

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. 1 (25). 2026
January–March

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

2020 жылдан бері шығарылады
Том 7. 1 (25). 2026
Қаңтар-Наурыз

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

Издается с 2020 г.
Том 7. 1 (25). 2026
Январь-Март

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)

Bakhtgerai 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 Sansyzbay — 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 Түлегеновна — 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)

Лучино 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 Мұхтарбай 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 Тоқсуажасвич — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)

Синчев Бахтгерей Кусанович — 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)

Абикаликова Замира Тұрсынбаевна — 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)

Дамеля Максұтовна Ескендірова — 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)

Тадеш Валлас — PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры

Мамырбаев Оркен Жұмажанович — 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 сулест 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)

Дамеуш Максустовна Ескендирова — 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 Сахиябай — кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алимсиреб 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)

Белошицкая 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 социо-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 SOCIOECONOMIC SYSTEMS

A.B. Zhalgas, Y.N. Kalpakov, B.Ye. Amirgaliyev MACHINE LEARNING-DRIVEN OPTIMIZATION OF LOGISTICS IN SMART CITIES: A CASE STUDY OF ASTANA	11
L. Kurmangazyieva, Sh. Kodanova, M. Urazgaliyeva, O. Findik, S. Iskakova INTEGRATING FUZZY LOGIC AND ARTIFICIAL INTELLIGENCE IN OPTIMIZING BUSINESS PROCESS AUTOMATION DECISIONS	26
Y. Mailybayev, U. Adilbayeva, R. Amanova ORGANIZATION OF AN ONLINE SURVEY OF PARTICIPANTS IN THE EDUCATIONAL PROCESS AND ANALYSIS OF THE RESULTS BASED ON THE MODIFIED DELPHI METHOD	48
V.A. Takizhanov, A.Z. Ibragimov, A. Shalakhmetov SIMULATION-BASED ROBUSTNESS ASSESSMENT OF ASTANA'S BUS NETWORK UNDER RANDOM AND TARGETED FAILURES	63
G. Hamada, M. Hamada LEVERAGING MACHINE LEARNING MODELS TO BOOST THE MANIPULATION OF UNSTRUCTURED DECISIONS AND UNSTRUCTURED DATA IN DIGITAL BUSINESS FIRMS	78

INFORMATION TECHNOLOGY

M. Zh. Aitimov, G. K. Muratova, Zh. K. Bissenbayeva, I.M. Bapiyev, M. Kassim SEMANTIC COMPLETENESS IN KAZAKH-LANGUAGE EXTRACTIVE QA THROUGH ONTOLOGY AND RETRIEVAL MECHANISMS	96
A.Aitim, A. Auyezova ENCODING NP-COMPLETE PROBLEM INSTANCES INTO NATURAL LANGUAGE FOR LLM PROCESSING	109
O.N. Akylbekov, Y.T. Dauletbek, A.N. Moldagulova, G.S. Zakariya, D.A. Gura MACHINE LEARNING METHODS FOR ANALYSING THREE-DIMENSIONAL SPATIAL DATA IN KAZAKHSTAN'S LAND USE PLANNING	126
S.Zh. Aliaskarov, R.K. Uskenbayeva, A. Razaque, A.B. Kassymova, A.M. Anartayeva TOWARDS EFFICIENT BIG DATA ANALYTICS IN REGIONAL SYSTEMS: PRACTICAL INSIGHTS FROM HYBRID ARCHITECTURE DEPLOYMENT	146
Zh. Bilyalova, N. Shazhdekeyeva, K. Uteulyieva, T.Akhmurzina, L. Nurgaliyeva A MATHEMATICAL MODEL FOR DECISION-MAKING UNDER UNCERTAINTY HYBRID BASED ON FUZZY LOGIC AND GENERATIVE ARTIFICIAL INTELLIGENCE	165
A. Ismailova, G. Yessenbayeva, K. Kadyrkulov, R. Moldasheva, A. Amangeldi DEVELOPMENT OF A HYBRID DEEP LEARNING MODEL FOR MULTICLASS CLASSIFICATION OF MICROSCOPIC IMAGES OF BACTERIA	186
G. Kalman, J. Kultan, A.N. Ismukamova, N.M. Ausilova, Y.V. Makhatova A DOMAIN-KNOWLEDGE-BASED MODEL FOR REFERENCE RESOLUTION IN LOW-RESOURCE LANGUAGES.....	199
Y. Kamen, Zh. Yessendauletova, L. Fazylova, M. Rakhimzhanova, A.M. Nedzved USING NEURAL NETWORKS FOR OBJECTIVE ASSESSMENT OF ATTENTION IN CHILDREN BASED ON EEG DATA	216
A.Ye. Kulakayeva, Ye.A. Bakhtiyarova, G.T. Jakanova, Sh. Nursultan COMPARATIVE ANALYSIS OF VARIOUS RADIO WAVE PROPAGATION MODELS FOR MOBILE NETWORK COVERAGE PREDICTION	231
A. Kalykulova, K. Bakirov, A. Shoman, A. Nugumanova FORMATION OF A UNIFIED DATA SET FOR ANALYZING THE RELATIONSHIP BETWEEN DOG BREEDS AND DISEASES	247
D. Kantayeva, A. Seisenbekkyzy, S. Biloshchytska COMPARATIVE EVALUATION MODEL BASED ON EXPERT SYSTEMS FOR ABSTRACTIVE SUMMARIZATION OF SCIENTIFIC TEXTS	269
N. Mordvintsev, Zh. Oralbekova, A. Bakiyeva DEEP LEARNING-BASED VEHICLE MAKE AND MODEL RECOGNITION: DATASET CHALLENGES	

AND INFLUENCING FACTORS IN URBAN TRAFFIC	269
M.B. Nurpeissova, Sh.K. Aitkazinova, A.M. Abenov, N.S. Donenbayeva METHODOLOGY FOR TRANSFORMING SATELLITE COORDINATES INTO A TOPOCENTRIC RECTANGULAR COORDINATE SYSTEM	315
L. Naizabayeva, G. Turken, R.A. Druba EFFECTIVE METHOD FOR OPTIMAL CLUSTERING IN UNSUPERVISED LEARNING	328
S. Nurgaliyeva, Zh. Seitkaliyev, A. Sakhipov, M. Kali, R. Hrmo YIELD FORECASTING FOR FRUITS AND VEGETABLES USING RANDOM FOREST ON LARGE- SCALE AGROCLIMATIC DATA	339
A. Nurgalykov, S. Sapakova, D. Zhalmukhanbetov, G. Nurgazy, A. Myrzabayev LANGUAGE ASYMMETRY IN LLMS: THE IMPACT OF NON-ENGLISH DATA ON ACCURACY AND RELIABILITY	355
A. Ospanov, P. Alonso-Jordá, A. Zhumadillayeva BLOCKCHAIN-ENABLED ERP WAREHOUSE INTEGRATION WITH IOT DIMENSIONERS AND MACHINE LEARNING–OPTIMIZED DIMENSIONAL WEIGHT RECONCILIATION	371
A.A. Sakhipov, R.B. Seıitbek EVENT-DRIVEN MICROSERVICES FOR INCIDENT DETECTION AND RESPONSE IN INTELLIGENT TRAFFIC SYSTEM	387
A. Tursynova, B. Omarov EVOLUTION OF THE EFFECTIVENESS OF CONDITIONAL GAN FOR THE SYNTHESIS OF CT IMAGES OF THE BRAIN IN THE DIAGNOSIS OF STROKE	400
G. Yusupova, K.S. Shadinova, D. Ussipbekova, Zh.Zh. Azhibekova, P. Schmidt DETERMINATION OF SOIL PROFILE STRATIFICATION AT 0–200 CM DEPTH USING A MULTILEVEL STACKING MODEL	414

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

S.A. Adilzhanova, M.Zh. Sakypbekova, L.Sh. Cherikbaeva, G.A. Tyulepberdinova, G.T. Zhubanysheva SYSTEMATIC ANALYSIS OF RISK ASSESSMENT METHODS AND MODELS IN INFORMATION SECURITY.....	427
T. K. Zhukabayeva, D.B. Baumuratova, E. Benkhelifa , N.A. Niyetbayeva EDGE COMPUTING-BASED TECHNIQUE FOR CONSTRUCTION OF ATTACK DETECTION MEANS IN CYBER-PHYSICAL SYSTEMS OF INDUSTRIAL INTERNET-OF-THINGS	453
N.E. Karabayev, S.K. Serikbayeva, Y.M. Mardenov, B. Tassuov, M. Fajkus DETECTION OF CYBER ATTACKS IN TRANSPORT NETWORKS BASED ON MACHINE LEARNING METHODS	475
B.A. Kumalakov, A.O. Dargulova A HYBRID FRAMEWORK FOR RESUME-JOB MATCHING SYSTEM	494
V. Makhatova, B. Dzhugembayeva, A. Gabdulova, L. Nurgaliyeva, A. Abdigaliyeva MATHEMATICAL MODEL FOR OPTIMAL SENSOR SELECTION IN SIEM SYSTEMS USING THE ANALYTIC HIERARCHY PROCESS	509

МАЗМУНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев АҚЫЛДЫ ҚАЛАЛАРДАҒЫ ЛОГИСТИКАНЫ МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН ОҢТАЙЛАНДЫРУ: АСТАНАНЫҢ ЖАҒДАЙЫН ЗЕРТТЕУ.....	11
Л.Қурманғазиева, Ш. Қоданова, М. Уразғалиева, О. Findik, С. Искакова ЖАСАНДЫ ИНТЕЛЛЕКТ ПЕН АЙҚЫН ЕМЕС ЛОГИКАНЫ БІРІКТІРУ АРҚЫЛЫ БИЗНЕС- ПРОЦЕСТЕРДІ АВТОМАТТАНДЫРУ ШЕШІМДЕРІН ОҢТАЙЛАНДЫРУ	26
Е. Майлыбаев, У. Адилбаева, Р. Аманова ҰЙЫМДАСТЫРЫЛҒАН ОНЛАЙН САУАЛНАМА АРҚЫЛЫ БІЛІМ БЕРУ ПРОЦЕСІНЕ ҚАТЫСУШЫЛАРДЫҢ ПІКІРЛЕРІН ЖИНАУ ЖӘНЕ НӘТИЖЕЛЕРІН МОДИФИКАЦИЯЛАНҒАН ДЕЛЬФИ ӘДІСІ НЕГІЗІНДЕ ТАЛДАУ	48
В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов	

МОДЕЛЬДЕУ НЕГІЗІНДЕ АСТАНАНЫҢ АВТОБУС ЖЕЛІСІНІҢ ТҮРАҚТЫЛЫҒЫН БАҒАЛАУ: КЕЗДЕЙСОҚ ЖӘНЕ МАҚСАТТЫ ІСТЕН ШЫҒУЛАР ЖАҒДАЙЫНДА	63
Г. Хамада, М. Хамада ЦИФРЛЫҚ БИЗНЕС-ҰЙЫМДАРЫНДА ҚҰРЫЛЫМДАЛМАҒАН ШЕШІМДЕР МЕН ҚҰРЫЛЫМДАЛ- МАҒАН ДЕРЕКТЕРМЕН ЖҰМЫС ІСТЕУ ТИІМДІЛІГІН АРТТЫРУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІН ПАЙДАЛАҢУ	78

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим ОНТОЛОГИЯ ЖӘНЕ ІЗДЕУ МЕХАНИЗМДЕРІ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ ЭКСТРАКЦИЯЛЫҚ ҚА-ДАҒЫ СЕМАНТИКАЛЫҚ ТОЛЫҚТЫҚ	96
Ә. Әйтгім, Ә. Әуезова LLM ӨНДЕУІ ҮШІН NP-ТОЛЫҚ ЕСЕПТЕРДІҢ ЭКЗЕМПЛЯРЛАРЫН ТАБИҒИ ТІЛДЕ КОДТАУ	109
О.Н. Ақылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура ҚАЗАҚСТАННЫҢ АУМАҚТЫҚ ЖОСПАРЛАУЫНДАҒЫ ҮШ ӨЛШЕМДІ КЕҢІСТІКТІК МӨЛІМЕТТЕРДІ ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ	126
С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымов, А.М. Анартаева АЙМАҚТЫҚ ЖҮЙЕЛЕРДЕГІ ҮЛКЕН ДЕРЕКТЕРДІ ТИІМДІ ТАЛДАУҒА ҚАРАЙ: ГИБРИДТІ АРХИТЕКТУРАНЫ ЕНГІЗУДІҢ ПРАКТИКАЛЫҚ ТҮСІНІКТЕР.....	146
Ж. Билялова, Н. Шаждекеева, К. Өтеулиева, Т. Ахмурзина, Л. Нұрғалиева АЙҚЫН ЕМЕС ЛОГИКА ЖӘНЕ ГЕНЕРАТИВТІ ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ БЕЛГІСІЗДІК ЖАҒДАЙЫНДА ШЕШІМ ҚАБЫЛДАУДЫҢ ГИБРИДТІ МАТЕМАТИКАЛЫҚ МОДЕЛІ	165
А.А. Исмаилова, Г.Р. Есенбаева, К.К. Кадиркулов, Р.Н. Молдашева, А. Амангелді РОСКОПИЯЛЫҚ БЕЙНЕЛЕРІН КӨПКЛАССТЫ ЖІКТЕУГЕ АРНАЛҒАН ГИБРИДТІ ТЕРЕҢ ОҚЫТУ МОДЕЛІН ӨЗІРЛЕУ	186
Г. Қалман, К. Ярослав, А.Н. Исмуханова, Н.М. Аусилова, В.Е. Махатова ПӘНДІК САЛА БІЛІМ НЕГІЗІНДЕ РЕУСРСТАРЫ АЗ ТІЛДЕРДЕГІ РЕФЕРЕНЦИЯНЫ ШЕШУДІҢ МОДЕЛІ.....	199
Е.Г. Кәмен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзьведь ЭЭГ ДЕРЕКТЕРІ БОЙЫНША БАЛАЛАРДЫҢ ЗЕЙІНІН ОБЪЕКТИВТІ БАҒАЛАУ ҮШІН НЕЙРОНДЫҚ ЖЕЛІЛЕРДІ ҚОЛДАНУ	216
А.Е. Құлакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нұрсұлтан ҰЯЛЫ БАЙЛАНЫС ЖЕЛІЛЕРІНІҢ ҚАМТУ АЙМАҒЫН БОЛЖАУҒА АРНАЛҒАН ӨРТҮРЛІ РАДИОТӨЛҚЫН ТАРАЛУ МОДЕЛЬДЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ	231
А. Калыкулова, К. Бакиров, А. Шоман, А. Нугуманова ИТ ТҰҚЫМДАРЫ МЕН АУРУЛАРЫ АРАСЫНДАҒЫ БАЙЛАНЫСТЫ ТАЛДАУҒА АРНАЛҒАН БЫРЫҒАЙ ДЕРЕКТЕР ЖИЫНТЫҒЫН ҚАЛЫПТАСТЫРУ	247
Д. Кантаева, А. Сейсенбекқызы, С. Билощницкая ҒЫЛЫМИ МӘТІНДЕРДІ АБСТРАКТИВТІ РЕФЕРАТТАУҒА АРНАЛҒАН ЭКСПЕРТТІК ЖҮЙЕЛЕРГЕ НЕГІЗДЕЛГЕН САЛЫСТЫРМАЛЫ БАҒАЛАУ МОДЕЛІ	269
Н. Мордвинцев, Ж. Оралбекова, А. Бакиева ТЕРЕҢ ОҚЫТУҒА НЕГІЗДЕЛГЕН КӨЛІК ҚҰРАЛДАРЫНЫҢ МАРКАСЫ МЕН МОДЕЛІН ТАҢУ: ҚА- ЛАЛЫҚ КӨЛІК ҚОЗҒАЛЫСЫНДАҒЫ ДЕРЕКТЕР ЖИЫНТЫҒЫНЫҢ МӘСЕЛЕЛЕРІ ЖӘНЕ ЫҚПАЛ ЕТУШІ ФАКТОРЛАР	285
М.Б. Нұрпейісова, Ш.Қ. Айтқазина, А.М. Абенов, Н.С. Дөненбаева СПУТНИКТІК КООРДИНАТТАРДЫ ТОПОЦЕНТРЛІК ТІК БҰРЫШТЫ КООРДИНАТТАР ЖҮЙЕСІНЕ ТҮРЛЕНДІРУДІҢ ӘДІСТЕМЕСІ	315
Л. Найзабаева, Г. Түркен, Р.А. Друба БАҚЫЛАНБАЙТЫН ОҚЫТУДАҒЫ ОҢТАЙЛЫ КЛАСТЕРЛЕУДІҢ ТИІМДІ ӘДІСІ	328
С. Нұрғалиева, Ж. Сенткалиев, А. Сахипов, М. Кали, Р. Грмо ҮЛКЕН КӨЛЕМДЕГІ АГРОКЛИМАТТЫҚ ДЕРЕКТЕР НЕГІЗІНДЕ RANDOM FOREST ӘДІСІН ПАЙДАЛАНА ОТЫРЫП, ЖЕМІС-КӨКӨНІС ДАҚЫЛДАРЫНЫҢ ӨНІМДІЛІГІН БОЛЖАУ	339
А. Нұрғалықов, С. Сапакова, Д. Жалмұханбетов, Г. Нұрғазы, А. Мырзабаев ҮЛКЕН ТІЛДІК МОДЕЛЬДЕРДЕГІ ТІЛДІК АСИММЕТРИЯ: АҒЫЛШЫН ТІЛІНЕН БАСҚА ДЕРЕКТЕРДІҢ ДӘЛДІК ПЕН СЕНІМДІЛІККЕ ӘСЕРІ	355
А. Оспанов, П. Алонсо-Хорда, А. Жұмаділлаева	

БЛОКЧЕЙН-ТЕХНОЛОГИЯСЫМЕН ЫҚПАЛДАС ERP ҚОЙМА ЖҮЙЕСІН ІОТ ДИМЕНСИОНЕРЛЕР ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ АРҚЫЛЫ ОПТИМИЗАЦИЯЛАНҒАН ӨЛШЕМДІ САЛМАҚ ЕСЕПТЕУМЕН ИНТЕГРАЦИЯЛАУ	371
А.А. Сахипов, Р.Б. Сейітбек ОҚИҒАҒА БАҒДАРЛАНҒАН МИКРОҚЫЗМЕТТЕР ЖҮЙЕСІ АРҚЫЛЫ АҚЫЛДЫ ТРАФИК ЖҮЙЕЛЕРІНДЕ ОҚИҒАЛАРДЫ АНЫҚТАУ ЖӘНЕ ШАРАЛАР ҚОЛДАНУ	387
А. Турсынова, Б. Омаров ИНСУЛЬТТИ ДИАГНОСТИКАЛАУ КЕЗІНДЕ МИ КТ-БЕЙНЕЛЕРІН СИНТЕЗДЕУГЕ АРНАЛҒАН ШАРТТЫ ҒАН ТИІМДІЛІГІНІҢ ЭВОЛЮЦИЯСЫ	400
Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, Р. Schmidt ТОПЫРАҚ ПРОФИЛІНІҢ 0–200 СМ ТЕРЕНДІКТЕГІ СТРАТИФИКАЦИЯСЫН КӨПДЕНГЕЙЛІ СТЕКИНГ-МОДЕЛІМЕН АНЫҚТАУ.....	414

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

С.А. Адилжанова, М.Ж. Сақыпбекова, Л.Ш. Черикбаева, Г.А. Түлепбердинова, Г.Т. Жубанышева АҚПАРАТТЫҚ ҚАУІПСІЗДІКТЕ ТӘУЕКЕЛДЕРДІ БАҒАЛАУ ӘДІСТЕРІ МЕН МОДЕЛЬДЕРІН ЖҮЙЕЛІ ТАЛДАУ.....	427
Т.К. Жукабаева, Д. Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниетбаева ШЕКАРАЛЫҚ ЕСЕПТЕУЛЕРДІ ҚОЛДАНА ОТЫРЫП, ЗАТТАРДЫҢ ӨНЕРКӘСІПТІК ИНТЕРНЕТІНІҢ КИБЕРФИЗИКАЛЫҚ ЖҮЙЕЛЕРІНДЕГІ ШАБУЫЛДАРДЫ АНЫҚТАУ ҚҰРАЛДАРЫН ҚҰРУ ӘДІСТЕМЕСІ.....	453
Н.Е. Қарабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файқус МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН КӨЛІК ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ	475
Б.А. Кумалаков, А.О. Даргулова ТҮЙІНДЕМЕЛЕР МЕН ВАКАНСИЯЛАРДЫ АВТОМАТТАНДЫРЫЛҒАН СӘЙКЕСТЕНДІРУГЕ НЕГІЗДЕЛГЕН ГИБРИДТІ ҮМІТКЕРЛЕРДІ ІРІКТЕУ ЖҮЙЕСІ	494
В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нургалиева, А. Абдигалиева ИЕРАРХИЯЛАРДЫ ТАЛДАУ ӘДІСІ НЕГІЗІНДЕ SIEM ЖҮЙЕЛЕРІНДЕ ОҢТАЙЛЫ СЕНСОРДЫ ТАҢДАУДЫҢ МАТЕМАТИКАЛЫҚ МОДЕЛІ	509

СОДЕРЖАНИЕ ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

А.Б. Жалғас, Е.Н. Калпаков, Б.Е. Амиргалиев ОПТИМИЗАЦИЯ ЛОГИСТИКИ В УМНЫХ ГОРОДАХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ: НА ПРИМЕРЕ АСТАНЫ	11
Л. Курмангазиева, Ш. Коданова, М. Уразғалиева, О. Финдик, С. Исакова ИНТЕГРАЦИЯ НЕЧЕТКОЙ ЛОГИКИ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОПТИМИЗАЦИИ РЕШЕНИЙ ПО АВТОМАТИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ	26
Е. Майлыбаев, У. Адилбаева, Р. Аманова СБОР МНЕНИЙ УЧАСТНИКОВ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПОСРЕДСТВОМ ОРГАНИЗОВАННОГО ОНЛАЙН-АНКЕТИРОВАНИЯ И АНАЛИЗ РЕЗУЛЬТАТОВ НА ОСНОВЕ МОДИФИЦИРОВАННОГО МЕТОДА ДЕЛЬФИ	48
В.А. Такижанов, А.Ж. Ибрагимов, А. Шалахметов ОЦЕНКА УСТОЙЧИВОСТИ АВТОБУСНОЙ СЕТИ АСТАНЫ НА ОСНОВЕ МОДЕЛИРОВАНИЯ ПРИ СЛУЧАЙНЫХ И ЦЕЛЕНАПРАВЛЕННЫХ ОТКАЗАХ	63
Г. Хамада, М. Хамада ИСПОЛЬЗОВАНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РАБОТЫ С НЕСТРУКТУРИРОВАННЫМИ РЕШЕНИЯМИ И НЕСТРУКТУРИРОВАННЫМИ ДАННЫМИ В ЦИФРОВЫХ БИЗНЕС-ОРГАНИЗАЦИЯХ	78

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

М.Ж. Айтимов, Г.К. Муратова, Ж.К. Бисенбаева, И.М. Бапиев, М. Кассим

СЕМАНТИЧЕСКАЯ ПОЛНОТА В КАЗАХСКОЯЗЫЧНОМ EXTRACTIVE QA ЧЕРЕЗ ОНТОЛОГИЮ И RETRIEVAL-МЕХАНИЗМЫ	96
А. Айтим, А. Ауезова КОДИРОВАНИЕ ЭКЗЕМПЛЯРОВ NP-ПОЛНЫХ ЗАДАЧ В ЕСТЕСТВЕННОМ ЯЗЫКЕ ДЛЯ ОБРАБОТКИ LLM	109
О.Н. Акылбеков, Е.Т. Даулетбек, А.Н. Молдагулова, Г.С. Закария, Д.А. Гура МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ТРЁХМЕРНЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ В ТЕРРИТОРИАЛЬНОМ ПЛАНИРОВАНИИ КАЗАХСТАНА	126
С.Ж. Алиаскаров, Р.К. Ускенбаева, А. Разак, А.Б. Касымова, А.М. Анартаева НА ПУТИ К ЭФФЕКТИВНОЙ АНАЛИТИКЕ БОЛЬШИХ ДАННЫХ В РЕГИОНАЛЬНЫХ СИСТЕМАХ: ПРАКТИЧЕСКИЕ ВЫВОДЫ ИЗ ВНЕДРЕНИЯ ГИБРИДНОЙ АРХИТЕКТУРЫ	146
Ж. Билялова, Н. Шаждекеева, К. Утеулиева, Т. Ахмурзина, Л. Нурғалиева ГИБРИДНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ НА ОСНОВЕ НЕЧЕТКОЙ ЛОГИКИ И ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	165
А.А. Исмаилова, Г.Р. Есенбаева, К.К. Кадиркулов, Р.Н. Молдашева, А. Амангелды РАЗРАБОТКА ГИБРИДНОЙ МОДЕЛИ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ БАКТЕРИЙ	186
Г. Калман, К. Ярослав, А.Н. Исмуханова, Н.М. Аусилова, В.Е. Махатова МОДЕЛЬ НА ОСНОВЕ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ РАЗРЕШЕНИЯ КОРЕФЕРЕНЦИИ В МАЛОРЕСУРСНЫХ ЯЗЫКАХ	199
Е.Г. Камен, Ж.Т. Есендаулетова, Л.С. Фазылова, М.Б. Рахимжанова, А.М. Недзведь ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБЪЕКТИВНОЙ ОЦЕНКИ ВНИМАНИЯ У ДЕТЕЙ ПО ДАННЫМ ЭЭГ	216
А.Е. Кулакаева, Е.А. Бахтиярова, Г.Т. Джаканова, Ш. Нурсултан СРАВНИТЕЛЬНЫЙ АНАЛИЗ РАЗЛИЧНЫХ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ РАДИОВОЛН ДЛЯ ПРОГНОЗИРОВАНИЯ ПОКРЫТИЯ СЕТЕЙ МОБИЛЬНОЙ СВЯЗИ	231
А. Калыкулова, К. Бакиров, А. Шоман, А. Нугуманова ФОРМИРОВАНИЕ ЕДИНОГО НАБОРА ДАННЫХ ДЛЯ АНАЛИЗА ВЗАИМОСВЯЗИ МЕЖДУ ПОРОДАМИ СОБАК И ЗАБОЛЕВАНИЯМИ	247
Д. Кантаева, А. Сейсенбеккызы, С. Билощицкая СРАВНИТЕЛЬНАЯ МОДЕЛЬ ОЦЕНКИ НА ОСНОВЕ ЭКСПЕРТНЫХ СИСТЕМ ДЛЯ АБСТРАКТИВНО-ГО РЕФЕРИРОВАНИЯ НАУЧНЫХ ТЕКСТОВ	269
Н. Мордвинцев, Ж. Оралбекова, А. Бакиева РАСПОЗНАВАНИЕ МАРКИ И МОДЕЛИ ТРАНСПОРТНЫХ СРЕДСТВ НА ОСНОВЕ ГЛУБОКОГО ОБУЧЕНИЯ: ПРОБЛЕМЫ НАБОРОВ ДАННЫХ И ФАКТОРЫ, ВЛИЯЮЩИЕ НА РАБОТУ СИСТЕМЫ В УСЛОВИЯХ ГОРОДСКОГО ДВИЖЕНИЯ	285
М.Б. Нурпеисова, Ш.К. Айтказинова, А.М. Абенов, Н.С. Доненбаева МЕТОДИКА ПРЕОБРАЗОВАНИЯ СПУТНИКОВЫХ КООРДИНАТ В ТОПОЦЕНТРИЧЕСКУЮ ПРЯМОУГОЛЬНУЮ СИСТЕМУ КООРДИНАТ	315
Л. Найзабаева, Г. Туркен, Р.А. Друба ЭФФЕКТИВНЫЙ МЕТОД ОПТИМАЛЬНОЙ КЛАСТЕРИЗАЦИИ В ОБУЧЕНИИ БЕЗ УЧИТЕЛЯ	328
С. Нурғалиева, Ж. Сенткалиев, А. Сахипов, М. Кали, Р. Грмо ПРОГНОЗИРОВАНИЕ УРОЖАЙНОСТИ ФРУКТОВ И ОВОЩЕЙ С ИСПОЛЬЗОВАНИЕМ RANDOM FOREST НА ОСНОВЕ КРУПНОМАСШТАБНЫХ АГРОКЛИМАТИЧЕСКИХ ДАННЫХ	339
А. Нурғалыков, С. Сапакова, Д. Жалмуханбетов, Г. Нурғазы, А. Мырзабаев ЯЗЫКОВАЯ АСИММЕТРИЯ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ: ВЛИЯНИЕ ДАННЫХ НА НЕАНГЛИЙСКИХ ЯЗЫКАХ НА ТОЧНОСТЬ И НАДЕЖНОСТЬ	355
А. Оспанов, П. Алонсо-Хорда, А. Жумадилаева ИНТЕГРАЦИЯ СКЛАДСКИХ МОДУЛЕЙ ERP-СИСТЕМ С ИСПОЛЬЗОВАНИЕМ БЛОКЧЕЙНА, ИОТ-ДИМЕНСИОНЕРОВ И ОПТИМИЗИРОВАННОГО МАШИНЫМ ОБУЧЕНИЕМ РАСЧЁТА ГАБАРИТНОГО ВЕСА	371
А.А. Сахипов, Р.Б. Сейитбек СОБЫТИЯ-ОРИЕНТИРОВАННЫЕ МИКРОСЕРВИСЫ ДЛЯ ОБНАРУЖЕНИЯ И РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ В ИНТЕЛЛЕКТУАЛЬНЫХ ТРАНСПОРТНЫХ СИСТЕМАХ	387
А. Тұрсынова, Б. Омаров ЭВОЛЮЦИЯ ЭФФЕКТИВНОСТИ УСЛОВНОГО GAN ДЛЯ СИНТЕЗА КТ-ИЗОБРАЖЕНИЙ ГОЛОВНО-	

ГО МОЗГА ПРИ ДИАГНОСТИКЕ ИНСУЛЬТА	400
Г.М. Юсупова, К.С. Шадинова, Д.И. Усипбекова, Ж.Ж. Ажибекова, П. Шмидт	
ОПРЕДЕЛЕНИЕ СТРАТИФИКАЦИИ ПОЧВЕННОГО ПРОФИЛЯ НА ГЛУБИНЕ 0–200 СМ С	
ИСПОЛЬЗОВАНИЕМ МОДЕЛИ МНОГОУРОВНЕВОГО НАЛОЖЕНИЯ	414

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

С.А. Адилжанова, М.Ж. Сакыпбекова, Л.Ш. Черикбаева, Г.А. Тюлепбердинова, Г.Т. Жубанышева	
СИСТЕМАТИЧЕСКИЙ АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННОЙ	
БЕЗОПАСНОСТИ	427
Т.К. Жукабаева, Д.Б. Баумуратова, Е. Бенкхелифа, Н.А. Ниембаева	
МЕТОДИКА ПОСТРОЕНИЯ СРЕДСТВ ОБНАРУЖЕНИЯ АТАК В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ	
ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАНИЧНЫХ	
ВЫЧИСЛЕНИЙ	453
Н.Е. Карабаев, С.К. Серикбаева, Е.М. Марденов, Б. Тасуов, М. Файкус	
ОБНАРУЖЕНИЕ КИБЕРАТАК В ТРАНСПОРТНЫХ СЕТЯХ НА ОСНОВЕ МЕТОДОВ МАШИННОГО	
ОБУЧЕНИЯ	475
Б.А. Кумалаков, А.О. Даргулова	
ГИБРИДНЫЙ ПОДХОД К АВТОМАТИЗИРОВАННОМУ ПОДБОРУ КАНДИДАТОВ НА ОСНОВЕ	
СОПОСТАВЛЕНИЯ РЕЗЮМЕ И ВАКАНСИЙ	494
В. Махатова, Б. Джугембаева, А. Габдулова, Л. Нургалиева, А. Абдигалиева	
МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЫБОРА ОПТИМАЛЬНОГО СЕНСОРА В SIEM-СИСТЕМАХ	
СРЕДСТВАМИ МЕТОДА АНАЛИЗА ИЕРАРХИЙ	509

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ISSN 2708–2032 (print)

ISSN 2708–2040 (online)

Vol. 7. Is.1. Number 25 (2026). Pp. 355–370

Journal homepage: <https://journal.iitu.edu.kz><https://doi.org/10.54309/IJICT.2026.25.1.021>

LANGUAGE ASYMMETRY IN LLMS: THE IMPACT OF NON-ENGLISH DATA ON ACCURACY AND RELIABILITY

A. Nurgalykov, S. Sapakova, D. Zhalmukhanbetov, G. Nurgazy, A. Myrzabayev*

International Information Technology University, Almaty, Kazakhstan.

E-mail: nurgalykovarman@gmail.com

Arman Nurgalykov — Master Student, International Information Technology University, Almaty, Kazakhstan

E-mail: nurgalykovarman@gmail.com, <https://orcid.org/0009-0004-9574-4797>;

Saya Sapakova — PhD, JSC “IITU”, Computer Engineering, Associate Professor, International Information Technology University, Almaty, Kazakhstan

<https://orcid.org/0000-0001-6541-6806>;

Dastan Zhalmukhanbetov — Master Student, International Information Technology University, Almaty, Kazakhstan

<https://orcid.org/0009-0007-0185-0081>;

Gani Nurgazy — Master Student, International Information Technology University, Almaty, Kazakhstan

<https://orcid.org/0009-0002-8570-1318>;

Amirzhan Myrzabayev — Master Student, International Information Technology University, Almaty, Kazakhstan

<https://orcid.org/0009-0002-0012-9296>.

© A. Nurgalykov, S. Sapakova, D. Zhalmukhanbetov, G. Nurgazy, A. Myrzabayev

Abstract. This paper investigates language asymmetry in large language models (LLMs) – the tendency for models trained primarily on English data to perform worse on other languages. We conduct a comparative evaluation using a university recommendation task to quantify how an English-centric training corpus impacts accuracy and reliability in non-English contexts. Our experiments confirm a stark disparity: the model’s recommendations for English-language (U.S.) queries were mostly accurate (≈ 83 % completely error-free, with only ~ 7 % containing critical factual errors), whereas recommendations for analogous Asian-context queries were often flawed (only ~ 46 % error-free, with ~ 21 % containing critical errors). Even when augmenting the system with a more powerful OpenAI model, the improvement was pronounced for English data (overall error rate dropping from ~ 30 % to ~ 12 %) but far more limited for Asian data (~ 60 % to ~ 42 %, still leaving ~ 16 % of responses with serious errors). These



findings illustrate that current LLMs remain significantly less reliable on non-English inputs, posing challenges for real-world applications (e.g. in education) that require region-specific knowledge.

Keywords: Large Language Models (LLM), language asymmetry, cross-lingual bias, hallucinations, information retrieval, multilingual data, applied NLP in education

For citation: Nurgalykov A., Sapakova S., Zhalmukhanbetov D., Nurgazy G., Myrzabayev A. (2026). Language asymmetry in llms: the impact of NON-english data on accuracy and reliability // International journal of information and communication technologies. Vol. 7. No. 25. Pp. 355-370. <https://doi.org/10.54309/IJICT.2026.25.1.021>. (In Eng.).

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

ҮЛКЕН ТІЛДІК МОДЕЛЬДЕРДЕГІ ТІЛДІК АСИММЕТРИЯ: АҒЫЛШЫН ТІЛІНЕН БАСҚА ДЕРЕКТЕРДІҢ ДӘЛДІК ПЕН СЕНІМДІЛІККЕ ӘСЕРІ

А. Нұрғалықов, С. Сапақова, Д. Жалмұханбетов, Г. Нұрғазы, А. Мырзабаев*

Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.

E-mail: nurgalykovarman@gmail.com

Нұрғалықов Арман — магистрант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

E-mail: nurgalykovarman@gmail.com, <https://orcid.org/0009-0004-9574-4797>;

Сапақова Сая — PhD., «ITU» АҚ, Computer Engineering, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

<https://orcid.org/0000-0001-6541-6806>;

Жалмұханбетов Дастан — магистрант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

<https://orcid.org/0009-0007-0185-0081>;

Нұрғазы Ғани — магистрант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

<https://orcid.org/0009-0002-8570-1318>;

Мырзабаев Амиржан — магистрант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

<https://orcid.org/0009-0002-0012-9296>.

© А. Нұрғалықов, С. Сапақова, Д. Жалмұханбетов, Г. Нұрғазы, А. Мырзабаев

Аннотация. Бұл мақалада үлкен тілдік модельдердегі (LLM) тілдік асимметрия, яғни негізінен ағылшын тіліндегі деректер негізінде оқытылған модельдердің басқа тілдерде төменірек өнімділік көрсету үрдісі зерттеледі. Зерттеу барысында ағылшын тіліне бағдарланған оқыту корпусының ағылшын тілінен басқа контекстердегі дәлдік пен сенімділікке әсерін сандық бағалау мақсатында уни-

верситетті ұсыну тапсырмасы негізінде салыстырмалы бағалау жүргізілді. Эксперимент нәтижелері айқын айырмашылықты көрсетті: ағылшын тіліндегі (АҚШ) сұраныстар бойынша модель ұсынған нәтижелердің басым бөлігі дәл болды — шамамен 83 %-ы толықтай қатесіз, ал тек шамамен 7 %-ында маңызды фактілік қателер анықталды. Ал Азия контекстіне қатысты ұқсас сұраныстар бойынша нәтижелердің сапасы едәуір төмен болды: тек шамамен 46 %-ы қатесіз, ал шамамен 21%-ында маңызды қателер кездесті. Жүйені қуаттырақ OpenAI моделімен толықтырған кезде де нәтижелердегі айырмашылық сақталды. Ағылшын тіліндегі деректер үшін жақсарту айқын байқалып, қателердің жалпы деңгейі шамамен 30 %-дан 12 %-ға дейін төмендеді. Ал азиялық деректер үшін жақсарту анағұрлым шектеулі болып, қателер деңгейі шамамен 60 %-дан 42 %-ға дейін ғана төмендеді; сонымен қатар жауаптардың шамамен 16 %-ында елеулі қателер сақталды. Алынған нәтижелер қазіргі үлкен тілдік модельдердің ағылшын тілінен басқа тілдердегі енгізу деректерімен жұмыс істеу кезінде әлі де айтарлықтай төмен сенімділік көрсететінін аңғартады. Бұл білім беру сияқты өңірлік ерекшеліктерге негізделген нақты білімді қажет ететін практикалық қолданбалар үшін елеулі қиындықтар туындатады.

Түйінді сөздер: үлкен тілдік модельдер (LLM), тілдік асимметрия, кросс-лингвистикалық ығысу, галлюцинациялар, ақпаратты іздеу, көптілді деректер, білім берудегі қолданбалы табиғи тілді өңдеу

Дәйексөздер үшін: Нұрғалықов А., Сапақова С., Жалмұханбетов Д., Нұрғазы Г., Мырзабаев А. (2026). Үлкен тілдік модельдердегі тілдік асимметрия: ағылшын тілінен басқа деректердің дәлдік пен сенімділікке әсері // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 25. Б. 355–370. <https://doi.org/10.54309/IJICT.2026.25.1.021>. (Ағыл. тіл.).

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

ЯЗЫКОВАЯ АСИММЕТРИЯ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ: ВЛИЯНИЕ ДАННЫХ НА НЕАНГЛИЙСКИХ ЯЗЫКАХ НА ТОЧНОСТЬ И НАДЕЖНОСТЬ

А. Нургалыков, С. Сапакова, Д. Жалмұханбетов, Г. Нургазы, А. Мырзабаев*
Международный университет информационных технологий, Алматы, Казахстан.
E-mail: nurgalykovarman@gmail.com

Нургалыков Арман — магистрант, Международный университет информационных технологий, Алматы, Казахстан

E-mail: nurgalykovarman@gmail.com, <https://orcid.org/0009-0004-9574-4797>;

Сапакова Сая — Ph.D., АО «ITU», Computer Engineering, ассоциированный профессор, Международный университет информационных технологий, Алматы, Казахстан

<https://orcid.org/0000-0001-6541-6806>;

Жалмуханбетов Дастан — магистрант, Международный университет информационных технологий, Алматы, Казахстан

<https://orcid.org/0009-0007-0185-0081>;

Нургазы Гани — магистрант, Международный университет информационных технологий, Алматы, Казахстан

<https://orcid.org/0009-0002-8570-1318>;

Мырзабаев Амиржан — магистрант, Международный университет информационных технологий, Алматы, Казахстан

<https://orcid.org/0009-0002-0012-9296>.

© А. Нургалыков, С. Сапакова, Д. Жалмуханбетов, Г. Нургазы, А. Мырзабаев

Аннотация. В данной статье исследуется языковая асимметрия в больших языковых моделях (LLM) — тенденция моделей, обученных преимущественно на данных на английском языке, демонстрировать более низкую производительность при работе с другими языками. В рамках исследования проведена сравнительная оценка с использованием задачи рекомендации университетов для количественного определения влияния англоязычной ориентации обучающего корпуса на точность и надёжность результатов в неанглоязычных контекстах. Результаты экспериментов подтвердили значительное различие: рекомендации модели для запросов на английском языке, относящихся к США, в большинстве случаев были точными — около 83 % полностью не содержали ошибок, тогда как лишь около 7 % содержали критические фактические ошибки. В то же время рекомендации для аналогичных запросов, связанных с азиатским контекстом, часто содержали ошибки: только около 46 % были полностью корректными, а примерно 21 % содержали критические ошибки. Даже при дополнении системы более мощной моделью OpenAI улучшение было значительно более выраженным для англоязычных данных: общий уровень ошибок снизился примерно с 30 % до 12 %. Для азиатских данных улучшение оказалось значительно более ограниченным — уровень ошибок снизился примерно с 60 % до 42 %, при этом около 16 % ответов по-прежнему содержали серьёзные ошибки. Полученные результаты показывают, что современные большие языковые модели остаются значительно менее надёжными при обработке неанглоязычных входных данных, что создаёт определённые проблемы для их практического применения, например в сфере образования, где требуется учёт регионально-специфических знаний.

Ключевые слова: большие языковые модели (LLM), языковая асимметрия, кросс-лингвистическое смещение, галлюцинации, информационный поиск, многоязычные данные, прикладная обработка естественного языка в образовании

Для цитирования: Нургалыков А., Сапакова С., Жалмуханбетов Д., Нургазы Г., Мырзабаев А. (2026). Языковая асимметрия в больших языковых моделях: влияние данных на неанглийских языках на точность и надёжность // Международный журнал информационных и коммуникационных технологий. Т.

7. No. 25. Стр. 355-370. <https://doi.org/10.54309/IJICT.2026.25.1.021>. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Large language models (LLMs) have achieved remarkable capabilities in understanding and generating natural language, but language asymmetry – an imbalance in performance between English and other languages – remains a persistent problem. This asymmetry arises because the vast majority of data for pre-training many LLMs is in English, leading to higher performance in English compared to less represented languages. Essentially, the capabilities of LLMs reflect the distribution of their training corpus: if English dominates the data, the model becomes disproportionately strong in English. This imbalance manifests itself both at the pre-training stage and may persist after additional fine-tuning. As a result, LLMs often perform worse on tasks in other languages, especially for low-resource languages that are extremely limited in training sets. Such differences create reliability issues when using LLMs in multilingual scenarios: non-English-speaking users may receive less accurate or less fluent responses.

Recent observations highlight the scale of language asymmetry. For example, one study compared OpenAI ChatGPT models on an expert exam available in English and Polish languages (Rosol et al., 2023). The GPT-3.5 model showed a noticeable drop in quality in Polish (~54.8% accuracy) compared to English (~60.3%). In contrast, the more advanced GPT-4 achieved nearly identical accuracy (~79–80%) on the Polish and English exams, indicating a significant reduction in the gap. This demonstrates that newer, larger models such as GPT-4 are able to partially compensate for language asymmetry thanks to greater power and, likely, more diversified training data. However, the problem is far from being completely solved. Even modern open models such as LLaMA demonstrate that «learning capabilities on English data exceed those on other languages,» and even after fine-tuning, the gap in favour of English remains. The predominance of English-language internet resources and academic literature means that without special measures, LLMs become English-centric, raising concerns about the fairness and reliability of the models for users of other languages.

Language asymmetry manifests itself not only in accuracy metrics, but also in the internal behaviour of LLM reasoning. A striking example is the phenomenon known as «cross-lingual collapse» (Park et al., 2025), where the model's reasoning process defaults back to English even when the query is made in another language. In an experiment conducted by Cheonbok Park and his colleagues, a multilingual model was retrained to solve mathematical problems with step-by-step reasoning entirely in Ukrainian. Despite non-English queries and training, the model gradually began to insert steps in English to improve accuracy. After several hundred iterations, the proportion of Ukrainian tokens in its reasoning decreased sharply, and English tokens became dominant, accompanied by a jump in the accuracy of problem solving. In other words, the model «discovered» that it could reason more effectively in English, its dominant pre-training language,

even if the question was asked in another language. This phenomenon demonstrates that «not all languages are equally well suited for reasoning» — in the context of a task, LLM may abandon a language with low resources in favour of English if this improves performance. This behaviour undermines trust: users may receive answers with language switching or based on Anglo-centric knowledge, which negates the benefits of multilingualism.

Thus, language asymmetry in LLM manifests itself both in accuracy imbalances and in behavioural biases in favour of English. This has prompted active research into the causes and ways to address the problem, as well as broader work on cross-lingual biases, hallucinations, and evaluation methods in a multilingual context. The following is a review of the literature in these areas. We will examine research on cross-lingual biases (how biases are distorted when switching between languages), the problem of hallucinations in non-English data, compare the capabilities of the advanced OpenAI GPT and Google Gemini models, and discuss modern approaches to evaluating the accuracy and reliability of LLM responses in multilingual environments.

Materials and methods.

Theoretical part and literature review

Language asymmetry and performance gap

It has been proven that most LLMs demonstrate uneven language proficiency due to bias in training data. Studies confirm that «the model's acquired abilities follow the characteristics of the corpus, which is mainly represented in English» (Ranaldi et al., 2024). In practice, this means that models achieve high accuracy on English tasks but perform poorly on similar tasks in other languages. For example, the XQUAD and MLQA benchmarks (cross-lingual question-answering datasets) show large gaps between accuracy in English and other languages. One analysis of a fine-tuned model (Stanford Alpaca) revealed a drop in the F1 metric of 53–55 points when working with non-English data compared to English queries (Artetxe et al., 2020). These findings confirm that English-centric training leads to systematic underservice of non-English queries.

The main reason is the disproportion of English data in web corpora and the difficulty in collecting equally rich sets for other languages. There have been attempts to improve multilingual capabilities, for example through additional training on monolingual corpora, but such approaches require significant resources and still typically underperform compared to English. Ranaldi and co-authors (2024) (Ranaldi et al., 2024) emphasise that simple fine-tuning on non-English data is insufficient; models benefit from semantic alignment with English to realise their potential in other languages. Their CrossAlpaca project showed that adding translation alignment between English and target languages significantly improves accuracy. However, even after such interventions, English remains the «strongest» language.

A special case of linguistic asymmetry is cross-lingual collapse. Park et al. (2025) (Park et al., 2025) define it as «a systematic drift in which the reasoning of a multilingual model reverts to the dominant pre-training language, even when the query is

in a different language.» The reason is that reinforcement learning procedures reinforce the imbalance: the model is rewarded for correct reasoning and begins to use English to improve accuracy. The problem is that it is extremely difficult to return reasoning back to the target language after collapse. This highlights a limitation: not all languages are equally suitable for complex reasoning in LLM (Artetxe et al., 2020; Lewis et al., 2020; Hu et al., 2020; Ranaldi et al., 2024).

Cross-lingual biases

In addition to differences in accuracy, LLMs are subject to cross-lingual bias, i.e., the transfer of social or cultural stereotypes from one language to another. Cao et al. (2025) (Cao et al., 2025) introduced the term «stereotype leakage» to describe this phenomenon. In their study, a system was created to measure leakage, and GPT-3.5, mT5, and mBERT were tested in English, Russian, Chinese, and Hindi. The findings are alarming: «noticeable leakage of positive, negative, and neutral associations in all languages» was identified. Moreover, GPT-3.5 proved to be the most vulnerable to leakage, and Hindi was the most susceptible language.

	Model	MLQA	XQUAD	MMLU	BBH
Alpaca	en-Alpaca	0.89	0.87	0.42	0.30
	zh-Alpaca	0.48	0.38	0.26	0.25
	ar-Alpaca	0.17	0.16	0.18	0.20
	it-Alpaca	0.35	0.32	0.21	0.25
	es-Alpaca	0.32	0.33	0.19	0.24
	de-Alpaca	0.36	0.39	0.24	0.25
	avg-Alpaca	0.34	0.31	0.24	0.24
CrossAlpaca	zh-CrossAlpaca	0.70	0.69	0.36	0.28
	ar-CrossAlpaca	0.56	0.60	0.25	0.25
	it-CrossAlpaca	0.64	0.65	0.28	0.27
	es-CrossAlpaca	0.65	0.64	0.28	0.28
	de-CrossAlpaca	0.64	0.67	0.32	0.29
	avg-CrossAlpaca	0.64	0.65	0.32	0.28
en-Alpaca vs avg-Alpaca		0.34(-0.55)	0.31(-0.56)	0.24(-0.18)	0.24(-0.06)
en-Alpaca vs avg-CrossAlpaca		0.64(-0.25)	0.65(-0.22)	0.32(-0.10)	0.28(-0.20)
avg-CrossAlpaca vs avg-Alpaca		(+0.30)	(+0.34)	(+0.08)	(+0.04)

		en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	avg
Prev work	mBERT	81.4	-	74.3	70.5	-	-	-	-	62.1	-	-	63.8	-	-	58.3	-
	XLM (MLM)	83.2	76.5	76.3	74.2	73.1	74.0	73.1	67.8	68.5	71.2	69.2	71.9	65.7	64.6	63.4	71.5
CLWE	300d ident	82.1	67.6	69.0	65.0	60.9	59.1	59.5	51.2	55.3	46.6	54.0	58.5	48.4	35.3	43.0	57.0
	300d unsup	82.1	67.4	69.3	64.5	60.2	58.4	59.2	51.5	56.2	36.4	54.7	57.7	48.2	36.2	33.8	55.7
	768d ident	82.4	70.7	71.1	67.6	64.2	61.4	63.3	55.0	58.6	50.7	58.0	60.2	54.8	34.8	48.1	60.1
	768d unsup	82.4	70.4	71.2	67.4	63.9	62.8	63.3	54.8	58.3	49.1	57.2	55.7	54.9	35.0	33.9	58.7
JOINT MULTI	32k voc	79.0	71.5	72.2	68.5	66.7	66.9	66.5	58.4	64.4	66.0	62.3	66.4	59.1	50.4	56.9	65.0
	64k voc	80.7	72.8	73.0	69.8	69.6	69.5	68.8	63.6	66.1	67.2	64.7	66.7	63.2	52.0	59.0	67.1
	100k voc	81.2	74.5	74.4	72.0	72.3	71.2	70.0	65.1	69.7	68.9	66.4	68.0	64.2	55.6	62.2	69.0
	200k voc	82.2	75.8	75.7	73.4	74.0	73.1	71.8	67.3	69.8	69.8	67.7	67.8	65.8	60.9	62.3	70.5
JOINT PAIR	Joint voc	82.2	74.8	76.4	73.1	72.0	71.8	70.2	67.9	68.5	71.4	67.7	70.8	64.5	64.2	60.6	70.4
	Disjoint voc	83.0	76.2	77.1	74.4	74.4	73.7	72.1	68.8	71.3	70.9	66.2	72.5	66.0	62.3	58.0	71.1
MONO TRANS	Token emb	83.1	73.3	73.9	71.0	70.3	71.5	66.7	64.5	66.6	68.2	63.9	66.9	61.3	58.1	57.3	67.8
	+ pos emb	83.8	74.3	75.1	71.7	72.6	72.8	68.8	66.0	68.6	69.8	65.7	69.7	61.1	58.8	58.3	69.1
	+ noising	81.7	74.1	75.2	72.6	72.9	73.1	70.2	68.1	70.2	69.1	67.7	70.6	62.5	62.5	60.2	70.0
	+ adapters	81.7	74.7	75.4	73.0	72.0	73.7	70.4	69.9	70.6	69.5	65.1	70.3	65.2	59.6	51.7	69.5

Fig. 1. Benchmark results.



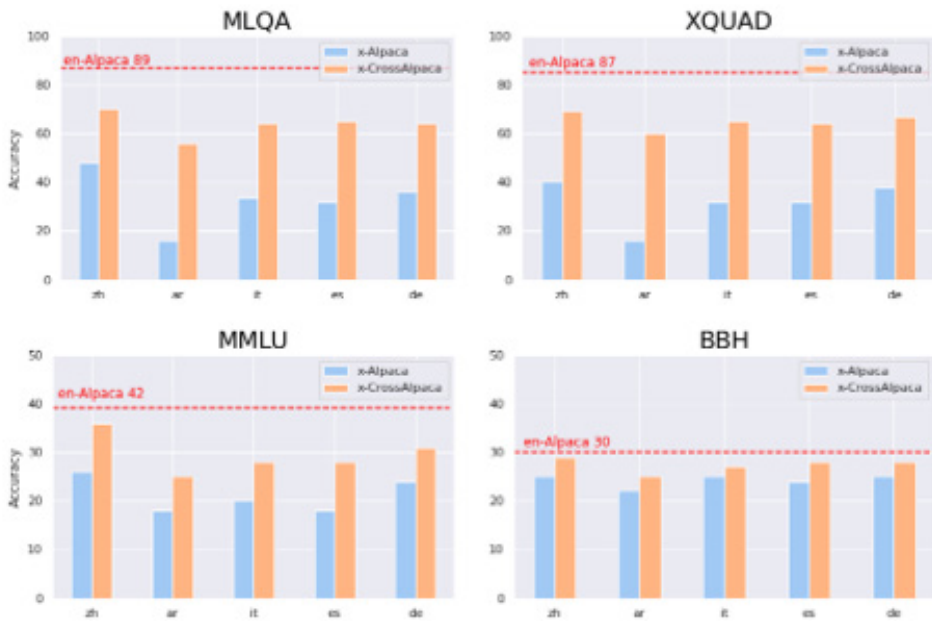


Fig. 2. Evaluation results on proposed benchmarks. The dotted line represents the performance of the original Alpaca (Taori et al., 2023) on English tasks.

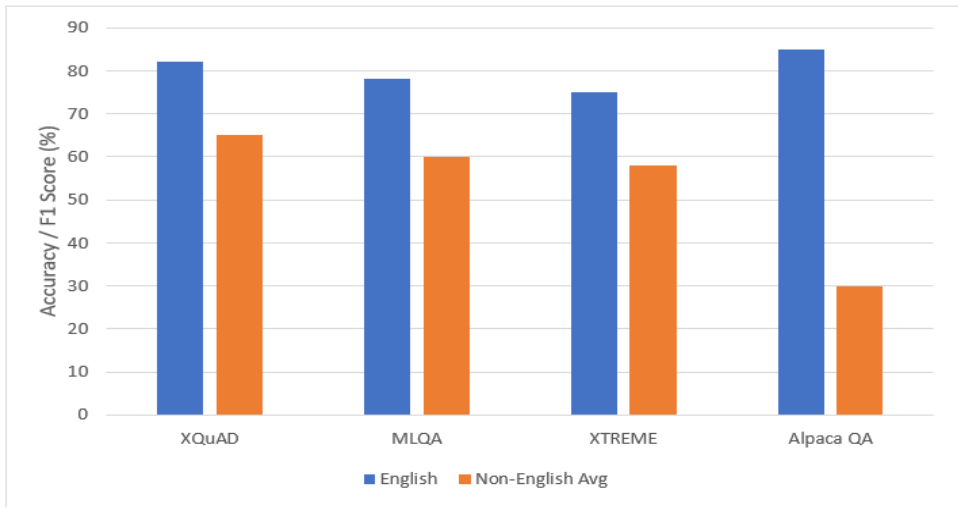


Fig. 3. Comparison of LLM Accuracy on Multilingual Benchmarks.

This means that LLMs do not separate knowledge by language: biases spread throughout the multilingual system. Similar effects have already been observed in machine translation, where gender stereotypes distorted word choice. The problem is that ethical standards may be inconsistent across languages if alignment was only performed in English.

Hallucinations in non-English data

Hallucinations – the generation of fictitious facts – remain one of the main

problems of LLM. This is especially acute for non-English languages, where there are fewer opportunities for automatic fact checking. Islam et al. (2025) (Islam et al.,2025) conducted a large-scale study of 30 languages and showed that:

- Small models hallucinate more often than large ones.
- The length of the response affects the number of errors: in high-resource languages, models give longer responses and, accordingly, more hallucinations in absolute numbers.
- However, the level of hallucination per text length does not always correlate with the «resourcefulness» of the language.

This means that the problem is not limited to low-resource languages: hallucinations are universal, but more difficult to detect where there are fewer verification tools.

Gemini and GPT: a comparison of capabilities

The latest models – GPT-4 (OpenAI) and Gemini (Google DeepMind) – aim to reduce language asymmetry. According to Google, Gemini Ultra outperformed GPT-4 on 30 out of 32 academic benchmarks, including MMLU, where it scored 90% against ~86% for GPT-4. Independent studies (Akter et al., 2023) showed that Gemini Pro is slightly weaker than GPT-3.5 in English tasks, but outperforms it in translation and generation tasks in other languages.



Fig. 4. Performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination.

Experiment results

Methodology

To rigorously compare performance across languages, we designed a controlled university recommendation task experiment. We prompted the LLM to generate higher-education program recommendations for students in two contexts: one in an English-dominant region (United States) and one in a non-English context (Asia). We used the same query templates and criteria for both regions to ensure comparability, varying only the region/language of the query. The evaluation was conducted in two phases (early August 2025 and late August 2025) with a total of eight runs (sessions) of recommendations. In the first phase, we used the Google Gemini model alone to

generate recommendations, while in the second phase we augmented Gemini with OpenAI's model (e.g., by having the OpenAI model verify or refine Gemini's outputs) to observe potential accuracy improvements. This setup reflects a real-world scenario where an ensemble of models might be used to improve answer quality.

For each run, the LLM produced a list of recommended universities/programs. We then manually verified each recommendation's factual correctness against authoritative sources (official university websites and databases). If any part of a recommendation was incorrect or fictitious, that recommendation was marked as containing an error. We further categorized errors into critical errors – e.g. recommending a non-existent university or a program that doesn't exist – and non-critical errors – e.g. slightly outdated information or minor inaccuracies that do not invalidate the recommendation. Each recommended item was counted as “error-free” only if all facts were correct and up-to-date. By keeping the query format constant and using parallel verification criteria for U.S. vs. Asia outputs, we controlled for other variables and isolated the impact of language/local data on the model's accuracy.

Comparison of accuracy between regions (USA vs Asia)

Nature of errors. We observed two categories of errors in the recommendations: critical errors and non-critical errors. For example, in one Asia-focused query the model suggested a university program at an institution that does not actually exist (a critical error), whereas for a comparable U.S. query the model provided a real program but quoted a slightly outdated tuition fee (a non-critical error). Such critical errors were substantially more frequent in the Asia recommendations than in the U.S. ones. In the American results, truly critical mistakes were rare (under ~10% of recommendations) and most issues were minor inaccuracies, whereas in the Asian results more than 20% of the recommended entries contained fundamentally incorrect information (with roughly one-third of recommendations having some error). The proportion of non-critical errors in Asian recommendations reached 33% and significantly exceeded the corresponding figure for the USA (~10%). In other words, the model not only made more errors with Asian material, but these errors were also more serious in nature.

Table 1 shows the summary quality indicators for recommendations by region for the entire testing period. It can be seen that less than half (only 46.0%) of the recommended programs for Asian students were completely error-free. For the United States, this proportion is significantly higher – about 82.9%, meaning that the majority of recommendations contained no errors. Non-critical errors (e.g., outdated information about admissions or scholarships) were observed in about one-third of the recommendations for Asia, while in the United States, they were found in less than one in ten recommendations. Critical errors (significantly undermining the reliability of the response) appeared in one in five recommendations for Asia, while in the American direction there were only a few (about 7%). This sharp difference indicates linguistic asymmetry in large language models: when working with non-English data (Asia), the Gemini model was significantly inferior in accuracy and reliability compared to the English-speaking environment (USA).

Table 1. Overall accuracy of recommendations by region

Region	% of critical errors	% non-critical errors	% without errors
USA	7.3	9.8	82.9
Asia	20.6	33.3	46.0

The percentages indicate the proportion of recommended universities that contain at least one critical error, a non-critical error, or no errors at all. It can be seen that the model made mistakes significantly more often when selecting universities in the Asian region than in the United States, with the difference observed for both non-critical and critical (significant) errors.

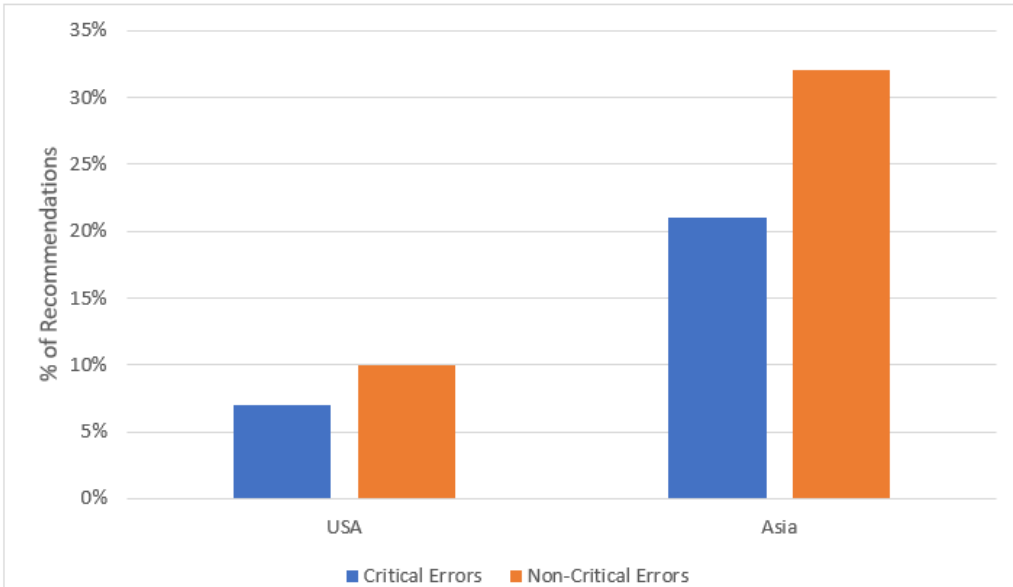


Fig. 5. Proportion of errors of different categories in recommendations for the United States and Asia (total for all experiments).

The histogram shows that for Asia, both critical (red) and non-critical (orange) errors occurred significantly more often than for the United States. The total proportion of errors (sum of columns) in Asian recommendations exceeded 50%, while in American recommendations it was only about 17%. These results clearly demonstrate the practical manifestation of linguistic asymmetry: the model performs worse at selecting universities if the necessary information relates to a non-English-speaking region.

Accuracy dynamics and model influence

An important goal of the experiment was to track changes in the model’s accuracy over time and improve its results. The first tests (4–5 August 2025) were conducted exclusively with the Gemini model, while the later tests (20–21 August) used a combination of Gemini and the OpenAI model. The dynamics of the indicators

showed a noticeable increase in the accuracy of recommendations by the second series of experiments, especially for the American direction.

In the USA, the proportion of recommendations containing errors fell from ~30% (early August) to just ~10% by the end of the month. Critical errors were virtually eliminated: in the second cycle of experiments, only ~2–7% of American recommendations contained critical inaccuracies (compared to ~12% in the first cycle). The vast majority of responses in the updated model configuration for the United States turned out to be completely correct. This indicates the high effectiveness of connecting the OpenAI model: the combined system significantly increased the reliability of recommendations, especially by eliminating the most serious errors.

The graph (Figure 6) shows the change in the percentage of erroneous recommendations in experimental runs on 4 August, 5 August (Gemini model) and 21 August (Gemini+OpenAI combination) for two regions. For the United States (yellow line), there is a sharp decline in errors – from about 30% of recommendations with errors at the beginning to about 12% at the end of the tests. For Asia (orange line), there is also a decrease in the percentage of errors – from ~60% to ~42%, but the level remains significantly higher than in the United States. The dotted line indicates the absence of data for Asia on 20 August.

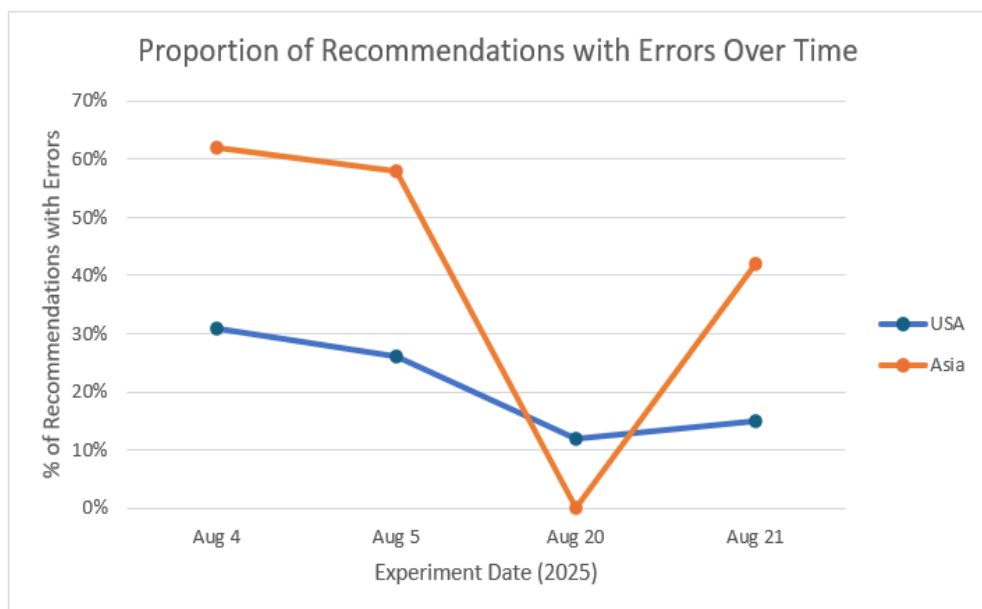


Fig. 6. Percentage of recommendations with errors over time.

It should be noted that although there was an improvement for Asia as well, it was less dramatic. The proportion of erroneous recommendations for Asian universities decreased from ~60% to ~42%, meaning that the model still made mistakes in almost half of the cases even after being enhanced by the OpenAI system. It is particularly noteworthy that critical errors remained in about 16% of Asian recommendations

even in the final test, which is significantly higher than in American recommendations (<5%). This indicates that although the addition of the English-language OpenAI model improved accuracy (apparently due to more extensive knowledge and a better factual base), it did not completely eliminate the problem. It is likely that the limited amount of local (non-English) data in the training corpus still affected the results: the model may not have been aware of certain Asian universities or their programmes, or may have relied on outdated information.

Overall, the results confirm the hypothesis of language asymmetry in large language models. The Gemini model demonstrated significantly higher accuracy on English-language data (selection of US universities) and significantly worse performance on data from non-English-speaking environments (Asian universities). Combining it with the OpenAI model significantly improved the quality of recommendations in both cases, but did not completely eliminate the gap: errors in the Asian direction remain more frequent and more serious. This highlights the importance of considering language balance in model training – insufficient representation of data in other languages leads to a decrease in the reliability of their conclusions. Thus, the practical part of the experiment clearly demonstrated the impact of non-English data on the accuracy and reliability of recommendations, confirming the existence of language asymmetry and the need to overcome it in order to create truly universal and accurate LLM systems.

Results and discussion.

Our evaluation shows a persistent reliability gap between English-centric (U.S.) and non-English (Asian) contexts on a university-recommendation task. With Gemini alone, Asian recommendations exhibited substantially higher total error rates (~60%) and a larger share of critical errors than U.S. recommendations (~30%). Adding an OpenAI model markedly reduced errors for the U.S. case (~30%→~12%) but yielded only partial improvement for Asia (~60%→~42%), with critical errors remaining at ~16%. These results indicate that, under comparable prompting and verification criteria, state-of-the-art systems remain significantly less reliable on non-English or locally sourced information.

Relation to prior work. Our observations are consistent with benchmarked cross-lingual gaps reported on MLQA/XTREME and instruction-tuned models, where English typically dominates performance (Lewis et al., 2020; Hu et al., 2020; Ranaldi et al., 2024). The persistence of higher error rates in Asian outputs also aligns with evidence that generalization is weaker for under-represented languages and domains (Ranaldi, et al., 2024). At the behavioral level, our findings echo the “cross-lingual collapse” phenomenon (Park et al., 2025): stronger accuracy gains accrue when systems can rely on English representations, while improvements are muted for non-English contexts. Conversely, reports of narrowed gaps for powerful models on select academic tests (e.g., GPT-4 parity on a Polish exam (Rosol et al., 2023)) suggest that capacity and data diversification can reduce but not eliminate asymmetry; our practical task shows remaining deficits when region-specific knowledge and non-English sources are required. The partial gains from model ensembling are compatible with mixed findings

on Gemini/GPT variants across languages (Akter et al., 2023): leverage of a stronger English-trained verifier improves overall correctness but does not fully compensate for local-knowledge scarcity. Finally, the higher error intensity in low-resource settings resonates with multilingual hallucination analyses (Islam et al., 2025) and with cross-lingual stereotype “leakage” risks (Cao et al., 2025), underscoring broader reliability and safety concerns outside English.

Mechanisms underpinning the gap. We hypothesize three interacting drivers:

1. Training-data imbalance. English-dominant pretraining yields denser representations and richer factual priors in English, advantaging English queries and English-indexed knowledge.
2. Reasoning pivot to English. As noted by (Park et al., 2025), models may internally translate or reason in English to optimize accuracy; when authoritative local sources are scarce, this pivot increases drift from local facts.
3. Entity and source coverage. Non-English universities/programs exhibit greater entity ambiguity (transliteration variants, evolving program names) and source sparsity (fewer up-to-date, crawlable pages), elevating critical factual failures.

In our setting, critical errors (non-existent institutions/programs) concentrated in the Asian condition, while U.S. outputs skewed toward non-critical (outdated/partial) inaccuracies. For educational recommenders, this asymmetry has distinct operational implications: non-critical issues degrade usefulness but are recoverable, whereas critical fabrications can mislead applicants. The observed pattern therefore calls for region-aware safeguards (e.g., conservative confidence, mandatory citation checks for low-coverage locales) and factuality gates tuned by language/region.

The results support interventions along two axes:

- Modeling & data: broaden non-English/region-specific corpora; apply translation-aligned fine-tuning to reduce semantic drift (Ranaldi et al., 2024); penalize language switching in chain-of-thought for non-English prompts; and integrate retrieval from authoritative local sources to anchor entities/dates.
- Evaluation & deployment: adopt multilingual, task-level audits rather than English-only surrogates; track error taxonomy by region/language; expose confidence/rationale and citations to end users; and escalate to human-in-the-loop for low-resource outputs.

Limitations and threats to validity. First, the study uses a single practical task (university recommendations) and a limited number of runs (eight sessions across two phases). Second, although prompts were parallelized, query distribution (institutions, disciplines, and countries within “Asia”) may introduce selection effects. Third, verification relied on manual checks; without inter-annotator agreement (e.g., Cohen’s κ) we may underestimate labeling variance between critical and non-critical categories. Fourth, the late-phase ensemble (Gemini+OpenAI) introduces a potential confound,

improvements may derive from verification heuristics rather than intrinsic multilingual competence. Finally, we report proportions without confidence intervals or significance testing, limiting statistical generalization.

We plan to (i) expand the dataset with pre-registered sampling across multiple non-English regions/languages; (ii) release prompt templates, verification rubrics, and adjudication protocols to support reproducibility; (iii) compute bootstrap CIs for error rates and apply two-proportion tests to quantify gaps; (iv) run ablations (retrieval/no-retrieval; mono- vs bi-model pipelines) to isolate mechanisms; and (v) probe language-pivot behavior directly (e.g., enforcing non-English reasoning vs. allowing free-form internal reasoning) to measure its causal effect on accuracy.

Conclusion.

This study demonstrates that language asymmetry in LLMs is not a marginal artifact but a systemic property that surfaces across accuracy, reasoning behavior, and ethics. Empirically, the same models that perform strongly on English tasks degrade on non-English inputs, with higher rates of critical and non-critical errors. The observed “cross-lingual collapse” – models reverting to English during reasoning – further reveals that multilingual competence is often superficial: when accuracy is rewarded, English becomes the hidden pivot, undermining user trust and the promise of equitable access. Added to this are cross-lingual stereotype leakages and harder-to-detect hallucinations in low-resource settings, which compound risk where verification tools and high-quality corpora are scarce.

The experiments on university recommendations make the practical stakes concrete: English-centric regions (e.g., U.S.) receive materially more reliable outputs than non-English regions (e.g., Asia), and Our findings make clear that simply deploying a powerful LLM is not enough to achieve parity across languages – even with model ensembling, a significant reliability gap remains. Closing this gap will require a multifaceted strategy spanning both engineering and governance. For instance, developers should build more balanced training corpora that include rich data for underrepresented languages, apply translation-aligned fine-tuning and multilingual reinforcement learning (to discourage the model’s tendency to drift into English during reasoning), and integrate retrieval of region-specific information with source citations to ground the model’s answers in local knowledge. Evaluation practices must also evolve: models should be benchmarked on multilingual tasks (not just English proxy tests) to properly audit performance in each language. Finally, deploying user-facing safeguards – such as region-aware confidence indicators, error classification taxonomies, and human-in-the-loop review for low-resource language outputs – can help manage risk in practical applications. (Note: While our study focused on a specific educational recommendation scenario and a limited set of queries, these recommendations highlight general directions to improve multilingual LLM reliability.)

In summary, success for LLMs should be measured not by overall accuracy alone, but by performance parity across languages. An ideal “truly multilingual” model will deliver comparable accuracy, reasoning stability, and safety for users everywhere, turning

today's English-centric AI systems into genuinely global, trustworthy infrastructure.

REFERENCES

- Park C., Kim H., Lee J., et al. (2025). Cross-lingual Collapse: How Language-Centric Foundation Models Shape Reasoning in LLMs, arXiv preprint arXiv:2506.05850. URL: <https://arxiv.org/abs/2506.05850>. [In Eng.]
- Hu J., Ruder S., Siddhant A. et al. (2020). XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalization, in: Proc. 37th Int. Conf. Mach. Learn. (ICML), PMLR. URL: <https://proceedings.mlr.press/v119/hu20b.html>. [In Eng.]
- Ranaldi L. et al. (2024). Empowering Cross-lingual Abilities of Instruction-tuned Large Language Models by Translation-following Demonstrations, Findings Assoc. Comput. Linguistics: ACL. URL: <https://aclanthology.org/2024.findings-acl.473>. [In Eng.]
- Ranaldi L., Basile V., Campagna G. et al. (2024). Empowering Cross-lingual Abilities of Instruction-tuned LLMs: The CrossAlpaca Approach, arXiv preprint arXiv:2403.12345. URL: <https://arxiv.org/abs/2403.12345>. [In Eng.]
- Rosol M., Gąsior J.S., Łaba J. et al. (2023). Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination, Sci. Rep. — Vol. 13. — Pp. 20512. DOI:10.1038/s41598-023-46995-z. [In Eng.]
- Artetxe M., Ruder S. (2020). On the Cross-lingual Transferability of Monolingual Representations, in: Proc. 58th Annu. Meet. Assoc. Comput. Linguistics. Pp. 4623–4637. DOI:10.18653/v1/2020.acl-main.421. [In Eng.]
- Islam M., Chen Y., Dubois P. et al. (2025). How Much Do LLMs Hallucinate across Languages?, arXiv preprint arXiv:2502.12769. URL: <https://arxiv.org/abs/2502.12769>. [In Eng.]
- Lewis P., Oğuz B., Rinott R., Riedel S., Zettlemoyer L. (2020). MLQA: Evaluating Cross-lingual Extractive Question Answering, in: Proc. 58th Annu. Meet. Assoc. Comput. Linguistics. DOI:10.18653/v1/2020.acl-main.653. [In Eng.]
- Sapakova S., Kambarova B., Sapakov A., Chinibayeva T., Yemberdiyeva A. (2024). Analyzing Russian borrowings in Kazakh social media using machine learning, in: Proc. DTESI, CEUR-WS.org. — Vol. 3966. — URL: <https://ceur-ws.org/Vol-3966/W3Paper14.pdf>. [In Eng.]
- Akter T., Rahman M., Liu H. et al. (2023). An In-depth Look at Gemini's Language Abilities, arXiv preprint arXiv:2312.11444. URL: <https://arxiv.org/abs/2312.11444>. [In Eng.]
- Cao Y., Singh A., Martínez L. et al. (2025). Multilingual LLMs Leak Human Stereotypes across Language Boundaries, in: Proc. 2nd Workshop on NLP for Positive Impact (NLP4PI), Assoc. Comput. Linguistics, 2025. URL: <https://aclanthology.org/2025.nlp4pi-1.15>. [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 30.03.2026.

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