

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

**ACADEMIC SCIENTIFIC
JOURNAL OF COMPUTER SCIENCE**

**№2
2026**

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



ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

3 (359)

JULY – SEPTEMBER 2026

**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, (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>

MAMYRBAEV 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>

USATOVA Olga Alexandrovna, PhD, Associate Professor, Chief Scientific Secretary of the Institute of Information and Computing Technologies of the National Academy of Sciences of the Republic of Kazakhstan (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

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>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: «Central Asian Academic Research Center» LLP (Almaty).

Certificate № KZ77VPY00121154 on the re-registration of the periodical printed and online publication of the information agency, issued on 05.06.2025 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.

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

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат 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>

БАЙГУНЧЕКОВ Жұмаіл Жаңабайұлы, 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>

УСАТОВА Ольга Александровна, PhD, қауымдастырылған профессор, ҚР ҒЖБМ "Ақпараттық және есептеу технологиялары институтының" бас ғалым хатшысы (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нұрсұлу Алдажарқызы, 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>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (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 **05.06.2025 ж.** берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

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

Мерзімділігі: *жылына 4 рет.*

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

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимканр Мутанович, доктор технических наук, профессор, академик НАН 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 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>

УСАТОВА Ольга Александровна, PhD, ассоциированный профессор, Главный ученый секретарь «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

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>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № **KZ77VPY00121154**. Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

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

Периодичность: *4 раза в год.*

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

© ТОО «Центрально-азиатский академический научный центр», 2026

CONTENTS

Akhmetova A.Zh., Kantureyeva M.A., Serimbetov B.A., Zhasuzakova M.Zh., Mustafayeva N.T. Identification of communities and active users in social networks.....	11
Akzhigitov D., Mukhtarkhanov M., Zhetpisbayeva A., Kulakayeva A., Zhetpisbayev K. Analysis of the mechanical structure of a 12U CubeSat nanosatellite and assessment of its mechanical stability.....	29
Alibekov A., Nurimanov A., Mukhametkali A., Kassym D., Varol H.A. Qolda-Math: a small vision-language model for mathematical reasoning in Kazakh.....	50
Alisheva S.S., Ospanova A.O., Shaimerdenova G.S., Shaimerdenova L., Zhidebayeva A.N. Detecting Kazakh language offensive comments in online content using deep learning algorithms.....	68
Askar G., Mutaikhan A., Aitkozhayeva M., Nauryzbaikyzy G., Makhataeva Z. Retrieval beats fine-tuning: adapting a Kazakh small language model for historical question answering.....	86
Beissenova G.I., Shynzhigit B.B., Beisebayeva A.Zh., Kaldarova B.S., Lesbayev A.U. Deep neural network-based speaker authentication for secure access control systems.....	106
Bekenuly B., Baimakhanova A., Baiturganova V., Bolatbekova D. Comparison of CNN architectures for automated ACL injury classification from knee MRI.....	123
Bekbolat A., Razak A., Alimbayeva Z., Ozhikenova A. Multi-task deep learning for knee MRI pathology classification.....	139
Issakhov A., Akylbekkyzy A., Abylkassymova A. Predicting the spread of airborne viral infections in enclosed spaces.....	158
Kisseleva O., Abdramanov K., Naumenko V., Mukanova Z., Maintser D. Evaluation of serverless cold start optimization techniques.....	177
Koshkarov Ye.Zh., Amirgaliyev B.Ye., Mulesa O. Physics-aware masked pretraining for cross-basin well log foundation models.....	197
Kurmetkan T., Turysbek Zh., Mukhitova A., Nauryzbaeva A. Research on methods for reducing knowledge conflicts in large language models in Kazakh question-and-answer systems.....	214
Makatov Ye.K., Abdul Razaque, Ismukanova A.N., Salykova A.K., Glock E.S. Machine learning-based model for phishing URL detection in online social networks.....	233
Mimenbayeva A.B., Naizagarayeva A.A., Ospanova T.T., Koikelova D.K., Kudabayeva R.N. Spatial and climate data analysis for machine learning-based crop yield prediction in Northern Kazakhstan.....	255
Mukhamediev R.I., Abdoldina F.N., Gertsen Y.A. Economic opportunities and constraints of artificial intelligence adoption in Kazakhstan's resource-oriented economy.....	274

Mutanov G.M., Karyukin V.I., Ussatova O.A., Begimbayeva Y.Y., Baispay G.B. The assessment of blockchain readiness for academic document verification.....	289
Mussiraliyeva Sh., Bolatbek M., Baisylbayeva K. Classification of cyber-extremist ideologies in the Kazakh language considering multi-annotator disagreement.....	306
Omarov B.S., Sultan D.R., Rakhymberdiev D.S., Abibullayeva A.A., Toktarova A.B. Deformable attention-guided Transformer for brain tumor segmentation.....	326
Oralbekova D., Bekarystankyzy A., Yerimbetova A., Berzhanova U. Development of a streaming Kazakh speech recognition system using end-to-end models.....	346
Serek A., Shoiynbek A., Zheludkov M., Kuanyshbay D., Meraliyev B. Comparative analysis of machine learning methods for detecting fraud in text messages.....	365
Smakhanova A.K., Uskenbayeva R.K., Kassymova A.B., Kalpeyeva Zh.B., Konyrbaev N.B. Experimental research using models and methods of the university digital ecosystem.....	384
Sultanali A.E., Kazbekova G.N. Machine learning-based detection of AI-generated content in students' written works.....	401
Tashenova Zh., Yergazin A., Nurlybaeva E., Abdugulova Zh., Amanzholova Sh. Experimental evaluation of a hybrid IDPS with adaptive learning for low-latency 5G networks using edge computing.....	418
Telman D., Yerimbetova A., Daiyrbayeva E. Leakage-resistant fake news detection for Kazakh: temporal evaluation of lexical, Transformer and hybrid models.....	441
Zhanegizov A., Mamyrbayev O., Abdigappar M. Application of YOLO12 for road defect detection in GPR images.....	457

МАЗМҰНЫ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасұзақова М.Ж., Мұстафаева Н.Т. Әлеуметтік желілердегі қауымдастықтар мен белсенді қолданушыларды анықтау.....	11
Ақжігітов Д.Ж., Мұхтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. 12U CubeSat үлгісіндегі наноспутниктің механикалық құрылымын талдау және оның механикалық орнықтылығын бағалау.....	29
Алибеков А., Нуриманов А., Мұхаметқали А., Қасым Д., Варол Х.А. Qolda-Math: қазақ тіліндегі математикалық пайымдауға арналған кіші көру-тілдік модель.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Онлайн-контенттегі қазақ тіліндегі бейәдеп пікірлерді терең оқыту алгоритмдері арқылы анықтау.....	68
Асқар Ғ., Мұтайхан А., Айтқожаева М., Наурызбайқызы Г., Махатаева Ж. Retrieval fine-tuning-тен басым: тарихи сұрақтарға жауап беру үшін қазақша шағын тілдік модельді бейімдеу.....	86
Бейсенова Г.И., Шынжігіт Б.Б., Бейсебаева А.Ж., Қалдарова Б.С., Лесбаев А.У. Терең нейрондық желіге негізделген қауіпсіз басқару жүйелерінде сөйлеушіні аутентификациялау.....	106
Бекенұлы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Алдыңғы крест тәрізді байлам жарақатын анықтауға арналған CNN архитектураларын салыстыру.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Тізе MPT патологиясын жіктеуге арналған көп тапсырмалы терең оқыту.....	139
Исахов А.А., Ақылбекқызы А., Абылкасымова А. Шектеулі кеңістіктерде әуе тамшыларымен вирустық инфекцияның таралуын сандық модельдеу.....	158
Киселева О., Абдраманов Қ., Науменко В., Мұқанова Ж., Майнцер Д. Serverless жүйелеріндегі суық іске қосуды оңтайландыру әдістерін эксперименттік бағалау.....	177
Кошкарров Е.Ж., Амиргалиев Б.Е., Мүлеса О. Бассейнаралық ұңғымалық диаграммалардың іргелі модельдерін физикалық тұрғыдан негізделген масқаланған алдын ала оқыту.....	197
Құрметқан Т., Тұрысбек Ж., Мұхитова А., Наурызбаева А. Қазақ тіліндегі сұрақ-жауап жүйелерінде үлкен тілдік модельдердегі білім қайшылықтарын азайту әдістерін зерттеу.....	214
Мақатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глюк Е.С. Онлайн әлеуметтік желілердегі фишингтік URL-дарды анықтауға арналған машиналық оқыту негізіндегі модель.....	233

Мименбаева А.Б., Найзағарова А.А., Оспанова Т.Т., Қойкелова Д.К., Құдабаева Р.Н. Солтүстік Қазақстандағы дақылдар өнімділігін машиналық оқыту арқылы болжауда кеңістіктік және климаттық деректерді талдау.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Қазақстанның ресурстық бағытталған экономикасына жасанды интеллектті енгізудің экономикалық мүмкіндіктері мен шектеулері.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Білім беру құжаттарының түпнұсқалығын тексеруге блокчейн технологиясының дайындығын бағалау.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Көпаннынотаторлы келіспеушілікті ескере отырып қазақ тіліндегі киберэкстремистік идеологияларды жіктеу.....	306
Омаров Б.С., Сұлтан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Тоқтарова А.Б. Ми ісіктерін сегменттеуде деформацияланатын зейін механизміне негізделген мультимодальды трансформер.....	326
Оралбекова Д.О., Бекарыстанқызы А., Еримбетова А.С., Бержанова У.Г. Интегралды модельдерді қолдану арқылы қазақ тіліндегі ағындық сөйлеуді тану жүйесін әзірлеу.....	346
Серек А., Шойынбек А., Желудков М., Қуанышбай Д., Мералиев Б. Мәтіндік хабарламалардағы алаяқтықты анықтауға арналған машиналық оқыту әдістерінің салыстырмалы талдауы.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Қоңырбаев Н.Б. Университеттің цифрлық экожүйесінің модельдері мен әдістерін қолдана отырып жүргізілген эксперименттік зерттеу.....	384
Сұлтанәлі А.Е., Казбекова Г.Н. Білім алушылардың жазбаша жұмыстарындағы AI-контентті машиналық оқыту арқылы анықтау.....	401
Ташенова Ж.М., Ергазин Ә., Нұрлыбаева Ә.Н., Абдуғұлова Ж.К., Аманжолова Ш.А. Edge Computing көмегімен төмен кідірісті 5G желілеріне арналған бейімделгіш оқытуы бар гибриді IDPS-ті эксперименттік бағалау.....	418
Тельман Д., Еримбетова А., Дайырбаева Ә. Қазақ тіліндегі жалған жаңалықтарды анықтау: лексикалық, трансформерлік және гибридті модельдерді уақыт бойынша бағалау.....	441
Жанегизов А.С., Мамырбаев Ә.Ж., Абдіғаппар М.Д. YOLO12 моделін георадар кескіндеріне қолдану.....	457

СОДЕРЖАНИЕ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасузакова М.Ж., Мустафаева Н.Т. Выявление сообществ и активных пользователей в социальных сетях.....	11
Акжигитов Д.Ж., Мухтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. Анализ механической конструкции наноспутника модели 12U CubeSat и оценка его механической устойчивости.....	29
Алибеков А., Нуриманов А., Мухаметкали А., Касым Д., Варол Х.А. Qolda-Math: маленькая мультимодальная модель зрения и языка для математических рассуждений на казахском языке.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Выявление оскорбительных комментариев на казахском языке в онлайн-контенте с помощью алгоритмов глубокого обучения.....	68
Аскар Г., Мутайхан А., Айткожаева М., Наурызбайкызы Г., Махатаева Ж. Retrieval эффективнее fine-tuning: адаптация малой языковой модели казахского языка для ответов на исторические вопросы.....	86
Бейсенова Г.И., Шынжигит Б.Б., Бейсебаева А.Ж., Калдарова Б.С., Лесбаев А.У. Аутентификация личности по голосу на основе глубокой нейронной сети для систем безопасного управления доступом.....	106
Бекенулы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Сравнение архитектур сверточных нейронных сетей для автоматической классификации разрыва ПКС по МРТ коленного сустава.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Многозадачное глубокое обучение для классификации патологий колена по МРТ.....	139
Исахов А.А., Акылбеккызы А., Абылкасымова А. Прогнозирование распространения вирусной инфекции воздушно-капельным путем в замкнутых помещениях.....	158
Киселева О., Абдраманов К., Науменко В., Муканова Ж., Майнцер Д. Экспериментальная оценка методов оптимизации холодного запуска в serverless-системах.....	177
Кошкарар Е.Ж., Амиргалиев Б.Е., Мулеса О. Физически-осознанное маскированное предобучение для межбассейновых фундаментальных моделей каротажных диаграмм.....	197
Курметкан Т., Турысбек Ж., Мухитова А., Наурызбаева А. Исследование методов снижения конфликтов знаний в больших языковых моделях казахских вопросно-ответных систем.....	214
Макатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глок Е.С. Модель обнаружения фишинговых URL в онлайн-социальных сетях на основе машинного обучения.....	233

Мименбаева А.Б., Найзагараева А.А., Оспанова Т.Т., Койкелова Д.К., Кудабаяева Р.Н. Анализ климатических данных для прогнозирования урожайности пшеницы методами машинного обучения в Северном Казахстане.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Экономические возможности и ограничения внедрения искусственного интеллекта в ресурсно-ориентированной экономике Казахстана.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Оценка готовности технологии блокчейн к проверке подлинности документов об образовании.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Классификация киберэкстремистских идеологий на казахском языке с учетом разногласий нескольких аннотаторов.....	306
Омаров Б.С., Султан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Токтарова А.Б. Мультимодальный трансформер на основе механизма деформируемого внимания для сегментации опухолей головного мозга.....	326
Оралбекова Д.О., Бекарыстанкызы А., Еримбетова А.С., Бержанова У.Г. Распознавание казахской речи с применением интегральных моделей.....	346
Серек А., Шойынбек А., Желудков М., Куанышбай Д., Мералиев Б. Сравнительный анализ методов машинного обучения для выявления мошенничества в текстовых сообщениях.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Конырбаев Н.Б. Экспериментальное исследование с использованием моделей и методов цифровой экосистемы университета.....	384
Султанали А.Е., Казбекова Г.Н. Обнаружение AI-контента в письменных работах обучающихся методами машинного обучения.....	401
Ташенова Ж.М., Ергазин А., Нурлыбаева Э.Н., Абдугулова Ж.К., Аманжолова Ш.А. Экспериментальная оценка гибридной системы обнаружения и предотвращения вторжений (IDPS) с адаптивным обучением для сетей 5G с низкой задержкой с использованием граничных вычислений.....	418
Тельман Д., Еримбетова А., Дайырбаева Э. Обнаружение фейковых новостей на казахском языке: временная оценка лексических, трансформерных и гибридных моделей.....	441
Жанегизов А.С., Мамырбаев О.Ж., Абдигаппар М.Д. Применение модели YOLOv12 для анализа данных георадара.....	457

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

ISSN 1991-346X

Volume 3.

Number 359 (2026). 68–85

<https://doi.org/10.32014/2026.2518-1726.452>

IRSTI 28.23.37

UDC 004.8

© Alisheva S.S.¹, Ospanova A.O.¹, Shaimerdenova G.S.¹, Shaimerdenova L.¹,
Zhidebayeva A.N.^{2*}, 2026.

¹ M. Auezov South Kazakhstan University, Shymkent, Kazakhstan;

² Peoples Friendship University named after Academician A. Kuatbekov,
Shymkent, Kazakhstan.

E-mail: aziza_68.kz@mail.ru

DETECTING KAZAKH LANGUAGE OFFENSIVE COMMENTS IN ONLINE CONTENT USING DEEP LEARNING ALGORITHMS

Alisheva Sandugash — Phd, Mukhtar Auezov South Kazakhstan University, Shymkent, Kazakhstan,
E-mail: sandu_alish@mail.ru, <https://orcid.org/0000-0001-6887-9208>;

Ospanova Aiman — professor, Mukhtar Auezov South Kazakhstan University, Shymkent, Kazakhstan,

E-mail: oaim09@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Shaimerdenova Guldana — Phd, associate professor, Mukhtar Auezov South Kazakhstan University, Shymkent, Kazakhstan,

E-mail: danel101kz@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Shaimerdenova Lazzat — Phd, Mukhtar Auezov South Kazakhstan University, Shymkent, Kazakhstan,

E-mail: Lizzat71@mail.ru, <https://orcid.org/0000-0002-5656-0918>;

Zhidebayeva Aziza — Candidate of Technical Sciences, Senior Lecturer, department of Computer science and Mathematics, Peoples Friendship University named after Academician A. Kuatbekov, Shymkent, Kazakhstan,

E-mail: aziza_68.kz@mail.ru, <https://orcid.org/0000-0002-3768-4835>.

Abstract. Nowadays, the rapid growth of social media and online communication platforms has led to a significant increase in offensive and abusive comments, creating serious psychological, social, and security challenges. Offensive language, cyberbullying, and hate speech negatively affect users' mental well-being and may result in anxiety, depression, social isolation, and reduced quality of life. According to recent surveys, approximately 75% of young people in Kazakhstan have experienced different forms of online bullying, highlighting the urgent need for intelligent automated content moderation systems for the Kazakh language. This study proposes a deep learning framework for detecting offensive comments in the Kazakh language by combining a Bidirectional Long Short-Term Memory network with a custom Kazakh stemming algorithm and linguistic preprocessing

techniques. To evaluate the proposed approach, a balanced dataset consisting of 40,000 manually annotated Kazakh comments (20,000 offensive and 20,000 neutral) was collected and processed using an expanding stemming dictionary. Experimental evaluation was performed using Accuracy, Precision, Recall, and F1-score metrics. Among all evaluated models, the proposed BiLSTM achieved the best overall performance with an Accuracy of 91.3%, Precision of 92.6%, Recall of 93.0%, and an F1-score of 92.7%. In contrast, the Naïve Bayes classifier produced the lowest performance, with an Accuracy of 60.2%, Precision of 52.4%, Recall of 58.5%, and an F1-score of 64.2%, demonstrating the effectiveness of deep learning methods for processing morphologically rich Kazakh texts. The proposed framework provides an effective solution for automatic offensive comment detection and can be integrated into social media platforms and online services to support cyberbullying prevention, intelligent content moderation, and safer digital communication environments for Kazakh-speaking users.

Keywords: Hate speech detection, deep learning, online bullying, Application Programming Interface, social media, text, database

For citations: Alisheva S.S., Ospanova A.O., Shaimerdenova G.S., Shaimerdenova L., Zhidebayeva A.N. Detecting kazakh language offensive comments in online content using deep learning algorithms. Academic Scientific Journal of Computer Science, 2026. — No.3. — P. 68–85. DOI: <https://doi.org/10.32014/2026.2518-1726.452>

© Алишева С.С.¹, Оспанова А.О.¹, Шаймерденова Г.С.¹, Шаймерденова Л.¹, Жидебаева А.Н.^{2*}, 2026.

¹ М. Өуезов атындағы Оңтүстік Қазақстан университеті,
Шымкент, Қазақстан;

² Академик А. Қуатбеков атындағы Халықтар достығы университеті,
Шымкент, Қазақстан.
E-mail: aziza_68.kz@mail.ru

ОНЛАЙН КОНТЕНТТЕ ҚАЗАҚ ТІЛДІ БЕЙӘДЕП ПІКІРЛЕРДІ ТЕРЕҢ ОҚЫТУ АЛГОРИТМДЕРІ АРҚЫЛЫ АНЫҚТАУ

Алишева Сандуғаш — PhD, М. Өуезов атындағы Оңтүстік Қазақстан университеті,
Шымкент, Қазақстан,

E-mail: sandu_alish@mail.ru, <https://orcid.org/0000-0001-6887-9208>;

Оспанова Айман — профессор, М. Өуезов атындағы Оңтүстік Қазақстан университеті,
Шымкент, Қазақстан,

E-mail: oaим09@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Шаймерденова Гүлдана — PhD, қауымдастырылған профессор, М. Өуезов атындағы
Оңтүстік Қазақстан университеті, Шымкент, Қазақстан,

E-mail: danel101kz@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Шаймерденова Ләззат — PhD, М. Өуезов атындағы Оңтүстік Қазақстан университеті,
Шымкент, Қазақстан,

E-mail: Lizzat71@mail.ru, <https://orcid.org/0000-0002-5656-0918>;

Жидебаева Азиза — техника ғылымдарының кандидаты, Информатика және математика кафедрасының аға оқытушысы, Академик А. Қуатбеков атындағы Халықтар достығы университеті, Шымкент, Қазақстан,

E-mail: aziza_68.kz@mail.ru, <https://orcid.org/0000-0002-3768-4835>.

Аннотация. Әлеуметтік желілер мен онлайн-коммуникация платформаларының қарқынды дамуы бейәдеп және қорлайтын пікірлер санының айтарлықтай өсуіне алып келіп, психологиялық, әлеуметтік және қауіпсіздікке қатысты бірқатар мәселелерді туындатуда. Бейәдеп тіл, кибербуллинг және өшпенділік тілі пайдаланушылардың психологиялық әлауқатына кері әсер етіп, мазасыздыққа, әлеуметтік оқшаулануға және онлайн-коммуникация сапасының төмендеуіне себеп болуы мүмкін. Зерттеуде қазақ тіліндегі бейәдеп пікірлерді автоматты түрде анықтауға арналған, екі бағытты ұзақ қысқа мерзімді жады желісін (BiLSTM), қазақ тіліне бейімделген стемминг алгоритмін және лингвистикалық алдын ала өңдеу әдістерін біріктіретін терең оқыту тәсілі ұсынылған. Зерттеу үшін қолмен белгіленген 40 000 қазақ тіліндегі пікірден тұратын теңгерімді деректер жиынтығы қалыптастырылды, оның 20 000-ы бейәдеп, 20 000-ы бейтарап пікірлерден тұрады. Деректер кеңейтілетін стемминг сөздігін пайдалану арқылы өңделді. Ұсынылған тәсіл Accuracy, Precision, Recall және F1-score көрсеткіштері негізінде бағаланды. Қарастырылған модельдердің ішінде BiLSTM ең жоғары нәтижелер көрсетті: Accuracy - 91,3%, Precision - 92,6%, Recall - 93,0%, F1-score - 92,7%. Ал Naïve Bayes классификаторы ең төмен нәтижелер көрсетті: Accuracy - 60,2%, Precision - 52,4%, Recall - 58,5%, F1-score - 64,2%. Алынған нәтижелер морфологиялық құрылымы күрделі қазақ тіліндегі мәтіндерді өңдеуде терең оқыту әдістерінің тиімділігін көрсетеді. Ұсынылған тәсілді әлеуметтік желілер мен онлайн-сервистерге интеграциялау автоматты контент модерациясын жетілдіруге, кибербуллингің алдын алуға және қазақтілді пайдаланушылар үшін қауіпсіз цифрлық коммуникация ортасын қалыптастыруға мүмкіндік береді.

Түйін сөздер: бейәдеп тілді анықтау; терең оқыту; кибербуллинг; қазақ тілі; әлеуметтік желілер; мәтінді жіктеу; BiLSTM

© Алишева С.С.¹, Оспанова А.О.¹, Шаймерденова Г.С.¹, Шаймерденова Л.¹, Жидебаева А.Н.^{2*}, 2026.

¹ Южно-Казахстанский университет имени М. Ауэзова,
Шымкент, Казахстан;

² Университет дружбы народов имени академика А. Куатбекова,
Шымкент, Казахстан.

E-mail: aziza_68.kz@mail.ru

ВЫЯВЛЕНИЕ ОСКОРБИТЕЛЬНЫХ КОММЕНТАРИЕВ НА КАЗАХСКОМ ЯЗЫКЕ В ОНЛАЙН-КОНТЕНТЕ С ПОМОЩЬЮ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ

Алишева Сандугаш — PhD, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан,

E-mail: sandu_alish@mail.ru, <https://orcid.org/0000-0001-6887-9208>;

Оспанова Айман — профессор, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан,

E-mail: oaimg09@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Шаймерденова Гулдана — PhD, ассоциированный профессор, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан,

E-mail: danel101kz@gmail.com, <https://orcid.org/0000-0001-8685-7125>;

Шаймерденова Лаззат — PhD, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан,

E-mail: Lizzat71@mail.ru, <https://orcid.org/0000-0002-5656-0918>;

Жидебаева Азиза — кандидат технических наук, старший преподаватель кафедры информатики и математики, Университет дружбы народов имени академика А. Куатбекова, Шымкент, Казахстан,

E-mail: aziza_68.kz@mail.ru, <https://orcid.org/0000-0002-3768-4835>.

Аннотация. Стремительное развитие социальных сетей и онлайн-платформ привело к значительному увеличению количества оскорбительных и агрессивных комментариев, что формирует психологические, социальные и связанные с безопасностью риски. Оскорбительная лексика, кибербуллинг и язык ненависти могут негативно влиять на психологическое состояние пользователей, способствовать тревожности, социальной изоляции и снижению качества онлайн-коммуникации. В исследовании предложен подход на основе глубокого обучения для выявления оскорбительных комментариев на казахском языке, объединяющий двунаправленную сеть долгой краткосрочной памяти (BiLSTM), разработанный для казахского языка алгоритм стемминга и методы лингвистической предварительной обработки. Для экспериментальной оценки была сформирована сбалансированная выборка из 40 000 вручную размеченных комментариев на казахском языке, включающая 20 000 оскорбительных и 20 000 нейтральных комментариев. Обработка данных осуществлялась с применением расширяемого словаря стемминга. Эффективность моделей оценивалась по показателям Accuracy, Precision, Recall и F1-score. Наилучшие результаты показала модель

BiLSTM: Accuracy - 91,3%, Precision - 92,6%, Recall - 93,0%, F1-score - 92,7%. Наивный байесовский классификатор продемонстрировал наиболее низкие показатели: Accuracy - 60,2%, Precision - 52,4%, Recall - 58,5%, F1-score - 64,2%. Полученные результаты подтверждают эффективность методов глубокого обучения при обработке морфологически сложных текстов на казахском языке. Предложенный подход может быть интегрирован в социальные сети и онлайн-сервисы для автоматической модерации контента, профилактики кибербуллинга и формирования более безопасной цифровой среды для казахоязычных пользователей.

Ключевые слова: выявление оскорбительной лексики; глубокое обучение; кибербуллинг; казахский язык; социальные сети; классификация текста; BiLSTM

Introduction. In 2025, 143 adolescents committed suicide, and 306 minors attempted suicide due to insulting or derogatory comments in online content. In general, 5% of adolescents have experienced the threat of using hateful comments online or may be exposed to online bullying 2-3 times a month or more. 12% of adolescents have been exposed to online bullying at least once. If their actions are classified as illegal, it can have negative social and professional consequences, including damage to their reputation, difficulties in making friends or securing employment (Alsubait, 2021). Statistics show that more than a third of those exposed to online bullying have received messages containing insults or discriminatory intent through private messages, 24% have seen posts about themselves with obvious discriminatory intent, and a third (31%) have encountered negative comments under their photos.

Educating people about the consequences of online bullying is one strategy to reduce online bullying. Educational webinars, guides and materials for educators and parents can help prevent online bullying from occurring (Dewani, 2021). We must do everything we can to create an environment where everyone treats each other with respect and compassion. We can make the world a better place for all people by creating a climate of mutual respect and a comfortable atmosphere (Hall, 2021). For example, according to statistics on the level of Internet use in the world in 2025, the state of Kazakhstan shows 15.47 million people, or 81.9% of the total population. Information on the Internet is available 24 hours a day, 7 days a week, and you cannot protect yourself from an "attack" on the Internet, because it can come at any time to your e-mail, mobile phone, or various personal social networking sites. This has a psychological impact on people, including teenagers. In addition, there is the possibility of being an anonymous user on the Internet, that is, you may not know who the "attack" is from (Arce-Ruelas, 2022). Psychologists say that emotional "attacks" are more likely to lead to psychological disorders than physical violence.

If we consider the detection of harmful information as the main task of

information tools, automation using machine learning methods is considered effective in implementing this task.

Literary review. Detecting profanity in online content and blocking it in online networks has been a current research topic, with many studies focusing on different methods and methodologies. This section briefly reviews the relevant literature and contextualizes current research in the broad field of profanity detection. Several studies have investigated the use of traditional machine learning algorithms in profanity detection. For example, (Ahmed, 2021) used support vector machines (SVMs) to classify profanity, highlighting the importance of feature engineering in achieving accurate results. Similarly (Toktarova, 2024) used random forests to detect hate speech, demonstrating the effectiveness of ensemble methods in solving this complex task.

In addition, deep learning models have gained a prominent place in hate speech detection research. RNNs, discussed in (Al-Marghilani, 2022), were used to detect sequential dependencies in text data, which showed promise in detecting hate speech patterns. Similarly, as studied in (Theng, 2021), convolutional neural networks (CNN) were used to extract relevant features from text data, which contributed to improving the accuracy of hate speech detection. In (Sarac, 2021), a study introduced a self-attention mechanism into LSTM networks, which effectively captures the important information in the content of hate speech. As discussed in (Sarac, 2021), transformer models were also used for this purpose, taking advantage of their ability to process long-term dependencies and contextual information (Table 1).

Table 1 — Analysis of scientific research works

Scientific research work	Language	Methods used	Database	Evaluation
Wulczyn (2017) and others	English	logistic regression, gradient boost, Deep neural networks	comments from Wikipedia	Accuracy, Precision, Recall, F1-Score
Djuric et al. (2015)	English	Feature-based approach (n gram, syntactic samples)	Comments from social networks	Accuracy, Precision, Recall, F1-Score
Fortuna and Nunes (2018)	Galician language	Transfer Learning, Pre-trained Embeddings	Comments from social networks	Accuracy, Precision, Recall, F1-Score
Chen et al. (2018)	China	BiLSTM	Comments from social networks	Accuracy, Precision, Recall, F1-Score
Nobata et al. (2016)	English	Attention based BiLSTM Networks	Internet community	Accuracy, Precision, Recall, F1-Score
Hassan et al. (2019)	Arabic language	Deep learning algorithms	Comments from social networks	Accuracy, Precision, Recall, F1-Score
Imran et al. (2018)	Urdu language	Exception-based approach, , SVM	Comments from Twitter	Accuracy, Precision, Recall, F1-Score
Choubey et al. (2019)	Hindi	Deep learning algorithms	Comments from social networks	Accuracy, Precision, Recall, F1-Score
Jha et al. (2018)	Bengali language	LSTM, Embeddings words	Comments from social networks	Accuracy, Precision, Recall, F1-Score

In addition to studying individual machine learning algorithms, researchers have also investigated ensemble methods. The (Gomez, 2022) used a combination of several models, such as LSTM and CNN, to achieve high performance in hate speech detection. Such ensemble approaches demonstrate the potential to combine the strengths of different algorithms. Another direction of research was the incorporation of domain-specific knowledge and linguistic features. The (Salawu, 2022) integrated linguistic features into a hate speech detection model, which highlights the importance of linguistic context in identifying hate speech instances. Furthermore, (Mladenović, 2021) proposed a graph-based approach to detect hate speech by considering semantic relationships between words.

In addition, transfer learning has become popular in the field. The (Sangwan, 2021) demonstrated the transferability of pre-trained models using large-scale language models for this purpose.

In conclusion, research in the field of hate speech detection encompasses a wide range of approaches. The reviewed literature highlights various strategies and methodologies that have been used to address the complex problem of hate speech detection in online environments (Toktarova, 2024). The present study aims to contribute to this ongoing discussion by investigating the effectiveness of BiLSTM networks in this important area (Yan, 2021).

Materials and methods. The proposed system consists of four modules: a web content collection module, a data parser module, a text corpus module for training, and a semantic analysis module for Kazakh-language texts.

Module 1: Web Content Collection Module. This module defines the "danger" criterion. One solution is to define a set of keywords. This method of identifying information types was used in the developed system. For this purpose, a set of keywords was compiled and used to analyze information on the VKontakte social network, using built-in methods for obtaining information from sources (API - Application Programming Interface) and from HTTP requests.

The web content collection module utilizes the VKontakte social network, which is popular in the Commonwealth of Independent States (Figure 1).

The web content collection module, in turn, consists of three separate modules:

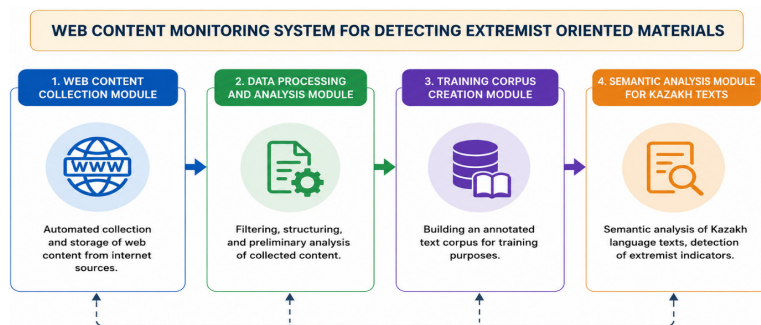


Figure 1 — Workflow of the proposed Kazakh text processing

The information collection module is responsible for receiving information from open sources and transmitting it for further processing. A Python framework was built to analyze data from the VKontakte social network. The official VK API (an interface that allows information to be retrieved from the vk.com database via HTTP requests to a dedicated server) was used. Kazakhstani profiles were partially analyzed, with the data stored in a Solr database (Figure 2).

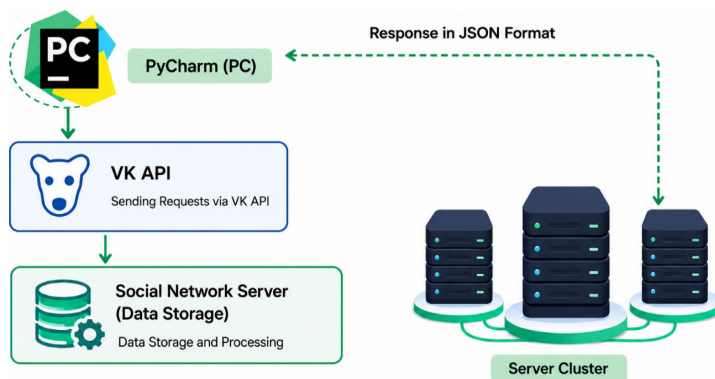


Figure 2 — Proposed Framework for VK API integration

a) The keyword search module is responsible for searching for keywords in large volumes of information;

b) The document ranking module is responsible for determining whether the information is dangerous. To rank documents by danger level, the module used a word2vec vectorizer and deep learning methods such as Long Short Term Memory (LSTM) and Bidirectional Long Short Term Memory (BiLSTM).

Module 2: Data Collection Parser Module. This module collects information from various sources, such as VKontakte, YouTube, Twitter, and news portals.



Figure 3 — parsing web interface

The figures 3 and 4 shows the main interface of the OnlineParser web application for automated collection of comments from social networks. The system consists of two main functional modules: VK Parser and YouTube Parser. The VK Parser module is designed to collect comments on the VK social network, and the YouTube Parser module is designed to automatically collect comments left on videos. In the upper part of the interface there is a navigation menu for going to the main page and the corresponding parser modules. In addition, the user can start the data collection process by entering a link to the desired web address. The collected text data is used for further processing, analysis, and use in machine learning models.

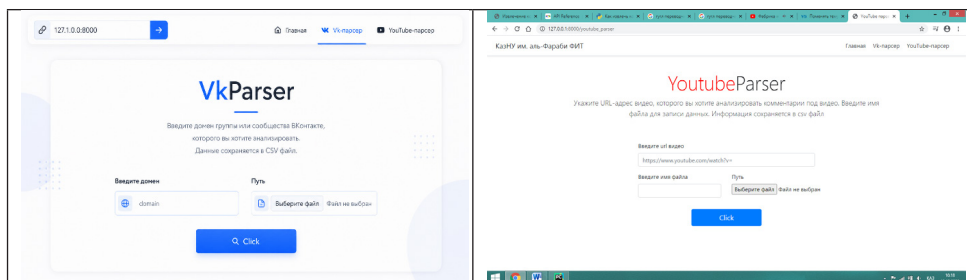


Figure 4 — An inside look at each social network parser

Module 3: Text Corpus Module for Training, contains texts with obscene and aggressive language in the Kazakh language, collected from open sources.

To collect data on obscene and aggressive language, comments from social media, open internet forums, user discussions, and general conversational texts were used. The data corpus contains over 20,000 sentences in the Kazakh language, including obscene and aggressive expressions.

Module 4: Semantic Text Analysis Module for Kazakh. The Semantic Text Analysis Module creates a machine learning model for detecting obscene and aggressive language in text, as well as a deep learning model.

At the initial stage of processing, the system checks whether the incoming text actually contains linguistic markers that make deeper analysis meaningful. In practice, not every document reaches the next phase. Some fragments are filtered out immediately because the content is too vague, while others simply do not contain the semantic patterns required for further examination. We noticed during testing that this preliminary filtering reduces unnecessary computational load, although occasionally it may skip texts that later turn out to be important.

The overall process operates through a server connected to multiple electronic devices over a data network. First, a collection of parameters describing potentially harmful content is formed. These parameters are not entirely static; some of them are adjusted manually during the configuration stage because fixed settings alone were not sufficient for stable detection. At one point the model became too

dependent on direct keywords, which created problems when the wording was intentionally altered.

The machine learning algorithm is then trained using different categories of examples, including both acceptable and problematic textual materials. The goal of the training phase is not only to recognize isolated words but also to identify the broader type of content based on combinations of features. In this case, particular attention is given to extremist materials. Interestingly, the system sometimes identifies suspicious content even when explicit terminology is absent, simply because several indirect indicators appear together within the same resource.

Text itself is not the only element involved in classification. The model also considers additional resource characteristics such as the URL address, the title of the webpage, HTML structure, embedded textual fragments, and screenshots of the resource. Initially this seemed somewhat excessive, yet later experiments showed that structural and visual properties often provide signals that plain text analysis misses. This became especially noticeable when certain resources attempted to disguise or distort textual information.

Once training is completed, the prepared machine learning model is distributed across connected electronic devices. Each device receives an already trained algorithm capable of comparing extracted resource features with previously learned patterns. The comparison process is multidimensional rather than linear. Sometimes the model reacts to secondary contextual clues rather than obvious indicators. At first this behavior looked inconsistent, though after additional evaluation it became clear that such responses occasionally improve overall detection accuracy.

Adapting machine learning techniques to the Kazakh language introduces several additional complications. Kazakh linguistic structure differs from many widely studied languages because of its extensive morphological variation. A single root word may appear in numerous grammatical forms, and separating stems from suffixes is not always straightforward. In several test cases the system interpreted different grammatical versions of the same word as unrelated tokens. That created instability during classification and increased noise inside the dataset.

Another issue involves stop words. These words occur constantly in both everyday communication and thematic publications. On one hand, they rarely carry strong semantic meaning. On the other hand, removing them entirely sometimes damages contextual relationships inside the sentence. During experimentation we observed that excessive repetition of stop words negatively affects model training and leads to additional classification errors. However, aggressive elimination of all stop words also produced weaker results in certain situations. Because of this, processing Kazakh-language texts requires a more flexible and less rigid preprocessing strategy than we originally expected.

Statistics – these are statistical data on the text, determined by the frequency of occurrence of a word or symbol. This data is effective for training texts in classification tasks. These features (Formulas R. 1 - R. 5):

$$feature_1 = \frac{\text{unique words}}{\text{word counts}} \quad (1);$$

$Feature_1$ determines the ratio of the number of unique words in the text to the total number of words and characterizes the lexical diversity of the text.

$$feature_2 = \text{word counts} \quad (2);$$

$Feature_2$ indicates the total number of words in the text and is used as a marker to characterize the size of the text.

$$feature_3 = \text{symbol counts} \quad (3);$$

$Feature_3$ determines the total number of special characters in the text and characterizes the symbolic structure of the text.

$$feature_{alphabet} = \sum_{i=0}^{n=41} X x_i \text{ in text} \quad (4);$$

$Feature_{alphabet}$ calculates the total number of characters belonging to a given alphabet that occur in the text

LIWC (Linguistic Inquiry and Word Count) was used—a text analysis program.

$$LIWC_k = \frac{f_k}{n} * 100 \quad (5).$$

$LIWC_k$ — percentage of the category k ;

f_k - the number of words belonging to category k ;

n - the number of all words in a text or sentence.

It calculates the extent to which various word categories are used in text and can process texts ranging from emails to speeches, poetry, and transcribed natural language in text or text formats.

Kazakh is an agglutinative language in which a single lexical root may generate numerous word forms through the addition of derivational and inflectional suffixes. Such morphological richness increases vocabulary sparsity and negatively affects text classification performance if different word forms are treated as independent lexical units. Therefore, a custom Kazakh stemming algorithm was developed to normalize word forms before feature extraction.

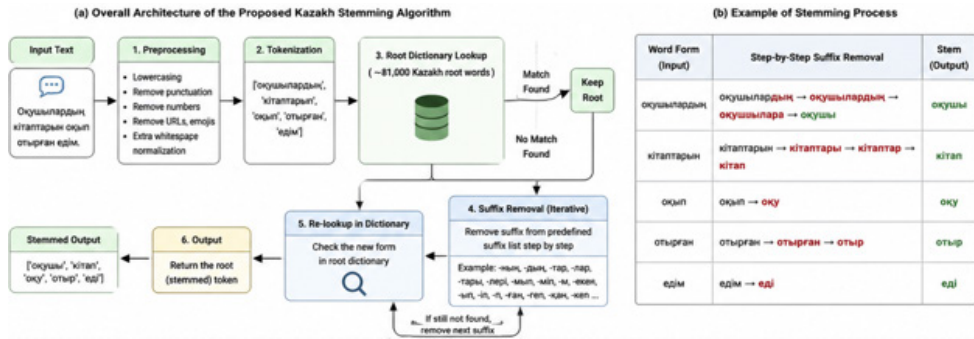


Figure 5 — Architecture and Workflow of the Proposed Kazakh Stemming Algorithm

Figure 5 shows the general architecture and word processing stages of the proposed Kazakh Stemming algorithm. The algorithm determines the word stem by preprocessing the text, tokenizing it, searching the stem dictionary, and removing suffixes step by step. If the stem is not found in the initial dictionary, the system iteratively removes suffixes and searches again after each step. This approach allows for effective processing of the agglutinative features of the Kazakh language, increasing the quality of text normalization and the accuracy of subsequent classification models.

Results. For this purpose, statistical characteristics, parts of speech (POS part of speech), linguistic inquiry and word count (LIWC), and term frequency (TF-IDF) characteristics are used.

For text classification, the following deep learning methods are used: convolutional neural network (CNN) and recurrent neural network (Figure 6) (LSTM).

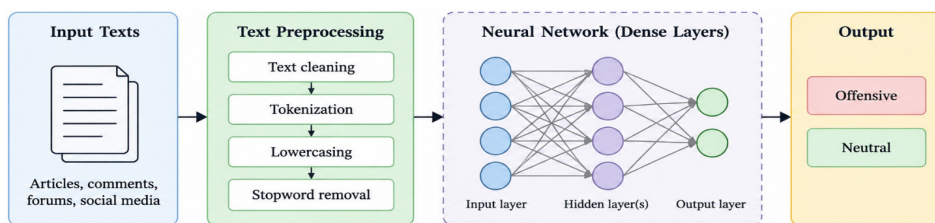


Figure 6 — Proposed model

In a neural network, the input layer, hidden layer, and output layer are known. Thus, the difference between the architecture of deep learning and neural networks lies in the specified number of hidden layers. A simple neural network has only one hidden layer, whereas deep learning has more than one hidden layer.

They start with a layer of input neurons, where feature vectors are input, and then the values are passed to hidden layers.

The output layer. With each connection, a value is passed forward, multiplied by the weight, and a bias is added to the value. This occurs with each connection, and at the end, the output layer value is obtained. The output layer consists of one or more output nodes. In the presented design, there is only one node, since the task is binary classification.

To calculate the values for each output node, each input node p is multiplied by the weight W and a bias b is added. All of this is then summed and passed to a function f . This function is considered the activation function. Typically, a rectified linear unit (ReLU) is used for hidden layers, a sigmoid function for the output layer in binary classification problems, or a softmax function for the output layer of multi-class classification problems. In the presented design, a sigmoid function is used. The algorithm begins by initializing the weights with random values, and then they are trained using a method called backpropagation. This is accomplished using optimization methods (also called optimizers), such as gradient descent, to reduce the error between the computed and desired outputs (also called target outputs).

The error is determined by a loss function, the loss of which is minimized using the Adam optimizer, and the cross-entropy loss function.

Database. The number of hate speech user comments included in the database corpus is about 20,000 and the number of neutral comments is about 20,000 sentence sequences.

The collected corpus was converted to .csv format and saved as a document. We divided the hate speech language corpus document into 4 columns: the ordinal number of comments in order, the comment itself, the clean version of the comment (cleaned from stop words, punctuation marks, emojis, etc.) and the class attributes "1/0".

Most of these stop words were taken from the nltk library. There are 300 words in the Kazakh language, as well as some of our own, on the topic of hate speech. As a result, they helped to better identify such thematic texts. Stemming removes unnecessary suffixes and prefixes, ultimately leading to the root. This part was developed using proprietary algorithms. The algorithm operates on two different approaches. The first is to search for unique words in the database and replace them in the text. If such words are not present in the database, it will be performed according to the rules of the Kazakh language. Stemming was written in Python (Figure 7).

According to the results, the BiLSTM model achieved the highest results in all main metrics (Accuracy – 91.3%, Precision – 92.6%, Recall – 93.0%, F1-score – 92.7%, AUC – 92%). While the Naïve Bayes model showed the lowest indicators. These results prove that deep learning models, especially the BiLSTM architecture, are more effective than traditional machine learning methods in detecting offensive comments in the Kazakh language.

When dealing with abusive language detection, the LSTM architecture turned out to be more flexible than many conventional classification approaches because it does not examine words as isolated elements only, but instead follows the sequence of language in a more natural way, preserving contextual relations between earlier and later parts of a sentence at the same time. During experimentation we noticed that offensive meaning is often hidden not in a single expression but in the interaction between phrases, tone, and surrounding context, and this is exactly where bidirectional processing becomes useful. The model analyzes textual information from both directions, which helps it capture subtle dependencies that simpler machine learning systems frequently ignore or misinterpret. In practice, some comments looked harmless when individual words were examined separately, although the complete sentence clearly carried aggressive intent. Another important point is that LSTM networks reduce dependence on manually constructed linguistic features, which honestly becomes a difficult and time-consuming process when datasets grow larger or language patterns begin changing rapidly, especially in online communication environments. Instead of relying entirely on predefined indicators, the system gradually learns internal representations directly from training examples, and this adaptive behavior makes the model more stable when users intentionally distort offensive vocabulary, replace characters, or create unusual sentence structures. At the same time, the model is not flawless either, since contextual interpretation sometimes becomes overly sensitive, but overall the ability to retain sequential information gives LSTM-based approaches a noticeable advantage in abusive language recognition tasks.

Figure 9 provides a comparative comparison of ten different algorithm models evaluated on four different standard performance metrics, namely accuracy, recall, precision, and F1-score, expressed as percentages.

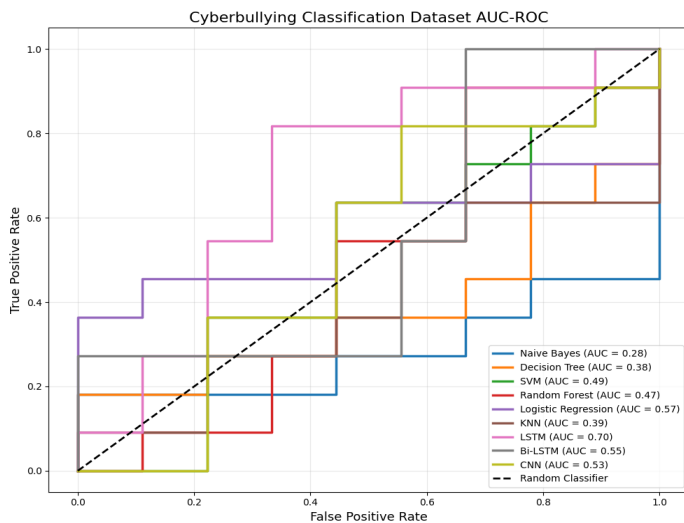


Figure 9 — AUC ROC evaluation metric indicator

Despite the many advantages of the proposed LSTM model for detecting abusive language words, it is worth noting that there are some limitations to consider. First, although the bidirectional model captures both old and new contexts, the complexity of the model and the computations require a considerable amount of time. This can be problematic when used in applications where speed is critical. Second, the performance of the model is largely dependent on the quality of the training data, i.e., if the training data does not sufficiently represent the diversity of the spoken language, the model may not be able to generalize to the unseen data. In addition, the model output requires close attention to hyperparameters, which require extensive tuning to achieve high performance. Finally, although the LSTM model can handle long sequences of text, it may be difficult to work with very long texts due to its fixed-size hidden state.

Discussion. The results of the experiments show the usefulness of deep learning approaches for recognizing objectionable comments in the Kazakh language. The best performance was obtained by the BiLSTM model, with an Accuracy of 91.3%, a Precision of 92.6%, a Recall of 93.0%, and an F1-score of 92.7%. These results outperformed the tested classical machine learning algorithms. The results show that the bidirectional contextual processing is beneficial for addressing the morphological complexity of the Kazakh language. The proposed stemming strategy is also further helpful to the text normalization process by decreasing the differences in the word forms. However, the success of the model is still dependent on the quality and diversity of the training data. Future study will focus on bigger Kazakh-language datasets and more powerful Transformer-based models, aiming for improved generalization and contextual comprehension.

Conclusion. In conclusion, although challenges and opportunities for further

improvement remain, the proposed LSTM model has shown significant progress and results in solving the problem of common language-based opinions on digital platforms. It has demonstrated the effectiveness of deep learning methods in understanding the complexity of human language by providing automated solutions to problems that cannot be effectively solved by traditional methods. This research work represents an important step towards harnessing the power of artificial intelligence to create secure and inclusive digital communication platforms. Future advances in this field are expected to not only produce reliable and effective models, but also provide new insights into language understanding and modeling in digital media.

References

- Alsubait T and D. Alfageh, (2021) Comparison of machine learning techniques for cyberbullying detection on youtube arabic comments,” *International Journal of Computer Science and Network Security*, vol. 21. — № 1. — P. 1–5. DOI: <https://doi.org/10.22937/IJCSNS.2021.21.1.1> (in Eng.)
- Dewani A., M. Memon and S Bhatti, (2021) Cyberbullying detection: Advanced preprocessing techniques & deep learning architecture for roman urdu data. *Journal of Big Data*, vol. 8, — № 17 — P. 1–20. DOI: <https://doi.org/10.1186/s40537-021-00550-7> (in Eng.)
- Hall D., Y. Silva, Y. Wheeler, L. Cheng and K. Baumel, (2021) Harnessing the power of interdisciplinary research with psychology-informed cyberbullying detection models. *International Journal of Bullying Prevention*, vol. 4. — №.1. — P. 47–54. DOI: <https://doi.org/10.1109/ICICCS48265.2020.9121046> (in Eng.)
- Arce-Ruelas K. (2022) Automatic cyberbullying detection: A Mexican case in high school and Higher Education Students, *IEEE Latin America Transactions*, vol. 20. — №. 5. — P. 770–779. DOI: <https://doi.org/10.1109/TLA.2022.9693561> (in Eng.)
- Ahmed T., M. Rahman, S. Nur, A. Islam and D. Das. (2021) Natural language processing and machine learning based cyberbullying detection for Bangla and romanized bangla texts. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20. — № 1. — P. 89–97. DOI: <http://doi.org/10.12928/telkomnika.v20i1.18630> (in Eng.)
- Toktarova A., Sultan D., & Azhibekova Z. (2024, May) Review of Machine Learning Models in Cyberbullying Detection Problem. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)). IEEE. — P. 233-238. DOI: <http://doi.org/10.1109/SIST61555.2024.10629223> (in Eng.)
- Al-Marghilani A. (2022) Artificial intelligence-enabled cyberbullying-free online social networks in smart cities. *International Journal of Computational Intelligence Systems*, vol. 15. — № 1. — P. 1–13, 2022. DOI: <https://doi.org/10.1007/s44196-022-00063-y> (in Eng.)
- Theng C., N. Othman, R. Abdullah, S. Anawar, Z. Ayop et al., (2021) Cyberbullying detection in twitter using sentiment analysis. *International Journal of Computer Science & Network Security*, vol. 21. — № 11. — P. 1-10, 2021. DOI: <https://doi.org/10.22937/IJCSNS.2021.21.11.1> (in Eng.)
- Sarac E. Essiz and M. Oturakci, (2021) Artificial bee colony–based feature selection algorithm for cyberbullying. *The Computer Journal*, vol. 64. — № 3. — P. 305–313. DOI: <https://doi.org/10.1093/comjnl/bxaa066> (in Eng.)
- Gomez MC.E.O. Sztainberg and R.E. Trana (2022) Curating cyberbullying datasets: a human-AI collaborative approach. *International journal of bullying prevention*, vol. 4. — № 1. — P. 35-46. (in Eng.)
- Salawu S., J. Lumsden and Y. He. (2022). A mobile-based system for preventing online abuse and cyberbullying,” *International Journal of Bullying Prevention.*, vol. 4. — № 1. — P. 66–88. (in Eng.)
- Mladenović M., V. Ošmjanski and S. V. Stanković, (2021) Cyber-aggression, cyberbullying, and cyber-grooming: a survey and research challenges,” *ACM Computing Surveys (CSUR)*, vol. 54. — №1. — P. 1–42. DOI: <https://doi.org/10.1145/3424246> (in Eng.)

Sangwan S.R. and M.P.S. Bhatia (2021) Denigrate comment detection in low-resource Hindi language using attention-based residual networks,” Transactions on Asian and Low-Resource Language Information Processing, vol. 21. — № 1. — P. 1–14. DOI: <https://doi.org/10.1145/3431729> (in Eng.)

Yan R., Y. Li, D. Li, Y. Wang, Y. Zhu et al., (2021) A Stochastic Algorithm Based on Reverse Sampling Technique to Fight Against the Cyberbullying. ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 15. — №. 4. — P. 1–22. DOI: <https://doi.org/10.1145/3441455> (in Eng.)

Toktarova A., Sultan D., & Azhibekova Z. (2024, May) Review of machine learning models in cyberbullying detection problem. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST). — P. 233-238. DOI: <https://doi.org/10.14569/ijacsa.2024.0151018> (in Eng.)

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

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 journals of the Central Asian Academic Research Center LLP 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 Central Asian Academic Research Center LLP 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 Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www.nauka-nanrk.kz

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

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*

Editors: *D.S. Alenov, T. Apendiev*

Computer layout: *G.D. Zhadyranova*

Signed for print: September 30, 2026

Format: 70×90 1/16. 26.5 printed sheets. Order No. 3.