

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN
ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
KAZAKHSTAN



**INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION
TECHNOLOGIES**

Published since 2020.
Volume 7. 2 (26). 2026
April–June

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

2020 жылдан бері шығарылады
Том 7. 2 (26). 2026
Сәуір-Маусым

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Издается с 2020 г.
Том 7. 2 (26). 2026
Апрель-Июнь

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

EDITOR-IN-CHIEF:

Kateryna Kolesnikova — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

DEPUTY EDITOR-IN-CHIEF:

Madina Ipalakova — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

EDITORIAL BOARD:

Abdul Razak — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Lucio Tommaso De Paolis — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)

Liz Bacon — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)

Michele Pagano — PhD, Professor, University of Pisa (Italy)

Mukhtarbay Otelbayev — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Bolatbek Rysbauly — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

Yevgeniya Daineko — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurzhan Duzbayev — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)

Bakhtgerci Sinchev — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Nurgul Seilova — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Ardak Mukhamediyeva — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)

Zamira Abdikalikova — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Yerlan Shildibekov — PhD, associate professor, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Damilya Yeskendirowa — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)

Aigul Niyazgulova — Candidate of Philological Sciences, Professor, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Altai Aitmagambetov — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Yelena Bakhtiyarova — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)

Kanibek Sansyrbay — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Sakhybay Tynymbayev — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)

Ali Abd Almisreb — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)

Yang Im Chu — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, Vice-Rector, Adam Mickiewicz University (Poland)

Orken Mamyrbayev — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)

Sergey Bushuyev — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association "UKRNET," Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)

Svetlana Beloshitskaya — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

MANAGING EDITOR

Raushan Mrzabayeva — Master of Science, editor, International Information Technology University (Kazakhstan)

International Journal of Information and Communication Technologies

Periodicity: 4 times a year.

Languages: Kazakh, Russian, English

DOI prefix: 10.54309

ISSN 2708-2032 (print)

ISSN 2708-2040 (online)

Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".

Distribution: Materials are distributed under the Creative Commons Attribution 4.0

Journal website: <https://journal.iitu.edu.kz>

Owner: International Information Technology University JSC (Almaty).

Copyright: © International Journal of Information and Communication Technologies, 2026

РЕДАКЦИЯ

БАС РЕДАКТОР:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

- Разак Абдул** — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)
Луччо Томмазо де Паолис — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры
Лиз Бэкон — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары
Микеле Пагано — PhD, Пиза Университетінің (Италия) профессоры
Өтелбаев Мухтарбай Өтелбайұлы — физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)
Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Қазақстан)
Дайнеко Евгения Александровна — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)
Дузаев Нуржан Токсулжаевич — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)
Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)
Сейлова Нургуль Абдуллаевна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)
Мухамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнес-медиа және басқару факультетінің деканы (Қазақстан)
Абдикаликова Замира Турсынбаевна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)
Шильдибеков Ерлан Жаржанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)
Дамелия Максумовна Ескендрова — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)
Ниязгулова Айгуль Аскарбековна — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)
Айтмағамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)
Бахтиярова Елена Ажибековна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)
Канибек Сансызбай — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)
Тынымбаев Сахибай — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)
Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)
Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)
Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)
Талеуш Валлас — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры
Мамырбаев Оркен Жумажанович — PhD, ҚР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)
Бушув Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сулет университеті жобаларды басқару кафедрасының меңгерушісі (Украина)
Белюшицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Мерзімділігі: жылына 4 рет.

Басылым тілі: қазақ, орыс, ағылшын.

Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".

Журнал сайты: <https://journal.iitu.edu.kz>

Тарату: материалдар Creative Commons Attribution 4.0 лицензиясы бойынша таратылады

Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).

Авторлық құқық: © Халықаралық ақпараттық және коммуникациялық технологиялар журналы, 2026

РЕДАКЦИЯ

ГЛАВНЫЙ РЕДАКТОР:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

Абдикаликова Замира Турсынбаевна — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шильдибеков Ерлан Жаржанович — PhD, ассоциированный профессор, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Дамелия Максуговна Ескендрова — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, профессор, заведующая кафедрой медиакоммуникации и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Бахтиярова Елена Ажибековна — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Канибек Сансызбай — PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

Тынымбаев Сахпай — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алимуралиев Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас — PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белошницкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Префикс DOI: 10.54309

Периодичность: 4 выпусков в год.

Язык издания: казахский, русский, английский.

Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социально-экономических систем".

Сайт журнала: <https://journal.iitu.edu.kz>

Распространение: материалы распространяются по лицензии Creative Commons Attribution 4.0

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Авторские права: © Международный журнал информационных и коммуникационных технологий, 2026

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

D. Abzhanova, A. Biloshchytski

A MODEL AND METHOD FOR MANAGING DATA ON EMISSIONS FROM STATIONARY SOURCES OF POLLUTION IN AN INTELLIGENT ENVIRONMENTAL MONITORING SYSTEM9

A. Slanbekova, M. Rakhimzhanova, A. Zhanibekova, A. Alimagambetova, M. Xudoyberganov

EARLY DETECTION OF HYDROLOGICAL HAZARDS BASED ON SPATIOTEMPORAL ANALYSIS25

INFORMATION TECHNOLOGY

F.N. Abdraimova, A.A. Kereibayeva, D.S. Dyussenova, D.A. Aliyeva, T.Zh. Toktarova

AI TECHNOLOGIES IN LANGUAGE EDUCATION: PRACTICAL ASPECTS AND CHALLENGES OF STUDENT USAGE36

G. Azieva, M. Yessenova, A. Abzhapparova, G. Abdikerimova, P. Schmidt

HYBRID STACKING FRAMEWORK FOR CROP CLASSIFICATION USING UAV DATA50

A.K. Aitim

JOINT MORPHOLOGICAL DISAMBIGUATION AND POS TAGGING FOR AGGLUTINATIVE LANGUAGES62

S.A. Yesniyazova, S.T. Kaimov

PREDICTIVE MAINTENANCE OF HEAVY-DUTY TRUCKS USING EXPLAINABLE MACHINE LEARNING78

T. Imanbekova, Zh. Ibrayeva, G. Jakanova, G. Askanbay

DATA COMPRESSION ALGORITHM BASED ON WAVELET TRANSFORMER; ANALYSIS AND IMPLEMENTATION IN MATLAB92

B.Z. Kenzhegulov, Zh.T. Bilyalova, K.N. Uteuliyeva, L. Nurgaliyeva, Sh.S. Nurzhanova

A MATHEMATICAL AND ALGORITHMIC APPROACH TO THE DEVELOPMENT OF AN INTELLIGENT TEXT-TO-SQL SYSTEM BASED ON LARGE LANGUAGE MODELS110

N.Sh. Maxutova, J.A. Tussupov, A.A. Shekerbek, Zh.E. Kenzhebayeva, Q.O. Rakhimov

MACHINE LEARNING FOR COMPREHENSIVE EVALUATION OF CARDIOVASCULAR DISEASE RISK AND BIOCHEMICAL ALTERATIONS: FOCUS ON ASPARTATE AMINOTRANSFERASE131

O.S. Salykova, V.A. Madin, B.R. Salykov, D.N. Komarov, N.V. Manuilov

INTEGRATION OF MEMS ACCELEROMETER SENSOR MODULES IN INDUSTRIAL MONITORING SYSTEMS146

R. Taberkhan, M.A. Sambetbayeva, G. Kalman

KAZCAUSAL: THE FIRST CORPUS-BASED ANNOTATION OF CAUSAL RELATIONSHIPS IN THE KAZAKH LANGUAGE160

S.Tynymbayev, S.E. Mamanova, R. Berdybayev, Zh.E. Temirbekova, T. Chinibayeva

DIVIDING DEVICES WITH PRELIMINARY PREPARATION OF MULTIPLES OF THE DIVISOR172

K.N. Uteuliyeva, B.Z. Kenzhegulov, T.A. Karazhigitova, H.İ. Bülbül, Z.Zh. Zhanuzakova

MATHEMATICAL AND ALGORITHMIC APPROACHES TO THE DEVELOPMENT OF A COLLABORATIVE FILTERING-BASED RECOMMENDER SYSTEM188

S. Sharmukhanbet, G. Turmukhanova, O. Findik, V. Makhatova, L. Kurmangazyeva

HIGH-PRECISION ROBOTIC ASSEMBLY UNDER VARIABLE ILLUMINATION: A ROBUST MECHATRONIC ARCHITECTURE FOR VISUAL SERVOING209

INFORMATION SECURITY AND COMMUNICATIONTECHNOLOGIES

A. Amirbay, Z. Amanbaikyzy, K. Maxutova, A. Mukhanova, M. Kassim

MACHINE LEARNING ALGORITHM FOR EARLY DETECTION OF AUTISM SPECTRUM DISORDERS IN CHILDREN BASED ON MULTIMODAL ANALYSIS OF EYE MOVEMENTS AND FACIAL EXPRESSIONS227

K. Baisylbayeva, Sh. Mussiraliyeva, Zh. Yeltay

DETECTION OF EXTREMIST IDEOLOGY IN THE KAZAKH LANGUAGE: ANNOTATION CHALLENGES AND DEEP LEARNING APPROACHES242

M.A. Bolatbek, A.M.Usmanova, K.B. Bagitova, G.B. Baispay

DEVELOPMENT AND RESEARCH OF A METHOD FOR ANALYZING NETWORK TRAFFIC TO IDENTIFY A CYBER THREAT	261
D.I. Prokopovych-Tkachenko, N.K. Zhumagalieva, D.N. Shchytyov, N.F. Mormul, D.A. Cherkaskyi FUZZY MODEL FOR EVALUATING INFORMATION SECURITY PARAMETERS OF INFORMATION SYSTEMS UNDER INCOMPLETE AND QUALITATIVE DATA: CONSTRUCTION METHODOLOGY, RULE BASE TUNING, AND DEMONSTRATION CASE FOR ORGANIZATIONS	279
E.A. Pustovoy, O.A. Pustovaya, A.N. Raushanova, I.S. Zaurbekov EVALUATION OF THE EFFECTIVENESS OF SYNTHESIS OF STOCHASTIC MODELS WITH CONTROLLED PROPERTIES	305
Y. Serzhan, T. Umarov, A. Abilbayeva FRAUD DETECTION IN CREDIT CARD TRANSACTIONS USING MACHINE LEARNING: A COMPARATIVE ANALYSIS	321

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

Д.Е. Абжанов, А.А. Белоощицкий ЭКОЛОГИЯЛЫҚ МОНИТОРИНГТІҢ ЗИЯТКЕРЛІК ЖҮЙЕСІНДЕГІ СТАЦИОНАРЛЫҚ ЛАСТАНУ КӨЗ-ДЕРІНІҢ ШЫҒАРЫНДЫЛАРЫ ТУРАЛЫ ДЕРЕКТЕРДІ БАСҚАРУДЫҢ МОДЕЛІ МЕН ӘДІСІ	9
А.Е. Сланбекова, М.Б. Рахимжанова, А.И. Жанибекова, А.З. Алимагамбетова, М. Худойбергенов КЕҢІСТІКТІК-УАҚЫТТЫҚ (SPATIOTEMPORAL) ТАЛДАУ НЕГІЗІНДЕ ГИДРОЛОГИЯЛЫҚ ҚАУІП-ҚАТЕРДІ ЕРТЕ АНЫҚТАУ	25

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

Ф.Н. Абдраимова, А.А. Керейбаева, Д.С. Дюсенова, Д.А. Алиева, Т.Ж. Токтарова ТІЛ БІЛІМІНДЕ ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫ: СТУДЕНТТЕР ҚОЛДАНУЫНЫҢ ПРАКТИКАЛЫҚ АСПЕКТІЛЕРІ МЕН МӘСЕЛЕЛЕРІ	36
Г.Т. Азиева, М.Б. Есенова, А.К. Абжаппарова, Г.Б. Абдикеримова, Р. Schmidt UAV ДЕРЕКТЕРІ НЕГІЗІНДЕ АУЫЛ ШАРУАШЫЛЫҒЫ DAҚЫЛДАРЫН ЖІКТЕУГЕ АРНАЛҒАН ГИБРИДТІ СТЕКИНГ МОДЕЛІ	50
Ә.Қ. Әйтiм АГГЛЮТИНАТИВТІ ТІЛДЕРГЕ АРНАЛҒАН МОРФОЛОГИЯЛЫҚ ДИЗАМБИГУАЦИЯ МЕН POS-ТАҢ-БАЛАУДЫ БІРЛЕСІП МОДЕЛЬДЕУ	62
С.А. Есниязова, С.Т. Каимов ТҮСІНДІРІЛЕТІН МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНА ОТЫРЫП АУЫР ЖҮК КӨЛІКТЕРІНЕ БОЛЖАМДЫ ТЕХНИКАЛЫҚ ҚЫЗМЕТ КӨРСЕТУ	78
Т.Д. Иманбекова, Ж.Б. Ибраева, Г.Т. Джаканова, Г.Т. Асқанбай МӨЛІМЕТТЕРДІ ВЕЙВЛЕТ-ТҮРЛЕНДІРГІШТІҢ НЕГІЗІНДЕ ҚЫСУ АЛГОРИТМІ; MATLAB ОРТАСЫНДА ТАЛДАУ ЖӘНЕ ІСКЕ АСЫРУ	92
Б.З. Кенжегулов, Ж.Т. Билялова, К.Н. Утеулиева, Л. Нурғалиева, Ш.С. Нуржанова ҮЛКЕН ТІЛДІК МОДЕЛЬДЕР НЕГІЗІНДЕ ИНТЕЛЛЕКТУАЛДЫ TEXT-TO-SQL ЖҮЙЕСІН ӨЗІРЛЕУДІҢ МАТЕМАТИКАЛЫҚ-АЛГОРИТМДІК ТӘСІЛІ	110
Н.Ш. Максұтова, Ж.А. Тусупов, А.Ә. Шекербек, Ж.Е. Кенжебаева, К.О. Рахимов ЖҮРЕК-ҚАН ТАМЫРЛАРЫ АУРУЛАРЫНЫҢ ҚАУІП-ҚАТЕРІН ЖӘНЕ БИОХИМИЯЛЫҚ ӨЗГЕРІСТЕРДІ КЕШЕНДІ БАҒАЛАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ: АСПАРТАМИНОТРАНСФЕРАЗАҒА ЕРЕКШЕ НАЗАР	131
О.С. Салықова, В.А. Мадин, Б.Р. Салықов, Д.Н. Комаров, Н.В. Мануилов ӨНЕРКӘСІПТІК МОНИТОРИНГ ЖҮЙЕЛЕРІНДЕГІ MEMS-АКСЕЛЕРОМЕТРЛЕРДІҢ СЕНСОРЛЫҚ МОДУЛЬДЕРІН ИНТЕГРАЦИЯЛАУ	146
Р. Таберхан, М.А. Самбетбаева, Г. Қалман KAZCAUSAL: ҚАЗАҚ ТІЛІНДЕГІ СЕБЕП-САЛДАРЛЫҚ ҚАТЫНАСТАРДЫҢ АЛҒАШҚЫ КОРПУСТЫҚ АННОТАЦИЯСЫ	160
С. Тынымбаев, С.Е. Маманова, Р. Бердібаев, Ж.Е. Темірбекова, Т. Чинибаева БӨЛГІШТІҢ ЕСЕЛІ МӘНДЕРІН АЛДЫН АЛА ДАЙЫНДАУМЕН ЖҮЗЕГЕ АСЫРЫЛАТЫН БӨЛУ ҚҰРЫЛҒЫЛАРЫ	172



К.Н. Утеулиева, Б.З. Кенжегулов, Т.А. Каражигитова, Х. Булбул, З.Ж. Жанузакова КОЛЛАБОРАТИВТІК СҮЗГІЛЕУ НЕГІЗІНДЕГІ ҰСЫНЫМДЫҚ ЖҮЙЕНІ ӨЗІРЛЕУДІҢ МАТЕМАТИКАЛЫҚ-АЛГОРИТМДІК ТӘСІЛДЕРІ	188
С. Шармуханбет, Г. Тұрмуханова, О. Финдик, В. Махатова, Л. Курмангазиева АЙНЫМАЛЫ ЖАРЫҚ ЖАҒДАЙЫНДАҒЫ ЖОҒАРЫ ДӘЛДІКТІ РОБОТТЫҚ ҚҰРАСТЫРУ: ВИЗУАЛДЫ СЕРВОТЕЖЕУДІҢ ТӨЗІМДІ МЕХАТРОНИКАЛЫҚ АРХИТЕКТУРАСЫ	209

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

А. Амирбай, З. Аманбайқызы, К. Максүтова, А. Муханова, М. Kassim КӨЗ ҚОЗҒАЛЫСТАРЫ МЕН БЕТ МИМИКА БЕЛГІЛЕРІН МУЛЬТИМОДАЛЬДЫ ТАЛДАУҒА НЕГІЗ- ДЕЛГЕН БАЛАЛАРДАҒЫ АУТИЗМ СПЕКТРІНІҢ БҰЗЫЛЫСТАРЫН ЕРТЕ АНЫҚТАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМІ	227
К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж. Елтай ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРЕМИСТІК ИДЕОЛОГИЯНЫ АНЫҚТАУ: АННОТАЦИЯЛАУ МӘСЕЛЕЛЕРІ ЖӘНЕ ТЕРЕҢ ОҚЫТУ ТӘСІЛДЕРІ	242
М.А. Болатбек, А.М. Усманова, Қ.Б. Багитова, Г.Б. Байспай КИБЕР ҚАУІПТІ АНЫҚТАУ ҮШІН ЖЕЛІЛІК ТРАФИКТІ ТАЛДАУ ӘДІСІН ӨЗІРЛЕУ ЖӘНЕ ЗЕРТТЕУ	261
Д.И. Прокопович-Ткаченко, Н.К. Жумагалиева, Д.Н. Щитов, Н.Ф. Мормуль, Д.А. Черкасский ТОЛЫҚ ЕМЕС ЖӘНЕ САПАЛЫҚ ДЕРЕКТЕР ЖАҒДАЙЫНДА АҚПАРАТТЫҚ ЖҮЙЕЛЕРДІҢ АҚПА- РАТТЫҚ ҚАУІПСІЗДІК ПАРАМЕТРЛЕРІН БАҒАЛАУДЫҢ БҰЛЫҢҒЫР МОДЕЛІ: ҚҰРУ ӘДІСТЕМЕСІ, ЕРЕЖЕЛЕР БАЗАСЫН БАПТАУ ЖӘНЕ ҰЙЫМДАРҒА АРНАЛҒАН ДЕМОНСТРАЦИЯЛЫҚ КЕЙС	279
Е.А. Пустовой, О.А. Пустовая, А.Н. Раушанова, И.С. Заурбеков БАСҚАРЫЛАТЫН ҚАСИЕТТЕРІ БАР СТОХАСТИКАЛЫҚ МОДЕЛЬДЕРДІ СИНТЕЗДЕУДІҢ ТИМДІЛІГІН БАҒАЛАУ	305
Е. Сержан, Т. Умаров, А. Әбілбаева МАШИНАЛЫҚ ОҚУ ӘДІСІ АРҚЫЛЫ КРЕДИТ КАРТА ОПЕРАЦИЯЛАРЫНДАҒЫ АЛАЯҚТЫҚТЫ АНЫҚТАУ: САЛЫСТЫРМАЛЫ ТАЛДАУ	321

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

Д.Е. Абжанова, А.А. Белошицкий МОДЕЛЬ И МЕТОД УПРАВЛЕНИЯ ДАННЫМИ О ВЫБРОСАХ СТАЦИОНАРНЫХ ИСТОЧНИКОВ ЗАГРЯЗНЕНИЯ В ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЕ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА	9
А.Е. Сланбекова, М.Б. Рахимжанова, А.И. Жанибекова, А.З. Алимагамбетова, М. Худойбергенов РАННЕЕ ВЫЯВЛЕНИЕ ГИДРОЛОГИЧЕСКИХ ОПАСНОСТЕЙ НА ОСНОВЕ ПРОСТРАНСТВЕННО- ВРЕМЕННОГО (SPATIOTEMPORAL) АНАЛИЗА	25

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Ф.Н. Абдраимова, А.А. Керейбаева, Д.С. Дюсенова, Д.А. Алиева, Т.Ж. Токтарова ТЕХНОЛОГИИ ИИ В ЯЗЫКОВОМ ОБРАЗОВАНИИ: ПРАКТИЧЕСКИЕ АСПЕКТЫ И ПРОБЛЕМЫ ПРИМЕНЕНИЯ СТУДЕНТАМИ	36
Г.Т. Азиева, М.Б. Есенова, А.К. Абжаппарова, Г.Б. Абдикеримова, Р. Schmidt ГИБРИДНАЯ МОДЕЛЬ СТЕКИНГА ДЛЯ КЛАССИФИКАЦИИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО ДАННЫМ UAV	50
Ә.Қ. Әйтiм СОВМЕСТНАЯ МОРФОЛОГИЧЕСКАЯ ДИЗАМБИГУАЦИЯ И POS-РАЗМЕТКА ДЛЯ АГГЛЮТИНАТИВНЫХ ЯЗЫКОВ	62
С.А. Есниязова, С.Т. Каимов ПРЕДИКТИВНОЕ ТЕХНИЧЕСКОЕ ОБСЛУЖИВАНИЕ ТЯЖЁЛЫХ ГРУЗОВИКОВ С ИСПОЛЬЗОВАНИ- ЕМ ОБЪЯСНИМОГО МАШИННОГО ОБУЧЕНИЯ	78
Т.Д. Иманбекова, Ж.Б. Ибраева, Г.Т. Джаканова, Г.Т. Асқанбай	

АЛГОРИТМ СЖАТИЯ ДАННЫХ НА ОСНОВЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАТЕЛЯ: АНАЛИЗ И РЕАЛИЗАЦИЯ В МАТЛАВ	92
Б.З. Кенжегулов, Ж.Т. Билялова, К.Н. Утеулиева, Л. Нургалиева, Ш.С. Нуржанова	
МАТЕМАТИКО-АЛГОРИТМИЧЕСКИЙ ПОДХОД К РАЗРАБОТКЕ ИНТЕЛЛЕКТУАЛЬНОЙ ТЕХТ-TO-SQL СИСТЕМЫ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ	110
Н.Ш. МаксUTOва, Д.А. Тусупов, А.А. Шекербек, Ж.Е. Кенжебаева, К.О. Рахмтов	
МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КОМПЛЕКСНОЙ ОЦЕНКИ РИСКА СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ И БИОХИМИЧЕСКИХ ИЗМЕНЕНИЙ: АКЦЕНТ НА АСПАРТАМИНОТРАНСФЕРАЗЕ ...	131
О.С. Салыкова, В.А. Мадин, Б.Р. Салыков, Д.Н. Комаров, Н.В. Мануйлов	
ИНТЕГРАЦИЯ СЕНСОРНЫХ МОДУЛЕЙ MEMS-АКСЕЛЕРОМЕТРОВ В СИСТЕМАХ ПРОМЫШЛЕННОГО МОНИТОРИНГА	146
Р. Таберхан, М.А. Самбетбаева, Г. Калман	
КАЗСАUSAL: ПЕРВАЯ КОРПУСНАЯ АННОТАЦИЯ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ НА КАЗАХСКОМ ЯЗЫКЕ	160
С. Тынымбаев, С.Е. Маманова, Р. Бердибаев, Ж.Е. Темирбекова, Т. Чинибаева	
УСТРОЙСТВА ДЕЛЕНИЯ ЧИСЕЛ С ПРЕДВАРИТЕЛЬНОЙ ПОДГОТОВКОЙ КРАТНЫХ ДЕЛИТЕЛЮ	172
К.Н. Утеулиева, Б.З. Кенжегулов, Т.А. Каражигитова, Х.Бюльбюль, З.Ж. Жанузакова	
МАТЕМАТИКО-АЛГОРИТМИЧЕСКИЕ ПОДХОДЫ К РАЗРАБОТКЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ	188
С. Шармуханбет, Г. Турмуханова, О.Финдик, В.Махатова, Л. Курмангазиева	
ВЫСОКОТОЧНАЯ РОБОТИЗИРОВАННАЯ СБОРКА ПРИ ПЕРЕМЕННОЙ ОСВЕЩЁННОСТИ: РОБАСТНАЯ МЕХАТРОННАЯ АРХИТЕКТУРА ВИЗУАЛЬНОГО СЕРВОУПРАВЛЕНИЯ	209

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

А. Амирбай, З. Аманбайкызы, К. МаксUTOва, А. Муханова, М. Kassim	
АЛГОРИТМ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАННЕГО ВЫЯВЛЕНИЯ РАССТРОЙСТВ АУТИСТИЧЕСКОГО СПЕКТРА У ДЕТЕЙ НА ОСНОВЕ МУЛЬТМОДАЛЬНОГО АНАЛИЗА ДАННЫХ ДВИЖЕНИЯ ГЛАЗ И МИМИЧЕСКИХ СИГНАЛОВ	227
К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж.Елтай	
ОБНАРУЖЕНИЕ ЭКСТРЕМИСТСКОЙ ИДЕОЛОГИИ НА КАЗАХСКОМ ЯЗЫКЕ: ПРОБЛЕМЫ АННОТИРОВАНИЯ И МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ	242
М.А. Болатбек, А.М. Усманова, К.Б. Багитова, Г.Б. Байспай	
РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА АНАЛИЗА СЕТЕВОГО ТРАФИКА ДЛЯ ВЫЯВЛЕНИЯ КИБЕРУГРОЗЫ	261
Д.И. Прокопович-Ткаченко, Н.К. Жумагалиева, Д.Н. Щитов, Н.Ф. Мормуль, Д.А. Черкасский	
НЕЧЕТКАЯ МОДЕЛЬ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ В УСЛОВИЯХ НЕПОЛНЫХ И КАЧЕСТВЕННЫХ ДАННЫХ: МЕТОДИКА ПОСТРОЕНИЯ, НАСТРОЙКА БАЗЫ ПРАВИЛ И ДЕМОСТРАЦИОННЫЙ КЕЙС ДЛЯ ОРГАНИЗАЦИЙ	279
Е.А. Пустовой, О.А. Пустовая, А.Н. Раушанова, И.С. Заурбеков	
ОЦЕНКА ЭФФЕКТИВНОСТИ СИНТЕЗА СТОХАСТИЧЕСКИХ МОДЕЛЕЙ С УПРАВЛЯЕМЫМИ СВОЙСТВАМИ	305
Е. Сержан, Т. Умаров, А. Абильбаева	
ВЫЯВЛЕНИЕ МОШЕННИЧЕСТВА С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ ПРИ ОПЕРАЦИЯХ С КРЕДИТНЫМИ КАРТАМИ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ	321



INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.2. Number 26 (2026). Pp. 242–260

Journal homepage: <https://journal.iitu.edu.kz>

<https://doi.org/10.54309/IJICT.2026.26.2.016>

УДК 81.93.29

DETECTION OF EXTREMIST IDEOLOGY IN THE KAZAKH LANGUAGE: ANNOTATION CHALLENGES AND DEEP LEARNING APPROACHES

K. Baisylbayeva^{1}, Sh. Mussiraliyeva¹, Zh. Yeltay²*

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

²University of Bonn, Germany.

E-mail: baisylbaeva.k@gmail.com

Baisylbayeva Kymbat — Senior lecturer of the Department of Cybersecurity and Cryptology, Faculty of information technologies and Artificial Intelligence, Al-Farabi Kazakh National University, Almaty, Kazakhstan

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

Mussiraliyeva Shynar — candidate of physico-mathematical Sciences, Faculty of information technologies and Artificial Intelligence, Research Professor of the department of Cybersecurity and Cryptology, Professor, Almaty, Kazakhstan

<https://orcid.org/0000-0001-5794-3649>;

Zhastay Yeltay — Researcher, University of Bonn, Germany

<https://orcid.org/0000-0002-9275-7582>.

© K. Baisylbayeva, Sh. Mussiraliyeva, Zh. Yeltay

Abstract. Destructive content currently distributed on social media platforms and instant messengers poses a serious threat to information security. The increase in the volume of content and the complexity of the methods of its hidden transmission reinforce the need for automatic detection of such materials. The lack of funds and methods, especially in the field of analyzing texts in the Kazakh language, makes this problem urgent. The main objective of the study is to establish a methodological framework for the automated analysis of extremist content in the Kazakh language and to systematize the capabilities of existing approaches. The paper analyzes the main ideological categories of extremist content-propaganda, recruitment and radicalization -and sets the task of classifying the text in these areas. The study provides for the compilation of a balanced data collection based on materials in the Kazakh language collected from the VK, YouTube and Telegram platforms. When annotating, it is noted that differences in expert opinions, the possibility of using soft labels and disparagement-based learning approaches are taken into account. Annotator disagreements are treated as informative signals, and three disagreement-based models-DART-4xBERT, DART-4xLLM, and DART-Classic-Ensemble-are compared. The



best result was achieved by DART-Classic-Ensemble (F1=0.9664), demonstrating the advantage of heterogeneous ensemble methods. Classical models and deep learning models are used to automatically identify extremist content. The effect of text preprocessing on the accuracy and accuracy of the model is also shown. The article offers a comprehensive methodology for analyzing the content of social networks in the Kazakh language and identifying extremist ideology using artificial intelligence and natural language processing technologies. The proposed approach is aimed at strengthening information security, as well as ensuring ideological stability in society.

Keywords: extremism, social networks, ideology, propaganda, recruitment, radicalization, artificial intelligence, machine learning, natural language processing

For citation: K. Baisylbayeva, Sh. Mussiraliyeva, Zh. Yeltay (2026). Detection of extremist ideology in the kazakh language: annotation challenges and deep learning approaches // International journal of information and communication technologies. Vol. 7. No. 26. Pp. 242–260. <https://doi.org/10.54309/IJICT.2026.26.2.016>. (In Kaz.).

Conflict of interest: The authors declare that there is no conflict of interest.

ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРЕМИСТІК ИДЕОЛОГИЯНЫ АНЫҚТАУ: АННОТАЦИЯЛАУ МӘСЕЛЕЛЕРІ ЖӘНЕ ТЕРЕҢ ОҚЫТУ ТӘСІЛДЕРІ

К.Д. Байсылбаева^{1}, Ш.Ж. Мусиралиева¹, Ж.Елтай²*

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

²Бонн Университеті, Германия.

E-mail: baisylbaeva.k@gmail.com

Байсылбаева Кымбат Данияровна — Ақпараттық технологиялар және жасанды интеллект факультеті, «Киберқауіпсіздік және криптология» кафедрасының аға оқытушысы, Алматы, Қазақстан

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

Мусиралиева Шынар Женисбековна — ф.-м.ғ.к., Ақпараттық технологиялар және жасанды интеллект факультеті, «Киберқауіпсіздік және криптология» кафедрасының профессор-зерттеушісі, профессор, Алматы, Қазақстан

<https://orcid.org/0000-0001-5794-3649>;

Жастай Елтай — Бонн Университетінің зерттеушісі, Германия

<https://orcid.org/0000-0002-9275-7582>.

© К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж. Елтай

Аннотация. Қазіргі уақытта әлеуметтік желі платформаларында және мессенджерлерде таралатын деструктивті мазмұн ақпараттық қауіпсіздікке елеулі қауіп төндіреді. Контент көлемінің артуы мен оның жасырын берілу әдістерінің күрделенуі мұндай материалдарды автоматты түрде анықтау қажеттілігін күшейте түседі. Әсіресе қазақ тіліндегі мәтіндерді талдау бағытында құралдар мен әдістердің

жетіспеушілігі бұл мәселені өзекті етеді. Жұмыстың негізгі мақсаты-қазақ тіліндегі экстремистік контентті автоматты түрде талдау үшін әдіснамалық негіз қалыптастыру және қолданыстағы тәсілдердің мүмкіндіктерін жүйелеу. Жұмыста экстремистік контенттің негізгі идеологиялық санаттары – пропаганда, рекрутмент және радикалдану-талданып, мәтінді осы бағыттар бойынша жіктеу міндеті қойылған. Зерттеу барысында VK, YouTube және Telegram платформаларынан жиналған қазақ тіліндегі материалдар негізінде баланс сақталған деректер жинағын құрастыру қарастырылған. Аннотациялау кезінде сарапшылардың пікір айырмашылықтары ескерілетіндігі, soft labels және disagreement-based learning тәсілдерін қолдану мүмкіндігі атап өтіледі. Аннотаторлар арасындағы пікір алшақтығы маңызды ақпараттық сигнал ретінде қарастырылып, disagreement-based learning тәсілі негізінде DART-4xBERT, DART-4xLLM және DART-Classic-Ensemble модельдері салыстырылады. Ең жоғары нәтиже DART-Classic-Ensemble жүйесінде тіркелді: F1=0.9664, бұл heterogeneous ансамбльдердің тиімділігін дәлелдейді. Экстремистік мазмұнды автоматты анықтау үшін классикалық және терең оқыту модельдері жайлы айтылады. Сондай-ақ мәтінді алдын ала өңдеудің модель дәлдігіне, нақтылығына әсері көрсетілген. Мақалада жасанды интеллект пен табиғи тілді өңдеу технологияларын қолдана отырып, қазақ тіліндегі әлеуметтік медиа контентін талдауға және экстремистік идеологияны анықтауға арналған кешенді әдістеме ұсынылады. Ұсынылған тәсіл ақпараттық қауіпсіздікті нығайтуға, сондай-ақ қоғамдағы идеологиялық тұрақтылықты қамтамасыз етуге бағытталған

Түйінді сөздер: экстремизм, әлеуметтік желілер, идеология, пропаганда, рекрутмент, радикалдану, жасанды интеллект, машиналық оқыту, табиғи тілді өңдеу

Дәйексөздер үшін: К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж. Елтай (2026). Қазақ тіліндегі экстремистік идеологияны анықтау: аннотациялау мәселелері және терең оқыту тәсілдері // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No. 26. Б. 242–260. <https://doi.org/10.54309/IJICT.2026.26.2.016>. (Қаз. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ОБНАРУЖЕНИЕ ЭКСТРЕМИСТСКОЙ ИДЕОЛОГИИ НА КАЗАХСКОМ ЯЗЫКЕ: ПРОБЛЕМЫ АННОТИРОВАНИЯ И МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ

К.Д. Байсылбаева^{1}, Ш.Ж. Мусиралиева¹, Ж.Елтай²*

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

²Боннский Университет, Германия.

E-mail: baysylbaeva.k@gmail.com

Байсылбаева Кымбат Данияровна — старший преподаватель кафедры «Кибербезопасность и криптология», факультет информационных технологий и искусственно-



го интеллекта, Алматы, Казахстан

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

Мусиралиева Шынар Женисбековна — к.ф.-м.н., факультет информационных технологий и искусственного интеллекта, профессор-исследователь кафедры «Кибербезопасность и криптология», профессор, Алматы, Казахстан

<https://orcid.org/0000-0001-5794-3649>;

Жастай Елтай — исследователь Боннского Университета, Германия

<https://orcid.org/0000-0002-9275-7582>.

© К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж. Елтай

Аннотация. Деструктивный контент, распространяемый в настоящее время на платформах социальных сетей и мессенджерах, представляет серьезную угрозу информационной безопасности. Увеличение объема контента и усложнение методов его скрытой передачи усиливают необходимость автоматического обнаружения таких материалов. Нехватка средств и методов, особенно в направлении анализа текстов на казахском языке, делает эту проблему актуальной. Цель настоящего исследования-разработка автоматизированной методики, позволяющей эффективно выявлять признаки экстремистской идеологии в текстах социальных сетей на казахском языке. В работе проанализированы основные идеологические категории экстремистского контента- пропаганда, рекрутмент и радикализация-и поставлена задача классификации текста по этим направлениям. В ходе исследования предусмотрено составление сборника данных с сохранением баланса на основе материалов на казахском языке, собранных с платформ VK, YouTube и Telegram. При аннотировании отмечается, что учитываются различия во мнениях экспертов, возможность применения подходов soft labels и disagreement-based learning. Расхождения в оценках аннотаторов рассматриваются как информативный сигнал, на основе которого сравниваются модели DART-4xBERT, DART-4xLLM и DART-Classic-Ensemble. Наилучший результат достигнут системой DART-Classic-Ensemble ($F1=0.9664$), что подтверждает эффективность гетерогенных ансамблей. Говорится о классических моделях и моделях глубокого обучения для автоматического определения экстремистского содержания. Также показано влияние предварительной обработки текста на точность, эффективность модели. В статье предлагается комплексная методика анализа контента социальных сетей на казахском языке и выявления экстремистской идеологии с использованием искусственного интеллекта и технологий обработки естественного языка. Предлагаемый подход направлен на укрепление информационной безопасности, а также обеспечение идеологической стабильности в обществе.

Ключевые слова: экстремизм, социальные сети, идеология, пропаганда, рекрутмент, радикализация, искусственный интеллект, машинное обучение, обработка естественного языка

Для цитирования: К.Д. Байсылбаева, Ш.Ж. Мусиралиева, Ж.Елтай (2026). Обнаружение экстремистской идеологии на казахском языке: проблемы аннотиро-

вания и методы глубокого обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 26. Стр. 242–260. <https://doi.org/10.54309/IJCT.2026.26.2.016>. (На каз.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе.

Қазіргі цифрлық дәуірде интернет кеңістігінің қарқынды дамуы мен әлеуметтік желілердің таралуы ақпарат алмасу процесін жылдамдатуда. Бірақ бұл айтып отырған жағдай қоғам үшін жаңа қауіптер туғызуда-әсіресе экстремистік және деструктивті идеологиялардың таралуы күшейе түсті. Әлеуметтік желілер мен мессенджерлер радикал идеялар мен зорлық-зомбылықты насихаттаудың, жаңа мүшелер тартудың және идеологиялық ықпал етудің негізгі алаңына айналды. Мұндай контент қоғамдағы тұрақтылыққа, ұлттық қауіпсіздікке және ақпараттық кеңістіктің тұтастығына тікелей қауіп төндіреді.

Субъективті NLP тапсырмаларында аннотаторлар арасындағы келіспеушіліктерді шешу және бірыңғай шынайы белгіні қалыптастыру үшін көпшілік дауысы мен орташа мәнге негізделген тәсілдер кеңінен қолданылады. Алайда аннотаторлар бір-бірімен жүйелі түрде келіспеуі мүмкін, бұл әсіресе аффектіні, агрессияны және жеккөрушілік тілін анықтау сияқты субъективті тапсырмаларда олардың жеке көзқарастары мен құндылықтарын көрсетеді. Мұндай келіспеушіліктер маңызды ерекшеліктерді қамтуы мүмкін, бірақ аннотацияларды бір ғана шынайы белгіге біріктіру кезінде бұл ақпарат жиі ескерілмей қалады. (Aida Mostafazadeh Davani, Mark Díaz, Vinodkumar Prabhakaran, 2022) Осы мәселені шешу мақсатында бірнеше аннотациямен жұмыс істейтін модельдердің тиімділігі қарастырылады.

Ғылыми зерттеулерге сәйкес, соңғы жылдары интернет кеңістігінде зиянды және радикалды контенттің көлемі айтарлықтай артқан, ал оны қолмен бақылау тиімсіз болып табылады. Осыған байланысты табиғи тілдерді өңдеу және машиналық оқыту әдістері экстремистік мәтіндерді автоматты түрде анықтаудың негізгі құралдары ретінде қарастырылуда. Сондықтан әлеуметтік желілердегі экстремистік мазмұнды талдау және автоматты түрде анықтау-қазіргі ғылым мен киберқауіпсіздік саласындағы өзекті бағыттардың бірі (Yelmen et al., 2023; Kibriya et al., 2024). Атап айтқанда, терең оқыту модельдері, соның ішінде трансформер архитектурасына негізделген BERT сияқты модельдер, сондай-ақ дәстүрлі машиналық оқыту әдістерімен бірге қолданылып, мәтіндік деректерден семантикалық белгілерді тиімді анықтауға мүмкіндік береді.

Экстремистік идеология үш негізгі формада көрініс табады: пропаганда, рекрутмент және радикалдану. Пропаганда-белгілі бір діни немесе этникалық топтарға қарсы өшпенділік пен төзімсіздікті тудыру; рекрутмент-адамдарды экстремистік ұйымдарға тарту мен қосылуға итермелеу; радикалдану-тұлғаның біртіндеп радикалды көзқарастарды қабылдап, әрекетке бейімдеу үдерісі. Әлеуметтік медиа осы үшеуін де жылдам тарату құралына айналып отыр.

Кесте 1 – Қазақ тіліндегі әлеуметтік желілердегі экстремистік идеология түрлері және олардың сипаттамалары

Идеология түрі	Анықтамасы	Мысал
Рекрутмент	Экстремистік ұйымға ашық не жасырын түрде қосылуға үндейтін, қатысуға немесе қолдауға шақыратын, топты тартымды етіп көрсететін үндеу.	«Біздің қозғалысқа қосылып, әділ қоғам орнатуға атсалысыңыз!», «Бізбен бірге болыңыз – сізді жаңа өмірмен шынайы еркіндік күтеді!», «Біздің ұйым – нағыз ерлер мен еріктілердің жолы!»
Пропаганда	Белгілі бір этникалық, діни немесе басқа да топтарға қарсы өшпенділік пен кемсітуге итермелеу, бірақ әрекетке тікелей шақырусыз. Басқа топтардың зорлық-зомбылық әрекеттерін сипаттау, дезинформация тарату.	«Олар біздің мәдениетімізге қауіп төндіріп отыр», «Бұл топ шындықты жасырып келеді», «Билік бәрін өз пайдасына бұрып отыр»
Радикалдану	Әлеуметтік мәселелерге қатысты радикалды көзқарастарды насихаттап, адамдарды әрекетке және қарсы топтарға өшпенділікке итермелейтін мазмұн.	«Әділет үшін күресетін уақыт келді!», «Халық үнсіз отырмауы керек, біз әрекет етуіміз қажет!», «Енді төзуге болмайды – шын өзгеріс қарсылықтан басталады!»

Әдебиеттерге шолу

Соңғы жылдары онлайн әлеуметтік желілердегі идеологиялық бағыттылық пен экстремизмді анықтау мәселесі ғылыми зерттеулерде ерекше назарға алынған. (Ram et al., 2025) кең ауқымды онлайн орталарда идеологияны автоматты түрде анықтау жүйелерін құруға арналған практикалық нұсқаулықтар ұсынды. Олар онлайн идеологияны анықтау зорлық-зомбылықтың алдын алу және пікірдің өзгерісін зерттеу үшін маңызды екенін атап өтті, сондай-ақ эталондық деректерді құрудың күрделілігі мен прокси-сигналдардың (мысалы, хэштегтер, медиа) контекстке тәуелділігін көрсетті. Зерттеушілер төрт компонентті құрылымды ұсынып, деректерді жинау, проксиерді таңдау, гомофильдік сипаттамаларды шығару және классификацияны жүзеге асыру арқылы пайплайнды құрудың тиімді әдістерін көрсетті.

Сонымен қатар, (Rav et al., 2024) жүйелі шолу арқылы онлайн әлеуметтік желілерде идеологиялық бағыттылық пен экстремизмді анықтау әдістерін талдады. Зерттеуде Twitter, Facebook, Reddit сияқты әлеуметтік медиа платформалардан алынған мәтіндік және желілік деректер жиі қолданылған. Қолданылған әдістер қатарында дәстүрлі классификаторлар мен қазіргі заманғы терең оқыту модельдері салыстырмалы түрде қарастырылған. Бағалау метрикалары ретінде дәлдік (accuracy), F1-score және ROC-AUC сияқты стандартты көрсеткіштер қолданылған. Авторлар зерттеулердің көпшілігі тек ағылшын тілді контентке бағытталғанын, бұл көптілді және мәдени әртүрлі контексттерде нәтижелердің тұрақсыз болуына алып келетінін атап өтеді.

(Erhard et al., 2024) PopBERT моделін пайдалана отырып, популизм мен оған байланысты идеологияларды зерттеген. Зерттеу 32 мыңнан астам депутаттық сөздер корпусына негізделген және антиэлитизм, халықшылдық бағыты бойынша аннотацияланған. Көп санатты (multi-label) классификацияға негізделген трансформер

модельдерін қолдану партиялар арасындағы популистік риториканың таралуын анықтауға және саяси дискурстың динамикасын жоғары дәлдікпен талдауға мүмкіндік беретіндігін көрсеткен.

(Ali et al., 2025) Израиль–Палестина қақтығысы контекстіндегі Reddit платформасындағы идеологиялық поляризацияны зерттеген. Пікірлер үш санатқа бөлініп аннотацияланған: Pro Israel, Pro Palestine және Neutral. Mixtral 8x7B тілдік моделі prompt engineering әдісімен бірге қолданылып, 74 % дәлдік пен шамамен 73 % макро F1 -score нәтижесін көрсетті. Бұл әдіс қазіргі ірі тілдік модельдердің қақтығысты әлеуметтік медиа кеңістіктерінде идеологиялық ұстанымдарды тиімді анықтай алатынын дәлелдеді. (Peng et al., 2024) зерттеуінде АҚШ-тағы 2020 жылғы президенттік сайлау аясында Twitter платформасындағы пайдаланушылардың саяси ұстанымдарын автоматты түрде анықтау мәселесі қарастырылған. Авторлар хэштегтерді саяси поляризация индикаторы ретінде пайдалана отырып, үлкен деректер жиынтығын (датасет) құрған. Талдау үшін пайдаланушы мен твит арасындағы өзара әрекеттестікті бейнелейтін екібөлікті граф моделі ұсынылған. Бұл тәсіл қолданушылардың мінез-құлықтық ерекшеліктерін ескеруге мүмкіндік береді. Ақпаратты тиімді пайдалану үшін Graph Convolutional Networks және skip aggregation деп аталатын жаңа механизм қолданылған. Эксперимент нәтижелері бұл модельдің дәстүрлі әдістерден жоғары өнімділігін және пайдаланушының мінез-құлқын ескерудің тиімділігін көрсетті.

(Mussiraliyeva et al., 2024) қазақ тіліндегі экстремистік мәтіндерді семантикалық талдау мәселесіне назар аударған. Зерттеу барысында діни экстремизмге қатысты мәтіндерді классификациялау үшін машиналық және терең оқытуға негізделген модельдер жасалған. Авторлар деректер теңгерімсіздігі мен қазақ тілінің ерекшеліктерін ескеріп, шешім ағашы, кездейсоқ орман (Random Forest), SVM, k-NN, логистикалық регрессия және аңғал Байес сияқты алгоритмдердің нәтижелерін салыстырған. Бұл тәсіл экстремистік контентті анықтаудың ең тиімді әдістерін айқындауға мүмкіндік берді.

(Shahin et al., 2024) Twitter-дегі араб тіліндегі қорлайтын сөйлеуді анықтау үшін бес санатты қамтитын көпклассты модель ұсынған. Модель ArabicBERT, 1D-CNN, BiLSTM және RBF әдістерін біріктіретін каскадты архитектураға негізделген. Ол жоғары дәлдікке (98,4%) және F1-score көрсеткішіне (98,4 %) қол жеткізіп, дәстүрлі классификаторлардан әлдеқайда жоғары нәтиже берді. Теңгерім жасалған деректер жиынтығы модельдің сенімді көпклассты және бинарлық классификация жасауына мүмкіндік берді.

(Ashraf et al., 2022) YouTube платформасындағы діни өшпенділік пен экстремизмге бағытталған пікірлерді анықтаумен айналысқан. Авторлар діни бейнелер астындағы пікірлерден жаңа деректер жиынтығын қалыптастырып, хабарламаларды сүзгіден өткізу және өшпенділік тудыратын бейнелерді анықтау үшін деректерді интеллектуалды талдау әдістеріне жүтінген. Негізгі классификация үшін SVM, логистикалық регрессия және k-NN алгоритмдері пайдаланылған. Нәтижелер жүйелі пікірталдаудың діни өшпенділік пен экстремизмді анықтауда тиімді болатынын

көрсетті.

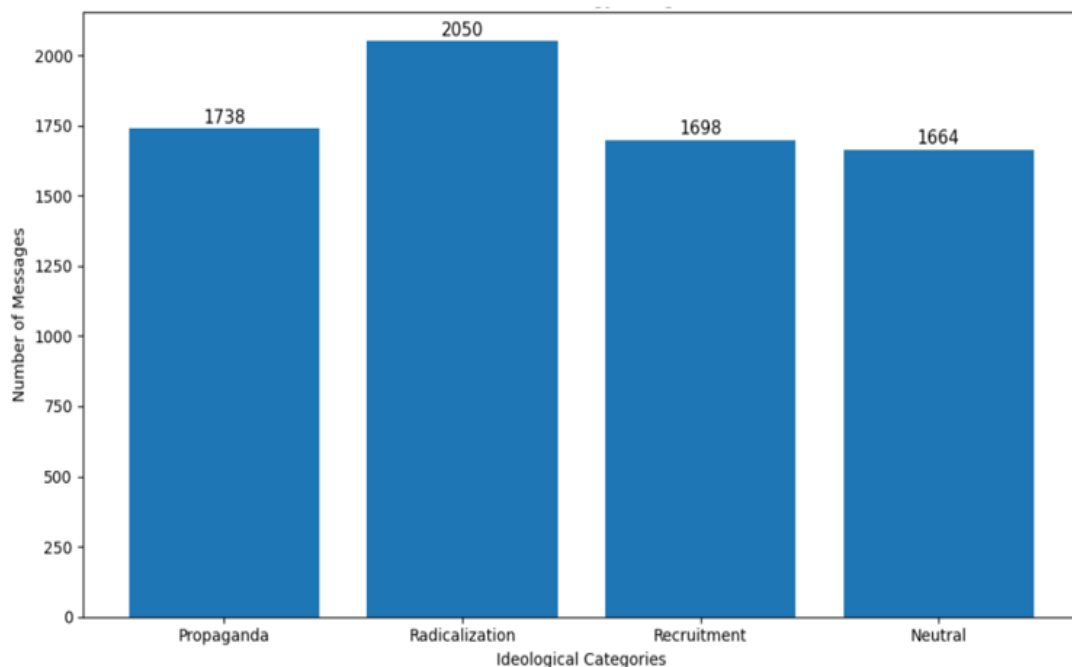
(Ramponi et al., 2022) әлеуметтік желілердегі діни өшпенділік сөзін автоматты түрде анықтау әдістерін әзірлеген. Авторлар ұсақталған аннотация схемасын ұсынып, ағылшын және итальян тілдерінде твиттер корпусын жасаған. Классификация үшін дәстүрлі машинамен оқыту және трансформер модельдері салыстырылған. Зерттеу модельдердің көптілі ортада бейімделу мүмкіндігін көрсетіп, шектеулі тілдік ресурстары бар тілдер үшін жаңа бағыттар ұсынады.

Талдаудан көрініп тұрғандай, қарастырылған зерттеулердің басым бөлігі ағылшын және басқа ресурсы мол тілдерге негізделген. Қазақ тілі сияқты агглютинативті және ресурсы шектелген тілдер үшін бұл әдістерді тікелей қолдану әркез тиімді бола бермейді. Сондықтан да қазақ тілінде экстремистік идеологияны анықтау кезінде тілдік құрылымға бейімделген және деректер тапшылығын ескеретін тәсілдер қажет.

Әдістер мен материалдар.

Деректер жинау және аннотациялау

Экстремистік контентті автоматты түрде анықтаудың негізі-сапалы және баланс сақталған датасет құру болып табылады. Ол үшін әлеуметтік желілерден мәтіндік жарияланымдар мен пікірлер парсинг арқылы жиналып, құрылымдалған форматқа (CSV, JSON) түрлендіріледі. Деректер анонимдендіріліп, идеологиялық мазмұнға қарай төрт санатқа бөлінеді: бейтарап, пропаганда, рекрутмент, радикалдану. Корпустың сапасын тексеру үшін жиі кездесетін сөздер мен тіркестер талданады, бұл модельдің тілдік үлгілерді тануына мүмкіндік береді.



Сур. 1. Идеологиялық бағыттар бойынша мәтіндер санының үлестірімі

Зерттеу барысында VK, YouTube және Telegram платформаларынан жиналған қазақ тіліндегі мәтіндерден тұратын корпус құрылды (1-сурет). Жалпы корпус көлемі 7150 мәтінді құрады. Деректер төрт санат бойынша аннотацияланды: propaganda (1738), radicalization (2050), recruitment (1698) және neutral (1664). Әр мәтін кемінде төрт аннотатор тарапынан бағаланды. Аннотаторлар арасындағы disagreement семантикалық екіұштылықтың индикаторы ретінде қарастырылды.

Кесте 2 – Қазақ тіліндегі экстремистік контент корпусындағы идеологиялық санаттар мен жиі қолданылатын сөздер

Идеология түрі	Жиі кездесетін сөздер мен тіркестер
Радикалдану (Radicalization)	«жарақат алу», «қаза табу», «шабуыл жасау», «жарылыс жасау», «атып өлтіру», «оқ жаудыру»
Рекрутмент (Recruitment)	«содырлардың қатарына қосылу», «жиһад жасауға бірігу», «қатарымызға қосылу»
Пропаганда (Propaganda)	«пропаганда», «қарсы тұру», «ДАИШ (ИШИМ)», «Ислам мемлекеті»
Бейтарап (Neutral)	«сөз», «кету», «адам», «болу», «күн»

Аннотациялау кезеңі ерекше маңызға ие болып табылады, себебі сарапшылардың пікірі субъективті болуы мүмкін. Сондықтан әр мәтін бірнеше сарапшы тарапынан бағаланып, олардың пікір айырмашылықтары жұмсақ белгі (soft label) форматында сақтап алады. Мұндай тәсіл келіспеушілікке негізделген оқыту (disagreement-based learning) әдісіне негізделген. Бұл жағдайда әрбір мәтін бір категорияға нақты бекітілмей, ықтималдық үлестірімімен сипатталады. Бұл тәсіл белгісіздік пен субъективті айырмашылықтарды ескеріп, модельге күрделі контентті дәлірек өңдеуге, анықтауға, ажыратуға мүмкіндік береді. (Uma et al., 2021) көрсеткендей, келіспеушіліктерді жою емес, олардан үйрену деректер сапасын және модельдің түсіндірілу қабілетін арттырады.

Қолданылған деректерді бағалау барысында класстар арасындағы теңсіздік, аннотация сапасы, соныменмен қатар, қазақ тіліне тән бірқатар ерекшеліктерді де ескеру қажет. Ең алдымен, субъективті немесе бағалау сипатындағы мәтіндерді талдауға арналған репрезентативті әрі ауқымды белгіленген корпусстың болмауы елеулі шектеу болып табылады. Қолда бар деректер жиынтығы, әдетте, көлемі жағынан шағын, жеке жобалар аясында жасалған және қазақтілді мәтіндердің жанрлық, өңірлік және социолінгвистикалық әртүрлілігін толық қамти бермейді, бұл өз кезегінде алынған нәтижелерді жалпылауды қиындатады.

Қазақ тілінің агглютинативті табиғаты мен мағыналардың жоғары контекстік тәуелділігі қосымша қиындықтар туындатады. Сөйлемнің семантикалық және прагматикалық салмағы көбіне жекелеген лексемалардан ғана емес, аффикстер жүйесі мен синтаксистік құрылымдар арқылы да қалыптасады. Бұл автоматтандырылған өңдеу процесін де, аннотация жүргізу ісін де күрделендіреді. Аталған ерекшеліктер ескерілмеген жағдайда класстардың бір-бірінен ажыратылуы бұрмаланып, белгілеу сапасының төмендеуіне алып келуі мүмкін.

Сонымен қатар, тілдің мәдени контексті де маңызды рөл атқарады, өйткені қазақ тіліндегі бағалау немесе әлеуетті токсикалық мазмұнды түсіндіру көп жағдай-

да қарым-қатынас нормалары мен әлеуметтік шарттылықтарға тәуелді. Әсіресе эвфемизмдер мен жасырын мағына беретін лексиканың кең қолданылуы назар аудартады, олар имплицитті негативті көзқарасты, ирония немесе астарлы бағалауды жеткізу үшін жиі пайдаланылады. Формалды түрде бейтарап көрінетін лексемалар белгілі бір контексте қорлайтын немесе кемсітушілік сипатқа ие болуы мүмкін, сондықтан оларды прагматикалық және мәдени факторларды ескермей бірмәнді түрде жіктеу қиынға соғады. Осылайша, қазақ тіліндегі деректерді дұрыс бағалау терең контекстуалдық талдау, мәтіндік корпус құрастыру мен аннотациялаудың мандандырылған тәсілдерін талап етеді.

Машиналық оқыту әдістері және модельдер

Экстремистік контентті анықтау үшін дәстүрлі және заманауи машиналық оқыту тәсілдері кеңінен қолданылады.

Классикалық әдістерге Naive Bayes, Logistic Regression, SVM және Random Forest т.с.с алгоритмдерді жатқызуға болады. Олар шағын деректерде жылдам және түсінікті нәтиже береді (Bolatbek et al., 2025).

Терең оқыту модельдері (LSTM, BERT, RoBERTa, CNN) мәтіннің семантикасы мен контекстін тереңірек түсінуге мүмкіндік береді (Hasib et al., 2023).

Ансамбльдік тәсілдер-бірнеше модельдің нәтижесін біріктіріп, болжам сапасын арттырады (мысалы, Stacking, Bagging, Boosting). Мұндай біріктіру әдістері күрделі идеологиялық мәтіндерде дәлдікті арттыруға айтарлықтай көмектеседі.

Мәтінді алдын ала өңдеу (Preprocessing)

Мәтіндік деректерді машиналық оқыту модельдеріне дайындау үшін бірнеше маңызды кезеңдер орындалады (Bolatbek et al., 2025).

1. Тазалау және қалыптандыру: артық белгілерді, сілтемелер мен эмодзилерді жою;

2. Токенизация

3. Stop-word сүзгілеу: мағынасыз сөздерді алып тастау және контексті сақтау.

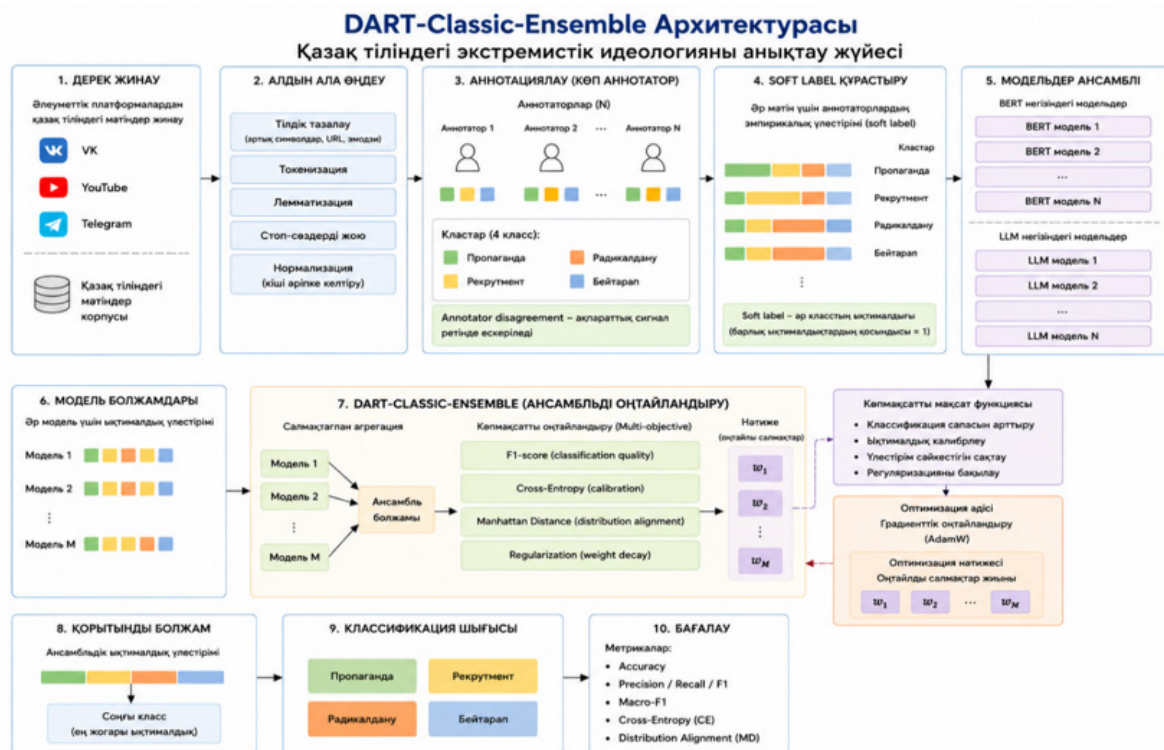
4. Стемминг немесе лемматизация: сөздердің түбірін сақтау;

Бұл қадамдар модельге мәтінді құрылымды түрде ұсынуға және семантикалық байланыстарды тиімді анықтауға мүмкіндік береді.

Қолданылған алгоритмдер мен алдын ала өңдеу тәсілдерінің тиімділігі қазақ тілінің морфологиялық ерекшеліктерімен тікелей байланысты. Классикалық машиналық оқыту модельдері векторизация сапасына тәуелді болғандықтан, мәтінді тазалау, нормализациялау және жиілікке негізделген ерекшеліктерді таңдау олардың нәтижелілігін едәуір арттырады; ал трансформер типіндегі модельдер контекстік ақпаратты жақсы меңгергенімен, дұрыс токенизация мен эмбеддинг сапасына тәуелді. Қазақ тілінің агглютинативті құрылымы стеммингті қиындатып, бір түбірден көптеген формалардың жасалуына әкеледі, бұл стеммерлердің дәлдігін төмендетуі және сөздік кеңістікті шамадан тыс ұлғайтуы мүмкін. Сондықтан алдын ала өңдеу кезеңінде агглютинацияға бейімделген тәсілдерді қолдану, ал модельдер деңгейінде контекстік эмбеддингтерді пайдалану тиімді болып табылады (Ahmed et al., 2024).

Жалпы тәсіл және архитектура

Экстремистік контентті талдау көпқадамды жүйе болып табылады. Алдымен VK, YouTube және Telegram сияқты платформалардан мәліметтер жиналып, алдын ала өңделеді. Кейін олар машиналық және терең оқыту модельдерінде сынақтан өтеді. Нәтижелер дәлдік (accuracy), F1-score, precision, recall сияқты көрсеткіштерімен бағаланып, ең тиімді модель таңдалады. Мұндай архитектура қазақ тіліндегі әлеуметтік желі мәтіндерін автоматты талдау мен идеологиялық классификациялауда әмбебап шешім бола алады.



Сур. 2. Киберэкстремизмнің идеологиялық бағыттарын анықтау жүйесінің архитектурасы

2- суретте киберэкстремизмнің идеологиялық бағыттарын анықтау жүйесінің жалпы архитектурасы көрсетілген. Онда әлеуметтік желілерден (VK, YouTube, Telegram) жиналған деректердің өңдеу және модель құру кезеңдері бейнеленген.

Нәтижелер және оларды талқылау.

Жүргізілген салыстырмалы эксперименттерде келіспеушілікке негізделген үш тікелей салыстырылатын жүйе бағаланды (DART- Disagreement-Aware Representation Transformer): DART-4xBERT, DART-4xLLM және DART-Classic-Ensemble. Бағалау F1, CE және MD метрикалары бойынша жүргізілді, мұнда F1 мәнінің жоғары болуы классификация сапасының артқанын, ал CE мен MD мәндерінің төмен болуы модель ықтималдықтарды дәлірек бағалайтынын және аннотатор үлестіріміне жақынырақ келгенін көрсетеді.

Ұсынылған DART тәсілінде әлсіз модельдердің нәтижелері салмақталған ан-

самбльдiк оқыту арқылы бiрiктiрiлдi. Ансамбльдiң соңғы болжамы әрбiр модельдiң ықтималдық нәтижелерiнiң салмақталған комбинациясы негiзiнде есептелдi. Мұндай тәсiл әртүрлi модельдердiң артықшылықтарын бiрiктiрiп, аннотаторлар арасындағы пiкiр айырмашылықтарын тиiмдi ескеруге мүмкiндiк бередi.

Модельдi оқыту барысында көпмақсатты оңтайландыру функциясы қолданылды. Негiзгi loss функциясы классификация сапасын арттыруға арналған (L_{F1}), soft label үлестiрiмiмен сәйкестiктi бағалайтын (L_{CE}), сондай-ақ аннотаторлар мен модель болжамдарының айырмашылығын өлшейтiн (L_{MD}) компоненттерiнен тұрды. Сонымен қатар тұрақтандыру мүшесi модельдiң шамадан тыс үйренуiн болдырмау үшiн енгiзiлдi.

(L_{F1}), функциясы теңгерiмсiздiк жағдайында классификация сапасын арттыруға және F1-score көрсеткiшiн оңтайландыруға бағытталды. Cross-Entropy функциясы модель болжамдарын аннотаторлардың soft label үлестiрiмiне жақындату үшiн қолданылды. Manhattan Distance функциясы модель ықтималдықтары мен аннотаторлар арасындағы disagreement деңгейiн бағалауға мүмкiндiк бередi.

Soft label үлестiрiмi әр класс бойынша аннотаторлар саны негiзiнде есептелдi. Ансамбльдiк оқыту барысында модель салмақтарына шектеулер қойылып, олардың қосындысы бiрге тең болатындай етiп қалыптандыру жүргiзiлдi.

3-кестеде келiспеушiлiкке негiзделген модельдер үшiн жүргiзiлген абляциялық зерттеу нәтижелерi келтiрiлген. Зерттеудiң мақсаты-ұсынылған көпмақсатты оңтайландыру функциясының құрамындағы әрбiр шығын компонентiнiң (L_{F1} , L_{CE} , L_{MD}) модель сапасына әсерiн бағалау.

Нәтижелер бойынша DART-4xBERT моделiнде ең жоғары нәтижелер жеке шығын функцияларын қолдану кезiнде емес, бiрнеше мақсаттық функцияларды бiрiктiру кезiнде байқалды. Атап айтқанда, L_{MD} only конфигурациясы ең жоғары F1 көрсеткiшiне (0.9450) және ең төмен MD мәнiне (0.3159) қол жеткiздi, бұл модельдiң мақсатты ықтималдық үлестiрiмiне жақсырақ сәйкестiгiн көрсетедi. Сонымен қатар, L_{CE} only нұсқасы ең төмен Cross-Entropy (0.5647) нәтижесiн көрсеттi.

DART-4xLLM моделiнде ең жоғары F1 нәтижесi бiрiктiрiлген шығын функцияларын қолдану кезiнде алынды. Атап айтқанда, $L_{F1}+L_{CE}$ және $L_{F1}+L_{CE}+L_{MD}$ конфигурациялары үшiн F1 көрсеткiшi 0.9562 деңгейiне жеттi. Бұл нәтижелер көпмақсатты оңтайландыру тәсiлiнiң үлкен тiлдiк модельдермен бiрге қолданылғанда жiктеу сапасын жақсарта алатынын көрсетедi. Сонымен қатар, тек Manhattan Distance қолданылған жағдайда MD көрсеткiшi айтарлықтай төмендегенiмен, жалпы классификация дәлдiгi әрдайым максималды мәнге жетпегенi байқалды.

DART-Classic-Ensemble моделi барлық конфигурациялар үшiн тұрақты нәтижелер көрсеттi. Барлық жағдайда F1 көрсеткiшi 0.9664 деңгейiнде сақталды, ал CE және MD көрсеткiштерi сәйкесiнше 0.5073 және 0.2960 шамасында өзгерiссiз қалды. Бұл классикалық ансамбльдiк модельдiң жеке шығын функцияларына сезiмталдығы төмен екенiн және тұрақты жұмыс iстейтiнiн көрсетедi.

Кесте 3 – Келіспеушілікке негізделген модельдер үшін абляциялық зерттеу нәтижелері

Source	Case	F1	CE	MD
DART-4xBERT	L_F1 only	0.942218	0.576611	0.321257
DART-4xBERT	L_CE only	0.944082	0.564686	0.316206
DART-4xBERT	L_MD only	0.945014	0.565900	0.315857
DART-4xBERT	L_F1+L_CE	0.944082	0.567213	0.317336
DART-4xBERT	L_F1+L_MD	0.943150	0.567156	0.317224
DART-4xBERT	L_CE+L_MD	0.944082	0.565305	0.316422
DART-4xBERT	L_F1+L_CE+L_MD	0.943150	0.566219	0.316496
DART-4xLLM	L_F1 only	0.955266	0.540463	0.366450
DART-4xLLM	L_CE only	0.951538	0.540254	0.357964
DART-4xLLM	L_MD only	0.953402	0.558867	0.317810
DART-4xLLM	L_F1+L_CE	0.956198	0.540289	0.358123
DART-4xLLM	L_F1+L_MD	0.955266	0.559567	0.318941
DART-4xLLM	L_CE+L_MD	0.951538	0.549266	0.326870
DART-4xLLM	L_F1+L_CE+L_MD	0.956198	0.549643	0.327918
DART-Classic-Ensemble	L_F1 only	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_CE only	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_MD only	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_F1+L_CE	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_F1+L_MD	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_CE+L_MD	0.966449	0.507252	0.296019
DART-Classic-Ensemble	L_F1+L_CE+L_MD	0.966449	0.507252	0.296019

Жалпы алғанда, абляциялық зерттеу нәтижелері әртүрлі шығын функцияларының модель өнімділігіне әсер ететінін көрсетті. DART-4xBERT және DART-4xLLM модельдері үшін көпмақсатты шығындарды біріктіру белгілі бір метрикалар бойынша нәтижелерді жақсартқанымен, DART-Classic-Ensemble моделі анағұрлым тұрақты және теңгерімді нәтижелер көрсетті.

4-кестеде әртүрлі модельдер үшін регуляризацияның (тұрақтандырудың) әсері көрсетілген. Бағалау көрсеткіштері ретінде F1, Cross-Entropy (CE) және Manhattan Distance (MD) метрикаларының өзгерісі (Δ) пайдаланылды. Мұнда оң мән көрсеткіштің өскенін, ал теріс мән оның төмендегенін білдіреді.

Нәтижелер бойынша DART-4xBERT моделінде регуляризация белгілі бір конфигурациялар үшін өнімділікті жақсартқанын көруге болады. Әсіресе L_MD only жағдайында F1 көрсеткіші +0.002796-ға артқан, ал CE (-0.002809) және MD (-0.001233) мәндері төмендеген. Сонымен қатар, L_F1+L_MD конфигурациясында CE көрсеткішінің ең үлкен төмендеуі (-0.009561) және MD мәнінің жақсаруы (-0.004065) байқалды. Бұл регуляризацияның модельдің тұрақтылығын арттырып, ықтималдық үлестірімдерін дәлірек сәйкестендіруге ықпал еткенін көрсетеді.

Кесте 4 – Регуляризация (тұрақтандыру) әсерінің салыстырмалы нәтижелері

Source	Case	$\Delta F1$	ΔCE	ΔMD
DART-4xBERT	L_F1 only	0.000000	-0.009044	-0.002399
DART-4xBERT	L_CE only	0.000000	-0.000612	-0.000448
DART-4xBERT	L_MD only	0.002796	-0.002809	-0.001233
DART-4xBERT	L_F1+L_CE	0.000000	-0.000089	-0.000052
DART-4xBERT	L_F1+L_MD	0.000932	-0.009561	-0.004065
DART-4xBERT	L_CE+L_MD	0.000932	-0.001044	-0.000615
DART-4xBERT	L_F1+L_CE+L_MD	0.000000	-0.001596	-0.001200
DART-4xLLM	L_F1 only	-0.000932	-0.007097	0.034436
DART-4xLLM	L_CE only	0.000000	0.000396	0.001354
DART-4xLLM	L_MD only	0.000932	0.000086	0.000217
DART-4xLLM	L_F1+L_CE	0.000000	-0.000018	0.000638
DART-4xLLM	L_F1+L_MD	0.000000	-0.000001	0.000000
DART-4xLLM	L_CE+L_MD	0.000932	0.000021	0.000429
DART-4xLLM	L_F1+L_CE+L_MD	0.000000	0.000005	0.000467
DART-Classic-Ensemble	L_F1 only	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_CE only	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_MD only	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_F1+L_CE	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_F1+L_MD	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_CE+L_MD	0.000000	-0.000768	0.000271
DART-Classic-Ensemble	L_F1+L_CE+L_MD	0.000000	-0.000768	0.000271

DART-4xLLM моделінде регуляризация әсері айтарлықтай тұрақсыз сипат көрсетті. Кейбір жағдайларда метрикалардың өзгерісі өте аз болды немесе нәтижелерге елеулі ықпал етпеді. Атап айтқанда, L_F1 only конфигурациясында CE көрсеткіші төмендегенімен (-0.007097), MD көрсеткіші керісінше нашарлап (+0.034436) өсті. Басқа конфигурациялар үшін өзгерістер нөлге жақын болды. Бұл үлкен тілдік модельдерде тұрақтандыру әсерінің шектеулі екенін көрсетеді.

DART-Classic-Ensemble моделінде барлық конфигурациялар бойынша бірдей өзгерістер байқалды: F1 өзгермеген ($\Delta F1 = 0.000000$), CE көрсеткіші аздап жақсарған ($\Delta CE = -0.000768$), ал MD шамалы артқан ($\Delta MD = 0.000271$). Бұл нәтижелер классикалық ансамбльдік тәсілдің бастапқыда тұрақты жұмыс істейтінін және регуляризация енгізілгеннен кейін көрсеткіштердің айтарлықтай өзгермейтінін көрсетеді.

Жалпы алғанда, жүргізілген талдау регуляризация әсерінің модель түріне тәуелді екенін көрсетті. DART-4xBERT моделі тұрақтандырудан көбірек пайда алған болса, DART-4xLLM моделінде әсер әлсіз байқалды, ал DART-Classic-Ensemble моделі жоғары тұрақтылық көрсетті.

5-кестеде идеологиялық бағыттарды классификациялау міндеті үшін әртүрлі агрегация стратегияларының тиімділігі салыстырылды. Зерттеуде үш тәсіл қарастырылды: көпшілік дауыс беру (Voting), қарапайым орташа мән арқылы біріктіру

(Averaging) және оңтайландырылған салмақталған агрегация (Optimised). Бағалау үшін F1, Cross-Entropy (CE) және Manhattan Distance (MD) метрикалары пайдаланылды.

Кесте 5 – Агрегация стратегиялары бойынша салыстырмалы нәтижелер

Source	Dataset	Method	F1	CE	MD
DART-4xBERT	Идеологиялық бағыттар	Voting	0.943150	0.565491	0.316266
DART-4xBERT		Averaging	0.943150	0.565491	0.316266
DART-4xBERT		Optimised	0.943150	0.566219	0.316496
DART-4xLLM		Voting	0.943150	0.613275	0.525520
DART-4xLLM		Averaging	0.945946	0.613275	0.525520
DART-4xLLM		Optimised	0.956198	0.549643	0.327918
DART-Classic-Ensemble		Voting	0.961789	0.492991	0.328111
DART-Classic-Ensemble		Averaging	0.959925	0.492991	0.328111
DART-Classic-Ensemble		Optimised	0.966449	0.507252	0.296019

DART-4xBERT моделі үшін барлық агрегация тәсілдері ұқсас нәтижелер көрсетті. Voting және Averaging әдістерінде көрсеткіштер толық сәйкес келіп, F1 = 0.943150, CE = 0.565491 және MD = 0.316266 мәндеріне жетті. Ал Optimised стратегиясы кезінде F1 көрсеткіші өзгермегенімен, CE (0.566219) және MD (0.316496) мәндері аздап өзгерді. Бұл жағдайда күрделі салмақталған агрегацияның айқын артықшылық бермегені байқалды.

DART-4xLLM моделінде агрегация тәсілдерінің әсері айқынырақ байқалды. Voting әдісі кезінде F1 көрсеткіші 0.943150 деңгейінде қалса, Averaging әдісі нәтижені жақсартып, F1 = 0.945946 мәніне жеткізді. Ең жоғары нәтиже Optimised тәсілінде алынды: F1 көрсеткіші 0.956198-ге дейін өсіп, CE мәні 0.549643-ке, ал MD көрсеткіші 0.327918-ге дейін төмендеді. Бұл нәтиже салмақтарды автоматты түрде оңтайландыру әрбір модельдің үлесін тиімді қайта бөлгенін көрсетеді. Оптимизация нәтижесінде алынған салмақтар тең емес болып қалыптасты, бұл кейбір модельдердің шешім қабылдауға көбірек үлес қосқанын білдіреді.

DART-Classic-Ensemble моделі үшін де агрегация стратегиялары арасындағы айырмашылық байқалды. Voting әдісі F1 = 0.961789 нәтижесін көрсетсе, Averaging тәсілінде көрсеткіш аздап төмендеді (F1 = 0.959925). Ал ең жоғары нәтиже Optimised стратегиясында алынды: F1 = 0.966449 және MD = 0.296019. Бұл ансамбль ішінде модельдердің әртүрлі салмаққа ие болуы жалпы сапаны жақсартатынын көрсетеді.

Жалпы алғанда, алынған нәтижелер қарапайым дауыс беру мен орташа мән арқылы біріктіру әдістеріне қарағанда оңтайландырылған агрегация стратегиясының тиімдірек екенін көрсетті. Әсіресе DART-4xLLM және DART-Classic-Ensemble модельдері үшін салмақтарды автоматты түрде бейімдеу классификация сапасын арттыруға және модельдер арасындағы келіспеушілікті тиімді пайдалануға мүмкіндік берді.

DART-4xBERT құрылымын кеңейтуге арналған қосымша эксперименттер

Негізгі DART нәтижелерін алғаннан кейін ұсынылған құрылымның тұрақтылығы мен кеңейтілу мүмкіндігін бағалау үшін үш қосымша эксперимент жүргізілді: optimization algorithm ауыстыру, temporal disagreement моделін енгізу және ensemble diversity metric негізіндегі қайта салмақтау. Бұл тәжірибелер негізгі DART 4xBERT конфигурациясын толық алмастыру үшін емес, оның сезімталдығын, орнықтылығын және қосымша жақсарту мүмкіндіктерін тексеру үшін орындалды.

Бірінші қосымша экспериментте ансамбль салмақтарын іздеудің балама тәсілі ретінде Particle Swarm Optimization (PSO) қолданылды. Бұл жерде weak predictor-лер өзгертілмей, тек бастапқы Differential Evolution алгоритмінің орнына жаңа оңтайландыру әдісі пайдаланылды. Нәтижелер бойынша PSO әдісі негізгі DART нәтижесін іс жүзінде қайталады. Сонымен қатар cross-entropy кішкене өсіп, ал manhattan distance аздап төмендеді. Бұл бастапқы Differential Evolution шешімінің жеткілікті тиімді болғанын және PSO әдісінің осы нәтижені тұрақты түрде қайталай алатынын көрсетті.

Екінші экспериментте уақыттық келіспеушілік (temporal disagreement) тәсілі зерттелді. Датасетте нақты уақыт белгілері болмағандықтан, уақыттық контекст мәтіннен алынған уақыттық-оқиғалық жанама топтар арқылы модельденді. Хабарламалар future_mobilization, ongoing_event, post_event, political_theme және atemporal_other санаттарына бөлініп, әр санат үшін жеке ансамбль салмақтары үйретілді. Нәтижесінде temporal-aware модель baseline нұсқасымен салыстырғанда жоғары нәтиже көрсетті. Ең жоғары келіспеушілік деңгейі political_theme санатында байқалса, ең айқын өсім future_mobilization тобында тіркелді. Бұл annotator disagreement көрсеткішінің мазмұнның уақыттық-оқиғалық контекстіне тәуелді өзгеруі мүмкін екенін көрсетті.

Үшінші экспериментте ансамбль ішіндегі модельдер әртүрлілігін бағалау үшін косинустық әртүрлілік (cosine diversity), Дженсен–Шеннон әртүрлілігі (Jensen–Shannon diversity) және ансамбльдік келіспеушілік коэффициенті (ensemble disagreement coefficient) қолданылды. Алынған нәтижелер бойынша disagreement coefficient ең тиімді тәсіл болды: accuracy және macro-F1 өзгермегенімен, cross-entropy кішкене көрсеткішке өсіп, ал Manhattan distance төмендеді. Ал қалған diversity-aware тәсілдері baseline нәтижесінен асып түспеді.

Жалпы, қосымша эксперименттер ішінде ең практикалық тиімді нәтиже temporal disagreement тәсілінде байқалды, себебі ол классификация сапасын нақты жақсартты. Ал PSO және diversity-aware weighting әдістері негізінен DART жүйесінің тұрақтылығын, ансамбль құрылымын және ықтималдық калибрлеуін талдау үшін пайдалы болды.

Қазақ тіліндегі психолингвистикалық мәтіндерді жіктеу үшін жүргізілген эксперименттердің нәтижелері де алынған болатын. Зерттеудің мақсаты-аннотаторлар арасындағы келіспеушіліктерді (Disagreement) ескеріп өңделген деректер негізінде ең тиімді модельдік комбинацияны анықтау болды. Модельдердің тиімділігі Precision, Recall, Accuracy, F1-score және ROC AUC метрикалары бойынша бағаланды.

Бірінші комбинацияда Disagreement өңдеуі мен PLAM белгілері TF-IDF векторизациясымен және дәстүрлі машиналық оқыту алгоритмдерімен бірге қолданылды.



Нәтижелер бойынша ең жоғары көрсеткіштерді SVC моделі көрсетті. Бұл модель барлық бағалау метрикалары бойынша басқа классикалық алгоритмдерден жоғары нәтиже беріп, дәстүрлі тәсілдер ішінде ең тиімдісі ретінде анықталды.

Екінші комбинацияда мәтіндерді ұсыну үшін BERT эмбедингтері пайдаланылып, олар PLAM белгілерімен біріктірілді. Бұл тәсіл классикалық модельдермен салыстырғанда жоғары нәтижелерге қол жеткізді. Accuracy көрсеткіші 95,1 %-ға жетіп, ROC AUC мәні 99,6%-дан асты. Бұл нәтиже контекстік тілдік модельдердің психоллингвистикалық ерекшеліктерді тиімді түрде бейнелей алатынын көрсетті.

Үшінші комбинацияда BERT эмбедингтері LSTM нейрондық желісі арқылы өңделіп, кейін PLAM белгілерімен біріктірілді. Алынған нәтижелер де өте жоғары болғанымен, негізгі метрикалар бойынша BERT + PLAM комбинациясынан сәл төмен болды. Бұл жағдайда BiLSTM қабатын қосу модельдің сапасын айтарлықтай арттырмағаны байқалды.

Жалпы алғанда, эксперимент нәтижелері бойынша ең жоғары көрсеткіштерді Disagreement + PLAM + BERT комбинациясы көрсетті. Сондықтан қазақ тіліндегі психоллингвистикалық мәтіндерді жіктеу міндетінде дәл осы тәсіл ең тиімді модель ретінде таңдалды.

Кесте 7 – DART құрылымды кеңейтуге арналған қосымша эксперименттердің нәтижелері

Эксперимент	Нұсқа	Accuracy		ROC AUC	CE	MD
Optimization algorithm	Differential Evolution	0.9432	0.9421	0.9961	0.5678	0.3177
Optimization algorithm	PSO	0.9432	0.9421	0.9961	0.5678	0.3177
Temporal disagreement	Global baseline	0.9441	0.9434	0.9962	0.5503	0.1906
Temporal disagreement	Temporal-aware DART	0.9459	0.9450	0.9962	0.5496	0.1902
Diversity metric	Baseline DART	0.9441	0.9434	0.9962	0.5503	0.1906
Diversity metric	Disagreement coefficient-aware	0.9441	0.9434	0.9962	0.5502	0.1906

Жүргізілген эксперименттік зерттеулер модельдер арасындағы әдіснамалық емес, нақты тәжірибелік айырмашылықтарды анықтауға мүмкіндік береді. Классикалық машиналық оқыту модельдері ішінде Logistic Regression+PLAM, ал трансформерлік арасында BERT+LSTM моделі барлық метрикалар бойынша жоғары нәтиже бергендігін (Mussiraliyeva, Bolatbek, & Baisylbayeva, 2025) зерттеу жұмысынан көруге болады. Аталған гибрид тәсіл қазақ тілінің агглютинативті құрылымы мен семантикалық күрделілігін ескергенде контекст пен реттілікті бір уақытта модельдеу тұрғысынан ең ұтымды тәсіл екенін айқындайды. Сондықтан қолданылған әдістердің тиімділігі теориялық емес, қазақ тіліндегі нақты корпусқа негізделген эксперименттік талдау арқылы дәлелденді.

Қорытынды.

Жүргізілген шолу нәтижесінде әлеуметтік желілер мен мессенджерлердегі экстремистік контентті анықтау мәселесі заманауи цифрлық қоғамның ең өзекті бағыттарының бірі екені анықталды. Интернеттегі ақпарат көлемінің қарқынды өсуі

мен пайдаланушылардың белсенділігі экстремистік және деструктивті идеялардың таралуына қолайлы орта туғызып жатыр. Мұндай мазмұн қоғамның тұрақтылығына, ұлттық қауіпсіздік пен ақпараттық кеңістікке тікелей қауіп төндіреді, сондықтан оны дер кезінде анықтап, автоматты түрде сүзгіден өткізу аса маңызды.

Экстремистік идеологияның негізгі түрлері-пропаганда, рекрутмент және радикалдану-мазмұны мен мақсаттары жағынан ерекшеленгенімен, үшеуі де қоғамға идеологиялық ықпал етуге бағытталған. Қазақ тіліндегі әлеуметтік медиа кеңістігінде мұндай контентті талдау мен классификациялау үшін сапалы датасет құру керек, аннотациялау сапасын арттыру керек және машиналық оқыту әдістерін тиімді қолдану қажет.

Деректерді жинау кезеңінде әлеуметтік желілерден алынған мәтіндік контент құрылымдалып, төрт негізгі классқа (бейтарап, пропаганда, рекрутмент, радикалдану) бөлінеді. Бұл корпус қазақ тіліндегі идеологиялық мазмұнды зерттеуге арналған маңызды база қалыптастырады. Ал аннотациялау процесінде сарапшылар арасындағы пікір айырмашылықтарын ескеретін disagreement-based learning тәсілі енгізілуі деректер сапасын жақсарта түседі. Мұндай әдіс мәтіндердің екіжақтылығын және белгісіздік деңгейін ескере отырып, модельге идеологиялық мазмұнды дәлірек әрі нақтырақ түсінуге мүмкіндік жасайды.

Машиналық оқыту тұрғысынан дәстүрлі әдістер мен терең оқыту архитектуралары біріге отырып, ансамбльдік модельдер арқылы нәтижелердің дәлдігі арттыруға болады.

Мәтінді алдын ала өңдеу кезеңдері-тазалау, қалыптандыру, стемминг, векторизация және токенизация-модельдің дәлдігі мен сенімділігін қамтамасыз ететін маңызды техникалық негіз болып табылады. Осы әдістердің кешені әлеуметтік желі деректерін автоматтандырылған түрде өңдеуге, идеологиялық ерекшеліктерін талдауға және экстремистік мазмұнды уақтылы әрі нақты анықтауға мүмкіндік туындатады.

Жұмыстың негізгі мақсаты-қазақ тіліндегі экстремистік контентті автоматты түрде талдау үшін әдіснамалық негіз қалыптастыру және қолданыстағы тәсілдердің мүмкіндіктерін жүйелеу болғандықтан, зерттеу бірінші кезеңде теориялық құрылымды құруға бағытталды. Әдебиеттерге сүйене отырып жасалған салыстырмалы талдау нәтижелері архитектура, аннотация стратегиялары және модельдік таңдау бойынша авторлық гипотезаларды анықтауға мүмкіндік берді.

Жалпы алғанда, зерттеу көрсеткендей, қазақ тіліндегі әлеуметтік желілер мен мессенджерлердегі экстремистік контентті анықтау үшін жасанды интеллект, табиғи тілді өңдеу, келіспеушілік негізіндегі оқыту әдістерін біріктірген кешенді жүйе тиімді шешім болып табылады. Мұндай тәсіл киберқауіпсіздік деңгейін арттыруға, қоғамдағы идеологиялық тұрақтылықты сақтауға және отандық ақпараттық кеңістікті қорғауға елеулі үлес қоса алады.

ӘДЕБИЕТТЕР

Ali H.J., Abrar A., Hossain S.M.H., Mridha M.F. (2025). Social media polarization during conflict: Insights from an ideological stance dataset on Israel-Palestine Reddit comments // arXiv. Article 2502.00414. DOI:10.48550/arXiv.2502.00414 (in Eng.)



Ashraf N., Rafiq A., Butt S., Shehzad H., Sidorov G., Gelbukh A. (2022). YouTube-based religious hate speech and extremism detection dataset with machine learning baselines // *Journal of Intelligent & Fuzzy Systems*. — Vol. 42. — No. 3. Pp. 1–9. DOI: 10.3233/JIFS-219264 (in Eng.)

Ahmed A., Refat R. (2024). Exploring the impact of stemming on text topic-based classification accuracy // *Journal of Linguistics, Culture and Communication*. — Vol. 2. — Pp. 204–224. DOI: 10.61320/jolcc.v2i2.204-224 (in Eng.)

Bolatbek M., Mussiraliyeva Sh., Sagynai M., Zhastay E. (2025). Machine learning methods for identifying destructive texts in the Kazakh language // *Proceedings of the International Conference on Algorithms, Data Mining, and Information Technology (ADMIT)*. DOI: 10.1145/3701100.3701121 (in Eng.)

Erhard L., Hanke S., Remer U., Falenska A., Heiberger R.H. (2024). PopBERT: Detecting populism and its host ideologies in the German Bundestag // *Political Analysis*. — Vol. 33. — No. 1. Pp. 1–17. DOI: 10.1017/pan.2024.12 (in Eng.)

Hasib K.M., Azam S., Karim A., Marouf A.A., Shamrat F.M.J.M., Montaha S., Yeo K.C., Jonkman M., Alhajib R., Rokne J. G. (2023). MCNN LSTM: Combining CNN and LSTM to classify multi-class text in imbalanced news data // *IEEE Access*. — Vol. 11. — Pp. 93048–93063. DOI: 10.1109/ACCESS.2023.3309697 (in Eng.)

Kibriya A., Siddiqa A., Khan W.Z., Khan M.K. (2024). Towards safer online communities: Deep learning and explainable AI for hate speech detection and classification // *Computer and Electrical Engineering*. — Vol. 116. — Article 109153. DOI: 10.1016/j.compeleceng.2024.109153 (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*. — Vol. 10. — Pp. 92–110. DOI:10.1162/tacl_a_00449 (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*. DOI:10.1109/ACCESS.2025.3596601 (in Eng.)

Mussiraliyeva S.Zh., Bolatbek M.A., Zhumakhanova A.N., Baispay G., Medetbek Zh. (2024). Creating a model of semantic analysis of extremist texts in the Kazakh language // *Journal of Mathematics, Mechanics and Computer Science*. — Vol. 121. — No. 1. DOI:10.26577/JMMCS2024121111 (in Eng.)

Peng X., Zhou Z., Zhang C., Xu K. (2024). Online social behavior enhanced detection of political stances in tweets // *Proceedings of the International AAAI Conference on Web and Social Media*. — Vol. 18. — No. 1. Pp. 1207–1219. DOI:10.1609/icwsm.v18i1.31383 (in Eng.)

Ram R., Thomas E., Kernot D., Rizoio M.A. (2025). Practical guidelines for ideology detection pipelines and psychosocial applications // *Proceedings of the International AAAI Conference on Web and Social Media*. — Vol. 19. — Pp. 1630–1648. DOI:10.1609/icwsm.v19i1.35892 (in Eng.)

Ravi K., Yuan J.S. (2024). Ideological orientation and extremism detection in online social networking sites: A systematic review // *Intelligent Systems with Applications*. — Vol. 24. — Article 200456. DOI:10.1016/j.iswa.2024.200456 (in Eng.)

Ramponi A., Testa B., Tonelli S., Jezek E. (2022). Addressing religious hate online: From taxonomy creation to automated detection // *PeerJ Computer Science*. — Vol. 8. — Article e1128. DOI: 10.7717/peerj-cs.1128 (in Eng.)

Shahin M.A., Nassif A.B., Elnagar A. (2024). Detection of Arabic offensive language in social media using machine learning models // *Intelligent Systems with Applications*. — Vol. 22. — Article 200376. DOI: 10.1016/j.iswa.2024.200376 (in Eng.)

Uma A.N., Fornaciari T., Paun S., Plank B., Poesio M. (2021). Learning from disagreement: A survey // *Journal of Artificial Intelligence Research*. — Vol. 72. — Pp. 1385–1470. DOI: 10.1613/jair.1.12752 (in Eng.)

Hasib K.M., Azam S., Karim A., Marouf A.A., Shamrat F.M.J.M., Montaha S., Yeo K.C., Jonkman M., Alhajib R., Rokne J.G. (2023). MCNN LSTM: Combining CNN and LSTM to classify multi-class text in imbalanced news data // *IEEE Access*. — Vol. 11. — Pp. 93048–93063. DOI:10.1109/ACCESS.2023.3309697 (in Eng.)

Xu J., Theune M., Braun D. (2024). Leveraging Annotator Disagreement for Text Classification // *Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024)*. Pp. 1–10. DOI:10.48550/arXiv.2409.17577 (in Eng.)

Yelmen A., Gunes A., Zontul M. (2023). Multi-class document classification using lexical ontology-based deep learning // *Applied Sciences*. — Vol. 13. — No. 10. Article 6139. DOI:10.3390/app13106139 (in Eng.)

**INTERNATIONAL JOURNAL OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Собственник:

АО «Международный университет информационных
технологий» (Казахстан, Алматы)

Главный редактор:

Колесникова Катерина Викторовна

Ответственный редактор:

Мрзабаева Раушан Жалиевна

Компьютерная верстка:

Калабай Замзагуль Ертугановна

Сайт журнала: <https://journal.iitu.edu.kz>

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Подписано в печать 30.06.2026.

050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).