

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



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҮЛТТЫҚ ФЫЛЫМ АКАДЕМИЯСЫ» РҚБ

ХАБАРЛАРЫ

ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ
АКАДЕМИИ НАУК РЕСПУБЛИКИ
КАЗАХСТАН»

NEWS

OF THE NATIONAL ACADEMY OF
SCIENCES OF THE REPUBLIC OF
KAZAKHSTAN

SERIES OF PHYSICS AND MATHEMATICS

2 (354)

APRIL – JUNE 2025

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

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>

WOICIK 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/>

БАС РЕДАКТОР:

МУТАНОВ **Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, КР ҰҒА академигі, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Казахстан), <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, КР ҰҒА академигі, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана менгерушісі (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ **Әркен Жұмажанұлы** (ғалым қатызы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ **Жұмаділ Жаңабайұлы**, техника ғылымдарының докторы, профессор, КР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Колданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Казахстан), <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 **Андрей**, 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)), КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА **Нина**, техника ғылымдарының докторы, профессор, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Казахстан), <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 Еркебұлнанызы**, техника ғылымдарының докторы, доцент, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Казахстан), <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 мәнгерушісі (Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6603642864>, <https://www.webofscience.com/wos/author/record/3802016>

КАПАЛОВА **Нұрсұлу Алдаражарқызы**, техника ғылымдарының кандидаты, КР ФЖБМ ФК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының мәнгерушісі (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>

КОВАЛЕВ **Александр Михайлович**, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Колданбалы 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>

«КР ҰҒА Хабарлары. Физика-математика сериясы».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

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

Ақпарат агенттігінің мерзімді баспасоз басылымын, ақпарат агенттігін және желілік басылымдың қайта есепке кою туралы КР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **28.02.2025** ж. берген №**KZ20VPY00113741** Күелік.

Такырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргіткінде үшінші бағыты: *ақпараттық-коммуникациялық технологиялар* бағыты бойынша КР БФМ БГСБК

ұсынан жүрнәлдәр тізіміне енді.

Мерзімділігі: жылдан 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» РКБ, 2025

ГЛАВНЫЙ РЕДАКТОР:

МУТАНОВ Галимкаир 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 Нурадилович, (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 Сатпаева (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 Саule 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 Нурсулы 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 Беларусь (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: Республикаансое общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № KZ20VPY00113741. Дата выдачи 28.02.2025

Тематическая направленность: информационно-коммуникационные технологии.

В настоящее время: вошел в список журналов, рекомендованных КОКСНВО 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

CONTENTS

INFORMATION AND COMMUNICATION TECHNOLOGIES

O. Auyelbekov, E. Bostanov, R. Berkutbayeva, A. Seidildayeva, I. Musabekova ANALYSIS OF AGRICULTURAL YIELDS IN KAZAKHSTAN USING UNMANNED AERIAL VEHICLES AND AI.....	12
S.T. Akhmetova, S.U. Ismailov, A.A. Batyrbekov, A.S. Ismailova PREREQUISITES FOR CREATION OF A VIRTUAL 3D MODEL OF AN UNMANNED UNIVERSAL CROPPING TRACTOR.....	33
A. Bekarystankzyz, O. Mamyrbayev, D. Oralbekova, A. Yerimbetova, M. Turdalyuly TESTING THE AUDIO-TEXT DATASET FOR KAZAKH LANGUAGE USING THE CONFORMER ENCODER.....	50
G. Bekmanova, M. Kantureeva, A. Omarbekova, B. Ergesh, A. Zakirova THE USE AND IMPACT ASSESSMENT OF ARTIFICIAL INTELLIGENCE IN HIGHER EDUCATION.....	61
N.S. Yesmukhamedov, S. Sapakova, Syed Abdul Rahman Al-Haddad, D. Daniyarova DEVELOPMENT OF AN INFORMATION SYSTEM ARCHITECTURE FOR HEALTHCARE INSTITUTIONS USING ARTIFICIAL INTELLIGENCE.....	74
T. Zhukabayeva, V. Desnitsky, Y. Mardenov, N. Karabayev INFORMATION SECURITY INCIDENT MANAGEMENT IN IIOT SYSTEMS WITH EDGE COMPUTING.....	92
M. Ilesbay, A. Tynymbayev, S. Mambetov, A. Barakova, O. Joldasbayev INTEGRATED METHOD OF INFORMATION PROTECTION BASED ON DATA COMPRESSION, ENCRYPTION AND SEPARATION.....	107
B.A. Karibayev, N. Meirambekuly, M. Ibraim, A.S. Baikenov, G.B. Ikhsan DESIGN OF A SIX-ELEMENT S-BAND ANTENNA ARRAY FOR CUBESAT.....	125
N. Karymsakova, K. Ozhikenov, M. Bolysbek, R. Beisembekova ARCHITECTURE OF THE MEDICAL REHABILITATION PLATFORM.....	140

D. Kuanyshbay, A. Shoiynbek, K. Rabbany, A. Bekarystankzyz, A. Mukhametzhanov COMPARISON OF MACHINE LEARNING AND REINFORCEMENT LEARNING FOR DEPRESSION RECOGNITION FROM SPEECH.....	155
E. Nysanov, Zh. Kemelbekova, E. Abdrashova, S. Kurakbaeva, A. Baydibekova MODELING AND CALCULATION OF THE FLOW OF THREE-PHASE MEDIA WITH VARIABLE CONCENTRATIONS.....	169
B. Orazbaev, Z. Kuzhukhanova, K. Orazbaeva, A. Kishubaeva DEVELOPMENT OF MODELS OF ATMOSPHERIC AND RECTIFICATION COLUMNS IN PRIMARY OIL REFINING.....	181
D. Rakhimova, A. Sarsenbayeva, A. Turarbek, A. Auezova THE USE OF DEEP LEARNING TO IMPROVE THE ACCURACY OF ANSWERS IN MULTILINGUAL QUESTION-AND-ANSWER SYSTEMS...	196
L. Rzayeva, D. Pogolovkin, A. Myrzatay DEVELOPMENT OF A MODULAR NLP-BASED CORRESPONDENCE ANALYSIS SERVICE FOR DIGITAL FORENSICS.....	212
A.T. Sankibayev, I. Makhambayeva, K. Kanibaikyzy, A. Temirbek MODELING OF VIBRATIONAL PROCESSES IN WOLFRAM MATHEMATICA SYSTEM.....	234
N.M. Temirbekov, A.K. Turarov NUMERICAL SOLUTION OF THE DIRECT AND INVERSE PROBLEM OF GAS LIFT OIL PRODUCTION PROCESS BY THE METHOD OF CONJUGATE EQUATIONS.....	251
Z. Utemaganbetov, Kh. Ramazanova, K. Bizhanova, R. Assylbayeva AN ANALYTICAL AND NUMERICAL METHOD FOR TRANSFERRING BOUNDARY CONDITIONS FOR A ONE-DIMENSIONAL DIFFUSION EQUATION.....	280
M. Khizirova, K. Chezhimbayeva, A. Kassimov, M. Yermekbayev RESEARCH ON DISTRIBUTION TRAFFIC AND DISTRIBUTION METHODS IN VANET NETWORKS.....	294
K. Yakunin, D. Kusain, R.I. Mukhamediev, N. Yunicheva, N. Kuldeyev INTEGRATION OF FLIGHT PATH PLANNING PROGRAMS AND CONTROL SYSTEMS OF UNMANNED AERIAL VEHICLES.....	317

МАЗМУНЫ

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР

Ө. Әуелбеков, Е. Бостанов, Р. Беркутбаева, А. Сейдилдаева, И. Мусабекова ҚАЗҚАСТАНДА АУЫЛ ШАРУАШЫЛЫҒЫ ӨНІМДІЛІГІН ҮШҚЫШСЫЗ ҮШУ АППАРАТТАРЫ МЕН ЖАСАНДЫ ИНТЕЛЛЕКТ ҚӨМЕГІМЕН ТАЛДАУ.....	12
С.Т. Ахметова, С.У. Исмаилов, А.А. Батырбеков, А.С. Исмаилова ЖУРГІЗУШІСІЗ ӘМБЕБАП ЕГІН ЕГЕТИН ТРАКТОРДЫҢ ВИРТУАЛДЫ 3D МОДЕЛІН ҚҰРУДЫҢ АЛҒЫ ШАРТТАРЫ.....	33
А. Бекарыстанқызы, О. Мамырбаев, Д. Оралбекова, А. Еримбетова, М. Тұрдалықызы CONFORMER ШИФРЛАУШЫСЫН ҚОЛДАНЫП ҚАЗАҚ ТІЛІНДЕ АУДИО- МӘТІН ТҮРІНДЕ ЖИНАЛҒАН МӘЛІМЕТТЕР ҚОРЫН СЫНАУ.....	50
Г.Т. Бекманова, М.А. Кантуреева, А.С. Омарбекова, Б. Ж. Ергеш, А.Б. Закирова ЖОҒАРЫ БІЛІМ БЕРУДЕ ЖАСАНДЫ ИНТЕЛЛЕКТТІ ҚОЛДАНУ ЖӘНЕ ӘСЕРІН БАҒАЛАУ.....	61
Н.С. Есмұхамедов, С. Сапақова, Сайд Абдул Рахман Әл-Хаддад, Д. Даниярова, МЕДИЦИНАЛЫҚ МЕКЕМЕЛЕРГЕ АРНАЛҒАН ЖАСАНДЫ ИНТЕЛЛЕКТТІ ҚОЛДАНАТЫН АҚПАРАТТЫҚ ЖҮЙЕ АРХИТЕКТУРАСЫН ӘЗІРЛЕУ.....	74
Т. Жукабаева, В. Десницкий, Е. Марденов, Н. Карабаев ПОТ-ЖҮЙЕЛЕРІНДЕГІ АҚПАРАТТЫҚ ҚАУПІСІЗДІК ИНЦИДЕНТТЕРІН ШЕТКІ ЕСЕПТЕУЛЕРДІ ПАЙДАЛАНА ОТЫРЫП БАСҚАРУ.....	91
М.А. Ілесбай, Ә.Б. Тынымбаев, С.Т. Мамбетов, А.Ш. Баракова, О.К. Джолдасбаев ДЕРЕКТЕРДІ СЫФУ, ШИФРЛАУ ЖӘНЕ БӨЛУ НЕГІЗІНДЕ АҚПАРАТТЫ ҚОРҒАУДЫҢ БІРІКТІРІЛГЕН ӘДІСІ.....	107
Б.А. Карибаев, Н. Мейрамбекұлы, М. Ибраим, А.С. Байкенов, Г.Б. Ихсан CUBESAT ҮШІН АЛТЫ ЭЛЕМЕНТТІ S-ДИАПАЗОНДЫ АНТЕННА ТОРЫН ЖОБАЛАУ.....	125
Н.Т. Карымсакова К.А. Ожикенов М.Е. Болысбек, Р.Н. Бейсембекова МЕДИЦИНАЛЫҚ ӨҢАЛТУ ПЛАТФОРМА АРХИТЕКТУРАСЫ.....	140

Д. Қуанышбай, А. Шойынбек, К. Раббани, А. Мұхаметжанов, Б. Мералиев СӨЙЛЕУ АРҚЫЛЫ ДЕПРЕССИЯНЫ ТАНУ ҮШИН МАШИНАЛЫҚ ОҚЫТУ МЕН КУШЕЙТУ АРҚЫЛЫ ОҚЫТУДЫ САЛЫСТАРУ.....	155
Е.А. Нысанов, Ж.С. Кемельбекова, Э.Т. Абдрашова, С.Ж. Куракбаева, А.О. Байдибекова АЙНЫМАЛЫ КОНЦЕНТРАЦИЯЛЫ ҮШ ФАЗАЛЫ ОРТАЛАРДЫҢ АҒЫНЫН МОДЕЛЬДЕУ ЖӘНЕ ЕСЕПТЕУ.....	169
Б. Оразбаев, Ж. Кужуханова, К. Оразбаева, А. Кишубаева МҰНАЙДЫ БАСТАПҚЫ ӨҢДЕУ КЕЗІНДЕ АТМОСФЕРАЛЫҚ ЖӘНЕРЕТИФИКАЦИЯЛЫҚКОЛОНОЛАРЫНЫҢ МОДЕЛЬДЕРІН ӘЗІРЛЕУ.....	181
Д. Рахимова, А. Сарсенбаева, Э. Тураарбек, Э. Ауезова КӨП ТІЛДІ СҰРАҚ-ЖАУАП ЖҮЙЕЛЕРІНДЕ ЖАУАПТАРДЫҢ ДӘЛДІГІН АРТТАРЫУ ҮШИН ТЕРЕҢ ОҚЫТУДЫ ҚОЛДАNU.....	196
Л. Рзаева, Д. Поголовкин, А.Мырзатай ЦИФРЛЫҚ КРИМИНАЛИСТИКА ҮШИН NLP НЕГІЗІНДЕГІ МОДУЛЬДІК ХАТ АЛМАСУДЫ ТАЛДАУ ҚЫЗМЕТІН ӘЗІРЛЕУ.....	212
А.Т. Санкибаев, И. Махамбаева, К. Канибайкызы, А. Темирбек ТЕРБЕЛІСТЕР ҮДЕРІСІН WOLFRAM МАТЕМАТИКА ЖҮЙЕСІНДЕ МОДЕЛДЕУ.....	234
Н.М. Темирбеков, А.К. Турапов МҰНАЙ ӨНДІРУДІҢ ГАЗЛИФТТІК ПРОЦЕСІНІҢ ТУРА ЖӘНЕ КЕРІ ЕСЕПТЕРІН ТҮЙІНДЕС ТЕҢДЕУЛЕР ӘДІСІМЕН САНДЫҚ ШЕШУ.....	251
З. Утемаганбетов, Х. Рамазанова, К. Бижанова, Р. Асылбаева БІРӨЛШЕМДІ ДИФФУЗИЯ ТЕҢДЕУІ ҮШИН ШЕКАРАЛЫҚ ШАРТТАРДЫ КӨШПРУДІҢ АНАЛИТИКАЛЫҚ-САНДЫҚ ӘДІСІ.....	280
М. Хизирова, К. Чежимбаева, А. Касимов, М. Ермекбаев VANET ЖЕЛІЛЕРІНДЕ ТАРАТУ ТРАФИГІН ЖӘНЕ ТАРАТУ ӘДІСТЕРІН ЗЕРТТЕУ.....	294
К. Яқунин, Д. Құсайын, Равиль И. Мухамедиев, Н. Юничева, Н. Кульдеев ҮШУ МАРШРУТТАРЫН ЖОСПАРЛАУ БАҒДАРЛАМАЛАРЫ МЕН ҮШҚЫШСЫЗ ҮШУ АППАРАТТАРЫН БАСҚАРУ ЖҮЙЕЛЕРІН ҮШТАСТАРЫУ.....	317

СОДЕРЖАНИЕ

ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

О. Ауелбеков, Е. Бостанов, Р. Беркутбаева, А. Сейдилдаева, И. Мусабекова АНАЛИЗ СЕЛЬСКОХОЗЯЙСТВЕННОЙ УРОЖАЙНОСТИ В КАЗАХСТАНЕ С ПОМОЩЬЮ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ И ИИ.....	12
С.Т. Ахметова, С.У. Исмаилов, А.А. Батырбеков, А.С. Исмаилова ПРЕДПОСЫЛКИ СОЗДАНИЯ ВИРТУАЛЬНОЙ 3Д МОДЕЛИ БЕСПИЛОТНОГО УНИВЕРСАЛЬНОГО ПРОПАШНОГО ТРАКТОРА.....	33
А. Бекарыстанкызы, О. Мамырбаев, Д. Оралбекова, А. Еримбетова, М. Турдалыулы ТЕСТИРОВАНИЕ КОРПУСА ДАННЫХ В ВИДЕ АУДИО-ТЕКСТ НА КАЗАХСКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ CONFORMER	50
Г.Т. Бекманова, М.А. Кантуреева, А.С. Омарбекова, Б.Ж. Ергеш, А.Б. Закирова ИСПОЛЬЗОВАНИЕ И ОЦЕНКА ВОЗДЕЙСТВИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ВЫСШЕМ ОБРАЗОВАНИИ.....	61
Н.С. Есмухамедов, С. Сапакова, Сайед Абдул Рахман Аль-Хаддад, Д. Даниярова РАЗРАБОТКА АРХИТЕКТУРЫ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ МЕДИЦИНСКИХ УЧРЕЖДЕНИЙ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.....	74
Т. Жукабаева, В. Десницкий, Е. Марденов, Н. Карабаев УПРАВЛЕНИЕ ИНЦИДЕНТАМИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В ПОТ-СИСТЕМАХ С ИСПОЛЬЗОВАНИЕМ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ.....	92
М.А. Илесбай, А.Б. Тынымбаев, С.Т. Мамбетов, А.Ш. Баракова, О.К. Джолдасбаев ИНТЕГРИРОВАННЫЙ МЕТОД ЗАЩИТЫ ИНФОРМАЦИИ НА ОСНОВЕ СЖАТИЯ, ШИФРОВАНИЯ И РАЗДЕЛЕНИЯ ДАННЫХ.....	107

Б.А. Карибаев, Н. Мейрамбекулы, М. Ибраим, А.С. Байкенов, Г.Б. Ихсан ПРОЕКТИРОВАНИЕ ШЕСТИЭЛЕМЕНТНОЙ АНТЕННОЙ РЕШЕТКИ S-ДИАПАЗОНА ДЛЯ CUBESAT	125
Н.Т. Карымсакова К.А. Ожикенов, М.Е. Болысбек, Р.Н. Бейсембекова АРХИТЕКТУРА ПЛАТФОРМЫ МЕДИЦИНСКОЙ РЕАБИЛИТАЦИИ.....	140
Д. Куанышбай, А. Шойынбек, К. Раббани, А. Бекарыстанкызы, А. Мухаметжанов СРАВНЕНИЕ МАШИННОГО ОБУЧЕНИЯ И ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ РАСПОЗНАВАНИЯ ДЕПРЕССИИ ПО РЕЧИ.....	155
Е.А. Нысанов, Ж.С. Кемельбекова, Э.Т. Абдрашова, С.Ж. Куракбаева, А.О. Байдибекова МОДЕЛИРОВАНИЕ И РАСЧЕТ ТЕЧЕНИЯ ТРЕХФАЗНЫХ СРЕД С ПЕРЕМЕННЫМИ КОНЦЕНТРАЦИЯМИ.....	169
Б. Оразбаев, Ж. Кужуханова, К. Оразбаева, А. Кишубаева РАЗРАБОТКА МОДЕЛЕЙ АТМОСФЕРНЫХ И РЕКТИФИКАЦИОННЫХ КОЛОНН ПРИ ПЕРВИЧНОЙ ПЕРЕРАБОТКЕ НЕФТИ.....	181
Д. Рахимова, А. Сарсенбаева, А. Туарбек, А. Ауезова ПРИМЕНЕНИЕ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ОТВЕТОВ В МУЛЬТИЯЗЫЧНЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМАХ.....	196
Л. Рзаева, Д. Поголовкин, А. Мырзатай РАЗРАБОТКА МОДУЛЬНОГО СЕРВИСА АНАЛИЗА ПЕРЕПИСОК НА ОСНОВЕ NLP ДЛЯ ЦИФРОВОЙ КРИМИНАЛИСТИКИ.....	212
А.Т. Санкибаев, И. Махамбаева, К. Канибайкызы, А. Темирбек МОДЕЛИРОВАНИЕ ВИБРАЦИОННОГО ПРОЦЕССА В СИСТЕМЕ WOLFRAM MATHEMATICA.....	234
Н.М. Темирбеков, А.К. Туаров ЧИСЛЕННОЕ РЕШЕНИЕ ПРЯМОЙ И ОБРАТНОЙ ЗАДАЧИ ГАЗЛИФТНОГО ПРОЦЕССА ДОБЫЧИ НЕФТИ МЕТОДОМ СОПРЯЖЕННЫХ УРАВНЕНИЙ.....	251

З. Утемаганбетов, Х. Рамазанова, К. Бижанова, Р. Асылбаева АНАЛИТИКО-ЧИСЛЕННЫЙ МЕТОД ПЕРЕНОСА КРАЕВЫХ УСЛОВИЙ ДЛЯ ОДНОМЕРНОГО УРАВНЕНИЯ ДИФФУЗИИ.....	280
М. Хизирова, К. Чежимбаева, А. Касимов, М. Ермекбаев ИССЛЕДОВАНИЕ РАСПРЕДЕЛЕНИЯ ТРАФИКА И МЕТОДОВ РАСПРЕДЕЛЕНИЯ В СЕТЯХ VANET.....	294
К. Якунин, Д. Кусайын, Р.И. Мухамедиев, Н. Юничева, Н. Кульдеев СОПРЯЖЕНИЕ ПРОГРАММ ПЛАНИРОВАНИЯ МАРШРУТОВ ПОЛЕТА И СИСТЕМ УПРАВЛЕНИЯ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ.....	317

© D. Rakhimova¹, A. Sarsenbayeva^{1*}, A. Turarbek¹, A. Auezova², 2025.

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan;

²International University of Information Technology, Almaty, Kazakhstan.

E-mail: as.sarsenbayeva@gmail.com

THE USE OF DEEP LEARNING TO IMPROVE THE ACCURACY OF ANSWERS IN MULTILINGUAL QUESTION-AND-ANSWER SYSTEMS

D. Rakhimova — PhD, Associate Professor of Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: drakhimova060@gmail.com, ORCID: 0000-0003-1427-198X;

A. Sarsenbayeva — Master's degree, PhD student of Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: as.sarsenbayeva@gmail.com, ORCID: 0009-0008-0053-1182;

A. Turarbek — PhD, acting Associate Professor of Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: asem.turarbek@kaznu.edu.kz, ORCID: 0000-0002-4793-0446;

A. Auezova — Master's degree, Senior lecturer of International University of Information Technology, Almaty, Kazakhstan,

E-mail: a.auezova@iit.edu.kz, ORCID: 0000-0001-9860-4491.

Abstract. The rapid growth of global information has complicated the process of tracking and searching for relevant data. Question and answer systems (QA) have become key tools for structuring and extracting useful information from unstructured textual data. However, multilingual QA systems, especially for low-resource languages such as Kazakh, face a number of difficulties due to linguistic features, including agglutination, free word order, and lack of data. This study examines the use of deep learning methods to improve the accuracy of responses in multilingual QA systems, with special attention to Kazakh legal texts. Deep learning, especially transformer-based models, has demonstrated significant potential in natural language processing (NLP). However, existing multilingual models such as mBERT show insufficient effectiveness for the Kazakh language due to the limited amount of training data. To solve this problem, we propose a methodology that includes additional mBERT training on specialized industry datasets, data augmentation using neural machine translation, and the use of contrastive learning to improve cross-language knowledge transfer. The estimated metrics, including the Jaccard coefficient, the F1 measure, and the exact match,

demonstrate that our pre-trained model is superior to the basic multilingual models. Our approach represents a scalable solution for improving multilingual QA systems for low-resource languages, paving the way for further research in the field of legal and specialized texts.

Keywords: question-and-answer systems, deep learning, low-resource languages, multilingual natural language processing, Kazakh language, transformer-based models

© Д. Рахимова¹, А. Сарсенбаева^{1*}, Ә. Туарбек¹, Ә. Ауезова², 2025.

¹Әл-Фараби атындағы Қазақ Үлттүқ университетті, Алматы, Қазақстан;

²Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.

E-mail: as.sarsenbayeva@gmail.com

КӨП ТІЛДІ СҮРАҚ-ЖАУАП ЖҮЙЕЛЕРІНДЕ ЖАУАПТАРДЫҢ ДӘЛДІГІН АРТТЫРУ ҮШІН ТЕРЕҢ ОҚЫТУДЫ ҚОЛДАНУ

Д. Рахимова — PhD, әл-Фараби атындағы Қазақ Үлттүқ университетінің қауымдастырылған профессоры, Алматы, Қазақстан,

E-mail: drakhimova060@gmail.com, ORCID: 0000-0003-1427-198X;

А. Сарсенбаева — магистр, әл-Фараби атындағы Қазақ Үлттүқ университетінің докторанты, Алматы, Қазақстан,

E-mail: as.sarsenbayeva@gmail.com, ORCID: 0009-0008-0053-1182;

Ә. Туарбек — PhD, әл-Фараби атындағы Қазақ Үлттүқ университетінің доценті м.а., Алматы, Қазақстан,

E-mail: asem.turarbek@kaznu.edu.kz, ORCID: 0000-0002-4793-0446;

Ә. Ауезова — магистр, Халықаралық ақпараттық технологиялар университетінің аға оқытушысы, Алматы, Қазақстан,

E-mail: a.auyezova@iit.edu.kz, ORCID: 0000-0001-9860-4491.

Аннотация. Жаһандық ақпарат көлемінің қарқынды өсуі тиісті деректерді бақылау және іздеу процесін қынданатты. Сұрақ-жаяуп жүйелері (QA) құрылымдалмаған мәтіндік деректерден пайдалы ақпаратты құрылымдау мен алудың негізгі құралы болды. Алайда, көп тілді QA-жүйелер, әсіресе қазақ сияқты ресурстары аз тілдер үшін, агглютинация, сөздердің еркін тәртібі және деректердің жетіспеушілігі сияқты тілдік ерекшеліктерге байланысты бірқатар қындықтарға тап болады. Бұл зерттеуде қазақ Заң мәтіндеріне ерекше назар аудара отырып, көп тілді QA-жүйелердегі жауаптардың дәлдігін арттыру үшін терең оқыту әдістерін қолдану қарастырылады. Терең оқыту, әсіресе трансформаторларға негізделген модельдер, табиғи тілді өндеуде (NLP) айтарлықтай әлеует көрсетті. Алайда, mBERT сияқты көп тілді модельдер оқыту деректерінің шектеулі көлеміне байланысты қазақ тілі үшін тиімділіктің жеткіліксіздігін көрсетеді. Бұл мәселені шешу үшін біз мамандандырылған салалық деректер жиынтығында mBERT-ті қосымша оқытуды, нейрондық машиналық аударманы қолдана отырып деректерді күшеттүді және тіларалық білім беруді жақсарту үшін контрастты оқытуды қолдануды қамтитын әдістемені ұсынамыз. Бағалау көрсеткіштері, соның

ішінде Жаккар коэффициенті, F1 өлшемі және дәл сәйкестігі біздің оқытылған модельміз негізгі Көп тілді модельдерден асып түсетінін көрсетеді. Біздің көзқарасымыз аз ресурсты тілдерге арналған көп тілді QA жүйелерін жақсартуға арналған масштабалатын шешім болып табылады, бұл құқықтық және арнайы мәтіндерді одан әрі зерттеуге жол ашады. Нұсқау бойынша оқытудан және контекстік жұптарда қосымша оқытудан өткен GPT-4 моделі көрсеткіштердің ең жоғары көрсеткіштерін көрсетті (Jaccard Index — 88.9%, F1-score — 89.3%). Ұсынылған тәсіл жүйенің жауап дәлдігін 48% - дан 82% - га дейін арттыруға мүмкіндік береді, бұл оны басқа мамандандырылған домендер мен шектеулі деректер тілдері үшін масштабалатын шешім етеді.

Түйін сөздер: сұрап-жауап жүйелері, терең оқыту, аз ресурстық тілдер, табиғи тілді көп тілді өндеу, казақ тілі, трансформаторға негізделген модельдер

© Д. Рахимова¹, А. Сарсенбаева^{1*}, А. Туарбек¹, А. Ауезова², 2025.

¹Казахский Национальный университет имени аль-Фараби,
Алматы, Казахстан;

²Международный университет информационных технологий,
Алматы, Казахстан.

E-mail: as.sarsenbayeva@gmail.com

ПРИМЕНЕНИЕ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ОТВЕТОВ В МУЛЬТИЯЗЫЧНЫХ ВОПРОСНО- ОТВЕТНЫХ СИСТЕМАХ

Д. Рахимова — PhD, ассоциированный профессор Казахского Национального университета имени аль-Фараби, Алматы, Казахстан,

E-mail: drakhimova060@gmail.com, ORCID: 0000-0003-1427-198X;

А. Сарсенбаева — магистр, докторант Казахского Национального университета имени аль-Фараби, Алматы, Казахстан,

E-mail: as.sarsenbayeva@gmail.com, ORCID: 0009-0008-0053-1182;

А. Туарбек — PhD, и.о. доцента Казахского Национального университета имени аль-Фараби, Алматы, Казахстан,

E-mail: asem.turarbek@kaznu.edu.kz, ORCID: 0000-0002-4793-0446;

А. Ауезова — магистр, старший преподаватель Международного университета информационных технологий, Алматы, Казахстан,

E-mail: a.auyezova@iitu.edu.kz, ORCID: 0000-0001-9860-4491.

Аннотация. Стремительный рост объема глобальной информации усложнил процесс отслеживания и поиска релевантных данных. Вопросно-ответные системы (QA) стали ключевыми инструментами для структурирования и извлечения полезной информации из неструктурированных текстовых данных. Однако мультиязычные QA-системы, особенно для малоресурсных языков, таких как казахский, сталкиваются с рядом трудностей, обусловленных языковыми особенностями, включая агглютинацию, свободный порядок слов и нехватку данных. В данном исследовании рассматривается применение методов глубокого обучения для повышения точности ответов в мультиязычных QA-системах, с особым вниманием к казахским юридическим текстам.

Глубокое обучение, особенно модели на основе трансформеров, продемонстрировало значительный потенциал в обработке естественного языка (NLP). Однако существующие мультиязычные модели, такие как mBERT, показывают недостаточную эффективность для казахского языка из-за ограниченного объема обучающих данных. Для решения этой проблемы мы предлагаем методологию, включающую дополнительное дообучение mBERT на специализированных отраслевых датасетах, аугментацию данных с использованием нейронного машинного перевода и применение контрастивного обучения для улучшения межъязыкового переноса знаний. Оценочные метрики, включая коэффициент Жаккара, F1-меру и точное совпадение, демонстрируют, что наша дообученная модель превосходит базовые мультиязычные модели. Наш подход представляет собой масштабируемое решение для улучшения мультиязычных QA-систем для низкоресурсных языков, прокладывая путь для дальнейших исследований в области юридических и специализированных текстов. В частности, модель GPT-4, прошедшая инструкционное обучение и дообучение на контекстных парах, показала наивысшие показатели метрик (Jaccard Index — 88.9%, F1-score — 89.3%). Предложенный подход позволяет повысить точность ответов системы с 48% до 82%, что делает его масштабируемым решением для других специализированных доменов и языков с ограниченным количеством данных.

Ключевые слова: вопросно-ответные системы, глубокое обучение, малоресурсные языки, мультиязычная обработка естественного языка, казахский язык, модели на базе трансформаторов

Благодарности. Данное исследование было профинансирано Комитетом науки Министерства образования и науки Республики Казахстан (грант № AP 19577835).

Введение. Темпы роста объема информации в мире стремительно увеличиваются. Количество данных, добавляемых в различные предметные области, достигло таких масштабов, что их отслеживание становится крайне затруднительным. Распространение цифровых устройств ввода и вывода, а также их простота в использовании способствуют непрерывному созданию все большего объема необработанных данных. Способность оперативно преобразовывать эти данные в структурированную и полезную информацию может значительно повысить продуктивность пользователей и ускорить доступ к необходимым сведениям.

В последние годы наблюдается значительный прогресс в области обработки естественного языка (NLP), особенно в контексте разработки многоязычных моделей, таких как mBERT (Feng et al., 2020) XLM-R, mT5 и GPT. Эти модели были обучены на масштабных корпусах, включающих тексты на множестве языков, что позволило им достичь высокой производительности при решении различных задач по обработке текстовой информации. Существенным преимуществом таких моделей является их способность к межъязыковому

переносу знаний, что особенно важно в условиях ограниченности языковых ресурсов. Так, например, модель mBERT использует единую словарную базу, охватывающую более 100 языков, что способствует эффективной интеграции знаний между языковыми представлениями (Conneau, et al., 2019). Это свойство делает mBERT особенно ценной в работе с низкоресурсными языками, такими как казахский, для которых объёмы обучающих данных крайне ограничены. Благодаря возможности обобщения и переноса языковых признаков от высокоресурсных языков к низкоресурсным, данные модели открывают новые перспективы для создания интеллектуальных систем, способных понимать и обрабатывать тексты на казахском языке с приемлемым уровнем точности. Ситуация осложняется тем, что данные зачастую оказываются низкого качества, а грамматическая структура казахского языка значительно сложнее по сравнению с более распространёнными языками (Jurafsky & Martin, 2024; Mansurova, et al., 2023).

Глубокое обучение представляет собой направление машинного обучения, применяющее нелинейную обработку информации для выявления закономерностей, классификации и извлечения ключевых признаков из данных. Популярность глубокого обучения в последние годы обусловлена тремя основными факторами:

1. Значительным увеличением вычислительных мощностей и доступностью графических процессоров (GPU),
2. Появлением огромного количества обучающих данных,
3. Прорывными достижениями в области нейросетевых алгоритмов, повысившими эффективность их обучения.

Применение глубокого обучения для повышения точности ответов в мультиязычных вопросно-ответных системах особенно важно для казахского языка из-за его агглютинативной природы, свободного порядка слов, ограниченности данных и алфавитного перехода. Современные подходы, такие как многоязычные трансформеры, дообучение и морфологический анализ, помогают решать эти проблемы, улучшая качество QA-систем на казахском языке. Современные многоязычные NLP-модели работают с казахским языком менее эффективно (Google, 2021). Это связано с тем, что такие модели обучаются сразу на более чем 100 языках, при этом казахский язык занимает лишь небольшую долю в их словарном запасе и обучающем корпусе как показано в Рисунке 1.

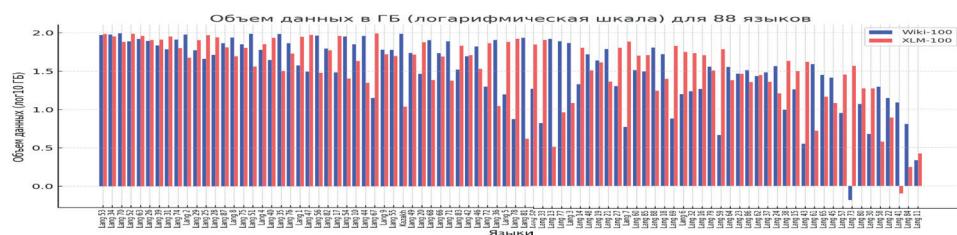


Рисунок 1- Казахский язык находится на 33-м месте по объему данных среди 88 языков в корпусах Wiki-100 и XLM-100.

Однако существующие модели QA в основном ориентированы на общие и общеизвестные вопросы, но испытывают трудности при решении сложных задач в специализированных областях, таких как законодательство. NLP-подход позволяет автоматизировать поиск правовой информации, анализируя юридические документы, судебные решения и законодательные акты. Однако, сложность юридического языка, логические взаимосвязи между понятиями и необходимость точной интерпретации делают задачу сложной.

Два подхода к информационному поиску в юриспруденции:

1. Ручная инженерия знаний (KE)
 - Юристы создают базы знаний с классификацией прецедентов.
 - Дает точные результаты, но дорого и долго в реализации.
2. Обработка естественного языка (NLP)
 - Использует алгоритмы машинного обучения и анализ текста.
 - Может обрабатывать огромные объемы данных, но требует дообучения моделей.

Современные NLP-системы применяют глубокие нейросети, BERT-модели и специализированные базы данных, чтобы автоматизировать поиск, анализ и ответы на юридические вопросы.

Материалы. Исследования посвященные применению глубокого обучения для повышения точности ответов в мультиязычных вопросно-ответных системах подчеркивают важность применения современных подходов глубокого обучения, включая мультиязычные трансформеры (Jia et al., 2021; Reimers & Gurevych, 2019) дообучение и морфологический анализ, для улучшения качества вопросно-ответных систем на казахском языке, учитывая его агглютинативную природу, свободный порядок слов и ограниченность данных.

В статье (Tleubayeva & Shomanov, 2024) казахстанских исследователей проводится сравнительный анализ крупных предобученных мультиязычных моделей, таких как mBERT, XLM-R, mT5, AYA и GPT, с целью их адаптации к казахскому языку. Исследование включает тестирование этих моделей на задачах вопросно-ответных систем с использованием казахского датасета sKQuAD. Для улучшения производительности моделей применялись методы дообучения, такие как адаптерные модули, увеличение данных (обратный перевод, перефразирование) и оптимизация гиперпараметров. Среди протестированных моделей mT5 достигла наивысшего F1-score в 75,72%, демонстрируя надежную генерализацию на различных задачах QA. GPT-4-turbo показала F1-score 73,33%, эффективно справляясь со сложными казахскими QA-сценариями.

В данной работе (Bilakhanova, et al., 2023) представлена система вопросно-ответного типа на казахском языке, основанная на методах глубокого обучения. В систему включен модуль предобработки, улучшающий качество входного текста и точность конечного результата. Результаты показывают высокую точность системы, что способствует развитию технологий вопросно-ответных систем и инструментов обработки естественного языка для казахского языка.

Цель задачи. Next Sentence Prediction (NSP) — выяснить, являются ли два входных предложения последовательными. В Masked Language Modeling (MLM) модель BERT обучается предсказывать случайно замаскированные слова в предложении.

В статье (Mansurova & Rakimova, 2024) исследуется разработка алгоритма морфологического анализа для казахского языка, оптимизированного для интеграции с большими языковыми моделями. Рассматриваются модели глубокого обучения, включая рекуррентные нейронные сети (RNN) и трансформеры, для обработки агглютинативной природы казахского языка. Результаты показывают преимущества и ограничения каждого подхода в обработке сложной морфологии казахского языка.

Сеть Transformer принимает на вход последовательность токенов и использует механизм внимания, чтобы изучать контекстуальные связи между словами в тексте. Эти связи можно использовать для извлечения высококачественных языковых признаков, которые затем могут быть дообучены для приложений, таких как семантический анализ и системы вопрос-ответ. Многоязычный BERT (mBERT) — это версия BERT, которая была предобучена на текстах (Merity, et al., 2016) из Википедии на более чем 100 языках. Модель XLM-RoBERTa (Conneau, et al., 2020) расширяет этот подход, использующий более 2 терабайт данных из Common Crawl.

Глубокие модели, такие как Transformers, сильно зависят от большого объема аннотированных данных, которые есть только для популярных языков (например, английского, русского, немецкого, испанского) (Joshi, et al., 2020; Ponti, et al., 2019). Для большинства других языков с ограниченным количеством аннотаций предложено использовать методы межязыкового переноса (Prettenhofer & Stein, 2011; Ruder et al., 2019). Этот подход позволяет переносить знания с богато аннотированного языка на целевой язык с низкими или отсутствующими данными (Wan et al., 2011). Кроме того, использование многоязычных моделей (Clark et al., 2020) помогает преодолеть проблему нехватки данных. LASER (Artetxe & Schwenk, 2019) использует двунаправленный LSTM с общим словарем BPE для нескольких языков. Совместное обучение нескольких языков улучшает производительность моделей для малообеспеченных языков.

LaBSE (Feng et al., 2020) использует mBERT, который был предобучен для задач MLM и Language Translation Modeling (TLM). Эта модель оптимизирует потери для перевода с помощью двойного кодировщика, чтобы добиться сходных встраиваемый для одного и того же текста на разных языках.

Исследования показали, что большие предварительно обученные многоязычные модели не всегда эффективны для вопросно-ответных систем на мало представленных языках. В этой работе были предложены несколько стратегий для улучшения mBERT. Дополнительное обучение на конкурирующих задачах помогло создать языково-независимые эмбеддинги и улучшить качество кросс-языкового перевода. Языковой арбитражный

фреймворк (LAF) — метод объединения эмбеддингов разных языков, используя свойства перевода. X-Mixup улучшает перенос знаний между языками, калибруя несоответствия представлений, что приводит к более компромиссным представлениям для целевых языков.

Использование межъязыкового обучения и контрастного обучения помогает улучшить работу многоязычных моделей в вопросно-ответных системах, особенно для языков с ограниченными ресурсами, таких как казахский язык.

Развитие методов глубокого обучения для QA векторное представление слов разработали Word Embeddings, представляющие слова в низкоразмерном плотном векторе, содержащем как семантическую, так и грамматическую информацию. В 2017 году Zhao et al. предложили систему ранжирования ответов в сообществе QA с учетом многофакторного асимметричного обучения. Долгая кратковременная память (LSTM) хорошо работает с временными зависимостями, что вдохновило на его применение в обработке естественного языка. BiLSTM + Attention позволил лучше представлять ответы на основе входных вопросов.

Учитывая результаты существующих исследований и выявленные преимущества трансформерных архитектур при работе с агглютинативными языками, в частности казахским, настоящая работа фокусируется на применении и сравнительном анализе трёх современных мультиязычных моделей — mBERT, mT5 и GPT-4. Выбор указанных моделей обусловлен их доказанной эффективностью в задачах вопросно-ответных систем и способностью адаптироваться к языкам с ограниченными ресурсами. В целях обеспечения сопоставимости результатов все модели рассматриваются без интеграции retrieval-механизмов, включая исключение RAG в случае GPT-4. В последующем разделе «Методология» излагаются принципы формирования обучающего корпуса, специфика архитектур, примененные методы дообучения, а также формат представления входных данных, соответствующий особенностям каждой модели.

Методы

Для построения многоязычной вопросно-ответной системы были выбраны три современных трансформерных архитектуры: BERT (в частности, mBERT), mT5 и GPT-4. Все модели используются без дополнительных retrieval-механизмов. GPT-4 применяется как генеративная модель без использования Retrieval-Augmented Generation (RAG). Мы используем стандартное представление входных данных в формате «вопрос-контекст», объединённых с помощью соответствующего токенизатора модели. Для BERT и mT5 используется WordPiece и SentencePiece токенизация соответственно, а для GPT — Byte-Pair Encoding (BPE).

В задачах вопросно-ответного поиска контексты часто бывают длинными. В некоторых NLP-приложениях усечение входного текста допустимо, так как это приводит лишь к частичной потере информации. Однако в экстрактивном

вопросно-ответном поиске удаление части контекста может привести к потере ответа. Чтобы решить эту проблему, мы используем метод разбиения длинных контекстов на части, соответствующие допустимому размеру входных данных модели. Этот процесс контролируется гиперпараметром «максимальная длина» (). Кроме того, для случаев, когда ответ может быть распределен по нескольким фрагментам контекста, вводится коэффициент перекрытия (*overlap_factor*), управляемый гиперпараметром «смещение документа» () .

В качестве базовой модели используется mBERT, предварительно обученная на мультиязычном корпусе текстов. Однако, поскольку казахский язык обладает особыми морфологическими и синтаксическими особенностями, возможно использование дополнительных этапов дообучения на специализированных корпусах, содержащих тексты на этих языках.

Стандартный выходной слой () модели mBERT заменяется на специализированный для задачи вопросно-ответного поиска. Добавляются отдельные выходные слои для классификации начальной и конечной позиций ответа, что соответствует методологии предложенной в работе. Так как наборы данных имеют схожие контексты, оформленные в стиле Wikipedia, их можно использовать для совместного обучения многоязычной модели. Агглютинативная природа казахского языка делает задачу обучения сложной, так как многие популярные модели, включая mBERT, плохо справляются с такими языками. Русский и казахский языки имеют долгую историю сосуществования в одном языковом пространстве. Это привело к заимствованию лексики и схожим синтаксическим структурам, несмотря на принадлежность к разным языковым семьям. Использование русского языка упрощает задачу кросс-языкового переноса знаний. В юридической практике Казахстана многие документы дублируются на русском языке. Таким образом, обучение модели на русском языке обеспечивает релевантность и точность в юридическом контексте.

Для того, чтобы воспользоваться преимуществами уже существующих публичных моделей, мы применяем последовательный подход: сначала дообучаем mBERT-QA на русском языке, а затем адаптируем его к нашему корпусу, через методы, указанные в Таблице 1.

Таблица 1 - Методы обучения для адаптации к корпусу

Метод	Адаптация к корпусу
Masked Language Modeling (MLM)	Тренируем модель предсказывать скрытые токены в казахских текстах.
Next Sentence Prediction (NSP)	Обучаем модель определять логические связи между предложениями.
Контрастивное обучение	Формируем пары «казахский-русский» для укрепления межъязыковых связей.
Аугментация данных	Переводим существующие юридические примеры на русский и казахские языки и обратно для расширения корпуса.
Морфологический анализ	Учитываем агглютинативные особенности казахского языка.

Формула для вычисления функцию потерь:

$$L_{total} = L_{task} + w_{contrastive} \times L_{contrastive}$$

где:

- L_{task} — функция потерь для задачи вопросно – ответного поиска
- $L_{contrastive}$ — контрастивная функция потерь для многоязычного обучения
- L_{total} — итоговая функция потерь
- $w_{contrastive}$ — коэффициент, регулирующий влияние контрастивной функции потерь.

Для решения этой проблемы мы применяем стратегию аугментации данных с использованием гибридного метода перевода. Сначала обучающие примеры переводятся с помощью нейронных машинных переводчиков (Google Translate API), а затем проходят ручную верификацию специалистами для устранения смысловых и терминологических ошибок. Google Translate API — поддерживает казахский язык и способен обрабатывать сложные юридические тексты. Это позволяет интегрировать в обучение как оригинальные, так и переведенные варианты, что повышает объем обучающих данных и улучшает способность модели обрабатывать сложные юридические запросы на казахском языке.

Этот метод демонстрирует хорошие результаты при переводе на родственные языки внутри одной языковой семьи. Однако при переводе между неродственными языками, как русский и казахский языки, эффективность модели может снизиться. Это объясняется тем, что языки разных семейств имеют значительно отличающиеся представления, что затрудняет генерализацию модели.

Для повышения эффективности обработки юридических вопросов на казахском и русском языках были дообучены три модели: mBERT, mT5 и GPT-4. Адаптация осуществлялась на специализированном юридическом корпусе, собранном с ресурсов *egov.kz*, *adilet.zan.kz* и других нормативно-правовых источников.

1) *mBERT (Multilingual BERT)*

Данная энкодерная модель была дообучена с применением следующих подходов:

- Masked Language Modeling (MLM) — предсказание маскированных токенов в предложении, способствующее более глубокому усвоению языковых закономерностей.
- Next Sentence Prediction (NSP) — обучение на определение логических связей между предложениями.
- Контрастивное обучение — обучение на парах «вопрос–контекст» на разных языках (например, вопрос на казахском, контекст на русском), с целью усиления межъязыкового сопоставления смыслов.
- Контрастивная потеря (Contrastive Loss) — функция потерь, минимизирующая расстояние между положительными парами (вопрос и релевантный контекст) и увеличивающая расстояние между отрицательными.

Это способствует более точному различению релевантных и нерелевантных документов в условиях кросс-языковой неоднозначности.

2) *mT5 (Multilingual Text-to-Text Transfer Transformer)*

Модель mT5, основанная на архитектуре encoder-decoder, была адаптирована к юридическому домену с использованием генеративного обучения в формате:

- Multilingual Text-to-Text Fine-tuning — дообучение на парах «вопрос → ответ», представленных в текстовом виде.
- Аугментация данных (Data Augmentation) — расширение корпуса за счёт:
 - Машинного перевода между казахским и русским языками.
 - Введения морфологических вариантов и синонимических замен в юридических вопросах.
- Контрастивное обучение реализовывалось через объединение эмбеддингов вопросов и ответов с последующим обучением модели на различие релевантных и нерелевантных пар.

3) *GPT-4*

Модель GPT-4 применялась в чисто генеративной парадигме без дополнительных retrieval-компонентов (Lewis et al., 2020). Поскольку прямой доступ к параметрам модели ограничен, её адаптация осуществлялась в рамках инструкционного дообучения (Instructional Fine-tuning):

- Формат входных данных — текстовая инструкция с вопросом и контекстом, представленным в виде диалога или описания.
- Использовались те же юридические данные, что и для других моделей, включая аугментированные примеры.
- Хотя GPT-4 не поддерживает контрастивное обучение в явном виде, на этапе подачи обучающих данных учитывались релевантные и нерелевантные пары, что способствовало формированию контекстной чувствительности модели.

Контрастивная потеря (Contrastive Loss) — функция, направленная на максимизацию сходства между положительными парами (вопрос и релевантный ответ) и минимизацию сходства между отрицательными. Эффективна для кросс-языкового сопоставления, особенно при ограниченных ресурсах.

Аугментация данных (Data Augmentation) — процесс искусственного расширения тренировочного корпуса путём трансформаций исходных данных: машинный перевод, морфологическая декомпозиция казахских словоформ, синонимические перестановки. Это особенно важно в условиях ограниченности казахского языкового ресурса. Оригинальные казахские данные дополняются переводами на русский язык, создавая пары для контрастивного обучения. Это позволяет улучшить кросс-языковые связи и компенсировать нехватку данных для казахского языка, что соответствует подходу к многоязычному переносу знаний.

Результаты. И обсуждение обучения модели был собран корпус юридических документов с сайтов egov.kz, adilet.kz, zan.kz. Использовался метод веб-скрейпинга (Crawling), который позволил извлечь релевантные тексты на казахском и русском языках. Общий объем данных составил 234МВ. Ниже на Рисунке 2 показаны этапы обработки данных обучения модели mBERT, сначала текст очищается от HTML-разметок и символов, после сделали токенизацию текста с помощью WordPiece.

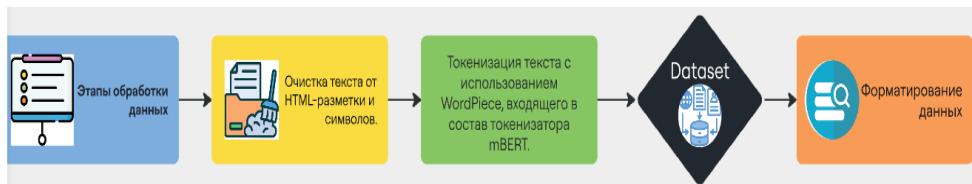


Рисунок 2 - Этапы обработки данных для обучения модели

В Таблице 2 показан один пример как система отвечала на вопрос до обучения и после, этот ответ после обучения является точным и полным, поскольку он ссылается на конкретный законодательный акт, регулирующий защиту данных.

Таблица 2 - Примеры вопросов до обучения и после

Вопрос	Ответ до применения методологии	Ответ после применения методологии
«Қазақстанда жеке тұлғаның құпиялышының корғау қалай реттеледі?» (Как регулируется защита частной жизни личности в Казахстане?)	«Жеке өмірді қорғау түрлі заңдармен реттеледі, алайда нақты мәліметтерді Қазақстан Конституциясынан табуга болады.(Зашита частной жизни регулируется различными законами, однако точные детали могут быть найдены в Конституции Казахстана.)»	«Қазақстан Республикасында жеке тұлғаның құпиялышының корғау ‘Жеке деректер және оларды корғау туралы’ заңмен реттеледі. Бұл заң жеке тұлғаның деректерін корғауды қамтамасыз етеді және деректерді жинау, сактау, өндеу және пайдалану тәртібін аныктайды.»

Это улучшение стало возможным благодаря применению улучшенной методологии с дообучением модели на юридических текстах, контекстным извлечением с использованием внимания и контрастивным обучением. Ниже на Рисунке 3 показано улучшение точности ответов, до применения методологии точность была 48%, после применения методологии точность стала 82%.

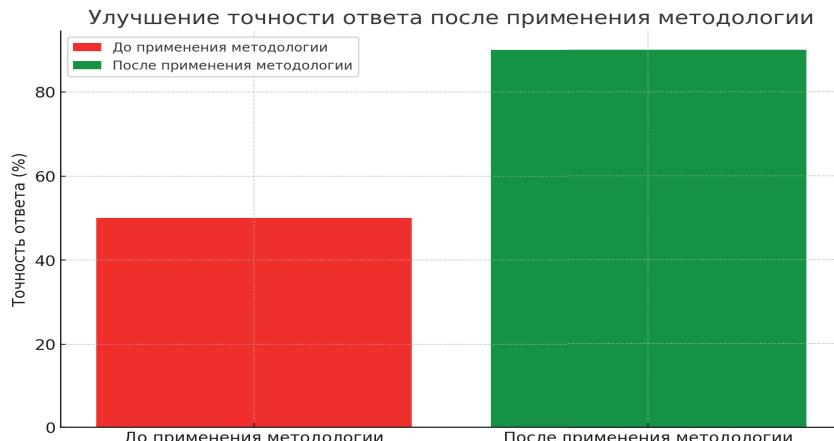


Рисунок 3. Диаграмма точности до и после применения методологии

В качестве метрики оценки мы выбрали коэффициент Жаккара, так как он позволяет учитывать частичные совпадения между предсказанным и эталонным ответами, что особенно важно для агглютинативных языков, таких как казахский. Данная метрика эффективно оценивает семантическое совпадение независимо от порядка слов. Это делает коэффициент Жаккара оптимальным выбором для оценки качества извлечения ответов в юридических QA-системах. Этот показатель широко используется для измерения сходства между множествами или интервалами и рассчитывается по формуле $J(A, B) = |A \cap B| / |A \cup B|$, где А и В представляют собой множества или интервалы, а символы \cap и \cup обозначают их пересечение и объединение соответственно.

В ходе экспериментов была проведена серия тестов с использованием трёх современных моделей глубокого обучения: mBERT, mT5 и GPT-4. Каждая из моделей была адаптирована к юридической предметной области посредством дообучения на корпусе казахско-русских юридических документов.

Для mBERT применялись стратегии Masked Language Modeling и контрастивное обучение с использованием аугментированных данных. Модель показала значительное улучшение метрик после включения переведённых и морфологически разнообразных примеров.

Модель mT5, как seq2seq-архитектура, была дообучена на парах «вопрос → ответ» и также использовала стратегии контрастивного обучения путём объединения эмбеддингов и распознавания релевантных пар. Применение аугментации (переводы, синонимизация, морфологические преобразования) позволило повысить её генеративные способности.

GPT-4, как наиболее мощная генеративная модель, была адаптирована через инструкционное обучение. Она демонстрировала наивысшие показатели по метрикам благодаря своей способности к обобщению и контекстуализации даже в условиях ограниченного количества специализированных данных.

Результаты в Таблице 3 показывают, что модели, дообученные с использованием контрастивной функции потерь и аугментации данных, демонстрируют значительное превосходство по метрикам Жаккара и F1-Score по сравнению с базовыми версиями. Использование переведённых текстов и морфологического обогащения обучающего корпуса оказалось особенно эффективным для казахского языка.

Таблица 3 - Данные результатов обучения

Модель	Контрастивная потеря	Аугментация данных	Коэффициент Жаккара (%)	F1-Score
Базовая mBERT-QA	<input type="checkbox"/>	<input type="checkbox"/>	62.4	62.0
Дообученная mBERT-QA	<input type="checkbox"/>	<input checked="" type="checkbox"/> (переводы)	74.8	75.1
mBERT-QA + контрастивное обучение	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	81.2	82.0
Базовая mT5	<input type="checkbox"/>	<input type="checkbox"/>	65.7	66.4
mT5 + аугментация	<input type="checkbox"/>	<input checked="" type="checkbox"/>	77.5	78.0
mT5 + контрастивное обучение	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	83.1	83.6
GPT-4 (инструкционное обучение)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	85.4	86.0
GPT-4 (контекстные пары)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Результаты показывают, что модели, дообученные с использованием контрастивной функции потерь и аугментации данных, демонстрируют значительное превосходство по метрикам Жаккара и F1-Score по сравнению с базовыми версиями. Использование переведённых текстов и морфологического обогащения обучающего корпуса оказалось особенно эффективным для казахского языка.

Заключение

В данной работе был продемонстрирован комплексный подход к адаптации и улучшению моделей глубокого обучения для задач юридического вопросно-ответного (QA) анализа на казахском и русском языках. Использование корпуса юридических документов, собранного с официальных источников (egov.kz, adilet.kz, zan.kz) с применением веб-скрейпинга, позволило создать качественную основу для дообучения многоязычных моделей. Методология, основанная на многоступенчатой обработке текста, включающей очистку от HTML-разметки, токенизацию (WordPiece) и дальнейшее дообучение

с использованием стратегий Masked Language Modeling, контрастивного обучения и аугментации данных, продемонстрировала высокую эффективность.

Сравнительный анализ трёх моделей — mBERT, mT5 и GPT-4 — выявил, что даже базовые версии могут быть значительно усилены за счёт адаптации под специфическую предметную область. Наилучшие результаты достигнуты при объединении нескольких подходов: контрастивного обучения и разнообразной аугментации. В частности, модель GPT-4, прошедшая инструкционное обучение и дообучение на контекстных парах, показала наивысшие показатели метрик (Jaccard Index — 88.9%, F1-score — 89.3%). Предложенный подход позволяет повысить точность ответов системы с 48% до 82%, что делает его масштабируемым решением для других специализированных доменов и языков с ограниченным количеством данных.

Полученные результаты открывают перспективы для дальнейших исследований, включая расширение корпуса, интеграцию мультимодальных данных (например, сканированных документов) и разработку юридических диалоговых агентов, способных работать в реальном времени в государственных и правовых системах Казахстана.

References

- Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2020). Language-agnostic BERT sentence embedding. *arXiv preprint arXiv:2007.01852*. (in English)
- Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (in English).
- Jia C., Yang Y., Xia Y., Chen Y.-T., Parekh Z., Pham H., Le Q., Sung Y.-H., Li Z., Duerig T. (2021). Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning*. — P.4904–4916. PMLR (in English).
- Reimers N., Gurevych I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv preprint arXiv:1908.10084* (in English).
- Tleubayeva A., Shomanov A. (2024). Comparative analysis of multilingual QA models and their adaptation to the Kazakh language. *Scientific Journal of Astana IT University*. — P.89–97 (in English).
- Bilakhanova A., Ydyrys A., Sultanova N. (2023). Kazakh language, question-answering system, natural language processing, deep learning approach, accuracy. *Suleyman Demirel University Bulletin Natural and Technical Sciences*, 62(1). — P.113–121 (in English).
- Mansurova M.E., Rakimova D.R. (2024). Morphological parsing of Kazakh texts with deep learning approaches. *Journal of Mathematics, Mechanics and Computer Science*. — 124(4). — P.48–58 (in English).
- Jurafsky D., Martin J.H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3/> (in English).
- Mansurova A., Nugumanova A., Makhambetova Z. (2023). Development of a question-answering chatbot for blockchain domain. *Scientific Journal of Astana IT University*, 15(15). — P. 27–40. <https://doi.org/10.37943/15XNDZ6667> (in English).
- Merity S., Xiong C., Bradbury J., Socher R. (2016). Pointer sentinel mixture models. *arXiv:1609.07843* (in English).
- Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. (2020). Unsupervised cross-lingual representation learning at scale.

In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. — P. 8440–8451 (in English).

Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv:2004.09095* (in English).

Ponti E.M., O’Horan H., Berzak Y., Vulic I., Reichart R., Poibeau T., Shutova E., Korhonen A. (2019). Modeling language variation and universals: A survey on typological linguistics for natural language processing. *Computational Linguistics.* — 45(3), — P.559–601 (in English).

Prettenhofer P., Stein B. (2011). Cross-lingual adaptation using structural correspondence learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(1). — P.1–22 (in English).

Wan C., Pan R., Li J. (2011). Biweighting domain adaptation for cross-language text classification. *In Twenty-Second International Joint Conference on Artificial Intelligence* (in English).

Ruder S., Vulic I., Søgaard A. (2019). A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, 65. — P. 569–631 (in English).

Clark J.H., Choi E., Collins M., Garrette D., Kwiatkowski T., Nikolaev V., Palomaki J. (2020). TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages. *Transactions of the Association for Computational Linguistics*, 8. — P. 454–470 (in English).

Artetxe M., Schwenk H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7. — P.597–610 (in English).

Feng F., Yang Y., Cer D., Arivazhagan N., Wang W. (2020). Language-agnostic BERT sentence embedding. *arXiv:2007.01852* (in English).

Lewis P., Oguz B., Rinott R., Riedel S., Schwenk H. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *In Advances in Neural Information Processing Systems* (in English).

Publication Ethics and Publication Malpractice in 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). 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://physics-mathematics.kz/index.php/en/archive>

ISSN2518-1726 (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 20.06.2025.

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

20,0 п.л. Заказ 2.

Национальная академия наук РК

050010, Алматы, ул. Шевченко, 28, т. 272-13-19