT в понимании контекста, подкрепленное механизмами самообслуживания, проявилось в превосходной обработке классификации текста, укрепляя его надежность при анализе текстовых нюансов в нашем наборе данных.

Авторегрессионная модель, известная своими возможностями генерации языков, столкнулась с проблемами при переориентации на классификацию. Несмотря на первоначальное улучшение показателей обучения, его производительность стабилизировалась, что указывает на потенциальное несоответствие между склонностью модели и спецификой задачи. Точность GPT, хотя и улучшалась с течением времени, оставалась ниже звездной производительности BERT, что соответствует внутренним различиям в конструкции между генеративными и дискриминативными моделями.

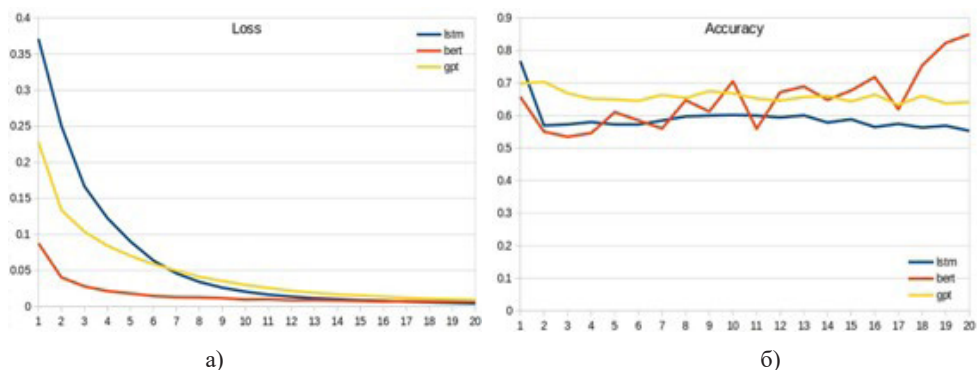


Рисунок 1 – Показатели трех моделей по а) loss и б) метрике accuracy

### Площадь под кривой (AUC)

Площадь под кривой (AUC) предоставила уникальное представление о производительности модели при различных пороговых значениях. Он продемонстрировал надежную различительную способность каждой модели. Оценки AUC BERT, подобные парящему орлу, подчеркнули его превосходное мастерство классификации при различных порогах принятия решения, что еще больше подтвердило его заслуги в нашей оценке (рис. 2).



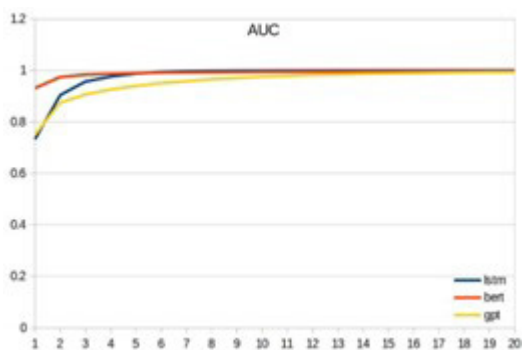


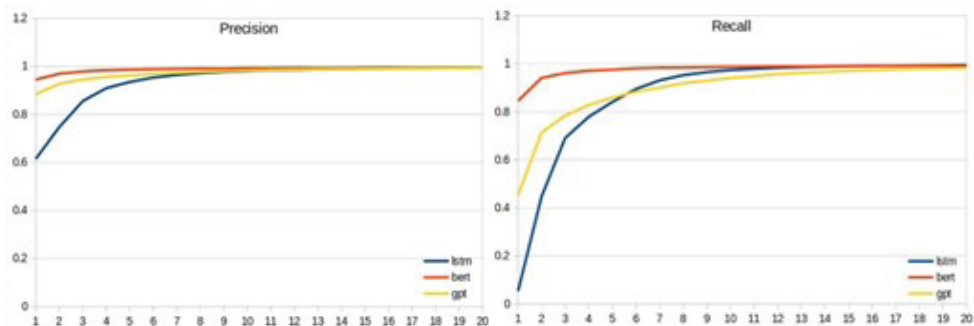
Рисунок 2 – Показатель AUC для трех моделей

### *Измерение точности (Precision) и полноты (Recall)*

По мере того, как мы копали глубже, точность и полнота стали дуэтом, отражающим баланс между специфичностью и чувствительностью модели. Precision (точность измерения) показывала, насколько хорошо каждая модель определяет соответствующие случаи. Recall - измерение полноты — отражало способность каждой модели идентифицировать все соответствующие экземпляры. Превосходство BERT было очевидно в обоих случаях, подчеркивая его тонкую дискриминацию соответствующих фрагментов текста среди моря данных (рис. 3).

Общей нитью среди моделей было расхождение между производительностью обучения и проверки. Этот раскол отражает вечное противоречие между усвоенными моделями модели и ее способностью к обобщению. Это баланс между запоминанием и адаптируемостью — ключ к развертыванию модели в реальных сценариях, где предсказуемость данных так же капризна, как ветер.

Все три модели (BERT, GPT, LSTM) работали над одной и той же задачей: классификацией по нескольким меткам на основе типов объектов во входном тексте. Основное различие заключается в том, как они обрабатывают входные данные и учатся на их основе, учитывая их архитектурные различия. BERT и GPT используют архитектуры преобразователей, способные двунаправленно понимать контекст, в то время как LSTM обрабатывает данные последовательно, что может повлиять на то, как они интерпретируют значение и отношения слов в предложении.



а) б)  
Рисунок 3 – Показатели трех моделей по метрикам а) precision и б) recall

Благодаря тщательной настройке модели и детальному процессу оценки наше исследование позволило получить значимые результаты. Архитектура BERT, оптимизированная для распознавания контекстуальных связей, демонстрирует превосходство над LSTM в стабильности и производительности. Модель GPT, хотя и не достигла высокой точности классификации, подчеркивает потенциал использования генеративных подходов в задачах дискриминации.

Таблица 1. Полученные результаты на основе тренировочных данных

Training Performance			
Metric / Model	LSTM	BERT	GPT
Loss	0.2093	0.0477	0.1073
Accuracy	0.4674	0.9255	0.5028
Precision	0.8453	0.9682	0.9414
Recall	0.8423	0.9631	0.9070
Auc	0.9495	0.9806	0.9515

Таблица 2. Полученные результаты на основе валидационных данных

Validation Performance			
Metric / Model	LSTM	BERT	GPT
Loss	0.1993	<b>0.0481</b>	0.1156
Accuracy	0.4681	<b>0.9202</b>	0.5111
Precision	0.8524	<b>0.9712</b>	0.9369
Recall	0.8483	<b>0.9585</b>	0.9105
Auc	0.9557	<b>0.9781</b>	0.9540

Сравнительный анализ показывает, что LSTM, благодаря своей способности эффективно обрабатывать последовательные данные, сохраняет актуальность в задачах анализа текста. Результаты BERT, основанные на преимуществах параллельной обработки с использованием архитектуры Transformer, подтверждают её эффективность в моделировании сложных языковых зависимостей. Наблюдения за работой GPT указывают на текущие ограничения

генеративных моделей в классификационных задачах, что расширяет наше понимание их применения в обработке естественного языка.

Модель LSTM показала заметное сокращение потерь при обучении, что свидетельствует об обучении, но показатели проверки выявили волатильность, предполагающую склонность к переобучению. Это отражается на точности и полноте модели, которые, несмотря на высокий уровень, не достигли оптимального баланса, о чем свидетельствуют другие модели. Показатели AUC для обучения и проверки LSTM были высокими, однако они оставались неизменно ниже показателей, достигнутых BERT, что сигнализировало о несоответствии в согласованности классификации при различных пороговых настройках.

Производительность BERT была отмечена быстрым и устойчивым снижением потерь при обучении, что согласуется с его архитектурным преимуществом в эффективном использовании контекста посредством двунаправленной обработки. Потеря валидации отражала профиль обучения, предполагая превосходный перенос обучения на невидимые данные. Точность, воспроизводимость, полнота и AUC модели были одинаково высокими, что подчеркивало ее надежность и превосходную производительность в задачах классификации. Примечательно, что показатель AUC, представляющий совокупную эффективность модели по всем пороговым значениям классификации, был самым высоким для BERT, что подтверждает его сильную дискриминационную способность.

Кривая обучения модели GPT, хотя и демонстрирует улучшение, вышла на плато раньше, чем кривая BERT, что указывает на насыщение обучения, которое можно объяснить генеративным характером модели, которая может не оптимально согласовываться с задачами классификации. Производительность модели GPT соответствовала ожиданиям от модели, в первую очередь предназначенной для генерации, а не дискриминации, о чем свидетельствуют ее более низкие показатели точности и AUC по сравнению с BERT.

Между этапами обучения и проверки было зафиксировано, что BERT сохранял наибольшую согласованность производительности, в то время как LSTM и GPT демонстрировали более существенные расхождения. Это говорит о том, что архитектура BERT и предварительное обучение на различных корпусах способствуют лучшему обобщению, что является критическим фактором для реального применения, где модели сталкиваются с разнообразными и непредсказуемыми данными.

Таким образом, данные подтверждают вывод о том, что архитектура BERT на основе Transformer обеспечивает существенные преимущества в задачах классификации, о чем свидетельствует ее высочайшая производительность по всем оцениваемым показателям. LSTM, хотя и полезен для прогнозирования последовательности, может потребовать дальнейшей оптимизации, чтобы уменьшить переобучение и улучшить обобщение. Более низкая производительность GPT в этом контексте усиливает предрасположенность

модели к генеративным задачам, указывая на необходимость тщательного рассмотрения при ее применении к задачам классификации.

По мере того, как мы намечаем путь вперед, наше внимание сосредотачивается на улучшении понимания этими моделями сложных повествовательных структур, особенно за счет интеграции распознавания сущностей в перекрестных предложениях. Эта инициатива направлена на то, чтобы расширить возможности моделей за пределы отдельных предложений и на целые документы, тем самым обогащая классификацию более глубоким пониманием текстового континуума. Позволяя нашим моделям различать и интерпретировать всеобъемлющее повествование, мы стремимся повысить точность задач классификации, тем самым более точно согласовывая модели с тонкостями понимания человеческого языка.

**Заключение.** В этом сравнительном анализе мы исследовали производительность моделей LSTM, BERT и GPT в контексте задачи классификации с четырьмя двоичными метками, задачи, которая расширяет границы классификации текста в области NLP. Благодаря тщательному исследованию, основанному на таких показателях, как потеря, точность, воспроизводимость и AUC в течение 20 эпох, мы обнаружили многогранный взгляд на динамику обучения каждой модели и ее способность к обобщению.

При оценке эффективности моделей NLP для классификации текста был проведен анализ данных, сравнивающий модели LSTM, BERT и GPT. Оценка основывалась на наборе показателей — потерях, точности, прецизионности, отзыве и AUC — в течение 20 периодов обучения, что давало полное представление о способности каждой модели к обучению и эффективности обобщения.

В целом несмотря на то, что структура BERT на основе трансформаторов доказала свою эффективность для поставленной задачи, обеспечивая превосходную производительность по всем оцениваемым показателям, остается еще большой потенциал в доработке LSTM для уменьшения переобучения и оптимизации GPT для задач классификации. Наши будущие усилия будут включать не только итеративные усовершенствования этих моделей, но и разработку передовых методов, таких как распознавание объектов в перекрестных предложениях, чтобы еще больше расширить границы возможностей классификации NLP.

#### References

Abibullayeva A., Çetin A. (2022) Keyword Extraction from Kazakh News Dataset with BERT. *El-Cezeri Journal of Science and Engineering*, vol. 9, no. 4, pp. 1193–1200, doi: 10.31202/ecjse.1131826 (in Eng.)

Bogdanchikov A., Ayazbayev D., Varlamis I. (2022) Classification of Scientific Documents in the Kazakh Language Using Deep Neural Networks and a Fusion of Images and Text. *Big Data and Cognitive Computing*. Vol. 6(4):123. <https://doi.org/10.3390/bdcc6040123> (in Eng.)

Bhowmik S., Sultana S., Sajid A.A. (2024) Robust multi-domain descriptive text classification

leveraging conventional and hybrid deep learning models. *Int. j. inf. tecnol.* 16, pp. 3219–3231. <https://doi.org/10.1007/s41870-023-01559-2>

Ezen-Can A. (2020) A Comparison of LSTM and BERT for Small Corpus. *ArXiv, abs/2009.05451*. (<https://arxiv.org/html/2009.05451>) (in Eng.)

Garrido-Merchan E.C., Gozalo-Brizuela R., Gonzalez-Carvajal S. (2023). Comparing BERT Against Traditional Machine Learning Models in Text Classification. *Journal of Computational and Cognitive Engineering*, 2(4), 352-356. <https://doi.org/10.47852/bonviewJCCE3202838> (in Eng.)

Jamshidi S., Mohammadi M., Bagheri S., Najafabadi H. E., Rezvanian A., Gheisari M., Ghaderzadeh M., Shahabi A.S., Wu Z. (2024). Effective text classification using BERT, MTM LSTM, and DT. *Data Knowl. Eng.* 151. <https://doi.org/10.1016/j.datak.2024.102306> (in Eng.)

Li J., Tang T., Zhao W. X., Nie J.-Y., Wen J.-R. (2024). Pre-Trained Language Models for Text Generation: A Survey. *ACM Comput. Surv.* 56, 9, Article 230 (September 2024), 39 pages. <https://doi.org/10.1145/3649449> (in Eng.)

Minaee S., Kalchbrenner N., Cambria E., Nikzad N., Chenaghlu M., Gao J. (2022) Deep Learning-based Text Classification: A Comprehensive Review. *ACM Comput. Surv.* Vol. 54, 3, Article 62, 40 p. <https://doi.org/10.1145/3439726>

Myrzakhetov B., Kozhirkbayev Z. (2018). Extended language modeling experiments for Kazakh. *CEUR Workshop Proceedings*, vol. 2303. <https://ceur-ws.org/Vol-2303/short2.pdf> (in Eng.)

Narejo K. Rani, Zan H., Oralbekova D., Parkash Dharmani K., Orken M., Mukhsina K. (2024) Enhancing Emoji-Based Sentiment Classification in Urdu Tweets: Fusion Strategies With Multilingual BERT and Emoji Embeddings. *IEEE Access*, vol. 12, pp. 126587-126600, doi: 10.1109/ACCESS.2024.3446897 (in Eng.)

Oralbekova D., Mamyrbayev O., Zhumagulova S., Zhumazhan N. (2024). A Comparative Analysis of LSTM and BERT Models for Named Entity Recognition in Kazakh Language: A Multi-classification Approach. In: Agarwal, N., Sakalauskas, L., Tukeyev, U. (eds) *Modeling and Simulation of Social-Behavioral Phenomena in Creative Societies. MSBC 2024. Communications in Computer and Information Science*, vol 2211. Springer, Cham. [https://doi.org/10.1007/978-3-031-72260-8\\_10](https://doi.org/10.1007/978-3-031-72260-8_10) (in Eng.)

Oralbekova D., Mamyrbayev O., Othman M., Alimhan K., Zhumazhanov B., Nuranbayeva B. (2022). Development of CRF and CTC Based End-To-End Kazakh Speech Recognition System. In: Nguyen, N.T., Tran, T.K., Tukayev, U., Hong, TP., Trawiński, B., Szczerbicki, E. (eds) *Intelligent Information and Database Systems. ACIIDS 2022. Lecture Notes in Computer Science*, vol. 13757, 2022. [https://doi.org/10.1007/978-3-031-21743-2\\_41](https://doi.org/10.1007/978-3-031-21743-2_41) (in Eng.)

Pandey R., Waghela H., Rakshit S., Rangari A., Singh A., Kumar R., Ghosal R., Sen J. (2024). Generative AI-Based Text Generation Methods Using Pre-Trained GPT-2 Model. *ArXiv, abs/2404.01786* (in Eng.)

Reusens M., Stevens A., Tonglet J., Smedt J. D., Verbeke W., Broucke S., Baesens B. (2024). Evaluating text classification: A benchmark study. *Expert Syst. Appl.* 254, C (Nov 2024). <https://doi.org/10.1016/j.eswa.2024.124302> (in Eng.)

Tikayat Ray A, Cole B., Pinon Fischer O., White R., Mavris D. (2023) *aeroBERT-Classifier: Classification of Aerospace Requirements Using BERT*. *Aerospace*. Vol. 10(3):279. <https://doi.org/10.3390/aerospace10030279> (in Eng.)

Yeshpanov R., Khassanov Y., Varol H.A. (2022) *KazNERD: Kazakh Named Entity Recognition Dataset*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 417–426. <https://aclanthology.org/2022.lrec-1.44.pdf> (in Eng.)

## CONTENTS

## INFORMATION AND COMMUNICATION TECHNOLOGIES

<b>A.Abdiraman, L.Aldasheva, A.Zakirova, B.Mukhametzhanova, I.Orman</b> GLOBAL ANALYSIS OF MOBILE BROADBAND NETWORK PERFORMANCE: INSIGHTS INTO 5G DEPLOYMENT AND FUTURE 6G CHALLENGES.....	5
<b>R. Abdualiyeva, L. Smagulova, A. Yelepbergenova</b> THE EFFECTIVENESS OF USING CHATGPT IN PROGRAMMING.....	17
<b>A.B. Aben, N.M. Zhunissov, G.N. Kazbekova, A.N. Amanov, A.A. Abibullayeva</b> DEEPPFAKE ARTIFICIAL VOICE DETECTION. COMPARISON OF THE EFFECTIVENESS OF THE LSTM AND CNN MODELS.....	32
<b>A.A. Aitkazina, N.O. Zhumazhan</b> DEVELOPMENT OF A BIOTECHNICAL SYSTEM FOR LASER TREATMENT OF SUNFLOWER SEEDS.....	49
<b>G. Aksholak, A. Bedelbayev, R. Magazov</b> SECURING KUBERNETES: AN ANALYSIS OF VULNERABILITIES, TOOLS, AND FUTURE DIRECTIONS.....	66
<b>A.T. Akynbekova, A.A. Mukhanova, Salah Al-Majeed, A.G. Altayeva</b> PROBLEMS OF IMPLEMENTATION OF FUZZY MODELS OF DECISION MAKING IN SOCIAL PROCESSES.....	78
<b>K.M. Aldabergenova, M.A. Kantureyeva, A.B. Kassekeyeva, A. Akhmetova, T.N. Esikova</b> FEATURES AND PROSPECTS FOR THE USE OF DIGITAL PLATFORMS AND INTERNET MARKETING IN THE DEVELOPMENT OF AGRICULTURAL PRODUCTION.....	93
<b>A. Yerimbetova, M. Sambetbayeva, E. Daiyrbayeva, B. Sakenov, U. Berzhanova</b> CREATING A MODEL FOR RECOGNIZING THE KAZAKH SIGN LANGUAGE USING THE DEEP LEARNING METHOD.....	108
<b>A.N. Zhidebayeva, S.T. Akhmetova, A.O. Aliyeva, B.O. Tastanbekova, G.S. Shaimerdenova</b> REVIEW OF DETECTION AND PREVENTION OF OFFENSIVE LANGUAGE VIA SOCIAL MEDIA DATA MINING.....	124

**K.S. Ivanov, D.T. Tulekenova**

ENSURING THE DETERMINABILITY OF MOTION OF AN ADAPTIVE SPACECRAFT DRIVE BY INTRODUCING AN ADDITIONAL VELOCITY CONSTRAINT FORCE.....136

**M.N. Kalimoldayev, Z.D. Ormansha, K.B. Begaliev, A.S. Ainagulova, A.O. Aukenova**

A BLOCKCHAIN MODEL FOR AGRICULTURAL PRODUCT TRACKING THAT SUPPORTS FEDERAL TRAINING.....151

**I. Massyrova, O. Joldasbayev, S. Joldasbayev, A. Bolysbek, S. Mambetov**  
AUTOMATION OF THE SYSTEM FOR INDUSTRIAL PRACTICE AND INTERNSHIPS FOR STUDENTS IN ORGANIZATIONS OUTSIDE OF THE UNIVERSITY.....168

**A.B. Mimenbayeva, G.O. Issakova, G.K. Bekmagambetova, A.B. Aruova, E.K. Darikulova**

DEVELOPMENT OF DEEP LEARNING MODELS FOR FIRE SOURCES PREDICTION.....185

**K. Momynzhanova, S.Pavlov, Sh. Zhumagulova**

MATHEMATICAL MODELS AND PRACTICAL IMPLEMENTATION OF AN OPTICAL-ELECTRONIC EXPERT SYSTEM FOR GLAUCOMA DETECTION.....202

**B.O. Mukhametzhanova, L.N. Kulbaeva, Z.B. Saimanova, E.K. Seipisheva, B.M. Sadanova**

OPTIMIZATION AND INTEGRATION OF DOCKER TECHNOLOGY IN MODERN INFORMATION SYSTEMS.....218

**A.R. Orazayeva, J.A. Tussupov, A.K. Shaikhanova, G.B. Bekeshova, A.D. Galymova**

FUZZY EXPERT SYSTEM FOR ASSESSING DYNAMIC CHANGES IN BIOMEDICAL IMAGES OF BREAST CANCER TUMORS.....227

**D. Oralbekova, O. Mamyrbayev, A. Akhmediyarova, D. Kassymova**  
USING KAZAKH NER DATASETS FOR MULTICLASS CLASSIFICATION IN THE LEGAL DOMAIN: A COMPARATIVE STUDY OF BERT, GPT, AND LSTM MODELS.....242

**A. Ospanov, A.J. Pedro, T. Turymbetov, K. Dyussekeyev, A. Zhumadillayeva**  
ADVANCEMENTS IN ERP SYSTEMS THROUGH EMERGING

TECHNOLOGIES, MACHINE LEARNING AND HYBRID OPTIMIZATION  
TECHNIQUES.....259

**K. Rabbany, A. Bekarystankyzy, A. Shoiynbek, D. Kuanyshbay,  
A. Mukhametzhano**  
DETECTION OF SUICIDAL TENDENCIES IN REDDIT POSTS  
USING MACHINE LEARNING.....270

**A. Taukenova**  
PERSONALIZED ARCHITECTURE: CREATING UNIQUE SPACES  
WITH DIGITAL TECHNOLOGIES.....283



**МАЗМҰНЫ**

**АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ  
ТЕХНОЛОГИЯЛАР**

<b>Ә. Әбдіраман, Л. Алдашева, А. Закирова, Б. Мухаметжанова, И. Орман</b> МОБИЛЬДІ КЕН ЖОЛАҚТЫ ЖЕЛІЛЕРДІҢ ТИІМДІЛІГІНІҢ ЖАҒАНДЫҚ ТАЛДАУ: 5G ЕНГІЗУ ЖӘНЕ 6G БОЛАШАҚ МӘСЕЛЕЛЕРІ.....	5
<b>Р.Е. Абдуалиева, Л.А. Смагулова, А.У. Елепбергенова</b> БАҒДАРЛАМАЛАУДА СНАТGPT ҚОЛДАНУ ТИІМДІЛІГІ.....	17
<b>А.Б. Абен, Н.М. Жунисов, Г.Н. Казбекова, А.Н. Аманов, А.А. Абибуллаева</b> DEEPFAKE ЖАСАНДЫ ДАУЫСТЫ АНЫҚТАУ. LSTM ЖӘНЕ CNN МОДЕЛЬДЕРІНІҢ ТИІМДІЛІГІ САЛЫСТЫРУ.....	32
<b>Ә.А. Айтқазина, Н.Ө. Жұмажан</b> КҮНБАҒЫС ТҰҚЫМДАРЫН ЛАЗЕРМЕН ӨНДЕУГЕ АРНАЛҒАН БИОТЕХНИКАЛЫҚ ЖҮЙЕНІ ДАМЫТУ.....	49
<b>Г.И. Ақшолақ, А.А. Бедельбаев, Р.С. Мағазов</b> KUBERNETES-ТІ ҚОРҒАУ: ОСАЛДЫҚТАРДЫ, ҚҰРАЛДАРДЫ ЖӘНЕ БОЛАШАҚ БАҒЫТТАРДЫ ТАЛДАУ.....	66
<b>А.Т. Ақынбекова, А.А. Муханова, Salah Al-Majeed, Г.С. Алтаева</b> ӘЛЕУМЕТТІК ПРОЦЕСТЕРДЕ ШЕШІМДЕР ҚАБЫЛДАУДЫҢ БҰЛДЫР МОДЕЛЬДЕРІН ЕНГІЗУ МӘСЕЛЕЛЕРІ.....	78
<b>К.М. Алдабергенова, М.А. Кантуреева, А.Б. Касекеева, А.Ж. Ахметова, Т.Н. Есикова</b> АУЫЛ ШАРУАШЫЛЫҒЫ ӨНДІРІСІН ДАМЫТУДА ЦИФРЛЫҚ ПЛАТФОРМАЛАР МЕН ИНТЕРНЕТ-МАРКЕТИНГТІ ҚОЛДАНУДЫҢ ЕРЕКШЕЛІКТЕРІ МЕН ПЕРСПЕКТИВАЛАРЫ.....	93
<b>А.С. Еримбетова, М.А. Сәмбетбаева, Э.Н. Дайырбаева, Б.Е. Сәкенов, У.Г. Бержанова</b> ТЕРЕҢ ОҚЫТУ ӘДІСІН ҚОЛДАНУ АРҚЫЛЫ ҚАЗАҚ ҰМ ТІЛІН ТАНУҒА АРНАЛҒАН МОДЕЛЬ ҚҰРУ.....	108

- А.Н. Жидебаева, С.Т. Ахметова, А.О. Алиева, Б.О. Тастанбекова,  
Г.С. Шаймерденова**  
ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕН DATA MINING АРҚЫЛЫ БЕЙӘДЕП  
СӨЗДЕРДІ АНЫҚТАУ ЖӘНЕ АЛДЫН АЛУҒА ШОЛУ.....124
- К.С. Иванов, Д.Т. Тулекенова**  
ЖЫЛДАМДЫҚ БАЙЛАНЫСЫНЫҢ ҚОСЫМША КҮШІН ЕНГІЗУ  
АРҚЫЛЫ ҒАРЫШ АППАРАТЫНЫҢ БЕЙІМДЕЛГЕН ЖЕТЕК  
ҚОЗҒАЛЫСЫНЫҢ АЙҚЫНДЫЛЫҒЫН ҚАМТАМАСЫЗ ЕТУ.....136
- М.Н. Калимолдаев, З.Д. Орманша, К.Б. Бегалиева, А.С. Айнагулова,  
А.О. Аукенова**  
ФЕДЕРАТИВТІ ОҚЫТУДЫ ҚОЛДАЙТЫН АУЫЛШАРУАШЫЛЫҚ  
ӨНІМДЕРІН БАҚЫЛАУҒА АРНАЛҒАН БЛОКЧЕЙН МОДЕЛІ.....151
- И. Масырова, О.К. Джолдасбаев, С.К. Джолдасбаев, А. Болысбек,  
С.Т. Мамбетов**  
УНИВЕРСИТЕТТЕН ТЫС ҰЙЫМДАРДА СТУДЕНТТЕРДІҢ  
ӨНДІРІСТІК ПРАКТИКАСЫ МЕН ТАҒЫЛЫМДАМАСЫН  
АВТОМАТТАНДЫРУ ЖҮЙЕСІ.....168
- А.Б. Мименбаева, Г.О. Исакова, Г.К. Бекмагамбетова, Ә.Б. Аруова,  
Е.Қ. Дәрікүлова**  
ӨРТ КӨЗДЕРІН БОЛЖАУ ҮШІН ТЕРЕҢ ОҚЫТУ МОДЕЛЬДЕРІН  
ӘЗІРЛЕУ.....185
- К.Р. Момынжанова, С.В. Павлов, Ш.П. Жұмағұлова, М.Т. Тұңғышбаев**  
ГЛАУКОМАНЫ АНЫҚТАУҒА АРНАЛҒАН ОПТИКАЛЫҚ-  
ЭЛЕКТРОНДЫҚ САРАПТАМАЛЫҚ ЖҮЙЕНІҢ МАТЕМАТИКАЛЫҚ  
МОДЕЛЬДЕРІ МЕН ПРАКТИКАЛЫҚ ІСКЕ АСЫРЫЛУЫ.....202
- Б.О. Мухаметжанова, Л.Н. Құлбаева, З.Б. Сайманова, Э.К. Сейпишева,  
Б.М. Саданова**  
ЗАМАНАУИ АҚПАРАТТЫҚ ЖҮЙЕЛЕРДЕГІ DOCKER  
ТЕХНОЛОГИЯСЫН ОҢТАЙЛАНДЫРУ ЖӘНЕ ИНТЕГРАЦИЯЛАУ.....218
- А.Р. Оразаева, Д.А. Тусупов, А.К. Шайханова, Г.Б. Бекешова,  
Ә.Д. Ғалымова**  
СҮТ БЕЗІ ҚАТЕРЛІ ІСІГІ КЕЗІНДЕ БИОМЕДИЦИНАЛЫҚ  
КЕСКІНДЕРІНДЕГІ ДИНАМИКАЛЫҚ ӨЗГЕРІСТЕРДІ БАҒАЛАУҒА  
АРНАЛҒАН АНЫҚ ЕМЕС САРАПТАМА ЖҮЙЕСІ.....227

<b>Д. Оралбекова, О. Мамырбаев, А. Ахмедиярова, Д. Қасымова</b> ҚАЗАҚ ТІЛІНДЕГІ NER ДЕРЕКТЕР ЖИНАҒЫН ҚҰҚЫҚТЫҚ САЛАДА КӨПСАНАТТЫ ЖІКТЕУ ҮШІН ПАЙДАЛАНУ: BERT, GPT ЖӘНЕ LSTM МОДЕЛЬДЕРІНІҢ САЛЫСТЫРМАЛЫ ЗЕРТТЕУІ.....	242
<b>А. Оспанов, П. Алонсо-Жорда, Т. Тұрымбетов, К. Дүйсекеев, А. Жұмаділлаева</b> ERP ЖҮЙЕЛЕРІНІҢ ЖЕТІЛДІРІЛУІ: ЗАМАНАУИ ТЕХНОЛОГИЯЛАР, МАШИНАЛЫҚ ОҚЫТУ ЖӘНЕ ГИБРИДТІ ОПТИМИЗАЦИЯ ӘДІСТЕРІ.....	259
<b>К. Раббани, А. Бекарыстанқызы, Д. Қуанышбай, А. Шойынбек, А. Мұхаметжанов</b> МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ АРҚЫЛЫ REDDIT ПОСТТАРЫНДАҒЫ СУИЦИДТІК ТЕНДЕНЦИЯЛАРЫН АНЫҚТАУ.....	270
<b>Ә. Таукенова</b> ЖЕКЕЛЕНДІРІЛГЕН АРХИТЕКТУРА: ДИДЖИТАЛ ТЕХНОЛОГИЯЛАРМЕН ЕРЕКШЕ КЕҢІСТІКТЕР ЖАРАТУ.....	283

## СОДЕРЖАНИЕ

ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ  
ТЕХНОЛОГИИ

<b>А. Абдираман, Л. Алдашева, А. Закирова, Б. Мухаметжанова, И. Орман</b> ГЛОБАЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ МОБИЛЬНОЙ ШИРОКОПОЛОСНОЙ СЕТИ: ВНЕДРЕНИЕ 5G И БУДУЩИЕ ЗАДАЧИ 6G.....	5
<b>Р.Е. Абдуалиева, Л.А. Смагулова, А.У. Елепбергенова</b> ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ SNATGPT В ПРОГРАММИРОВАНИИ.....	17
<b>А.Б. Абен, Н.М. Жунисов, Г.Н. Казбекова, А.Н. Аманов, А.А. Абибуллаева</b> ОБНАРУЖЕНИЕ ИСКУССТВЕННОГО ГОЛОСА DEEPFAKE. СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ LSTM И CNN.....	32
<b>А.А. Айтказина, Н.О. Жумажан</b> РАЗРАБОТКА БИОТЕХНИЧЕСКОЙ СИСТЕМЫ ДЛЯ ЛАЗЕРНОЙ ОБРАБОТКИ СЕМЯН ПОДСОЛНЕЧНИКА.....	49
<b>Г.И. Акшолок, А.А. Бедельбаев, Р.С. Магазов</b> ЗАЩИТА KUBERNETES: АНАЛИЗ УЯЗВИМОСТЕЙ, ИНСТРУМЕНТОВ И НАПРАВЛЕНИЙ НА БУДУЩЕЕ.....	66
<b>А.Т. Акынбекова, А.А. Муханова, Salah Al-Majeed, Г.С. Алтаева</b> ПРОБЛЕМЫ РЕАЛИЗАЦИИ НЕЧЕТКИХ МОДЕЛЕЙ ПРИНЯТИЯ РЕШЕНИЙ В СОЦИАЛЬНЫХ ПРОЦЕССАХ.....	78
<b>К.М. Алдабергенова, М.А. Кантуреева, А.Б. Касекеева, А.Ж. Ахметова, Т.Н. Есикова</b> ОСОБЕННОСТИ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ ЦИФРОВЫХ ПЛАТФОРМ И ИНТЕРНЕТ-МАРКЕТИНГА В РАЗВИТИИ СЕЛЬСКОХОЗЯЙСТВЕННОГО ПРОИЗВОДСТВА.....	93
<b>А.С. Еримбетова, М.А. Самбетбаева, Э.Н. Дайырбаева, Б.Е. Сакенов, У.Г. Бержанова</b> СОЗДАНИЕ МОДЕЛИ ДЛЯ РАСПОЗНАВАНИЯ КАЗАХСКОГО ЖЕСТОВОГО ЯЗЫКА С ИСПОЛЬЗОВАНИЕМ МЕТОДА ГЛУБОКОГО ОБУЧЕНИЯ.....	108

- А.Н. Жидебаева, С.Т. Ахметова, А.О. Алиева, Б.О. Тастанбекова,  
Г.С. Шаймерденова**  
ОБЗОР ОБНАРУЖЕНИЯ И ПРЕДОТВРАЩЕНИЯ ОСКОРБИТЕЛЬНОЙ  
ЛЕКСИКИ С ПОМОЩЬЮ DATA MINING В СОЦИАЛЬНЫХ СЕТЯХ....124
- К.С. Иванов, Д.Т. Тулеkenова**  
ОБЕСПЕЧЕНИЕ ОПРЕДЕЛИМОСТИ ДВИЖЕНИЯ АДАПТИВНОГО  
ПРИВОДА КОСМИЧЕСКОГО АППАРАТА С ПОМОЩЬЮ ВВЕДЕНИЯ  
ДОПОЛНИТЕЛЬНОЙ СИЛЫ СКОРОСТНОЙ СВЯЗИ.....136
- М.Н. Калимолдаев, З.Д. Орманша, К.Б. Бегалиева, А.С. Айнагулова,  
А.О. Аукенова**  
БЛОКЧЕЙН-МОДЕЛЬ ДЛЯ ОТСЛЕЖИВАНИЯ  
СЕЛЬСКОХОЗЯЙСТВЕННОЙ ПРОДУКЦИИ С ПОДДЕРЖКОЙ  
ФЕДЕРАТИВНОГО ОБУЧЕНИЯ.....151
- И. Масырова, О.К. Джолдасбаев, С.К. Джолдасбаев, А. Болысбек,  
С.Т. Мамбетов**  
АВТОМАТИЗАЦИЯ СИСТЕМЫ ДЛЯ ПРОИЗВОДСТВЕННОЙ  
ПРАКТИКИ И СТАЖИРОВКИ СТУДЕНТОВ В ОРГАНИЗАЦИЯХ  
ВНЕ ВУЗА.....168
- А. Мименбаева, Г. Исакова, Г.К. Бекмагамбетова, А.Б. Аруова,  
Е.К. Дарикулова**  
РАЗРАБОТКА МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ  
ПРОГНОЗИРОВАНИЯ ИСТОЧНИКОВ ПОЖАРОВ.....185
- К.Р. Момынжанова, С.В. Павлов, Ш.П. Жумагулова, М.Т. Тунгушбаев**  
МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ  
ОПТИКО-ЭЛЕКТРОННОЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ  
ВЫЯВЛЕНИЯ ГЛАУКОМЫ.....202
- Б.О. Мухаметжанова, Л.Н. Кулбаева, З.Б. Сайманова, Э.К. Сейпишева,  
Б.М. Саданова**  
ОПТИМИЗАЦИЯ И ИНТЕГРАЦИЯ ТЕХНОЛОГИИ DOCKER В  
СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ.....218
- А.Р. Оразаева, Д.А. Тусупов, А.К. Шайханова, Г.Б. Бекешова,  
А.Д. Галымова**  
НЕЧЕТКАЯ ЭКСПЕРТНАЯ СИСТЕМА ДЛЯ ОЦЕНКИ ДИНАМИЧЕСКИХ  
ИЗМЕНЕНИЙ В БИМЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ ОПУХОЛЕЙ  
ПРИ РАКЕ МОЛОЧНОЙ ЖЕЛЕЗЫ.....227

<b>Д. Оралбекова, О. Мамырбаев, А. Ахмедиярова, Д. Касымова</b> ИСПОЛЬЗОВАНИЕ НАБОРОВ ДАННЫХ NER НА КАЗАХСКОМ ЯЗЫКЕ ДЛЯ МУЛЬТИКЛАССИФИКАЦИИ В ПРАВОВОЙ СФЕРЕ: СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МОДЕЛЕЙ VERT, GPT И LSTM.....	242
<b>А. Оспанов, П. Алонсо-Жорда, Т. Турымбетов, К. Дюсекеев, А. Жумадилаева</b> ПРОДВИЖЕНИЕ ERP СИСТЕМ С ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ ТЕХНОЛОГИЙ, МАШИННОГО ОБУЧЕНИЯ И ГИБРИДНЫХ МЕТОДОВ ОПТИМИЗАЦИИ.....	259
<b>К. Раббани, А. Бекарыстанкызы, Д. Куанышбай, А. Шойынбек, А. Мухаметжанов</b> ОБНАРУЖЕНИЕ СУИЦИДАЛЬНЫХ ТЕНДЕНЦИЙ В ПУБЛИКАЦИЯХ НА REDDIT С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ.....	270
<b>А. Таукенова</b> ПЕРСОНАЛИЗИРОВАННАЯ АРХИТЕКТУРА: СОЗДАНИЕ УНИКАЛЬНЫХ ПРОСТРАНСТВ С ПОМОЩЬЮ ЦИФРОВЫХ ТЕХНОЛОГИЙ.....	283

**Publication Ethics and Publication Malpractice  
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ([http://publicationethics.org/files/u2/New\\_Code.pdf](http://publicationethics.org/files/u2/New_Code.pdf)). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

**[www.nauka-nanrk.kz](http://www.nauka-nanrk.kz)**

**<http://physics-mathematics.kz/index.php/en/archive>**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Редакторы: *Д.С. Аленов, Ж.Ш. Әден*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 20.03.2025.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

20,0 п.л. Заказ 1.