

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

**ACADEMIC SCIENTIFIC
JOURNAL OF COMPUTER SCIENCE**

**№2
2026**

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



ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

3 (359)

JULY – SEPTEMBER 2026

**PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR**

ALMATY, NAS RK

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

MAMYRBAEV Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

BAIGUNCHEKOV Zhumadil Zhanabaevich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

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

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

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

Certificate № KZ77VPY00121154 on the re-registration of the periodical printed and online publication of the information agency, issued on 05.06.2025 by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan

Subject area: *information and communication technologies*.

Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies»*.

Periodicity: 4 times a year.

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

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

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

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жұмаділ Жаңабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохаммед, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

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

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025 ж.** берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

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

Мерзімділігі: *жылына 4 рет.*

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

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимканир Мутанович, доктор технических наук, профессор, академик НАН РК, (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сатпаева (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

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

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

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

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

Тематическая направленность: *информационно-коммуникационные технологии.*

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

Периодичность: *4 раза в год.*

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

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

CONTENTS

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

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

МАЗМҰНЫ

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

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

СОДЕРЖАНИЕ

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

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

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

ISSN 1991-346X

Volume 3.

Number 359 (2026). 346–364

<https://doi.org/10.32014/2026.2518-1726.467>

IRSTI 28.23.00; 28.23.37; 28.23.25

UDC 004.912; 004.93; 004.8; 004.5; 004.6

© **Oralbekova D.¹, Bekarystankyzy A.¹, Yerimbetova A.^{1,2},
Berzhanova U.^{1,3,*}, 2026.**

¹Institute of Information and Computational Technologies of the Committee
of Science of the Ministry of Science and Higher Education of the Republic of
Kazakhstan, Almaty, Kazakhstan;

²META University, Almaty, Kazakhstan;

³Farabi University, Almaty, Kazakhstan.

E-mail: berzhanovaulmekenn@gmail.com

DEVELOPMENT OF A STREAMING KAZAKH SPEECH RECOGNITION SYSTEM USING END-TO-END MODELS

Oralbekova Dina — PhD, Leading Researcher, Institute of Information and Computational
Technologies, Almaty, Kazakhstan,

E-mail: dinaoral@mail.ru, ORCID ID: <https://orcid.org/0000-0003-4975-6493>;

Bekarystankyzy Akbayan — PhD, Senior Researcher of Institute of information and computational
technologies, Almaty, Kazakhstan,

E-mail: akbayan.b@gmail.com, ORCID ID: <https://orcid.org/0000-0003-3984-2718>;

Yerimbetova Aigerim — PhD, Candidate of Technical Science, Associate Professor, 2META
University; Leading Researcher of Institute of Information and Computational Technologies CS
MSHE RK, Almaty, Kazakhstan,

E-mail: aigerian8888@gmail.com, ORCID ID: <https://orcid.org/0000-0002-2013-1513>;

Berzhanova Ulmeken — doctoral student of the 2nd year of specialty 8D06101-Information
systems Al-Farabi Kazakh National University, Almaty, Kazakhstan,

E-mail: berzhanovaulmekenn@gmail.com, ORCID ID: <https://orcid.org/0009-0000-2467-5721>.

Abstract. This paper presents an overview and comparative analysis of online speech recognition architectures that incorporate attention mechanisms, including the Recurrent Neural Transducer, Neural Transducer, and Monotonic Chunkwise Attention. The models were adapted for the Kazakh language and designed to operate in real-time. Training was conducted on a corpus containing 200 hours of annotated Kazakh speech, which was further expanded to 260 hours using the Voice Conversion method. Mel-frequency cepstral coefficients were used for feature extraction, and the architecture was built using bidirectional recurrent neural networks. Experimental results demonstrated that the Monotonic Chunkwise Attention model achieved the highest accuracy (WER 14.8%), outperforming both

Recurrent Neural Transducer and standard Neural Transducer despite requiring more complex tuning. All three models showed high efficiency even under limited-resource conditions, confirming their applicability to low-resource languages. The tuning process is complex, but all three models demonstrated high performance even with limited resources. Experimental evaluation was conducted using two corpora of different sizes. Speech recognition accuracy improved as the dataset size increased. Comparative analysis showed that MoChA achieved reduced WER by 0.7 and 0.5 percentage points in absolute terms, outperforming RNN-T and NT respectively. It is worth noting that tuning MoChA requires additional time because its architecture involves numerous parameters and requires complex optimization. The results of the study demonstrate that the approaches discussed are effective for developing streaming speech recognition systems for resource-constrained languages such as Kazakh. Because these models do not require pre-recorded audio files and can operate in real time, their integration into various technology platforms and applications is anticipated.

Keywords: Speech recognition, monotonic chunkwise attention, recurrent neural transducer, integral models, neural transducer

For citations: Oralbekova D., Bekarystankyzy A., Yerimbetova A., Berzhanova U. Development of a streaming kazakh speech recognition system using end-to-end models. Academic Scientific Journal of Computer Science, 2026. — No.3. — P. 346–364. DOI: <https://doi.org/10.32014/2026.2518-1726.467>

© Оралбекова Д.О.¹, Бекарыстанқызы А.¹, Еримбетова А.С.^{1,2},
Бержанова У.Г.^{1,3*}, 2026.

¹ ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институты, Алматы, Қазақстан;

² META University, Алматы, Қазақстан;

³ Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.

E-mail: berzhanovaulmekenn@gmail.com

ИНТЕГРАЛДЫ МОДЕЛЬДЕРДІ ҚОЛДАНУ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ АҒЫНДЫҚ СӨЙЛЕУДІ ТАҢУ ЖҮЙЕСІН ӘЗІРЛЕУ

Оралбекова Дина — PhD, жетекші ғылыми қызметкер, Ақпараттық және есептеу технологиялары институты, Алматы, Қазақстан,

E-mail: dinaoral@mail.ru, <https://orcid.org/0000-0003-4975-6493>;

Бекарыстанқызы Ақбаян — PhD, Ақпараттық және есептеу технологиялары институтының аға ғылыми қызметкері, Алматы, Қазақстан,

E-mail: akbayan.b@gmail.com, <https://orcid.org/0000-0003-3984-2718>;

Еримбетова Айгерім — PhD, техника ғылымдарының кандидаты, қауымдастырылған профессор, META University; ҚР ҒЖБМ Ғылым комитетінің Ақпараттық және есептеу технологиялары институтының жетекші ғылыми қызметкері, Алматы, Қазақстан,

E-mail: aigerian8888@gmail.com, <https://orcid.org/0000-0002-2013-1513>;

Бержанова Үлмекен — Әл-Фараби атындағы Қазақ ұлттық университетінің 8D06101 «Ақпараттық жүйелер» білім беру бағдарламасының 2-курс докторанты; ҚР ҒЖБМ Ғылым комитетінің Ақпараттық және есептеу технологиялары институтының кіші ғылыми қызметкері, Алматы, Қазақстан,
E-mail: berzhanovaulmekenn@gmail.com, <https://orcid.org/0009-0000-2467-5721>.

Аннотация. Мақалада назар аудару механизмдеріне негізделген ағындық сөйлеуді тану архитектураларына шолу жасалып, салыстырмалы талдау ұсынылады. Зерттеуде рекуррентті нейрондық түрлендіргіш (RNN-T), нейрондық түрлендіргіш (NT) және бөліктер бойынша монотонды назар аудару (MoChA) модельдері қарастырылды. Модельдер қазақ тіліне бейімделіп, нақты уақыт режимінде жұмыс істеуге икемделді. Оқыту үшін 200 сағаттық аннотацияланған қазақша сөйлеу корпусы пайдаланылып, Voice Conversion әдісі арқылы оның көлемі 260 сағатқа дейін кеңейтілді. Акустикалық белгілерді бөліп алу үшін мел-жиіліктік кепстралдық коэффициенттер қолданылып, модель архитектураларында екібағытты рекуррентті нейрондық желілер пайдаланылды. Эксперименттік нәтижелер бөліктер бойынша монотонды назар аудару моделінің ең жоғары дәлдікке қол жеткізгенін көрсетті: WER = 14,8%. Бұл көрсеткіш рекуррентті нейрондық түрлендіргіш пен стандартты нейрондық түрлендіргіш нәтижелерінен жоғары болды, алайда MoChA моделін баптау күрделірек болды. Барлық үш модель ресурстары шектеулі жағдайда да жоғары тиімділік көрсетті. Эксперименттік бағалау көлемі әртүрлі екі корпус негізінде жүргізілді және деректер жиынтығы ұлғайған сайын сөйлеуді тану дәлдігінің артатыны анықталды. Салыстырмалы талдау MoChA моделінің WER көрсеткішін RNN-T және NT модельдерімен салыстырғанда тиісінше 0,7 және 0,5 пайыздық тармаққа төмендеткенін көрсетті. Сонымен қатар MoChA архитектурасында параметрлер саны көп және оңтайландыру үдерісі күрделі болғандықтан, оны баптауға қосымша уақыт қажет болды. Зерттеу нәтижелері қарастырылған тәсілдердің қазақ тілі сияқты ресурстары шектеулі тілдерге арналған ағындық сөйлеуді тану жүйелерін әзірлеуде тиімді екенін көрсетеді. Бұл модельдер алдын ала жазылған аудиофайлдарды қажет етпей, нақты уақыт режимінде жұмыс істей алатындықтан, оларды әртүрлі технологиялық платформалар мен қолданбаларға интеграциялау мүмкіндігі жоғары.

Түйін сөздер: сөйлеуді тану; бөліктер бойынша монотонды назар аудару; рекуррентті нейрондық түрлендіргіш; интегралды модельдер; нейрондық түрлендіргіш

© Оралбекова Д.О.¹, Бекарыстанкызы А.¹, Еримбетова А.С.^{1,2},
Бержанова У.Г.^{1,3*}, 2026.

¹ Институт информационных и вычислительных технологий Комитета науки
Министерства науки и высшего образования Республики Казахстан,

Алматы, Казахстан;

² META University, Алматы, Казахстан;

³ Казахский национальный университет имени аль-Фараби,
Алматы, Казахстан.

E-mail: berzhanovaulmekenn@gmail.com

РАСПОЗНАВАНИЕ КАЗАХСКОЙ РЕЧИ С ПРИМЕНЕНИЕМ ИНТЕГРАЛЬНЫХ МОДЕЛЕЙ

Оралбекова Дина — PhD, ведущий научный сотрудник, Институт информационных и вычислительных технологий, Алматы, Казахстан,

E-mail: dinaoral@mail.ru, <https://orcid.org/0000-0003-4975-6493>;

Бекарыстанкызы Акбаян — PhD, ведущий научный сотрудник Института информационных и вычислительных технологий Комитета науки Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан,

E-mail: akbayan.b@gmail.com, <https://orcid.org/0000-0003-3984-2718>;

Еримбетова Айгерим — PhD, кандидат технических наук, ассоциированный профессор, META University; ведущий научный сотрудник Института информационных и вычислительных технологий Комитета науки Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан,

E-mail: aigerian8888@gmail.com, <https://orcid.org/0000-0002-2013-1513>;

Бержанова Улмекен — докторант 2-го курса образовательной программы 8D06101 «Информационные системы» Казахского национального университета имени аль-Фараби; младший научный сотрудник Института информационных и вычислительных технологий Комитета науки Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан,

E-mail: berzhanovaulmekenn@gmail.com, <https://orcid.org/0009-0000-2467-5721>.

Аннотация. В статье представлен обзор и сравнительный анализ архитектур потокового автоматического распознавания речи, использующих механизмы внимания, включая рекуррентный нейронный преобразователь (RNN-T), нейронный преобразователь (NT) и монотонное внимание по фрагментам (MoChA). Модели адаптированы для казахского языка и предназначены для работы в режиме реального времени. Обучение проводилось на корпусе, содержащем 200 часов размеченной казахской речи, объем которого был увеличен до 260 часов с помощью метода Voice Conversion. Для извлечения акустических признаков использовались мел-частотные кепстральные коэффициенты, а в архитектурах моделей применялись двунаправленные рекуррентные нейронные сети. Экспериментальные результаты показали, что модель Monotonic Chunkwise Attention достигла наилучшей точности при WER = 14,8%, превзойдя RNN-T и стандартный Neural Transducer, несмотря на более высокую сложность настройки. Все три модели

продemonстрировали высокую эффективность даже в условиях ограниченных ресурсов. Экспериментальная оценка проводилась с использованием двух корпусов различного объема, при этом увеличение размера набора данных сопровождалось повышением точности распознавания речи. Сравнительный анализ показал, что MoChA снизила WER по сравнению с RNN-T и NT на 0,7 и 0,5 процентного пункта соответственно. Вместе с тем настройка MoChA требует дополнительного времени из-за большого числа параметров и более сложной процедуры оптимизации. Полученные результаты подтверждают эффективность рассмотренных подходов для разработки потоковых систем распознавания речи для языков с ограниченными ресурсами, включая казахский язык. Возможность работы в реальном времени без необходимости предварительной записи аудиофайлов создает предпосылки для интеграции таких моделей в различные технологические платформы и приложения.

Ключевые слова: распознавание речи; монотонное внимание по фрагментам; рекуррентный нейронный преобразователь; интегральные модели; нейронный преобразователь

©Оралбекова Д.О.¹, Бекарыстанқызы А.¹, Еримбетова А.С.^{1,2},
Бержанова У.Г.^{1,3,*}, 2026.

¹Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан;

²META University, Алматы, Қазақстан;

³Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.

E-mail: berzhanovaulmekenn@gmail.com

ИНТЕГРАЛДЫ МОДЕЛЬДЕРДІ ҚОЛДАНУ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ АҒЫНДЫҚ СӨЙЛЕУДІ ТАҢУ ЖҮЙЕСІН ӘЗІРЛЕУ

Оралбекова Дина — PhD, жетекші ғылыми қызметкер, Ақпараттық және есептеуіш технологиялары институты, Алматы, Қазақстан,

E-mail: dinaoral@mail.ru, ORCID ID: <https://orcid.org/0000-0003-4975-6493>;

Бекарыстанқызы Ақбаян — PhD, Ақпараттық және есептеуіш технологиялар институтының аға ғылыми қызметкері, Алматы, Қазақстан,

E-mail: akbayan.b@gmail.com, ORCID ID: <https://orcid.org/0000-0003-3984-2718>;

Еримбетова Айгерим — PhD, техн. ғылым. кандидаты, қауымдастырылған профессор, META University; ҚР ҒЖБМ ҒК Ақпараттық және есептеуіш технологиялар институтының жетекші ғылыми қызметкері, Алматы, Қазақстан,

E-mail: aigerian8888@gmail.com, ORCID ID: <https://orcid.org/0000-0002-2013-1513>;

Бержанова Улмекен — Әл-Фараби атындағы Қазақ ұлттық университетінің 8D06101-Ақпараттық жүйелер мамандығының 2-курс докторанты; ҚР ҒЖБМ ҒК Ақпараттық және есептеуіш технологиялар институтының кіші ғылыми қызметкері, Алматы, Қазақстан,
E-mail: berzhanovaulmekenn@gmail.com, ORCID ID: <https://orcid.org/0009-0000-2467-5721>.

Аннотация. Бұл мақалада назар аудару механизмдеріне негізделген ағынды сөйлеуді тануға арналған архитектураларға шолу жасалып, салыстырмалы талдау ұсынылды. Зерттеуде рекуррентті нейрондық түрлендіргіш, нейрондық түрлендіргіш және бөліктер бойынша монотонды назар аудару модельдері

қарастырылды. Бұл модельдер қазақ тіліне бейімделіп, нақты уақыт режимінде жұмыс істеуге икемделген. Модельдерді оқыту үшін 200 сағаттық аннотацияланған қазақша сөйлеу корпусы пайдаланылып, Voice Conversion әдісі арқылы ол 260 сағатқа дейін кеңейтілді. Акустикалық белгілерді шығару үшін мел-жиіліктік кепстрал коэффициенттері қолданылып, модель архитектурасында екібағытты рекуррентті нейрондық желілер пайдаланылды. Эксперименттік нәтижелер бойынша бөліктер бойынша монотонды назар аудару моделі ең жоғары дәлдікке (WER = 14.8%) қол жеткізіп, рекуррентті нейрондық және стандартты нейрондық түрлендіргіш модельдерінен басым түсті. Баптау үдерісінің күрделілігіне қарамастан, барлық үш модель шектеулі ресурстар жағдайында да жоғары тиімділік көрсетті. Тәжірибелік бағалау әртүрлі өлшемдегі 2 корпусы пайдалану арқылы жүргізілді. Деректер жиынтығының көлемі ұлғайған сайын, сөйлеуді тану дәлдігі жақсарды. Салыстырмалы талдау MoChA моделі WER көрсеткішін RNN-T және NT өнімділігінен сәйкесінше 0,7 және 0,5 пайыздық тармаққа төмендеткенін көрсетті. Дегенмен, MoChA-ны баптау әлдеқайда көп уақытты алды, себебі бұл архитектура көптеген параметрлерді қамтиды және күрделі оңтайландыруларды қажет етеді. Зерттеу нәтижелері қарастырылған тәсілдердің қазақ тілі сияқты ресурсы шектеулі тілдер үшін ағындық сөйлеуді тану жүйелерін әзірлеуде тиімді екенін көрсетеді. Бұл модельдер аудио файлдарды алдын ала жазуды қажет етпейтіндіктен және нақты уақыт режимінде жұмыс істей алатындықтан, оларды әртүрлі технологиялық платформалар мен қолданбаларға біріктіру күтілуде.

Түйін сөздер: сөйлеуді тану, монотонды бөліктік назар аудару, рекуррентті нейрондық түрлендіргіш, интегралды модельдер, нейрондық түрлендіргіш

Алғыс білдіру. Бұл зерттеу Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (Грант № BR24992875).

**Оралбекова Д.О.¹, Бекарыстанқызы А. ¹, Еримбетова А.С.^{1,2},
Бержанова У.Г.^{1,3,*}, 2026.**

¹Институт информационных и вычислительных технологий КН МНВО РК,
Алматы, Казахстан;

²META University, Алматы, Казахстан;

³Казахский Национальный университет им. аль-Фараби,
Алматы, Казахстан.

E-mail: berzhanovaulmekenn@gmail.com

РАСПОЗНАВАНИЯ КАЗАХСКОЙ РЕЧИ С ПРИМЕНЕНИЕМ ИНТЕГРАЛЬНЫХ МОДЕЛЕЙ

Оралбекова Дина — PhD, ведущий научный сотрудник, Институт информационных и вычислительных технологий, Алматы, Казахстан,

E-mail: dinaoral@mail.ru; ORCID ID: <https://orcid.org/0000-0003-4975-6493>;

Бекарыстанкызы Акбаян — PhD, ведущий научный сотрудник Института информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан,

E-mail: madina_jgtu@mail.ru; ORCID ID: <https://orcid.org/0000-0003-3984-2718>;

Еримбетова Айгерим — PhD, к.т.н., ассоциированный профессор, META University; ведущий научный сотрудник Института информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан,

E-mail: aigerian8888@gmail.com, ORCID ID: <https://orcid.org/0000-0002-2013-1513>;

Бержанова Улмекен — докторант 2-го курса по специальности 8D06101 - Информационные системы Казахского национального университета имени Аль-Фараби; младший научный сотрудник Института информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан,

E-mail: berzhanovaulmekenn@gmail.com; ORCID ID: <https://orcid.org/0009-0000-2467-5721>.

Аннотация. В данной статье представлен обзор и сравнительный анализ онлайн-архитектур автоматического распознавания речи, использующих механизмы внимания, такие как рекуррентный нейронный преобразователь, нейронный преобразователь и монотонное внимание по фрагментам. Модели адаптированы для казахского языка и предназначены для работы в режиме реального времени. Обучение проводилось на корпусе, содержащем 200 часов размеченной казахской речи, увеличенном до 260 часов с помощью метода Voice Conversion. Для извлечения признаков использовались мел-частотные кепстральные коэффициенты, а в архитектурах применялись двунаправленные рекуррентные нейронные сети. Эксперименты показали, что монотонное внимание по фрагментам достигает наилучших результатов (WER 14.8%), опережая рекуррентный нейронный и стандартный нейронный преобразователь, несмотря на более высокую сложность настройки. Сложность процесса настройки высока, но однако все три модели показали высокую эффективность даже при ограниченных ресурсах. Экспериментальная оценка проводилась с использованием двух корпусов разного размера. По мере увеличения размера набора данных точность распознавания речи улучшалась. Сравнительный анализ показал, что MoChA снизила WER, обходя RNN-T и NT в абсолютном выражении на 0,7 и 0,5 процентного пункта соответственно. Стоит отметить, что настройка MoChA занимает больше дополнительного времени, потому что данная архитектура включает множество параметров и нуждается в сложной оптимизации. Результаты исследования показывают, что рассмотренные подходы эффективны при разработке потоковых систем распознавания речи для языков с ограниченными ресурсами, таких как казахский. Поскольку эти модели не требуют предварительной записи аудиофайлов и могут работать в реальном времени, ожидается их интеграция в различные технологические платформы и приложения.

Ключевые слова: распознавание речи, монотонное внимание по фрагментам, рекуррентный нейронный преобразователь, интегральные модели, нейронный преобразователь

Кіріспе. Соңғы жылдары автоматты сөйлеуді тану технологиясы тез дамып, машиналық оқыту тапсырмалары арасында маңызды орын алды. Ертедегі автоматты сөйлеуді тану жүйелерінің негізгі мақсаты мүмкіндігі шектеулі адамдарға мәтінді қолмен енгізумен байланысты қиындықтарды жеңуге көмектесу болды. Дегенмен, бүгінгі таңда бұл технологиялар күнделікті өмірде дауыспен енгізу және дауыстық командалық хабарлама жасау сияқты күнделікті тапсырмалардың тиімділігін арттыру үшін кеңінен қолданылады.

Ақпараттық жүйелер мен есептеу технологияларының жетістіктері көптеген машиналық оқыту технологияларын, әсіресе сөйлеуді тануды жақсартуға әкелді. Сөйлеуді тану дегеніміз әртүрлі жүйелерді, бағдарламаларды және алгоритмдерді пайдаланатын машиналарға адам сөйлеуін тануға және түсінуге және оны мәтінге түрлендіруге мүмкіндік беретін технологияны білдіреді. Автоматты сөйлеуді тану жүйелері әртүрлі салаларда, соның ішінде автомобильдерді, үйлерді және тұрмыстық техниканы дауыспен басқаруда, сондай-ақ әртүрлі қолданбаларда, навигациялық жүйелерде және мүгедектерге арналған қызметтерде дауыспен енгізуде кеңінен қолданылды. Бұлар тек автоматты сөйлеуді тану жүйелерінің қалай қолданылатынының бірнеше мысалы ғана. Дәстүрлі сөйлеуді тану жүйесі әдетте үш негізгі тәуелсіз элементтен тұрады: дыбыстан контекстке тәуелді субфонема күйлерін болжауға арналған акустикалық модельдер, тілдік модельдер және фонемаларды сөздермен салыстыруға арналған лексикон. Мұндай жүйелер тәуелсіз оқытылатын әртүрлі элементтерден тұрады. Классикалық акустикалық модельді Гаусс үлестірімдері мен жасырын Марков модельдерінің қоспаларын пайдаланып оқытуға болады, ал тілдік модельдерді n-граммдарды пайдаланып оқытуға болады. Жасырын Марков модельдері сөйлеуді тануда ұзақ уақыт бойы кеңінен қолданылып келеді, негізгі технология ретінде қызмет етеді. НММ негізінен кадр деңгейінде динамикалық уақытты бұрмалау үшін қолданылады, ал Гаусс ықтималдық тығыздығының таралуының қоспалары әдетте бір айтылымға сәйкес келетін белгіленген, қысқа уақыт аралығындағы сигнал таралуын көрсету үшін қолданылады. Ұзақ уақыт бойы НММ-GMM моделі сөйлеуді танудың жалпы негізі болды. Жакында терең оқыту көптеген зерттеулерді және сөйлеуді танудың дамуын айтарлықтай жақсартты. Классикалық сөйлеуді тану жүйесінің сценарийінің әрбір элементінде жасанды нейрондық желілерді белсенді пайдалану оның тиімділігін арттырады, бұл көптеген ғылыми мақалаларда көрініс тапты. Терең оқыту технологияларының дамуымен терең нейрондық желілер акустикалық модельдеу үшін сөйлеуді тануда қолданыла бастады. DNN рөлі дәстүрлі GMM бақылау ықтималдығын ауыстыра отырып, НММ күйлерінің апостериорлық ықтималдығын есептеу болып табылады. Демек, Гаусс таралуының қоспасы бар НММ моделі бәсекеге қабілетті нәтижелерге қол жеткізетін және автоматты сөйлеуді танудың танымал тәсіліне айналатын терең нейрондық желісі бар НММ-ге айналады. Терең нейрондық желілер әдетте тиімді акустикалық модель алу

үшін пайдаланылды, ал рекуррентті нейрондық желілер және ұзақ мерзімді және қысқа мерзімді жад желілері тиісінше тілдік модельдер мен сөздік қорын құру үшін пайдаланылды. Конволюциялық нейрондық желілер сөйлеу сигналынан ерекшеліктерді алу үшін де қолданылды. Осылайша, көптеген жарияланған нәтижелер ұсынылған тәсілдің барлық заманауи сөйлеуді тану жүйелерінің ішіндегі ең жақсы өнімділікті көрсететінін көрсетеді. Сөйлеуді тану жүйелерінің барлық кезеңдерінде әртүрлі жасанды нейрондық желі архитектураларын қолданудың негізін құрайды.

Адамдар үшін ауызша сөздерді мәтінге айналдыру табиғи процесс, бірақ компьютерлік жүйелер үшін бұл әлі де күрделі міндет және үлкен есептеу ресурстарын қажет етеді. Соңғы жылдары терең оқыту мен жоғары өнімді есептеу технологияларының дамуы сөйлеуді тану жүйелерінің дәлдігі мен тиімділігін айтарлықтай жақсартты. Қазіргі уақытта ең танымал жүйелер интегралдық модельдер мен назар аудару механизмдеріне негізделген (Mamutbayev et al., 2022). Дегенмен, ресурстары аз тілдерді енгізу қиын, себебі олар үлкен көлемдегі деректерді қажет етеді (көбінесе мыңдаған сағаттық аудиожазба). Деректердің жетіспеушілігін деректерді кеңейту арқылы жоюға болады, бірақ олардың тиімділігі әрқашан көрсетілмеген (Wang et al., 2024).

Жақында машиналық оқытуды пайдаланатын интегралдық сөйлеуді тану әдістері кең тарала бастады. Мұндай жүйелерде модель тек бір ғана нейрондық желіні пайдаланып жүзеге асырылады. Модельдің интеграцияланған іске асырылуы көбінесе сөйлеуді тану жылдамдығы мен дәлдігі тұрғысынан жоғары өнімділікті көрсетеді. Көптеген зерттеулер интегралдық жүйенің өнімділігін жақсарту желіні оқыту үшін пайдаланылатын оқыту деректерінің көлемін арттыруға байланысты екенін көрсетті. Қазіргі уақытта VTM, Baidu Deep Speech және басқалары сияқты танымал қолданбалар интегралдық тәсілді қолданады. Негізгі қағида - заманауи интегралдық модельдер үлкен деректерді пайдаланып оқытылады. Әсіресе қазақ, қырғыз және түрік тілдері сияқты шектеулі оқыту деректері бар тілдерді танудың негізгі мәселесін ашады. Мұндай ресурстары аз тілдер үшін оқыту деректерінің үлкен корпусары жок. Сондай-ақ, ресурстары аз тілдерге арналған интегралдық архитектура модельдері мен әдістері әзірленбегенін немесе зерттелмегенін атап өткен жөн. Қазіргі уақытта қазақ тілін интегралдық тануға арналған тиімді алгоритмдер немесе бағдарламалық құралдар жок.

Интегралдық модельдің негізгі артықшылықтары - жоғары дәлдік пен сөйлеуді жылдам тану, сондай-ақ жергілікті құрылғыда жұмыс істеу мүмкіндігі. Нақты уақыттағы сөйлеуді өңдеуге деген сұраныстың артуымен сөйлеу сигналдарын тікелей мәтіндік көріністерге түрлендіретін автоматты сөйлеуді тану (ASR) модельдеріне назар артып келеді. Бұл мәселені шешу үшін назар аудару механизмін қолданатын онлайн модель қолданылады. Ең көп таралған әдістерге монотонды бөлікке негізделген назар аудару (MoChA), қайталанатын нейрондық желі трансформаторлары (RNN-T) және стандартты нейрондық желі трансформаторлары (NT) жатады.

Бұл мақалада біз онлайн интеграция моделін қарастырамыз және қазақ тілін мәтінге түрлендіру әдісін ұсынамыз.

Әдебиеттерге шолу. RNN-T архитектурасы Connected Time Series классификациясының (СТС) жақсартылған нұсқасы ретінде енгізілді (Graves, 2012). Бұл құрылым кіріс және мақсатты белгілер арасындағы байланыс нақты анықталмаған жағдайларда тізбекті белгілеу мәселесін шешеді. СТС архитектурасы туралау мәселесін арнайы бос белгіні енгізу арқылы шешеді, бұл модельге шығыс белгісінің ықтималдығын бағалауға мүмкіндік береді (Li et al., 2019). Дегенмен, бұл әдістің негізгі кемшілігі - ағымдағы циклдің шығыс белгісі алдыңғы шығыс белгісінен тәуелсіз деген болжам (Eomet al., 2024).

Нәтижесінде, көптеген зерттеулер СТС моделінің агглютинативті тілдер үшін тілдік модельдерсіз жақсы жұмыс істейтінін көрсетті, мысал ретінде қазақ тілін қолданды. Дегенмен, ResNet ең жақсы нәтижелерді көрсетті, тіпті тілдік модельмен де таңбаларды дұрыс тану деңгейі 11,52% және сөздерді дұрыс тану деңгейі 19,57% болды. Назар аудару механизмі бар кодтаушы-декодер архитектурасын пайдаланатын екі бағытты LSTM желісі де ең жақсы CER 8,01% және WER 17,91% көрсетті.

СТС тек кіріс акустикалық тізбегіне тәуелді фонема тізбектері бойынша таралуды анықтайды. RNN-T СТС тәрізді желіні алдыңғыларына негізделген әрбір фонеманы болжайтын бөлек RNN-мен біріктіреді. Осылайша бірлесіп оқытылған акустикалық және тілдік модель жасайды. СТС әрбір кіріс уақыт қадамында шығыс таралуын анықтаса, RNN-T әрбір кіріс және шығыс уақыт қадамдарының тіркесімі үшін бөлек ықтималдық таралуын анықтайды. Модель назар аудару модельдеріне қарағанда кейбір артықшылықтарын сақтай отырып, СТС-мен байланысты қиындықтарды шешеді. Түрлендіргіш авторегрессивті. Ол алдыңғы шығыстарды кіріс ретінде қабылдайды және стандартты тіл моделі сияқты келесі шығысты болжау үшін пайдаланылуы бойынша мүмкіндіктерді жасайды.

Әдетте шығыс шығару үшін бүкіл кіріс тізбегін өңдеуі керек seq2seq модельдерінің көпшілігінен айырмашылығы, RNN-T кіріс модельдерін үздіксіз өңдейді және шығыс таңбаларын шығарады. Іске асыруда шығыс таңбалары алфавиттік таңбалар болып табылады. RNN-T таңбаларды бір-бірлеп, қажет болған жерде бос орындармен шығарады. Ол мұны келесі таңбаларды болжау үшін модель болжаған таңбаларды кері қайтаратын кері байланыс циклін пайдаланып жасайды. Желі деректер тізбегін онлайн шығарады, ағынды сөйлеуді тануды орындайды, белгі күйіне негізделген кодтаушы мен болжау желісін жаңартады. Бос белгі пайда болған кезде деректер шығысы тоқтайды.

Шығару кезіндегі ең ықтимал белгі тізбегі сәулелік іздеуді пайдаланып есептеледі, алгоритмге модельдің жұмысына нұқсан келтірмей аз ресурстарды пайдалануға мүмкіндік береді. Ағынды сөйлеуді тану үшін қолданылатын Neural Transducer-ден айырмашылығы, RNN-T-де болжау желісі кодтаушыдан шыққан шығысты ескермейді. Бұл артықшылық RNN-T-ге декодерді тілдік

модель ретінде мәтіндік деректер бойынша алдын ала оқытуға мүмкіндік береді. Модель архитектурасына өзгеріс енгізбестен, кез келген сөйлеушілер санына жалпылайтын қабаттасуға негізделген нысана орналастыру стратегиясы бар жаңа көп айналымды RNN-T моделі жасалды. Зерттеу LibriCSS тест жинағында жүргізілді және нәтижесінде WER 28% жақсарды. Атап айтқанда, бұл жүйелердің әрқайсысы сыртқы айтылым лексиконын немесе бөлек тілдік модельді пайдаланбай жазбаша домендегі графемаларды тікелей болжайды. Бұл модельдер диктант тест жинақтарындағы дәстүрлі заманауи тәсілдермен бәсекеге қабілетті екені анықталды. RNN-T өнімділігін ұзын контекстті қамтитын және алмастыру қателерін айтарлықтай азайтатын ішкі сөз бірліктерін пайдалану арқылы жақсартуға болады. Шығыс нысаналары ретінде отыз мың сөз үзіндісін пайдаланып оқытылған 2-қабатты LSTM декодері бар 12- қабатты LSTM кодтаушысы RNN-T дауысты қалпына келтіру үшін 8,5% және дауысты диктант үшін 5,2% сөз қателіктерінің жақсырақ деңгейіне қол жеткізеді. RNN-T оқытуын жеделдетудің бірнеше әдістері бар, соның ішінде оқу жылдамдығын төмендету стратегиясын жетілдіру, CNN қабаттарын қосу арқылы кодтаушыны алдын ала оқытуды жою және тіл моделін тиісті түрде кішірейту және инициализациялау арқылы оқытуды жеделдету. Жалпы алғанда, нәтижелер 2%-ға жақсарды. RNN-T өнімділігін жақсартуға болады. Біз бірнеше үлкен тензорларды жадта сақтамау үшін кодтаушы мен болжау желілерінің шығыстарын тиімді біріктіру арқылы RNN-T оқытуының жад тұтынуын азайттық. Сондай-ақ, біз бірнеше жаңа құрылымдарды ұсыну арқылы LSTM бірліктерін пайдаланатын RNN-T моделінің негізгі құрылымын жақсарттық. Барлық ұсынылған құрылымдар қабат траекториясы тұжырымдамасын пайдаланады. Ол жіктеу тапсырмасын уақытша модельдеу тапсырмасынан LSTM/GRU тереңдік бірліктері мен LSTM/GRU уақыт бірліктерін пайдаланып бөледі. Бұл негізгі RNN-T моделімен салыстырғанда Cortana, Conversation және DMA тест жинақтарында WER сәйкесінше 6,6%, 7,5% және 11,8% төмендеуіне әкелді.

Фонеманы тану бойынша зерттеулерде трансформатор ретінде қолданылатын 3 сатылы энкодерлерді және 2 LSTM-ді пайдаланып NT (нейрондық желі) енгізді (Sathyendra et al., 2022). Ұсынылған қосымша мүмкіндіктер фонеманы тану өнімділігін айтарлықтай жақсартады, мұны PER мәндерінің төмендеуінен және озық бір бағытты модельдермен салыстыруға болатын нәтижелерден көруге болады. Сонымен қатар, эксперименттік нәтижелер жақсартылған туралау сапасы жүйенің дәлдігін жақсартуда маңызды рөл атқаратынын көрсетеді.

Өзгертілген бөліктік зейін (MoChA) механизмі қатты және жұмсақ монотонды зейін арасындағы алшақтықты тиімді түрде толтыратынын және онлайн сөйлеуді тану өнімділігін жақсартатынын көрсетті (Li et al., 2021). Атап айтқанда, автоматтандырылған құжаттарды қысқаша сипаттау тапсырмалары үшін MoChA пайдалану тек монотонды зейінге негізделген модельмен салыстырғанда сапаның 20%-ға жақсаруына әкелді. Сонымен

қатар, параметрлер санының және есептеу шығындарының өсуі елеусіз болды. Зерттеулер (Kim et al., 2020), MoChA онлайн сөйлеуді тануға қолданылған кезде модельдің WER көрсеткіші 9,95% болды және оның тиімділігі расталды. Сонымен қатар, эксперименттік зерттеулерде (Hou et al., 2020) MoChA ағынды сөйлеуді тануда ең жақсы нәтижелерге қол жеткізгені, қатты монотонды зейін моделінен айтарлықтай асып түсетіні көрсетілді. Бастапқыда, кідірістерді басқару мүмкіндігі бар бейімделгіш монотонды топтық зейін механизмін пайдалана отырып, онлайн LAS моделі жасалды. Кодер жағында тікелей есептеудің кідірісін азайту үшін екі бағытты кідіріспен басқарылатын құрылым пайдаланылды, ал зейін салмағының таралуын есептеу үшін бейімделгіш монотонды топтық зейін механизмі ұсынылды. Бұл модель LAS базалық деңгейімен салыстырғанда өнімділіктің 3,5% салыстырмалы түрде төмендеуін көрсетті.

Келтірілген зерттеулер сөйлеуді дәл танудың жақсы деңгейіне қол жеткізді, бірақ бұл жүйелер әлі де сөздерді тану қателерін адам деңгейіне дейін азайту және оқыту деректерінің көлемін арттыру бойынша жетілдіруді қажет етеді. Үміткер нәтижелерді көрсететін зейінге негізделген модельге негізделген қазақ тілінде сөйлеуді танудың интеграцияланған жүйелері әзірленбеген. Зейінге негізделген модель тілдік модельдермен де, онсыз да жақсы жұмыс істей алады, сөйлеуді танудың төмен қателерін көрсетеді. Бұл тәсіл перспективалы бағыт болып табылады және шектеулі оқыту жиынтығы бар сөйлеуді тану жүйелерін әзірлеу үшін пайдаланылуы мүмкін.

Материалдар мен әдістер. RNN-T құрылымдары кодтаушыларға, болжамды желілерге және байланыстырылған желілерге негізделген (Gong et al., 2024). (Cheng et al., 2024) көрсеткендей, декодерді біріктіру желісі мен болжамды желінің тіркесімі ретінде сипаттаған кезде, RNN-T архитектурасы басқа интеграциялық модельдерге, мысалы, тізбектен тізбекке (seq2seq) моделіне ұқсас. Қайталанатын нейрондық желіге (RNN) негізделген кодтаушылар кіріс акустикалық сигналдарын дәстүрлі сөйлеуді тану жүйелеріндегі акустикалық модельдерге ұқсас аралық, жоғары деңгейлі көріністерге түрлендіреді. Шығыс белгілері бір-бірінен тәуелсіз CTC-тен айырмашылығы, RNN-T болжам белгілерінің тарихын ескеретін және оған негізделген шығыс ықтималдықтарын жасайтын болжам желісін қосады. Мысалы, болжамды желі кейінгі таңбалардың ықтималдық таралуын бағалау үшін кіріс ретінде ең соңғы бос емес болжам белгілерін пайдаланады.

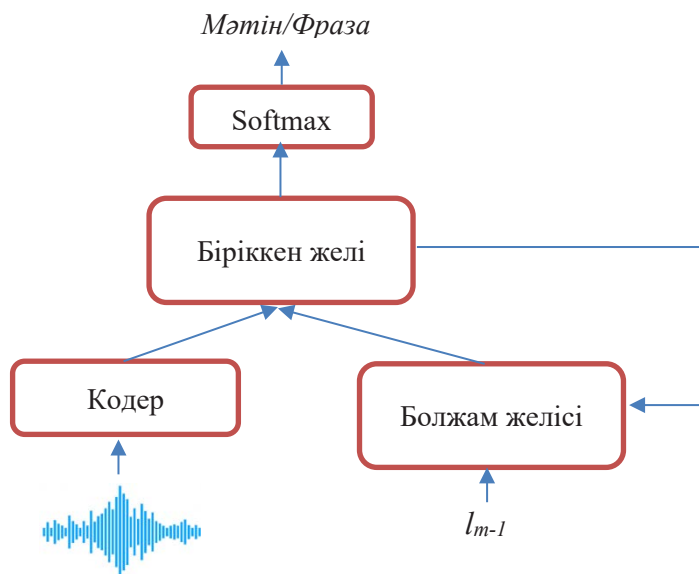
Сурет 1-де көрсетілгендей, кодтаушы шығысы болжау желісінің шығысымен ағындық қабатты қамтитын пул желісі арқылы біріктірілген.

Алынған logit softmax қабаты арқылы өтеді және шығыс таңбасы немесе сөзі үшін ықтималдық үлестірімін жасайды.

RNN-T моделін іске қосу кезінде әрбір жаңа акустикалық фрагмент $A = (a_1, \dots, a_m)$ бұрын болжанған b_{m-1} белгісін ескере отырып, қорытынды фазасында өңделеді. Модель бос емес белгіні шығарған кезде, бұл шығыс болжау желісіне қайтарылады, оның күйі $P(b|m, n+1)$ ықтималдығын бағалау

үшін жаңартылады. Екінші жағынан, егер бос белгі шығарылса, тек кодтаушы кейінгі аудио кадрларды өңдейді және болжау желісі алдыңғы күйін сақтайды. Нәтижесінде, ықтималдықтың таралуы $P(b|m+1, n)$ -ге жаңартылады. Осылайша, RNN-T ағымдағы шығыс күйіне негізделген ағынды тануды жүзеге асырады, кодтаушы мен болжау желісінің әрекетін икемді түрде өзгертеді (Liang et al., 2022).

Соңғы нәтиже ретінде ең ықтимал тізбекті табу үшін модель (Cheng et al., 2024) есептеу күрделілігін азайту үшін алгоритмге кейбір өзгерістер енгізілді. Егер бірнеше гипотезаның префиксі бірдей болса, олардың ықтималдықтарын қосу процесі алынып тасталады, бұл нәтиже сапасын төмендетпей, процестің тиімділігін арттырады.



Сурет 1 – RNN-T жалпы архитектурасы

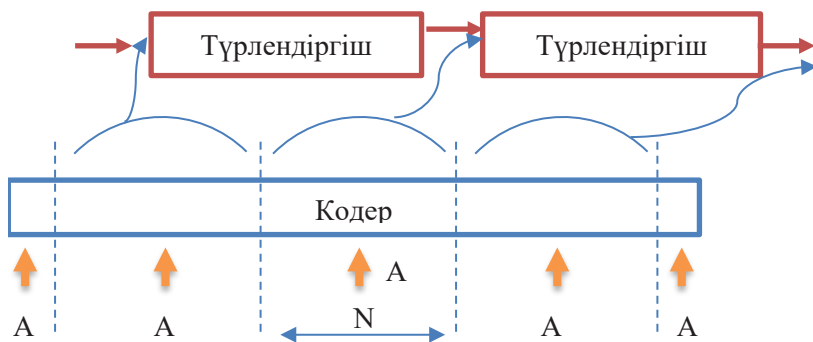
RNN-T жүйесінің 3-ші элементі болып табылатын болжау желісі кодтаушыдан алынған ақпаратқа сүйенбей автономды түрде жұмыс істейді. Бұл мүмкіндік RNN-T жүйесін басқа екі жақты онлайн әдістерден ерекшелендіреді. Бұл мүмкіндік декодерді мәтіндік деректерге негізделген тілдік модель ретінде оқытуға мүмкіндік береді, осылайша шығыс тізбектерін дәлірек жасауға ықпал етеді.

Нейрондық түрлендіргіштер. Кіріс деректерін блоктармен өңдеу арқылы NT онлайн тану талаптарына сай келетін шығыс болжамдарын кезең-кезеңімен жасай алады.

Жалпы алғанда, сөйлеу процессоры сөйлеу сигналын жоғары деңгейлі көрініске түрлендіретін кодтаушыдан және сөздер немесе жеке сөз тіркестері түрінде семантикалық шығыс деректерін жасайтын декодерден тұрады.1.

Бұл модельдің ерекшеліктерінің бірі - кіріс акустикалық тізбектің ұзындығы мен шығарылған шығыс тізбегінің ұзындығы арасында ықтимал сәйкессіздік болуы мүмкін. Сондықтан нейрондық процессор 2 мәселені шешуі керек: акустикалық ерекшеліктерді мәтіндік болжамдарға жіктеу және кіріс және шығыс тізбектері арасындағы сәйкестікті орнату. NT модельдерінің әртүрлі түрлері жіктеу және туралау механизмдерінің қалай жүзеге асырылатыны бойынша ерекшеленеді (Sainath et al., 2018).

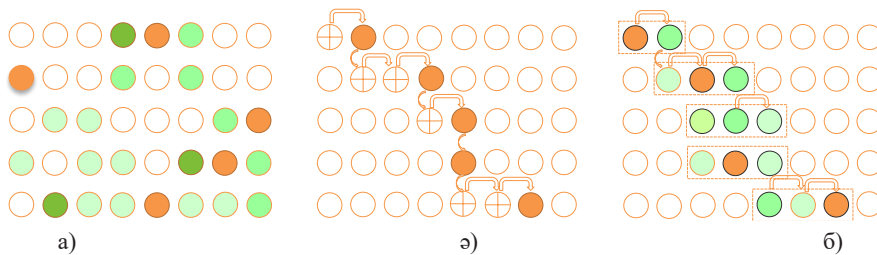
Модель логикасында өңделген деректер ағынында $(a_1, \dots, a_M)$, $b = (b_1, \dots, b_N)$ деректер ағыны болып табылады, әрбір b_n элементі W -өлшемді вектормен көрсетіледі, ал нейрондық желі түрлендіргіш стохастикалық деректерді сәйкестендіру моделі $P(b | a)$ -ға жуықтайды. Кіріс деректерін a жоғары деңгейлі $c = (c_1, \dots, c_{M'})$ көрсетуге түрлендіру кодтаушы арқылы жүзеге асырылады және уақытша іріктеуді пайдаланады, сондықтан шығыс тізбегінің M' ұзындығы кірістен ($M' \leq M$) қысқа болуы мүмкін. Архитектураға байланысты кодтаушыларды CNN және RNN сияқты әртүрлі нейрондық желі құрылымдарын пайдаланып жүзеге асыруға болады (сурет 2). Декодер кіріс және шығыс тізбектерін туралауға және оларды мәтінге түрлендіруге жауапты (Liu et al., 2021).



Сурет 2 – Нейрондық түрлендіргіш архитектурасы

Фрагменттік ақпаратқа біркелкі назар аудару. MoChA-ны талқыламас бұрын, тізбектен тізбекке архитектураның негізгі принциптерін және оның негізгі жұмсақ назар аудару механизмін қарастырған жөн. Бұл тәсілді (Azim et al., 2023) еңбегінде ұсынылған.

MoChA моделі контекстік көріністерді қалыптастыру үшін қолданылатын 2 түрлі зейін механизмін біріктіреді. Біріншісі - болжам жасау кезінде кіріс тізбегінің қай бөлігін ескеру керектігін анықтайтын қатты монотонды зейін, ал екіншісі - бөлік ішіндегі жұмсақ зейін, бұл модельге қатты монотонды зейін таңдаған позициядан бұрын жадтың шағын бөліктерін қарастыруға мүмкіндік береді (сурет 3) (Mundotiya et al., 2022).



Сурет 3 – Monotonic Chunkwise Attention моделінің қолданылуы: а) жұмсақ зейін, ә) қатты монотонды зейін, б) бөлікке негізделген монотонды зейін

MoChA моделі қатты монотонды зейін шоғырланған аймақтың алдындағы шағын жад сегментіне жұмсақ зейінді қолданады. Бұл тәсіл есептеу тиімділігін сақтай отырып, тізбекті өңдеудің икемділігін арттырады. Сонымен қатар, MoChA қолданыстағы seq2seq модельдеріне біріктірілуі немесе стандартты кері таралуды қолдана отырып үйретілуі мүмкін бейімделгіш оқыту процедураларын қамтиды.

Нәтижелер. Онлайн модель сөйлеу корпусы мен мәтін корпусын пайдаланып оқытылды. Корпустың жалпы сыйымдылығы 200 сағат таза, алдын ала белгіленген аудио деректер. Бұл корпус көпшілікке жабық және Ақпараттық есептеу технологиялары институтында қолжетімді (Renetal., 2022). Корпуста әртүрлі жастағы және жыныстағы қазақ тілінде сөйлейтіндерден жиналған аудио деректер бар, сонымен қатар, жаңалықтар порталдары мен көркем әдебиет аудиокітаптарынан алынған транскрипцияланған телефон сөйлесулері мен аудио материалдары қосымша түрде пайдаланылады.

Корпусты әзірлеу кезінде аймақтық екпіндер, әртүрлі жанрлар және кіріспе сөздер сияқты барлық мүмкін акустикалық деректерді қамту қажет. Әзірленген корпусты бағалау үшін фонетикалық тұрғыдан бай мәтін жасау қажет. Алынған үлкен корпусты спикерді анықтау және тексеру, тілді тану және жасанды интеллектті дамыту үшін пайдалануға болады. Қазақ тілі корпусын құру үшін қазақ тіліндегі мәтіндік материалдар ашық интернет көздерінен жиналды, сондай-ақ басқа да қолжетімді электрондық деректер де пайдаланылды. Барлық жиналған мәтіндік материалдар ана тілі қазақ тілінде сөйлейтіндермен айтылып, оқытылды. Сөйлеуді тану жүйесінің сапасы көп жағдайда оқыту деректерінің көлеміне байланысты. Сөйлеу және мәтіндік деректерді жинау үшін жаңалықтар веб-сайттары, аудиокітап веб-сайттары, бейнеарналар және қазақ тіліндегі пайдалы веб-сайттар қарастырылды. Қазақ тіліндегі аудиокітаптардың кең ауқымын интернеттен табуға болады. Дегенмен, олардың мәтіндік деректері әрқашан қолжетімді бола бермейді, сондықтан оны басқа қолжетімді дереккөздерден іздеу қажет. Осыған қарамастан, көптеген аудиокітаптардың жазбалары анық және таза айтылуы қажет. Жаңалықтар веб-сайттары мен бейнеарналар тек аудио деректерді ұсынады, бұл сөйлеуді қолмен транскрипциялау үшін қосымша күш-жігерді

қажет етеді. Дегенмен, мұндай арналардың артықшылығы - әртүрлі шулы ортада жазылған дауыстардың әртүрлілігі.

Сипатталған сөйлеуді түрлендіру (VC) әдісін қолдана отырып, деректер жиынтығын кеңейту үшін (Wang et al., 2024). Бұл тәсілде дыбыстық тректің мазмұнын өзгертпей жылдамдық пен темпті өзгерту арқылы корпустың сыйымдылығын 260 сағаттық аудио деректерге дейін арттыруға болады.

Оқыту және тестілік жиынтықтар аудиожазбалар деңгейінде бөлінді. Кейбір дикторлардың жазбалары екі жиынтықта да кездеседі.

Белгілерді алу кепстральды талдау арқылы сөйлеудің нақты дыбыс параметрлерін алу үшін Mel-жиілік кепстральды коэффициент әдісін қолдану арқылы жүргізілді. Кіріс сигналы алдымен кадр ретінде қалыптастырылып, өңделеді, содан кейін Фурье түрлендіруі арқылы есептеледі, ал алынған спектр мәндері Mel шкаласына қабаттастырылады. Әрбір кадр үшін энергия есептеледі, нәтижесінде Mel-жиілік кепстральды коэффициент мәндері алынады.

Тәжірибе жүргізу. Сөйлеуді өңдеу тапсырмаларында сөйлеу деректерінен тиісті ерекшеліктерді алу үшін Mel-жиілік cepcтpум коэффициенті (MFCC) жиі қолданылады (Mamyrbayev et al., 2021). RNN-T моделі 6 қабатты қайталанатын нейрондық желіні (LSTM және BLSTM) пайдаланады, ал ұқсас BLSTM архитектуралары NT және MoChA модельдеріне қолданылды (Mamyrbayev et al., 2021).

Оқудың гиперпараметрлері:

- Оқыту кезінде оқу жылдамдығы $10e-3$ дейін инициализацияланды..
- Әрбір қайталанатын қабаттың шығысындағы төмендеу (реттеу) 0.5-ке орнатылды..

- Оңтайландыру үшін Adam алгоритмі (Paul et al., 2024).

Тану дәлдігін бағалау үшін негізгі индекс ретінде сөз қателігінің деңгейі (WER) пайдаланылды. Ол қате танылған сөздердің жалпы санына қатынасын білдіреді.

Кесте 1 – Бастапқы және кеңейтілген корпустардағы онлайн түрлендіргіш модельдерінің өнімділігі

Модель атауы	WER, %	Деректер көлемі, сағат
RNN-T	16.3	200
	15.5	260
NT	16.1	200
	15.3	260
MoChA	15.4	200
	14.8	260

Талқылау. Нәтижелер MoChA сынақтан өткен әдістердің ішіндегі ең дәл қазақ тілін тану өнімділігін көрсеткенін көрсетті.

Тәжірибелік бағалау әртүрлі өлшемдегі 2 корпусты пайдалану арқылы жүргізілді. Деректер жиынтығының көлемі ұлғайған сайын, сөйлеуді тану

дәлдігі жақсарды. Салыстырмалы талдау MoChA ең жоғары тану дәлдігіне қол жеткізгенін, RNN-T және NT өнімділігінен сәйкесінше 4% және 3%-ға асып түскенін көрсетті. Дегенмен, MoChA-ны баптау әлдеқайда көп уақытты алды, себебі бұл архитектура көптеген параметрлерді қамтиды және күрделі оңтайландыруларды қажет етеді.

Тәжірибелік жағдайларда тілдік модельдер мен фонема сөздіктері сияқты көмекші ресурстар алынып тасталды. Сондықтан, бұл ресурстарды жүйеге біріктіру арқылы тану өнімділігін одан әрі жақсартуға болады.

Жалпы алғанда, 3 модель арасындағы нәтижелердегі айырмашылық аз, бұл үш тәсілдің де шектеулі деректермен де сөйлеуді тануға қабілетті екенін көрсетеді.

RNN-T, NT және MoChA архитектурасы екі бағытты қайталанатын желіге негізделген және ағынды сөйлеуді тануға бейімделуі мүмкін. Бұл модельдер аудио файлдарды алдын ала жазуды қажет етпейтіндіктен және нақты уақыт режимінде жұмыс істей алатындықтан, оларды әртүрлі технологиялық платформалар мен қолданбаларға біріктіру күтілуде.

Қорытынды. Бұл мақалада қазақ тіліне оңтайландырылған онлайн автоматты сөйлеуді тану моделі салыстырылып, талданады. Тәжірибелер бұл модельдердің жоғары тиімділігін көрсетеді.

Осы зерттеуде қарастырылған архитектуралардың ішінде MoChA қазақ сөйлеуін тану үшін ең жақсы өнімділікті көрсетті. Бұл тұжырым модельді шектеулі тілдік ресурстары бар тілдерде сөйлеуді автоматты түрде тануға қолдануға болатынын дәлелдеді.

Бұл зерттеудің ерекшелігі модель қосымша тілдік модельдерді немесе сөздіктерді пайдаланбай, бірыңғай тәсілмен зерттелді. Осылайша, алынған нәтижелер ресурстар шектеулі ортада ерекше құнды. Сондай-ақ, корпус өлшемін ұлғайту таңдалған архитектураға қарамастан дәлдікті жақсартатыны анықталды.

Мұндай архитектураның болашақ перспективаларына дәстүрлі ASR модельдерімен салыстырғанда жоғары өнімділікті сақтай отырып, минималды есептеу ресурстарымен нақты уақыт режимінде үздіксіз сөйлеуді тану сапасын жақсарту кіреді. Зерттеу нәтижелері көрсетілген онлайн архитектуралардың қазақ тіліндегі ағындық сөйлеуді тану міндетіне тиімді бейімделетінін көрсетті. Деректер көлемін ұлғайту барлық модельдердің тану дәлдігін арттыратыны тәжірибелік нәтижелер арқылы расталды. Сонымен қоса, деректер сапасы да оқыту және тану дәлдігіне тікелей байланысты. Болашақ зерттеулерге ресурстары шектеулі тілдер үшін оңтайландырылған басқа модельдермен эксперименттер және шектеулі деректер жағдайында оқу мен өнімділікті оңтайландыру жолдарын зерттеу кіреді.

References

Azim M.A., Hussein W., Badr N.L. (2023) Using Character-Level Sequence-to-Sequence Model for Word Level Text Generation to Enhance Arabic Speech Recognition. IEEE Access, vol. 11. — P. 91173–91183. <https://doi.org/10.1109/ACCESS.2023.3302257> (in Eng.)

- Cheng L., Zhu H., Hu Z., Luo B. (2024) A Sequence-to-Sequence Model for Online Signal Detection and Format Recognition. *IEEE Signal Processing Letters*, vol. 31. — P. 994–998. <https://doi.org/10.1109/LSP.2024.3384015> (in Eng.)
- Eom S., Yoon E., Yoon H. S., Kim C., Hasegawa-Johnson M., Yoo C. D. (2024) AdaMER-CTC: Connectionist Temporal Classification with Adaptive Maximum Entropy Regularization for Automatic Speech Recognition. *ICASSP 2024 – IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Seoul, Korea. — P. 12707–12711. <https://doi.org/10.1109/ICASSP48485.2024.10446721> (in Eng.)
- Gong X., Wu Y., Li J., Liu S., Zhao R., Chen X., Qian Y. (2024) Advanced Long-Content Speech Recognition With Factorized Neural Transducer. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 32. — P. 1803–1815. <https://doi.org/10.1109/TASLP.2024.3350893> (in Eng.)
- Graves A. (2012) Sequence Transduction with Recurrent Neural Networks. *arXiv preprint arXiv:1211.3711* (accessed 02.02.2025) (in Eng.)
- Hou J., Guo W., Song Y., et al. (2020) Segment Boundary Detection Directed Attention for Online End-to-End Speech Recognition. *J. Audio Speech Music Proc.* (in Eng.)
- Kim C., et al. (2020) End-to-End Training of a Large Vocabulary End-to-End Speech Recognition System. *IEEE ASRU 2019*, doi: 10.1109/ASRU46091.2019.9003976 (in Eng.)
- Li Q., et al. (2021) Confidence Estimation for Attention-Based Sequence-to-Sequence Models for Speech Recognition. *ICASSP 2021 – IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Toronto, Canada. — P. 6388–6392. <https://doi.org/10.1109/ICASSP39728.2021.9414920> (in Eng.)
- Li J., Zhao R., Hu H., Gong Y. (2019) Improving RNN Transducer Modeling for End-to-End Speech Recognition. *IEEE ASRU 2019*, Singapore. — P. 114–121. (in Eng.)
- Liang S., Yan W.Q. (2022). A Hybrid CTC+Attention Model Based on End-to-End Framework for Multilingual Speech Recognition. *Multimedia Tools and Applications*, vol. 81. — P. 41295–41308. <https://doi.org/10.1007/s11042-022-12136-3> (in Eng.)
- Liu S., Cao Y., Wang D., Wu X., Liu X., Meng H. (2021). Any-to-Many Voice Conversion With Location-Relative Sequence-to-Sequence Modeling. *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 29. — P. 1717–1728. <https://doi.org/10.1109/TASLP.2021.3076867> (in Eng.)
- Mamyrbayev O., Kydyrbekova A., Alimhan K., Oralbekova D., Zhumazhanov B., Nuranbayeva B. (2021) Development of Security Systems Using DNN and I and x-vector Classifiers. *Eastern-European Journal of Enterprise Technologies*, vol. 4(9). — P. 32–45. <https://doi.org/10.15587/1729-4061.2021.239186> (in Eng.)
- Mamyrbayev O., Oralbekova D., Kydyrbekova A., Turdalykyzy T., Bekarystankyzy A. (2021). End-to-End Model Based on RNN-T for Kazakh Speech Recognition. *3rd International Conference on Computer Communication and the Internet (ICCCI)*. — P. 163–167. <https://doi.org/10.1109/ICCCI51764.2021.9486811> (in Eng.)
- Mamyrbayev O., Oralbekova D., Alimhan K., Othman M., Turdalykyzy T. (2022) A Study of Transformer-Based End-to-End Speech Recognition System for Kazakh Language. *Scientific Reports*, vol. 12. — 8337 p. <https://doi.org/10.1038/s41598-022-12260-y> (in Eng.)
- Mundotiya R.K., Mehta A., Baruah R., Singh A.K. (2022) Integration of Morphological Features and Contextual Weightage Using Monotonic Chunk Attention for Part of Speech Tagging. *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34(9). — P. 7324–7334. <https://doi.org/10.1016/j.jksuci.2021.08.023> (in Eng.)
- Paul S., Abul Md., Walid Md., Abul A., Paul R., Uddin Md, Rana Md, Devnath M.K., Dipu I., Haque Md M. (2024) An Adam-Based CNN and LSTM Approach for Sign Language Recognition in Real Time for Deaf People. *Bulletin of Electrical Engineering and Informatics*. (in Eng.)
- Ren Z., Yolwas N., Slamun W., Cao R., Wang H. (2022) Improving Hybrid CTC/Attention Architecture for Agglutinative Language Speech Recognition. *Sensors*, vol. 22(19). — 7319 p. <https://doi.org/10.3390/s22197319> (in Eng.)
- Sainath T.N., et al. (2018) Improving the Performance of Online Neural Transducer Models. *ICASSP 2018*. — P. 5864–5868. (in Eng.)

Sathyendra K.M., Muniyappa T., Chang F., Liu J., Su J., Strimel G. P., Mouchtaris A., Kunzmann S. (2022) Contextual Adapters for Personalized Speech Recognition in Neural Transducers. ICASSP 2022 – IEEE Int. Conf. on Acoustics, Speech and Signal Processing. — P. 8537–8541. (in Eng.)

Wang H., Jin Z., Geng M., Hu S., Li G., Wang T., Xu H., Liu X. (2024). Enhancing Pre-Trained ASR System Fine-Tuning for Dysarthric Speech Recognition Using Adversarial Data Augmentation. ICASSP 2024 – IEEE Int. Conf. on Acoustics, Speech and Signal Processing. — P. 12311–12315. (in Eng.)

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

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

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

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

www.nauka-nanrk.kz

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

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*

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

Computer layout: *G.D. Zhadyranova*

Signed for print: September 30, 2026

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