

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., Aitkozhasyeva 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. Tize MPT патологиясын жіктеуге арналған көп тапсырмалы терең оқыту.....	139
Исahов 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
Кошкарлов 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., Найзағараева А.А., Оспанова Т.Т., Қойкелова Д.К., Қудабаева 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., Абдуғулова 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). 306–325

<https://doi.org/10.32014/2026.2518-1726.465>

IRSTI 81.96.00

UDC 004.056

© **Mussiraliyeva Sh., Bolatbek M., Baisylbayeva K.***, 2026.

Al-Farabi Kazakh National University, Almaty, Kazakhstan.

E-mail: baisylbaeva.k@gmail.com

CLASSIFICATION OF CYBER-EXTREMIST IDEOLOGIES IN THE KAZAKH LANGUAGE CONSIDERING MULTI-ANNOTATOR DISAGREEMENT

Mussiraliyeva Shynar — Candidate of Physical and Mathematical Sciences, Research Professor of the Department of Cybersecurity and Cryptology, professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: mussiraliyevash@gmail.com, <http://orcid.org/0000-0001-5794-3649>;

Bolatbek Milana — PhD, Acting Associate Professor of the Department of Cybersecurity and Cryptology, Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: bolatbek.milana@gmail.com, <http://orcid.org/0000-0002-2153-180X>;

Baisylbayeva Kymbat — Senior Lecturer of the Department of Cybersecurity and Cryptology, Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: baisylbaeva.k@gmail.com, <https://orcid.org/0000-0001-9753-0398>.

Abstract. The spread of cyber-extremist content on digital platforms affects information security and social stability; therefore, its automatic detection and classification by ideological directions represents a relevant scientific problem. In the automatic classification of Kazakh-language cyber-extremist texts, the content of a single text may be close to several classes. This is reflected in differences between the judgments of several annotators when the semantic boundaries of the propaganda, recruitment, and radicalization categories overlap. A traditional majority-vote approach combines such differences into a single hard label, which may result in the loss of information about annotation uncertainty. The aim of this study is to compare strategies that preserve multi-annotator assessments in the classification of Kazakh-language cyber-extremist texts and enable a comprehensive evaluation of their modeling quality. The study uses a multiclass corpus of texts collected from the open digital platforms VK, Telegram, and YouTube. The corpus includes the propaganda, recruitment, radicalization, and neutral classes. The annotation process was based on independent assessments provided by several experts. The following strategies for processing multi-annotator information were compared: majority voting, probability-based

processing, ensemble integration, and an instruction-based strategy. Final-class prediction quality was evaluated using Macro-F1, while the correspondence between the model's probability distributions and the annotators' label distribution was assessed using Cross-Entropy (CE), Kullback–Leibler divergence (KL), Jensen–Shannon divergence (JS), and Manhattan Distance (MD). The ensemble integration strategy achieved the highest observed classification performance among the approaches considered. It also demonstrated better probability-distribution matching according to the CE, KL, and JS metrics. The majority-voting method achieved the best result in terms of MD. Compared with majority voting, ensemble integration improved classification quality and produced lower values for the metrics measuring probabilistic agreement. Within the DART framework, the $L_{CE} + L_{KL} + L_{MD}$ combination preserved classification quality while bringing the model predictions closer to the annotators' distribution of assessments. The findings indicate that a single hard label is insufficient for evaluating multi-annotator ideological texts, whereas modeling differences between annotators' opinions is an effective approach.

Keywords: cyber-extremism, ideological classification, multi-annotator annotation, disagreement-based learning, soft labels, ensemble learning, Macro-F1, Cross-Entropy, Jensen–Shannon divergence

For citations: Mussiraliyeva Sh., Bolatbek M., Baisylbayeva K. Classification of cyber-extremist ideologies in the kazakh language considering multi-annotator disagreement. Academic Scientific Journal of Computer Science, 2026. — No.3. — P. 306–325. DOI: <https://doi.org/10.32014/2026.2518-1726.465>

© Мусиралиева Ш., Болатбек М., Байсылбаева К.*, 2026.

Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.

E-mail: baisylbaeva.k@gmail.com

КӨПАННОТОРАЛЫ КЕЛІСПЕУШІЛІКТІ ЕСКЕРЕ ОТЫРЫП ҚАЗАҚ ТІЛІНДЕГІ КИБЕРЭКСТРЕМИСТІК ИДЕОЛОГИЯЛАРДЫ ЖІКТЕУ

Мусиралиева Шынар — физика-математика ғылымдарының кандидаты, профессор, «Киберқауіпсіздік және криптология» кафедрасының профессор-зерттеушісі, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,

E-mail: mussiraliyevash@gmail.com, <https://orcid.org/0000-0001-5794-3649>;

Болатбек Милана — PhD, «Киберқауіпсіздік және криптология» кафедрасының қауымдастырылған профессорының міндетін атқарушы, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,

E-mail: bolatbek.milana@gmail.com, <https://orcid.org/0000-0002-2153-180X>;

Байсылбаева Кымбат — «Киберқауіпсіздік және криптология» кафедрасының аға оқытушысы, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,

E-mail: baisylbaeva.k@gmail.com, <https://orcid.org/0000-0001-9753-0398>.

Аннотация. Цифрлық платформаларда киберэкстремистік контенттің таралуы ақпараттық қауіпсіздік пен қоғамдық тұрақтылыққа әсер етеді, сондықтан мұндай контентті автоматты түрде анықтау және идеологиялық бағыттар бойынша жіктеу өзекті ғылыми міндет болып табылады. Қазақ тіліндегі киберэкстремистік мәтіндерді автоматты жіктеу кезінде бір мәтіннің мазмұны бірнеше класқа мағыналық жағынан жақын болуы мүмкін. Бұл пропаганда, рекрутмент және радикализация санаттарының мағыналық шекаралары қабаттасқан кезде бірнеше аннотатордың бір мәтінге әртүрлі белгі беруінен көрінеді. Көпшілік дауысқа негізделген дәстүрлі тәсіл мұндай айырмашылықтарды бір ғана қатаң белгіге біріктіреді, соның салдарынан аннотациядағы белгісіздік туралы ақпарат жоғалуы мүмкін. Зерттеудің мақсаты — қазақ тіліндегі киберэкстремистік мәтіндерді жіктеу кезінде көпаннотаторлы бағаларды сақтайтын стратегияларды салыстыру және оларды модельдеу сапасын кешенді бағалау. Зерттеуде VK, Telegram және YouTube ашық цифрлық платформаларынан жиналған мәтіндерден тұратын көпкласты корпус пайдаланылды. Корпус пропаганда, рекрутмент, радикализация және бейтарап кластарды қамтиды. Аннотациялау бірнеше сарапшының тәуелсіз бағалауларына негізделді. Көпаннотаторлы ақпаратты өңдеудің төрт стратегиясы салыстырылды: көпшілік дауысы, ықтималдыққа негізделген өңдеу, ансамбльдік біріктіру және нұсқаулыққа негізделген стратегия. Қорытынды класты болжау сапасы Macro-F1 метрикасы арқылы, ал модельдің ықтималдық үлестірімдерінің аннотаторлар белгілерінің үлестіріміне сәйкестігі Cross-Entropy (CE), Kullback–Leibler divergence (KL), Jensen–Shannon divergence (JS) және Manhattan Distance (MD) метрикалары арқылы бағаланды. Қарастырылған тәсілдердің ішінде ансамбльдік біріктіру стратегиясы жіктеудің ең жоғары нәтижесін көрсетті. CE, KL және JS метрикалары бойынша да осы стратегия ықтималдық үлестірімдерінің жақсырақ сәйкестігін қамтамасыз етті. Көпшілік дауыс әдісі MD көрсеткіші бойынша ең жақсы нәтижеге қол жеткізді. Көпшілік дауыс әдісімен салыстырғанда ансамбльдік біріктіру жіктеу сапасын жақсартып, ықтималдық сәйкестігін бағалайтын метрикалар бойынша төмен мәндер көрсетті. DART шеңберіндегі LCE + LKL + LMD комбинациясы жіктеу сапасын сақтай отырып, модель болжамдарын аннотаторлар бағаларының үлестіріміне жақындатуға мүмкіндік берді. Нәтижелер көпаннотаторлы идеологиялық мәтіндерді модельдеу үшін бір ғана қатаң белгі жеткіліксіз екенін, ал аннотаторлар арасындағы пікір айырмашылықтарын ескеру неғұрлым тиімді әрі ақпаратты тәсіл екенін көрсетті.

Түйін сөздер: киберэкстремизм; идеологиялық жіктеу; көпаннотаторлы таңбалау; келіспеушілікке негізделген оқыту; жұмсақ белгілер; ансамбльдік оқыту; Macro-F1; Cross-Entropy; Jensen–Shannon divergence

© Мусиралиева Ш., Болатбек М., Байсылбаева К.*, 2026.

Казахский национальный университет имени аль-Фараби,

Алматы, Казахстан.

E-mail: baisylbaeva.k@gmail.com

КЛАССИФИКАЦИЯ КИБЕРЭКСТРЕМИСТСКИХ ИДЕОЛОГИЙ НА КАЗАХСКОМ ЯЗЫКЕ С УЧЕТОМ РАЗНОГЛАСИЙ НЕСКОЛЬКИХ АННОТАТОРОВ

Мусиралиева Шынар — кандидат физико-математических наук, профессор, профессор-исследователь кафедры «Кибербезопасность и криптология», Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,
E-mail: mussiraliyevash@gmail.com, <https://orcid.org/0000-0001-5794-3649>;

Болатбек Милана — PhD, и.о. ассоциированного профессора кафедры «Кибербезопасность и криптология», Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,
E-mail: bolatbek.milana@gmail.com, <https://orcid.org/0000-0002-2153-180X>;

Байсылбаева Кымбат — старший преподаватель кафедры «Кибербезопасность и криптология», Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,
E-mail: baisylbaeva.k@gmail.com, <https://orcid.org/0000-0001-9753-0398>.

Аннотация. Распространение киберэкстремистского контента на цифровых платформах влияет на информационную безопасность и общественную стабильность, поэтому его автоматическое выявление и классификация по идеологическим направлениям являются актуальной научной задачей. При автоматической классификации киберэкстремистских текстов на казахском языке содержание одного текста может быть семантически близким к нескольким классам. Это проявляется в различиях между оценками нескольких аннотаторов при пересечении смысловых границ категорий пропаганды, рекрутмента и радикализации. Традиционный подход, основанный на большинстве голосов, сводит подобные расхождения к одной жесткой метке, что может приводить к потере информации о неопределенности аннотации. Цель исследования заключается в сравнении стратегий, сохраняющих многоаннотаторные оценки при классификации киберэкстремистских текстов на казахском языке и позволяющих комплексно оценивать качество их моделирования. В исследовании использован многоклассовый корпус текстов, собранных на открытых цифровых платформах VK, Telegram и YouTube. Корпус включает классы пропаганды, рекрутмента, радикализации и нейтрального содержания. Аннотирование проводилось на основе независимых оценок нескольких экспертов. Сравнивались четыре стратегии обработки многоаннотаторной информации: большинство голосов, вероятностная обработка, ансамблевое объединение и стратегия, основанная на инструкциях. Качество определения итогового класса оценивалось с помощью Macro-F1, а соответствие вероятностных распределений модели распределению меток аннотаторов - с использованием Cross-Entropy (CE), Kullback–Leibler divergence (KL), Jensen–Shannon

divergence (JS) и Manhattan Distance (MD). Среди рассмотренных подходов стратегия ансамблевого объединения продемонстрировала наивысшее качество классификации. По метрикам CE, KL и JS она также обеспечила более высокое соответствие вероятностных распределений. Метод большинства голосов показал лучший результат по MD. По сравнению с большинством голосов ансамблевое объединение повысило качество классификации и обеспечило более низкие значения метрик вероятностного расхождения. В рамках DART комбинация LCE + LKL + LMD позволила сохранить качество классификации и приблизить прогнозы модели к распределению оценок аннотаторов. Полученные результаты показывают, что одной жесткой метки недостаточно для моделирования многоаннотаторных идеологических текстов, тогда как учет различий между мнениями аннотаторов является более эффективным и информативным подходом.

Ключевые слова: киберэкстремизм; идеологическая классификация; многоаннотаторная разметка; обучение с учетом разногласий; мягкие метки; ансамблевое обучение; Macro-F1; Cross-Entropy; Kullback–Leibler divergence; Jensen–Shannon divergence

Кіріспе. Әлеуметтік желілер мен мессенджерлер ақпарат алмасудың негізгі арналарының біріне айналғанымен, олар экстремистік идеяларды насихаттау, радикал көзқарастарды тарату және жаңа қатысушыларды тарту үшін де пайдаланылуы мүмкін. Онлайн экстремизмді автоматты анықтауға арналған зерттеулерде машиналық оқыту, терең оқыту, семантикалық талдау және трансформерлік архитектуралардың әртүрлі нұсқалары қолданылып келеді (Aldera et al., 2021; Berjawi et al., 2023; Gaikwad et al., 2021; Gaikwad et al., 2023; Govers et al., 2023). Дегенмен жарияланған жұмыстардың едәуір бөлігі ресурсы мол тілдерге немесе экстремистік/экстремистік емес түріндегі бинарлық классификацияға бағытталған. Қазақ тіліндегі осы бағыттағы зерттеулер бірнеше кезеңнен өтті. Алғашқы кезеңде интернет кеңістігіндегі деструктивті мәтіндерді автоматты жинауға арналған бағдарламалық құрал әзірленіп, мәтіндік деректерді жүйелі түрде жинақтаудың техникалық негізі қалыптастырылды (Bolatbek et al., 2023). Кейін машиналық оқыту алгоритмдерін қолдана отырып, мәтіндерді пропаганда, радикализация және рекрутмент бағыттары бойынша ажырату мүмкіндігі зерттелді (Mussiraliyeva et al., 2024). Төрт класты – пропаганда, рекрутмент, радикализация және бейтарап мәтіндерді – қамтитын қазақ тіліндегі арнайы датасет құру жұмысы жеке зерттеу ретінде ұсынылды (Mussiraliyeva et al., 2025b).

Келесі кезеңде көпкласты идеологиялық жіктеуге психоллингвистикалық сипаттамалар енгізілді. PLAMkz (psycho-linguistic analysis module for Kazakh language texts) модулін мәтіндік векторизация және машиналық оқыту алгоритмдерімен біріктіру, сондай-ақ BERT және BERT+LSTM архитектураларын пайдалану жіктеу сапасын жоғарылатуға мүмкіндік берді (Mussiraliyeva et al., 2025a). Бұдан бөлек, CNN, BiLSTM, GRU, CNN+BiLSTM

және трансформерлік модельдер қазақ тіліндегі идеологиялық мәтіндерде жеке салыстырылды (Bolatbek et al., 2025). Аннотациялау мәселелері мен терең оқыту модельдерінің мүмкіндіктері кейінгі жұмыста арнайы қарастырылып, көп-аннотаторлы бағалау қажеттілігі айқындалды (Baisylbayeva et al., 2026).

Халықаралық зерттеулерде де бұл мәселе соңғы жылдары жеке бағыт ретінде қалыптасты. Көпшілік дауыс аннотаторлардың бірнеше бағасын бір ғана қатаң белгіге біріктіреді, бірақ субъективті NLP міндеттерінде пікір айырмашылығы деректің көпмәнділігі немесе әртүрлі интерпретациялардың көрінісі болуы мүмкін (Aroyo and Welty, 2015; Dumitrache et al., 2018; Fornaciari et al., 2021; Mostafazadeh Davani et al., 2022; Uma et al., 2021). Сондықтан аннотаторлар келіспеушілігін қате ретінде алып тастаудың орнына, оны оқыту нысанасы немесе қосымша сигнал ретінде пайдалану ұсынылады.

Зерттеудің негізгі мақсаты – қазақ тіліндегі киберэкстремистік идеологияларды жіктеуде аннотаторлық келіспеушілікті төрт түрлі стратегия арқылы модельдеп, олардың тиімділігін бір мезгілде қатаң және жұмсақ бағалау критерийлері бойынша салыстыру.

Әдебиеттерге шолу. *Онлайн экстремизмді көп-класты жіктеу.* Онлайн экстремизмді анықтауға арналған жүйелі шолуларда датасет тапшылығы, класстар теңгерімсіздігі және бинарлық классификацияға басымдық берілуі негізгі мәселелердің қатарына жатады (Aldera et al., 2021; Gaikwad et al., 2021). Сонымен қатар экстремизм, радикализация және саяси өшпенділік тілі арасындағы шекараның әрдайым анық болмауы күрделі дискурстық құбылыстарды бинарлық белгімен сипаттауды шектейді (Govers et al., 2023). Онлайн радикализацияны автоматты анықтауға арналған 68 зерттеуді қамтыған шолуда машиналық және терең оқыту модельдері, мәтіндік белгілер мен контекстік ақпаратты қолдану тәсілдері жүйеленген (Berjawi et al., 2023).

Бірнеше идеологиялық дереккөзді біріктіру арқылы экстремистік контент пропаганда, радикализация және рекрутмент санаттарына бөлінген мультиклассты датасет қалыптастырылған. Оны жіктеуде SVM, Naive Bayes, Random Forest және XGBoost алгоритмдері салыстырылып, бірнеше идеологиялық бағытты бір деректер жиынында қамтудың маңыздылығы көрсетілген. Бұл тәсіл киберэкстремистік мәтіндерді бинарлық жіктеумен шектемей, олардың идеологиялық бағыттарын жеке кластар бойынша анықтауға мүмкіндік береді (Gaikwad et al., 2023).

Аннотаторлар келіспеушілігі: қателіктен ақпарат көзіне дейін

Аннотаторлар арасындағы келіспеушілікті дәстүрлі түрде аннотациядағы шу ретінде қарастыру кең таралған. Aroyo және Welty адам аннотациясы жағдайында бір мәтінге қатысты бірнеше негізді түсіндірудің болуы қалыпты құбылыс екенін көрсетті (Aroyo and Welty, 2015). CrowdTruth әдістемесі де келіспеушілікті аннотатор сапасы, бастапқы деректің көпмәнділігі және белгі таңдаудағы келіспеушілік туралы қосымша ақпарат көзі ретінде пайдалануды ұсынады (Dumitrache et al., 2018).

Көпаннотаторлы деректермен жұмыс істеуге арналған тәсілдерді жүйелеу барысында біріктірілген белгілер, жұмсақ белгілер, жекелеген аннотатор пікірлерін модельдеу және бірнеше көзқарасты қатар сақтау қағидалары өзара салыстырылған. Мұндай салыстырулар сапалы бірнеше аннотациясы бар деректер жиынында барлық пікірді бір ғана қорытынды белгіге келтіру әрдайым тиімді бола бермейтінін көрсетеді. Кейбір жағдайларда аннотаторлар таңдауының үлестірімін сақтайтын жұмсақ белгілер негізінде оқыту біріктірілген немесе эталондық белгіге негізделген оқытудан жоғары нәтиже береді (Fornaciari et al., 2021).

Аннотаторлар арасында қарама-қайшы бағалары бар деректерден модель үйрету кезінде келіспеушілікті жеке оқыту сигналы ретінде сақтау ұсынылады (Mostafazadeh Davani et al., 2022). Сонымен қатар келіспеушіліктің барлығын бір ғана себеппен түсіндіру дұрыс емес: ол деректің көпмәнділігінен, таңбалаудағы шудан немесе өзара мағыналық жағынан қабаттасатын класстардан туындауы мүмкін.

SemEval-2023 LeWiDi аясында аннотаторлар арасындағы келіспеушілікті бағалауға ерекше назар аударылып, жұмсақ бағалау негізгі өлшемдердің бірі ретінде қолданылды (Leonardelli et al., 2023). Көпшілік дауысына негізделген тәсілде нысаналы әлеуметтік топ өкілдері аннотаторлар арасында азшылық болған жағдайда, олардың пікірі ескерілмей қалуы мүмкін екені көрсетілген (Fleisig et al., 2023). Келіспеушілікті ескеретін табиғи тілді өңдеу тәсілдері болжам нысанасы, аннотаторлық ақпаратты біріктіру тәсілі және бағалау өлшемдері бойынша жүйеленген (Xu and Jurgens, 2026). MO-WEL көпмақсатты ансамбльдік әдісінде F1, кросс-энтропия және Манхэттен қашықтығы бірлесіп оңтайландырылып, ансамбль салмақтары аннотаторлық үлестірімге сәйкес таңдалады. Бұл тәсіл жіктеу сапасын сақтай отырып, аннотаторлар пікірлерінің әртүрлілігін ескеруге мүмкіндік береді (Cui et al., 2026).

Халықаралық зерттеулерде онлайн экстремизмді анықтау көбіне ресурсы мол тілдердегі корпустарға және бинарлық немесе аз класты жіктеу схемаларына негізделеді (Aldera et al., 2021; Berjawi et al., 2023; Gaikwad et al., 2021; Gaikwad et al., 2023; Govers et al., 2023). Бұл жұмыстарда машиналық оқыту, трансформерлік модельдер және мәтіндік белгілерді біріктіру тәсілдері кеңінен қолданылған. Алайда әр зерттеудегі корпус көлемі, кластардың мазмұны, деректерді бөлу протоколы және бағалау өлшемдері әртүрлі болғандықтан, олардың сандық нәтижелерін осы жұмыстың қазақ тіліндегі төрт класты корпусымен тікелей салыстыру әдістемелік тұрғыдан дұрыс емес. Сондықтан салыстыру жалпы әдістемелік бағыттар, яғни көпкласты жіктеу, трансформерлік модельдер, психолингвистикалық белгілер және аннотаторлар келіспеушілігін сақтау деңгейінде жүргізілді.

Материалдар мен әдістер. Деректер жиыны. Зерттеуде қазақ тіліндегі киберэкстремистік мәтіндердің кейінгі нақтыланған көпаннотаторлы нұсқасы пайдаланылды. Нақтыланған корпус 7152 мәтіннен тұрады. Кластардың

таралуы мынадай: пропаганда – 1738 мәтін (24,30%), рекрутмент – 1698 (23,74%), радикализация – 2052 (28,69%) және бейтарап – 1664 (23,27%). Аннотациялау процедурасы төрт сарапшының тәуелсіз бағалауына негізделді. Бастапқы келісім Cohen к бойынша шамамен 0,92, ал нұсқаулықты нақтылау мен калибрлеуден кейін шамамен 0,98 болды. Мәтіндер VK, Telegram және YouTube сияқты ашық цифрлық платформалардан жинақталған. Деректер тазалау, нормализациялау, токенизациялау және мәтіндік формаға келтіру кезеңдерінен өтті. Пайдаланушы идентификаторлары мен жеке деректер модельдеу міндетіне енгізілген жоқ; зерттеу мәтіндік-лингвистикалық белгілерді талдауға бағытталды (Mussiraliyeva et al., 2025b). Деректерді жинау кезеңінде ашық веб-ресурстардан мәтіндерді автоматты алуға арналған бұрын әзірленген парсерді қолдану тәжірибесі ескерілді (Bolatbek et al., 2023). Деректер жиынының бастапқы құрылымы мен идеологиялық кластар жүйесі бұрын жарияланған көпидеологиялық корпус және көпкласты жіктеу жөніндегі жұмыстармен сабақтастықта қалыптастырылды (Mussiraliyeva et al., 2024; Mussiraliyeva et al., 2025b). Осы мақалада корпус көпаннотаторлы қайта бағалау, 92 мәтінді қайта қарау және жұмсақ үлестірімдерді есептеу арқылы нақтыланған нұсқа ретінде пайдаланылды.

Зерттеуде тек ашық цифрлық платформаларда жарияланған мәтіндік материалдар пайдаланылды. Пайдаланушы идентификаторлары, байланыс деректері және жеке тұлғаны тікелей анықтауға мүмкіндік беретін басқа да дербес мәліметтер талдауға енгізілмеді. Деректер тек ғылыми мақсатта, экстремистік мазмұнды қайта тарату үшін емес, оның тілдік және идеологиялық белгілерін автоматты түрде анықтау әдістерін зерттеу үшін қолданылды.

Салыстырылатын стратегиялар. Зерттеуде көпаннотаторлы ақпаратты өңдеудің төрт тәсілі қарастырылды.

Көпшілік дауысы (Majority)

Бұл тәсілде әр мәтін үшін аннотаторлар арасында ең көп дауыс алған класс қорытынды белгі ретінде қабылданады. Әдістің басты артықшылығы – қарапайымдылығы және дәстүрлі жіктеу модельдеріне оңай енгізілуі.

Ықтималдықтық тәсіл (Probability)

Бұл тәсілде аннотаторлар берген белгілер бір ғана қорытынды класқа біріктірілмей, әр класс бойынша ықтималдық үлестірімі түрінде сақталады. Модель осы үлестірімге жақындауға үйретіледі. Нәтижесінде аннотаторлар арасындағы пікір айырмашылығы оқыту барысында жоғалмай, мәтіннің мағыналық көпмәнділігі мен белгісіздік деңгейін сипаттайтын қосымша ақпарат ретінде пайдаланылады (Fornaciari et al., 2021; Uma et al., 2021). Probability стратегиясында төрт аннотатордың жауаптарынан әр класс бойынша ықтималдық үлестірімі есептеліп, модель осы жұмсақ нысанаға жақындауға үйретілді.

Ансамбльдік тәсіл (Ensemble)

Ансамбльдік тәсіл бірнеше модельден алынған ықтималдық болжамдарын

бір қорытынды шешімге біріктіруге негізделеді. Әр модель мәтінді өзінше бағалайтындықтан, олардың болжамдарын біріктіру жеке модельдің кездейсоқ немесе шамадан тыс сенімді шешімдерінің әсерін азайтуға мүмкіндік береді. Ensemble стратегиясында әр аннотатордың таңбалау нәтижелері бойынша жеке оқытылған BERT модельдерінің ықтималдық болжамдары біріктіріліп, мәтіннің төрт идеологиялық класқа тиесілілігін сипаттайтын қорытынды ықтималдық үлестірімі қалыптастырылды.

Нұсқаулыққа негізделген тәсіл (Instruction)

Бұл тәсілде әр мәтінге оны белгілі бір аннотатордың көзқарасы тұрғысынан бейтарап, пропаганда, рекрутмент немесе радикализация кластарының біріне жіктеуді сұрайтын мәтіндік нұсқаулық қосылды. Төрт аннотатордың әрқайсысы үшін оның жеке таңбалары негізінде көптілді BERT моделі бөлек оқытылды.

Көпанныотаторлы деректер жағдайында модель сапасын тек қорытынды классификация нәтижесімен бағалау жеткіліксіз. Сондықтан бұл жұмыста модель болжамдарының аннотаторлар қалыптастырған ықтималдық үлестірімдерге сәйкестігін бағалау үшін CE, KL, JS және MD метрикалары қолданылды.

Модель нәтижелерін жан-жақты бағалау үшін бірнеше метрика қатар қолданылды. Бұл тәсіл модельдің тек дұрыс классты анықтау қабілетін ғана емес, сонымен бірге аннотаторлар арасындағы пікір айырмашылығын қаншалықты дұрыс бейнелейтінін бағалауға мүмкіндік берді.

Масго-F1 төрт кластың әрқайсысы үшін есептелген F1-көрсеткіштердің арифметикалық орташа мәні ретінде алынды. Бұл көрсеткіш модельдің көпшілік дауысы негізінде алынған қатаң таңбаларды барлық кластар бойынша қаншалықты теңгерімді анықтайтынын бағалау үшін қолданылды. Масго-F1 мәні неғұрлым жоғары болса, модельдің жіктеу сапасы соғұрлым жақсы деп есептелді.

Кросс-энтропия (CE) аннотаторлар бағасынан алынған жұмсақ ықтималдықтық үлестірім мен модель болжаған ықтималдықтар арасындағы сәйкестікті бағалау үшін қолданылды. CE мәнінің төмен болуы модель болжамының аннотаторлар пікірлерінің үлестіріміне жақын екенін көрсетті.

Куллбек–Лейблер дивергенциясы (KL) аннотаторлар қалыптастырған жұмсақ таңба үлестірімі мен модель болжамы арасындағы ақпараттық айырмашылықты бағалау үшін есептелді. Бұл көрсеткіш бағытталған болғандықтан, модельдік ықтималдықтардың аннотаторлық үлестірімнен ауытқу дәрежесін сипаттады. KL мәні неғұрлым төмен болса, үлестірімдер арасындағы сәйкестік соғұрлым жоғары деп қабылданды.

Дженсен–Шеннон дивергенциясы (JS) аннотаторлық және модельдік ықтималдық үлестірімдерінің симметриялық айырмашылығын бағалау үшін қолданылды. Бұл көрсеткіш екі үлестірімнің жалпы ұқсастығын тұрақты түрде салыстыруға мүмкіндік берді. JS мәнінің төмен болуы үлестірімдердің бір-біріне жақын екенін көрсетті.

Манхэттен қашықтығы (MD) әр класс бойынша аннотаторлық және модельдік ықтималдықтар айырмасының абсолюттік мәндерін қосу арқылы есептелді. Бұл көрсеткіш модель болжамының аннотаторлар бағасына қаншалықты жақын екенін тікелей сипаттау үшін пайдаланылды. MD мәні неғұрлым төмен болса, модельдік және аннотаторлық үлестірімдер арасындағы сәйкестік соғұрлым жоғары деп бағаланды.

Манхэттен қашықтығы (MD) модельдік және аннотаторлық ықтималдықтардың абсолюттік айырмаларын қосу арқылы есептеледі. Сондықтан олекі ықтималдық үлестірімі арасындағы жалпы айырмашылықты өлшейді, бірақ аннотаторлар қолдаған нақты класқа берілген төмен ықтималдықты жеке қате ретінде арнайы бөліп көрсетпейді және оны логарифмдік түрде күшейтпейді. Кросс-энтропия (CE) мен Куллбек–Лейблер дивергенциясы (KL) аннотаторлар қолдаған класқа модель тарапынан төмен ықтималдық берілген жағдайда қатаңырақ өседі. Ал Дженсен–Шеннон дивергенциясы (JS) екі ықтималдық үлестірімін симметриялық түрде салыстырады. Осы себепті Majority әдісі MD бойынша ең төмен мәнге ие болғанымен, CE, KL және JS көрсеткіштері бойынша нашар нәтиже көрсетуі мүмкін.

Аталған көрсеткіштер модельдік және аннотаторлық үлестірімдер арасындағы айырмашылықты әртүрлі қырынан сипаттайды (1-формула):

$$L = \alpha L_{F1} + \beta L_{CE} + \gamma L_{KL} + \delta L_{JS} + \eta L_{MD} + \lambda \|w\|_2^2 \quad (1)$$

Жіктеу сапасын тікелей оңтайландыру үшін көпшілік дауысы негізінде алынған белгілерге қатысты масго-F1 көрсеткішінің дифференциалданатын жуықтауы пайдаланылады (2-формула):

$$L_{F1} = 1 - \frac{2 \sum_{i=1}^N \hat{y}_i y_i}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i + \varepsilon} \quad (2)$$

Модель ықтималдықтарының аннотаторлар қалыптастырған нақты үлестірімге сәйкестігін бағалау үшін кросс-энтропия қолданылады (3-формула):

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C p_i(c) \log q_i(c) \quad (3)$$

Мұндағы, N -мәтіндер саны, C -класстар саны, $p_i(c)$ -і мәтін үшін аннотаторлар қалыптастырған c -класының эмпирикалық ықтималдығы, $q_i(c)$ -модель болжаған ықтималдық.

Екі ықтималдық үлестірім арасындағы ақпараттық айырмашылықты бағалау үшін KL алшақтығы пайдаланылады (4-формула):

$$L_{KL} = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C p_i(c) \log \left(\frac{p_i(c)}{q_i(c)} \right) \quad (4)$$

Аннотаторлық және модельдік үлестірімдердің симметриялы сәйкестігін бағалау үшін JS алшақтығы қолданылады (5-формула):

$$L_{JS} = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} [KL(p_i \| m_i) + KL(q_i \| m_i)] \quad (5)$$

Аннотаторлар арасындағы пікір айырмашылығы мен модель болжамының тікелей абсолюттік ауытқуын бағалау үшін Манхэттен қашықтығына негізделген орташа қашықтық көрсеткіші пайдаланылады (6-формула):

$$L_{MD} = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C |p_i(c) - q_i(c)| \quad (6)$$

Осы метрикаларды қатар пайдалану модельдің тек дұрыс класты таңдау қабілетін ғана емес, аннотаторлар арасындағы пікір айырмашылығын қаншалықты сақтайтынын да бағалауға мүмкіндік береді. CE, KL, JS және MD мәндерінің төмендеуі модельдік болжамдардың аннотаторлық үлестірімдерге жақындағанын білдіреді. Осылайша, аталған өлшемдер көпаннотаторлы деректер негізіндегі модельдерді кешенді бағалаудың негізгі көрсеткіштері ретінде пайдаланылды.

Осы жұмыста DART жеке трансформер архитектурасы ретінде емес, базалық модельдердің ықтималдық болжамдарын біріктіретін көпмақсатты ансамбльдік оңтайландыру қабаты ретінде қарастырылды. DART кірісіне таңдалған Ensemble әдісінің бірнеше ішкі нұсқасы қалыптастырған ықтималдық векторлары берілді. DART осы болжамдардың салмақтарын валидациялық жиында көпмақсатты функция бойынша оңтайландырып, оларды қорытынды ықтималдық үлестіріміне біріктірді. Әрбір модельдің үлесі w_m арқылы белгіленіп, салмақтарға $w_m \geq 0$ және $\sum_m w_m = 1$ шарттары қойылады. Қорытынды ықтималдық үлестірімі Ensemble отбасының нұсқалары қалыптастырған ықтималдық болжамдарының DART арқылы оңтайландырылған салмақталған қосындысы ретінде есептеледі.

Негізгі модель архитектурасының сипаттамасына келетін болсақ, bert-base-multilingual-uncased, максималды ұзындық – 128 токен, batch – 64, 20 эпоха, learning rate – 2×10^{-5} , төрт класты softmax қабаты көрсетіледі. BERT+LSTM үшін BiLSTM өлшемі және dropout мәні беріледі.

Эксперименттік нәтижелер. Төрт стратегияның нәтижелері 1-кестеде көрсетілген. Кесте нәтижелері бойынша Ensemble тәсілі жалпы ең тиімді нәтиже көрсетті: Macro-F1 – 0.907909, CE – 0.530665, KL – 0.276017, JS – 0.063951. Majority әдісі MD көрсеткіші бойынша ең жақсы мәнге ие болды – 0.319204. Probability және Instruction тәсілдерінің көрсеткіштері салыстырмалы түрде төмен болды. Жалпы, Ensemble жіктеу сапасы мен ықтималдық үлестірімдерінің сәйкестігі бойынша ең теңгерімді тәсіл ретінде ерекшеленді

Кесте 1 – Көпаннотаторлы ақпаратты өңдеу стратегияларының салыстырмалы нәтижелері

Әдіс	Macro-F1 ↑	CE ↓	KL ↓	JS ↓	MD ↓
Majority	0.903085	0.694143	0.439496	0.065156	0.319204
Probability	0.899954	0.530719	0.276072	0.066590	0.355420
Ensemble	0.907909	0.530665	0.276017	0.063951	0.343187
Instruction	0.900568	0.535533	0.280886	0.066684	0.356443

Төмендегі кестеде DART (Disagreement-Aware Representation Transformer) тәсіліндегі көпмақсатты оңтайландыру функциясының жеке компоненттері мен олардың әртүрлі комбинацияларының жіктеу сапасына және аннотаторлық үлестірімге сәйкестігіне әсері салыстырылды. Нәтижелер бір ғана мақсаттық функцияны оңтайландыру барлық көрсеткіштер бойынша бірдей артықшылық бермейтінін көрсетті. Атап айтқанда, L_{JS} жеке қолданылған жағдайда JS көрсеткіші бойынша ең төмен мән – 0.060796, ал L_{MD} кезінде MD бойынша ең төмен мән – 0.307109 алынды. Алайда бұл конфигурацияларда CE және KL көрсеткіштері нашарлаған.

Бірнешемақсаттыбіргеескергенжағдайданеғұрлымтеңгерімдінәтиже $L_{CE} + L_{KL} + L_{MD}$ комбинациясында байқалды. Бұл конфигурация Macro-F1=0.907040 нәтижесін сақтай отырып, CE=0.528543 және KL=0.273896 бойынша кестедегі ең төмен мәндерге қол жеткізді. Сонымен қатар төрт бағалау көрсеткіші бойынша жақсару тіркелді. $L_{CE} + L_{KL} + L_{JS} + L_{MD}$ комбинациясы да төрт көрсеткіш бойынша жақсару бергенімен, CE және KL мәндері тиісінше 0.528695 және 0.274048 болып, үш компонентті нұсқадан сәл төмен нәтиже көрсетті.

ОсынәтижелерDARTаясындамақсаттықфункциякомпоненттерініңсанын жай ғана көбейту автоматты түрде тиімділікті арттырмайтынын көрсетеді. Киберэкстремистік мәтіндердің идеологиялық бағыттарын көпаннотаторлы деректер негізінде жіктеу үшін жіктеу сапасы мен аннотаторлық үлестірімге сәйкестікті қатар ескеретін $L_{CE} + L_{KL} + L_{MD}$ комбинациясы қарастырылған нұсқалардың ішіндегі ең теңгерімді шешім ретінде анықталды.

Кесте 2 – DART көпмақсатты оңтайландыруындағы шығын функциялары комбинацияларының абляциялық талдау нәтижелері

Жағдай	Macro-F1 ↑	CE ↓	KL ↓	JS ↓	MD ↓	L ↓
L_{FI}	0.906054	0.530867	0.276220	0.063708	0.342186	0.083768
L_{CE}	0.906103	0.528614	0.273967	0.064052	0.344253	0.512194
L_{KL}	0.906103	0.528614	0.273967	0.064052	0.344253	0.263197
L_{JS}	0.906054	0.581183	0.326535	0.060796	0.318299	0.058901
L_{MD}	0.904092	0.687783	0.433136	0.061531	0.307109	0.294330

$L_{F1} + L_{CE}$	0.907006	0.529252	0.274605	0.064196	0.344793	0.596902
$L_{F1} + L_{KL}$	0.907006	0.529252	0.274605	0.064196	0.344793	0.347904
$L_{F1} + L_{JS}$	0.905117	0.580961	0.326314	0.060944	0.319042	0.142447
$L_{F1} + L_{MD}$	0.904092	0.687783	0.433136	0.061531	0.307109	0.378791
$L_{CE} + L_{KL}$	0.907006	0.528666	0.274019	0.064073	0.344356	0.775107
$L_{CE} + L_{JS}$	0.906103	0.528833	0.274186	0.064006	0.343983	0.573777
$L_{CE} + L_{MD}$	0.907040	0.532337	0.277690	0.063053	0.338780	0.839948
$L_{KL} + L_{JS}$	0.906103	0.528833	0.274186	0.064006	0.343983	0.324779
$L_{KL} + L_{MD}$	0.907040	0.532337	0.277690	0.063053	0.338780	0.590950
$L_{JS} + L_{MD}$	0.904092	0.687783	0.433136	0.061531	0.307109	0.353508
$L_{F1} + L_{CE} + L_{KL}$	0.907006	0.529277	0.274630	0.064203	0.344818	0.860039
$L_{F1} + L_{CE} + L_{JS}$	0.907006	0.529266	0.274619	0.064185	0.344741	0.658548
$L_{F1} + L_{CE} + L_{MD}$	0.906054	0.533217	0.278570	0.063016	0.338442	0.923791
$L_{F1} + L_{KL} + L_{JS}$	0.907006	0.529266	0.274619	0.064185	0.344741	0.409551
$L_{F1} + L_{KL} + L_{MD}$	0.906054	0.533217	0.278570	0.063016	0.338442	0.674794
$L_{F1} + L_{JS} + L_{MD}$	0.904092	0.687783	0.433136	0.061531	0.307109	0.437970
$L_{CE} + L_{KL} + L_{JS}$	0.906103	0.528703	0.274056	0.064028	0.344120	0.836729
$L_{CE} + L_{KL} + L_{MD}$	0.907040	0.528543	0.273896	0.063782	0.342897	1.103729
$L_{CE} + L_{JS} + L_{MD}$	0.907040	0.534967	0.280320	0.062698	0.336613	0.900497
$L_{KL} + L_{JS} + L_{MD}$	0.907040	0.534967	0.280320	0.062698	0.336613	0.651499
$L_{F1} + L_{CE} + L_{KL} + L_{JS}$	0.907006	0.529292	0.274645	0.064198	0.344794	0.921708
$L_{F1} + L_{CE} + L_{KL} + L_{MD}$	0.906103	0.529296	0.274649	0.063792	0.342758	1.188437
$L_{F1} + L_{CE} + L_{JS} + L_{MD}$	0.906054	0.536411	0.281764	0.062539	0.335604	0.984338
$L_{F1} + L_{KL} + L_{JS} + L_{MD}$	0.906054	0.536411	0.281764	0.062539	0.335604	0.735341
$L_{CE} + L_{KL} + L_{JS} + L_{MD}$	0.907040	0.528695	0.274048	0.063747	0.342687	1.165173
$L_{F1} + L_{CE} + L_{KL} + L_{JS} + L_{MD}$	0.906103	0.529311	0.274663	0.063787	0.342738	1.249891

Кестеде әртүрлі мақсаттық функциялар мен олардың комбинацияларының базалық Ensemble нәтижесіне қатысты өзгерістері көрсетілген. Мұнда Macro-F1 үшін оң өзгеріс жақсы нәтиже білдірсе, CE, KL, JS және MD метрикалары үшін теріс өзгеріс көрсеткіштің төмендеп, жақсарғанын білдіреді.

Нәтижелер бойынша $L_{CE} + L_{KL} + L_{MD}$ комбинациясы ең теңгерімді нұсқалардың бірі болды. Бұл конфигурацияда Macro-F1 небәрі 0.000869-ға төмендегенімен, CE және KL көрсеткіштері тиісінше 0.002121-ге жақсарды, сонымен қатар JS 0.000169-ға және MD 0.000290-ға төмендеді, яғни жіктеу сапасының өте аз төмендеуі есебінен аннотаторлық және модельдік ықтималдық үлестірімдерінің сәйкестігін сипаттайтын төрт көрсеткіштің барлығы жақсарды.

$L_{CE} + L_{KL} + L_{JS} + L_{MD}$ комбинациясында да Macro-F1 өзгерісі -0.000869 деңгейінде қалып, CE, KL, JS және MD көрсеткіштерінің барлығы жақсарды. Бұл нұсқа JS және MD бойынша $L_{CE} + L_{KL} + L_{MD}$ комбинациясынан жақсырақ болғанымен, CE және KL бойынша жақсару шамасы салыстырмалы түрде төмен болды.

Жеке L_{JS} немесе L_{MD} компоненттерін оңтайландыру сәйкесінше JS және MD көрсеткіштерін едәуір жақсартқанымен, CE және KL мәндерінің нашарлауына және Macro-F1 көрсеткішінің төмендеуіне алып келді. Осыған ұқсас үрдіс JS және MD компоненттері басым бірқатар комбинацияларда да байқалды.

Жалпы нәтижелер бір ғана метриkanı оңтайландыру басқа көрсеткіштердің нашарлауына әкелуі мүмкін екенін көрсетті. Ал $L_{CE} + L_{KL} + L_{MD}$ комбинациясы Macro-F1 деңгейін іс жүзінде сақтай отырып, ықтималдық үлестірімдерінің сәйкестігін сипаттайтын CE, KL, JS және MD көрсеткіштерінің барлығын бір мезгілде жақсартқандықтан, DART көпмақсатты оңтайландыруы үшін теңгерімді шешім ретінде қарастырылды.

Кесте 3 – DART көпмақсатты оңтайландыруындағы метрикалар өзгерістерінің салыстырмалы талдауы

Жағдай	$\Delta \text{Macro-F1} \uparrow$	$\Delta \text{CE} \downarrow$	$\Delta \text{KL} \downarrow$	$\Delta \text{JS} \downarrow$	$\Delta \text{MD} \downarrow$
L_{F1}	-0.001855	0.000202	0.000202	-0.000244	-0.001001
L_{CE}	-0.001806	-0.002050	-0.002050	0.000101	0.001066
L_{KL}	-0.001806	-0.002050	-0.002050	0.000101	0.001066
L_{JS}	-0.001855	0.050518	0.050518	-0.003155	-0.024888
L_{MD}	-0.003817	0.157118	0.157118	-0.002420	-0.036078
$L_{F1} + L_{CE}$	-0.000903	-0.001412	-0.001412	0.000245	0.001606
$L_{F1} + L_{KL}$	-0.000903	-0.001412	-0.001412	0.000245	0.001606
$L_{F1} + L_{JS}$	-0.002792	0.050297	0.050297	-0.003008	-0.024145
$L_{F1} + L_{MD}$	-0.003817	0.157118	0.157118	-0.002420	-0.036078
$L_{CE} + L_{KL}$	-0.000903	-0.001998	-0.001998	0.000121	0.001169
$L_{CE} + L_{JS}$	-0.001806	-0.001832	-0.001832	0.000055	0.000796
$L_{CE} + L_{MD}$	-0.000869	0.001672	0.001672	-0.000899	-0.004407
$L_{KL} + L_{JS}$	-0.001806	-0.001832	-0.001832	0.000055	0.000796
$L_{KL} + L_{MD}$	-0.000869	0.001672	0.001672	-0.000899	-0.004407
$L_{JS} + L_{MD}$	-0.003817	0.157118	0.157118	-0.002420	-0.036078
$L_{F1} + L_{CE} + L_{KL}$	-0.000903	-0.001388	-0.001388	0.000251	0.001631
$L_{F1} + L_{CE} + L_{JS}$	-0.000903	-0.001398	-0.001398	0.000234	0.001554
$L_{F1} + L_{CE} + L_{MD}$	-0.001855	0.002553	0.002553	-0.000935	-0.004745
$L_{F1} + L_{KL} + L_{JS}$	-0.000903	-0.001398	-0.001398	0.000234	0.001554

$L_{F1} + L_{KL} + L_{MD}$	-0.001855	0.002553	0.002553	-0.000935	-0.004745
$L_{F1} + L_{JS} + L_{MD}$	-0.003817	0.157118	0.157118	-0.002420	-0.036078
$L_{CE} + L_{KL} + L_{JS}$	-0.001806	-0.001961	-0.001961	0.000077	0.000933
$L_{CE} + L_{KL} + L_{MD}$	-0.000869	-0.002121	-0.002121	-0.000169	-0.000290
$L_{CE} + L_{JS} + L_{MD}$	-0.000869	0.004302	0.004302	-0.001253	-0.006574
$L_{KL} + L_{JS} + L_{MD}$	-0.000869	0.004302	0.004302	-0.001253	-0.006574
$L_{F1} + L_{CE} + L_{KL} + L_{JS}$	-0.000903	-0.001373	-0.001373	0.000247	0.001607
$L_{F1} + L_{CE} + L_{KL} + L_{MD}$	-0.001806	-0.001369	-0.001369	-0.000159	-0.000429
$L_{F1} + L_{CE} + L_{JS} + L_{MD}$	-0.001855	0.005746	0.005746	-0.001413	-0.007583
$L_{F1} + L_{KL} + L_{JS} + L_{MD}$	-0.001855	0.005746	0.005746	-0.001413	-0.007583
$L_{CE} + L_{KL} + L_{JS} + L_{MD}$	-0.000869	-0.001969	-0.001969	-0.000205	-0.000500
$L_{F1} + L_{CE} + L_{KL} + L_{JS} + L_{MD}$	-0.001806	-0.001354	-0.001354	-0.000165	-0.000449

Осылайша, оңтайландырылған тәсіл жіктеу сапасын жоғары деңгейде сақтай отырып, аннотаторлық және модельдік үлестірімдердің сәйкестігін жақсартты. Алынған нәтижелер көпаннотаторлы деректермен жұмыс істеуде бірнеше бағалау көрсеткішін қатар ескерудің тиімділігін көрсетеді.

Төмендегі кестеде идеологиялық бағыттарды жіктеу үшін қолданылған үш біріктіру тәсілінің салыстырмалы нәтижелері көрсетілген. Дауыс беру CE, KL және JS бойынша бұл тәсілдің нәтижелері қалған әдістермен салыстырғанда төмен болды. Ал ықтималдықтарды орташалау әдісі Macro-F1=0,907909 мәнімен жіктеу сапасы бойынша ең жоғары нәтижені көрсетті.

Кесте 4 – DART көпмақсатты оңтайландыруындағы метрикалар өзгерістерінің салыстырмалы талдауы

Деректер жиыны	Әдіс	Macro-F1 ↑	CE ↓	KL ↓	JS ↓	MD ↓
Ideology	Voting	0.903085	0.694143	0.439496	0.065156	0.319204
Ideology	Averaging	0.907909	0.530665	0.276017	0.063951	0.343187
Ideology	Optimised ($L_{CE} + L_{KL} + L_{MD}$)	0.907040	0.528543	0.273896	0.063782	0.342897

DART әдісі арқылы $L_{CE} + L_{KL} + L_{MD}$ мақсаттық функциясы бойынша оңтайландырылған ансамбльде Macro-F1 көрсеткіші 0,907040-қа тең болып, орташалау әдісінен бар болғаны 0,000869-ға төмен болды. Соған қарамастан, бұл тәсілде кросс-энтропия 0,528543-ке, KL дивергенциясы 0,273896-ға, JS дивергенциясы 0,063782-ге дейін төмендеді. MD көрсеткіші 0,342897 болды.

Алынған нәтижелер DART негізіндегі оңтайландырудың жіктеу сапасын елеулі төмендетпей, модельдің болжамдық үлестірімін аннотаторлар қалыптастырған ықтималдықтық үлестірімге жақындатуға мүмкіндік

беретінін көрсетті. Осы тұрғыдан алғанда, ұсынылған тәсіл тек қатаң классты дұрыс анықтауға емес, сонымен қатар аннотаторлар пікіріндегі белгісіздік пен айырмашылықты модель болжамында сақтауға бағытталған.

Сонымен қатар қазақ тіліндегі мәтіндерді психолингвистикалық белгілер негізінде жіктеуге арналған жеке эксперименттер жүргізілді. Бұл кезеңде аннотаторлар арасындағы келіспеушіліктер ескеріліп, нақтыланған деректер жиыны пайдаланылды. Негізгі мақсат – өңделген деректер негізінде мәтін кластарын ажыратуда ең тиімді модель немесе модельдер комбинациясын анықтау. Эксперименттік нәтижелер Precision, Recall, Accuracy, F1-score және ROC-AUC метрикалары арқылы бағаланып, әр модельдің жіктеу сапасы өзара салыстырылды.

Кесте 5 – Disagreement және PLAMkz (psycho-linguistic analysis module) негізіндегі мультиклассификациялау модельдерінің салыстырмалы нәтижелері

	Алгоритм	Вектор- лау әдісі	Precision	Recall	Accuracy	F1-Score	R O C - AUC
Disagreement+ PLAMkz	SVC	TF-IDF	0.945350	0.945159	0.946191	0.944719	0.995233
	Logistic Regression	TF-IDF	0.918157	0.917524	0.919637	0.917377	0.990805
	Naive Bayes	TF-IDF	0.889656	0.889653	0.892383	0.889625	0.983923
	Gradient Boosting	TF-IDF	0.889892	0.886223	0.888190	0.885991	0.983603
	Extra Trees	TF-IDF	0.893095	0.885343	0.886792	0.885337	0.981851
	Random Forest	TF-IDF	0.880841	0.868897	0.870720	0.869525	0.979048
	Decision Tree	TF-IDF	0.833143	0.833121	0.836478	0.833038	0.889516
	AdaBoost	TF-IDF	0.805355	0.787692	0.791055	0.789301	0.931845
	K-Nearest Neighbors	TF-IDF	0.679528	0.676808	0.687631	0.673754	0.873464
Disagreement+ PLAMkz	BERT	BERT	0.950302	0.950577	0.951083	0.950433	0.996074
Disagreement+ PLAMkz	BERT+ LSTM	BERT	0.947460	0.945405	0.946890	0.946075	0.995724

Бірінші эксперименттік конфигурацияда аннотаторлар арасындағы келіспеушіліктерді ескере отырып өңделген деректерге PLAMkz психолингвистикалық белгілері қосылып, мәтіндер TF-IDF тәсілі арқылы векторланды. Алынған белгілер жиыны дәстүрлі машиналық оқыту алгоритмдеріне берілді. Салыстыру нәтижесінде SVC моделі классикалық алгоритмдер арасында ең жоғары көрсеткіштерге қол жеткізді. Оның бағалау метрикалары бойынша тұрақты нәтиже көрсетуі аталған конфигурацияның дәстүрлі модельдер ішіндегі тиімді нұсқа екенін көрсетті.

Келесі экспериментте мәтіннің контекстік ерекшеліктерін көрсету үшін BERT моделінен алынған эмбеддингтер пайдаланылды. Бұл белгілер

PLAMkz модулі арқылы есептелген психолінгвистикалық сипаттамалармен біріктірілді. Мұндай біріктіру нәтижесінде модельдің жіктеу сапасы классикалық тәсілдермен салыстырғанда жоғарылады. Accuracy мәні 95,1%-ға жетіп, ROC-AUC көрсеткіші 99,6%-дан жоғары болды. Алынған мәндер контекстік мәтіндік белгілер мен психолінгвистикалық сипаттамаларды бір белгілер кеңістігінде қарастырудың тиімді екенін көрсетті.

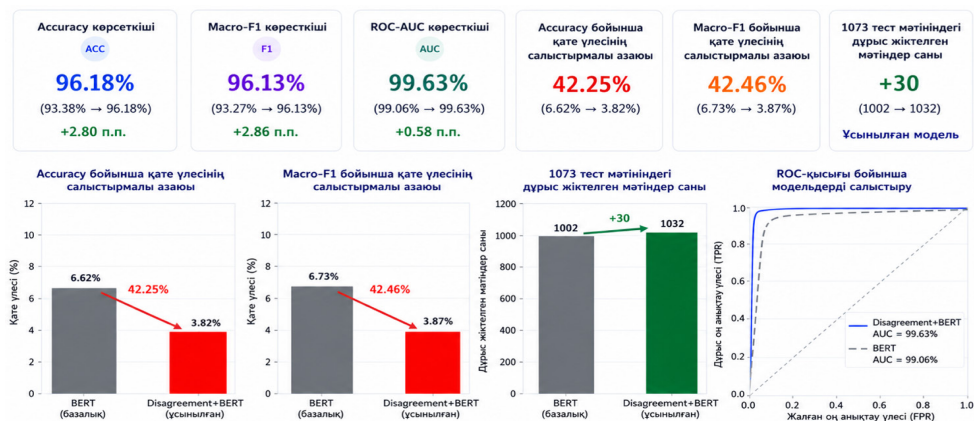
Үшінші конфигурацияда BERT эмбедингтері алдымен LSTM желісі арқылы өңделіп, одан кейін PLAMkz белгілерімен біріктірілді. Бұл модель де жоғары нәтижелер көрсетті, алайда негізгі бағалау көрсеткіштері бойынша BERT және PLAMkz белгілерін тікелей біріктірген конфигурациядан аздаған айырмашылық байқалды. Демек, осы эксперимент жағдайында LSTM қабатын енгізу модель сапасына елеулі қосымша өсім берген жоқ. Жүргізілген салыстырулар PLAMkz белгілері енгізілген модельдердің арасынан Disagreement + BERT + PLAMkz конфигурациясының нәтижесі жоғары болғанын көрсетті. Сондықтан зерттеу аясында қазақ тіліндегі мәтіндердің психолінгвистикалық ерекшеліктерін ескере отырып жіктеу үшін аталған тәсіл негізгі конфигурация ретінде қарастырылды.

Талқылау. Аннотаторлар арасындағы келіспеушілікті ескеру негізінде нақтыланған деректер жиынын пайдалану жіктеу нәтижелерінің жақсаруымен қатар жүрді. Базалық BERT моделінде Accuracy көрсеткіші 93.38% болса, келіспеушілік ақпараты ескерілген деректермен оқытылған Disagreement+BERT моделінде бұл мән 96.18%-ға дейін өсті. Macro-F1 көрсеткіші де 93.27%-дан 96.13%-ға жоғарылады. Сонымен қатар ROC-AUC 99.06%-дан 99.63%-ға жетіп, модельдің кластарды ажырату қабілетінің сақталғанын және біршама жақсарғанын көрсетті.

Ұсынылған BERT нәтижелері әртүрлі эксперименттік конфигурациялар негізінде алынған. 5-кестеде PLAMkz белгілерімен біріктірілген модельдердің көрсеткіштері берілген. BERT моделі үшін Precision мәні – 0,950302, Recall – 0,950577, ал Accuracy – 0,951083.

Ал 1-суреттегі 96,18% Accuracy PLAMkz белгілері пайдаланылмаған абляциялық эксперимент нәтижесін сипаттайды. Бұл экспериментте базалық BERT және Disagreement+BERT конфигурациялары салыстырылған.

Қателер үлесін салыстыру да осы үрдісті айқындайды. Accuracy бойынша қате үлесі 6.62%-дан 3.82%-ға дейін төмендеп, салыстырмалы азаю 42.25%-ды құрады. Macro-F1 тұрғысынан қате үлесі 6.73%-дан 3.87%-ға дейін азайып, салыстырмалы төмендеу 42.46% болды. 1073 тест мәтінінің ішінде базалық модель 1002 мәтінді дұрыс жіктесе, келіспеушілікке негізделген тәсіл қолданылғаннан кейін дұрыс жіктелген мәтіндер саны 1032-ге дейін артты, яғни қосымша 30 мәтін дұрыс анықталды (1-сурет).



Сурет 1 – Модельдердің абляциялық әсерін бағалау

Бұл нәтижелер көпнотаторлы таңбалаудағы қайшылықты мысалдарды жай ғана қате ретінде алып тастамай, оларды анықтап, қайта қарау деректер жиынындағы класс шекараларын нақтылауға ықпал еткенін көрсетеді. Әсіресе мағыналық жағынан жақын пропаганда, радикализация және рекрутмент кластары үшін мұндай нақтылау модельге бір-біріне ұқсас мәтіндерді жақсырақ ажыратуға мүмкіндік берген болуы мүмкін. Осылайша, келіспеушілікке негізделген оқыту бұл зерттеуде тек аннотация сапасын бақылау құралы ретінде емес, деректер жиынын жетілдіру және кейінгі классификация сапасын арттыруға ықпал ететін әдістемелік кезең ретінде қарастырылды.

Мәтіндік мысалдар мен типтік жіктеу қателерінің талдауы (Mussiraliyeva et al., 2025a) жұмыста келтірілген. Аталған еңбекте мәтіндердің кластарға бөліну ерекшеліктері және модельдердің қате жіктеу жағдайлары талданған. Сонымен қатар пропаганда, рекрутмент және радикализация кластары арасындағы қателердің семантикалық жақындықпен, әрекетке шақыру белгілерімен, зорлық-зомбылыққа қатысты лексикамен және идеологиялық мазмұнмен байланысты екені түсіндірілген.

Қорытынды. Зерттеу нәтижелері қазақ тіліндегі киберэкстремистік мәтіндерді жіктеуде аннотаторлар арасындағы келіспеушілікті ескерудің тиімді екенін көрсетті. Көпшілік дауысы, ықтималдықтық тәсіл, ансамбльдік біріктіру және нұсқаулыққа негізделген стратегияларды салыстыру барысында Ensemble әдісі Macro-F1=0.907909 көрсеткішімен ең жоғары нәтижеге қол жеткізіп, CE, KL және JS метрикалары бойынша да жақсы нәтиже көрсетті. DART аясындағы көпмақсатты оңтайландыруда $L_{CE} + L_{KL} + L_{MD}$ комбинациясы жіктеу сапасы мен аннотаторлық үлестірімге сәйкестік арасындағы теңгерімді сақтады. Нақтыланған көпнотаторлы деректер жиынын пайдалану BERT негізіндегі жіктеу сапасын арттырып, қате үлесін төмендетті. PLAMkz психолінгвистикалық белгілерін BERT-пен біріктіру де

жоғары нәтижелер көрсетті. Жалпы алғанда, аннотаторлар келіспеушілігін қосымша ақпарат көзі ретінде пайдалану қазақ тіліндегі идеологиялық мәтіндерді дәлірек жіктеуге және модель нәтижелерінің сенімділігін арттыруға мүмкіндік берді.

References

- Aldera S., Emam A., Al-Qurishi M., Alrubaihan M., Alothaim A. (2021) Online Extremism Detection in Textual Content: A Systematic Literature Review. *IEEE Access*, 9. — P. 42384–42396. DOI: 10.1109/ACCESS.2021.3064178 (in Eng.).
- Aroyo L., Welty C. (2015) Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation. *AI Magazine*, 36(1). — P. 15–24. DOI: 10.1609/aimag.v36i1.2564 (in Eng.).
- Baisylbayeva K., Mussiraliyeva Sh., Yeltay Zh. (2026) Қазақ тіліндегі екстремистік идеологияны анықтау: аннотациялау мәселелері және терең оқыту тәсілдері [Identification of extremist ideology in the Kazakh language: problems of annotation and approaches to deep learning]. *International Journal of Information and Communication Technologies*, 7(2). — P. 242–260. DOI: 10.54309/IJICT.2026.26.2.016 (in Kaz.).
- Berjawi O., Fenza G., Loia V. (2023) A Comprehensive Survey of Detection and Prevention Approaches for Online Radicalization: Identifying Gaps and Future Directions. *IEEE Access*, 11. — P. 120463–120491. DOI: 10.1109/ACCESS.2023.3326995 (in Eng.).
- Bolatbek M.A., Baisylbayeva K.D., Sagynai M., Mussiraliyeva Sh.Zh., Zhumakhanova A.N. (2023) Internet кеңістігіндегі зһастарға бартыталан деһруктивті мәтіндерді зһинақтауға қазһетті парсер бардартласын әзірлеу [Development of a parser program necessary for the collection of destructive texts aimed at young people in the internet space]. *Academic Scientific Journal of Computer Science*, 4. — P. 117–131. DOI: 10.32014/2023.2518-1726.223 (in Kaz.).
- Bolatbek M., Mussiraliyeva S., Baisylbayeva K. (2025) Detection and Classification of Ideological Texts in the Kazakh Language Using Machine Learning and Transformers. *Research in Language*, 23. — P. 341–353. DOI: 10.18778/1731-7533.23.21 (in Eng.).
- Cui X., Huang Z., Abeynayake N.R. (2026) Learning from Annotator Disagreement Via Weighted Ensemble Optimisation for Subjective Text Classification. *Datenbank-Spektrum*, 26. — P. 39–48. DOI: 10.1007/s13222-026-00529-9 (in Eng.).
- Dumitrache A., Inel O., Aroyo L., Timmermans B., Welty C. (2018) CrowdTruth 2.0: Quality Metrics for Crowdsourcing with Disagreement. *Joint Proceedings of SAD 2018 and CrowdBias 2018*. — P. 11–18. arXiv:1808.06080 (in Eng.).
- Fleisig E., Abebe R., Klein D. (2023) When the Majority Is Wrong: Modeling Annotator Disagreement for Subjective Tasks. *Proceedings of EMNLP 2023*. — P. 6715–6726. DOI: 10.18653/v1/2023.emnlp-main.415 (in Eng.).
- Fornaciari T., Uma A., Paun S., Plank B., Hovy D., Poesio M. (2021) Beyond Black & White: Leveraging Annotator Disagreement via Soft-Label Multi-Task Learning. *Proceedings of NAACL-HLT 2021*. — P. 2591–2597. DOI: 10.18653/v1/2021.naacl-main.204 (in Eng.).
- Gaikwad M., Ahirrao S., Phansalkar S., Kotecha K. (2021) Online Extremism Detection: A Systematic Literature Review With Emphasis on Datasets, Classification Techniques, Validation Methods, and Tools. *IEEE Access*, 9. — P. 48364–48404. DOI: 10.1109/ACCESS.2021.3068313 (in Eng.).
- Gaikwad M., Ahirrao S., Phansalkar S., Kotecha K., Rani S. (2023) Multi-Ideology, Multiclass Online Extremism Dataset, and Its Evaluation Using Machine Learning. *Computational Intelligence and Neuroscience*, 2023. — Article 4563145. DOI: 10.1155/2023/4563145 (in Eng.).
- Govers J., Feldman P., Dant A., Patros P. (2023) Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech. *ACM Computing Surveys*, 55(14s). — Article 319. DOI: 10.1145/3583067 (in Eng.).
- Leonardelli E., Abercrombie G., Almanea D., Basile V., Fornaciari T., Plank B., Rieser V., Uma A., Poesio M. (2023) SemEval-2023 Task 11: Learning with Disagreements (LeWiDi). *Proceedings*

of the 17th International Workshop on Semantic Evaluation. — P. 2304–2318. DOI: 10.18653/v1/2023.semeval-1.314 (in Eng.).

Mostafazadeh Davani A., Diaz M., Prabhakaran V. (2022) Dealing with Disagreements: Looking Beyond the Majority Vote in Subjective Annotations. *Transactions of the Association for Computational Linguistics*, 10. — P. 92–110. DOI: 10.1162/tacl_a_00449 (in Eng.).

Mussiraliyeva S., Baisylbayeva K., Bolatbek M., Yeltay Z. (2024) Decoding Ideology: Machine Learning-Based Detection of Extremist Content. 2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS). — P. 114–120. DOI: 10.1109/ICCNS62192.2024.10776480 (in Eng.).

Mussiraliyeva S., Bolatbek M., Baisylbayeva K. (2025) Extremist Ideology Classification in Kazakh: A Multi-Class Approach Using Machine Learning and Psycholinguistic Analysis. *IEEE Access*, 13. — P. 140500–140518. DOI: 10.1109/ACCESS.2025.3596601 (in Eng.).

Mussiraliyeva S., Bolatbek M., Baisylbayeva K., Yeltay Z. (2025) Multi-ideology Dataset for Detecting Online Extremism in Kazakh Language. *Proceedings of the International Conference on Information Technology and Applications (ICITA 2024). Lecture Notes in Networks and Systems*, 1248. — P. 499–509. DOI: 10.1007/978-981-96-1758-6_41 (in Eng.).

Uma A.N., Fornaciari T., Hovy D., Paun S., Plank B., Poesio M. (2021) Learning from Disagreement: A Survey. *Journal of Artificial Intelligence Research*, 72. — P. 1385–1470. DOI: 10.1613/JAIR.1.12752 (in Eng.).

Xu Y., Jurgens D. (2026) Beyond Consensus: Perspectivist Modeling and Evaluation of Annotator Disagreement in NLP. *arXiv preprint*. — arXiv:2601.09065 (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.