Жобаның қысқаша мәліметі

Атауы	BR24993001 «Қазақ тілін қолдау және технологиялық ілгерілеу мақсатында үлкен тілдік модель (LLM) жасау»
Өзектілігі	Бағдарламаны іске асыру қазақ тілінің мемлекеттік бірегейліктің, мәдени мұраның және мәдениетаралық өзара әрекеттестіктің негізі ретінде мәртебесін нығайтуға ықпал етеді. Бағдарламаның маңызды аспектілерінің бірі – Қазақстанда алғаш рет жасанды интеллект міндеттерін шешуге арналған жоғары өнімді деректер орталығын құру. Бұл деректер орталығы ұжымдық пайдалануға арналған зертхана ретінде жұмыс істеп, еліміздің ғылыми-зерттеу қауымдастығына мүмкіндік береді.
Мақсаты	Қазақ тілін мемлекеттік және Қазақстандағы мәдениетаралық қатынас тілі ретінде қолдауға, технологияларды дамытуға, деректер қауіпсіздігін қамтамасыз етуге, білім беру мен ғылыми зерттеулерді жетілдіруге ықпал ететін заманауи үлкен тілдік модельді (LLM) жасау.
Міндеттері	1. Үлкен тілдік модельдерді (Large Language Model, LLM) жасауға арналған қолданыстағы әдістер мен технологияларды талдау. Бұл мәселені шешу Бағдарламаны іске асырудың одан арғы стратегиясын нақтылау мен жаңартуда шешуші рөл атқарады. Стратегияны жаңарту тек LLM әзірлеуге түбегейлі жаңа тәсілдер табылған жағдайда ғана мүмкін болады. 2. LLM оқытуға арналған мәліметтер корпусын жинау және талдау. Бұл мәселені шешу Бағдарламаның негізгі нәтижесіне тікелей әсер етеді, оған әзірленетін қазақ тілді LLM-нің сапасы, заңдылығы және этикалық тұрғыдан дұрыстығы кіреді. 3. LLM моделінің архитектурасын әзірлеу және өнімділігін оңтайландыру. Бұл мәселені шешу модельдің жоғары өнімділігін және ұзақмерзімді қолданылуын қамтамасыз ету үшін маңызды. 4. LLM моделін GPU кластеріне орналастыру. Бұл мәселені шешу де модельдің өнімділігі мен ұзақмерзімді пайдаланылуын қамтамасыз ету үшін негізгі рөл атқарады. 5. Әзірленген тілдік модельді әртүрлі жүйелер мен платформаларға біріктіру. Бұл мәселені шешу Бағдарламаның мақсатына қол жеткізу үшін аса маңызды, себебі ол қазақ тіліндегі жасанды интеллектіні экономика мен мемлекеттік басқарудың түрлі салаларында кеңінен қолдануға мүмкіндік береді.

Күтілетін және қол жеткізілген нәтижелер	1. Retrieval-Augmented Generation (RAG) платформасының прототипін жасау және 1,94 миллиард параметрлі LLaMA архитектурасындағы көмекші-модельдің пилоттық нұсқасын іске қосу. 2. LLM компоненттері, іздеу жүйесі (retrieval system) және деректер қоймасы арасындағы өзара әрекеттестікті қамтамасыз ететін базалық инфрақұрылымды әзірлеу және орнату. 3. Қазақтілді мәтіндер корпусының 400 миллион сөзден тұратын көлемін жинақтау және қалыптастыру, түрлі жанрлар, стильдер мен тақырыптарды қамту. 4. Мәтіндік деректерді алдын ала өңдеу: тазалау, нормализациялау, аннотациялау және әрі қарайғы лингвистикалық таңбалауға дайындау. 5. Ашық бастапқы кодты LLM модельдерінің архитектурасын (8 және 70 миллиард параметр) әзірлеуді бастау және олардың өнімділігін оңтайландыру. 6. LLM оқыту үдерісін параллельдеу бойынша гибридті тәсілді зерттеу және іске асыру (тензорлық, модельдік және конвейерлік схемалар). 7. Есептеу жүктемелерін үлестіру және GPU-кластерін тиімді пайдалану шешімдерін әзірлеу. 8. Ғылыми мақалалар дайындау және жобаны әрі қарай ауқымдандыруға арналған әдістемелік негіз қалыптастыру.
Зерттеу тобы мүшелерінің аты жөні, олардың идентификаторлары (Scopus Author ID, Researcher ID, ORCID, бар болса) және сәйкес профильдерге сілтемелері	Мансурова Мадина Есимханқызы – жобаның ғылыми жетекшісі, физика-математика ғылымдарының кандидаты, доцент, әл-Фараби атындағы ҚазҰУ-дың Жасанды интеллект және Big Data кафедрасының меңгерушісі, жетекші ғылыми қызметкер. Ғылыми қызығушылықтары: жоғары өнімді есептеулер, табиғи тілдерді өңдеу (NLP), жасанды интеллект. 90-нан астам ғылыми мақала, 4 монография, ҚР БҒМ грифі берілген 2 оқулықтың авторы. Scopus ID: 56617164900, Researcher ID: жоқ, ORCID ID: 0000-0002-9680-2758. Хирш индексі: Scopus – 6, Web of Science – 4. Төкеев Уәлихан Әбдікәрімұлы – орындаушы, техникалық ғылымдар докторы, әл-Фараби атындағы ҚазҰУ профессоры. Формалды грамматикалар мен машиналық аударма саласының маманы, көптеген гранттық және халықаралық жобалардың қатысушысы әрі жетекшісі, 20 жылдан астам тәжірибесі бар. Scopus ID: 55701639900, Researcher ID (WoS): AAR-3668-2020, ORCID ID: 0000-0001-9878-981X.

	Хирш индексі: 5. Рахимова Динара Рыскалиевна – орындаушы, PhD, ҚазҰУ-дың аға оқытушысы. Мәтіндерді алдын ала өңдеудің инновациялық әдістерін әзірлеуге қатысады, есептік лингвистика бойынша магистрлік бағдарламаның үйлестірушісі. Scopus-та индекстелген 27 жарияланымның авторы. Scopus ID: 57191430929, Researcher ID: P-4962-2017, ORCID ID: 0000-0003-0140-4181. Хирш индексі: 4. Кәрібаева Аружан Серікқызы – орындаушы, PhD, ҚазҰУ-дың аға оқытушысы. Морфологиялық талдау саласының маманы, халықаралық жобаларға қатысу және Эдинбургте тағылымдамадан өту тәжірибесі бар. Scopus ID: 57196004542, ORCID ID: 0000-0002-2023-1573. Хирш индексі: 3. Шормақова Айжан Нұрланқызы – орындаушы, магистр, ҚазҰУ-дың аға оқытушысы, PhD докторант. Қазақ тіліне арналған үлкен тілдік модельдер (LLM) мен табиғи тілді өңдеумен айналысады. Scopus ID: 55786942400, ORCID ID: 0000-0002-1637-4643. Хирш индексі: 2. Карюкин Владислав – орындаушы, PhD, әл-Фараби атындағы ҚазҰУ-дың Ақпараттық жүйелер кафедрасының аға оқытушысы. Қазақ тіліндегі мәтіндерді алдын ала өңдеу саласына маманданған. Scopus ID: 57218952479, ORCID ID: 0000-0002-8768-0349. Хирш индексі: 2. Жұманов Жандос Мұратұлы – кіші ғылыми қызметкер, магистр, «Alem Research» ЖШС-нің бағдарламашысы. Синтаксистік талдау, автоматты сөйлеуді тану және синтездеу саласының маманы. Scopus ID: 55701348900, Researcher ID: B-2733-2015, ORCID ID: 0000-0002-0395-9730. Хирш индексі: 4. Дүйсенбекқызы Жансая – кіші ғылыми қызметкер, техника ғылымдарының магистрі, ҚазҰУ-дың «Ақпараттық жүйелер» мамандығы бойынша 1-курс PhD докторанты. Қазақ балаларының сөйлеуін тану мәселесімен айналысады.

	Researcher ID: JMQ-5368-2023, ORCID ID: 0000-0001-7743-0795. Нұрмағамбет Даурен – орындаушы, техник, әл-Фараби атындағы ҚазҰУ-дың «Есептік лингвистика» бағыты бойынша 1-курс магистранты. Синтаксистік талдауға арналған бағдарламалық қамтамасыз етуді әзірлеуге қатысады. Алиев Р. – орындаушы, техник, ҚазҰУ-дың «Ақпараттық жүйелер» мамандығының студенті. Морфологиялық талдауға арналған бағдарламалық қамтамасыз етуді әзірлеуге қатысады. Scopus ID: 00000000000, Researcher ID: JPA-3963-2023, ORCID ID: 0009-0003-5432-0480. Хирш индексі: 0. Әбдіахметова Зухра Мұратқызы – орындаушы, компьютерлік ғылымдар PhD, магистр, әл-Фараби атындағы ҚазҰУ-дың қауымдастырылған профессоры (в.а. доцент). Деректерді өңдеу, машиналық оқыту және Big Data талдау саласының маманы. Мәтіндерді алдын ала өңдеу мен Цифрлық егіз жасау бойынша зерттеулерге белсенді қатысады. Scopus ID: 57193647310, ORCID ID: 0000-0001-7158-0995. Хирш индексі: 4. Жайсанова Динара Сайлауқызы – орындаушы, PhD, әл-Фараби атындағы ҚазҰУ-дың «Жасанды интеллект және Big Data» кафедрасының аға оқытушысы. Қазақ тіліне арналған LLM архитектурасын оңтайландыру бойынша эксперименттерге қатысады, нәтижелерді талдап, құжаттайды. Scopus ID: 57204395807, ORCID ID: 0000-0002-8116-6111, Researcher ID: CJV-4009-2022. Хирш индексі: 2. Исабаева Даража Нағашыбайқызы – орындаушы, педагогика ғылымдарының кандидаты, қауымдастырылған профессор, ҚазҰУ-дың «Жасанды интеллект және Big Data» кафедрасының профессоры (в.а.). Білім беру саласында машиналық оқытуды қолдану және оқыту үдерісін оңтайландыру бойынша маман. Scopus ID: 56366568600, ORCID ID: 0000-0002-9979-3121, Researcher ID: EZV-5601-2022. Хирш индексі: 3. Қадырбек Нұрғали Қазбекұлы – орындаушы, магистр, әл-Фараби атындағы ҚазҰУ-дың аға ғылыми қызметкері. LLM архитектурасын

	жетілдіру, модельдің өнімділігін тестілеу және оңтайландыру жұмыстарына қатысады. Scopus ID: 57220749433, ORCID ID: 0000-0002-5461-8899, Researcher ID: AEK-7760-2022. Хирш индексі: 3. Қайратұлы Бауыржан – орындаушы, магистр, әл-Фараби атындағы ҚазҰУ-дың «Деректер ғылымы» мамандығы бойынша 1-курс докторанты. LLM-ді оқыту мен тестілеуге арналған корпустық деректерді жасау және таңбалау жұмыстарымен айналысады. Scopus ID: 00000000000, ORCID ID: 0000-0002-0341-8899. Хирш индексі: 0. Қырғызбаева Маржан Ескендерқызы – орындаушы, магистр, әл-Фараби атындағы ҚазҰУ-дың «8D06103 – Компьютерлік ғылымдар» мамандығы бойынша 3-курс докторанты. LLM әзірлеуге, архитектураларды таңдауға және біріктіруге, модельді оқыту мен оның тиімділігін бағалауға қатысады. Scopus ID: 57220744243, ORCID ID: 0009-0004-5333-3252. Хирш индексі: 2. Оспан Әсел Ғалымжанқызы – орындаушы, магистр, әл-Фараби атындағы ҚазҰУ-дың аға оқытушысы. LLM өнімділігін оңтайландырады, тар орындарды жояды, жетілдірулер енгізіп, тестілеу жүргізеді. Scopus ID: 57238489800, ORCID ID: 0000-0002-1860-6997. Хирш индексі: 2. Сарсембаева Талшын Сағдатбекқызы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың докторанты. LLM сапасын бағалау жүйесін, метрикаларды, тестілеу процедураларын әзірлейді және нәтижелерді талдайды. Scopus ID: 57224454827, ORCID ID: 0000-0001-7668-2640. Хирш индексі: 3. Шоманов Адай Сәкенұлы – орындаушы, PhD, Назарбаев Университетінің инструкторы. LLM қауіпсіздік әдістерін зерттейді, қорғаныс механизмдерін әзірлейді, аудиттер мен осалдыққа қарсы тестілеу жүргізеді. Scopus ID: 57195543732, ORCID ID: 0000-0001-8253-7474, Researcher ID: ABG-7891-2021. Хирш индексі: 5. Жиенбаев Мейран Мұратұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың докторанты. Арнайы чат-боттар үшін LLM-ді қайта оқыту,

	деректерді жинау және дайындау, модельдерді бағалау және оңтайландыру жұмыстарымен айналысады. ORCID ID: 0009-0005-0497-0965. Хирш индексі: 0. Байбурин Ержан Мұхамбетқалиевич – орындаушы, Цифрлық технологиялар және модельдеу зертханаларының ғылыми қызметкері. LLM-пен өзара әрекеттесудің инновациялық әдістерін әзірлейді, интерфейстерді зерттейді және олардың тиімділігін бағалайды. Scopus ID: 56111999400, ORCID ID: 0000-0002-1583-9912, Researcher ID: GRT-5436-2022. Хирш индексі: 4. Әбішев Қуаныш Шұрабатырұлы – орындаушы, PhD (математика), қауымдастырылған профессор, Алматы Менеджмент Университетінің Цифрлық технологиялар мектебінің деканы. Data Science, Machine Learning және Computer Vision салаларының сарапшысы. Machine Learning және Computer Vision курстарын оқытады, Association for Symbolic Logic (ASL) мүшесі. Scopus ID: 56124436100, ORCID ID: 0000-0003-1140-7431, Researcher ID: N-9739-2014. Хирш индексі: 5. Тұрар Олжас Нұрқонысұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың Ақпараттық технологиялар факультетінің деканы. Математикалық модельдеу және машиналық оқыту саласының маманы. GPU-кластерде LLM архитектурасымен айналысады. Scopus ID: 56523413700, ORCID ID: 0000-0002-6720-0045. Хирш индексі: 4. Дарибаев Бейімбет Серікұлы – орындаушы, әл-Фараби атындағы ҚазҰУ кафедра меңгерушісі. Параллель алгоритмдер саласының маманы, ғылыми-зерттеу жұмыстарында 10 жылдан астам тәжірибесі бар. GPU құрылғыларын таңдауға және баптауға жауапты. Scopus ID: 57191892577, ORCID ID: 0000-0003-1313-9004. Хирш индексі: 4. Лебедев Данил Владимирович – орындаушы, «Astana IT University» ЖШС маманы. Таратылған және параллель есептеулер саласында 20 жылдан астам тәжірибесі бар. Модельдерді қысу әдістерін (прунинг, квантау) зерттейді және LLM үшін стратегия әзірлейді.

	Scopus ID: 56522634900, ORCID ID: 0000-0002-5186-6483. Хирш индексі: 4. Ермек Алимжанов – орындаушы, әл-Фараби атындағы ҚазҰУ маманы. Автоматтандыру және таратылған есептеулер саласында 10 жылдан астам тәжірибесі бар. GPU-кластерге бағдарламалық жасақтаманы орналастыруға арналған скрипттерді әзірлеу және күйге келтіру үшін жауапты. Scopus ID: 57191433356, ORCID ID: 0000-0002-8758-2220. Хирш индексі: 2. Мәткерім Базаргүл – орындаушы, әл-Фараби атындағы ҚазҰУ-дың аға оқытушысы. Таратылған жүйелер бойынша маман, GPU кластерлерінде ресурстарды жоспарлау моделін әзірлейді. Scopus ID: 55898795100, ORCID ID: 0000-0002-5336-687X. Хирш индексі: 3. Аубакиров Санжар – орындаушы, әл-Фараби атындағы Қазақ ұлттық университеті. Web/мобильді қосымшаларды әзірлеу және ақпараттық жүйелер архитектурасы саласының маманы. GPU кластеріндегі тапсырмалар кезегін басқару алгоритмдерін әзірлеп, оларды қолданыстағы инфрақұрылыммен біріктіреді. Scopus ID: 57190230321, ORCID ID: 0000-0002-8416-527X. Хирш индексі: 3. Ғылыми жобаларда 15 жылдан астам тәжірибесі бар. Нурахов Едил Сергазиевич – орындаушы, «Astana IT University» ЖШС-нің цифрлық технологиялар департаментінің директоры. Дайындалған LLM моделін жүктеу мен сақтау үшін инфрақұрылымды дайындайды және оны GPU кластерінде орналастыру параметрлерін баптайды. Scopus ID: 57204717628. Хирш индексі: 3. IT жобалар мен ақпараттық жүйелерді әзірлеу саласында 5 жылдан астам тәжірибесі бар. Кенжебек Ержан Ғалымжанұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың аға оқытушысы. GPU өнімділігін бақылау құралдарын орнатып, метрикалар мен ескерту жүйесін әзірлейді. Scopus ID: 57221598108, ORCID ID: 0000-0002-6492-8292. Хирш индексі: 2. 6 жылдан астам тәжірибесі бар, 15 жарияланымы, оның 3-еуі Scopus-та.

	Нейрожелілер мен мұнай қайтарымын арттыру салаларында тәжірибесі бар. Қасымбек Нұрислам Мұратбекұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың аға оқытушысы. Адаптивті оқыту әдістерін зерттеп, GPU кластерінде LLM моделінің бейімделген оқыту прототипін іске асырады. Scopus ID: 57217825275, ORCID ID: 0000-0001-5663-2267. Хирш индексі: 1. 7 жылдан астам тәжірибесі бар, 13 жарияланым, оның 4-еуі Scopus-та. Параллельді есептеулер (MPI, OpenMP) саласының маманы. Мустафин Мақсат Бейбітұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың аға оқытушысы. LLM архитектурасын талдап, GPU ресурстары шектеулі жағдайында оны оңтайландыру әдістерін әзірлейді. Scopus ID: 58693756000, ORCID ID: 0000-0002-3655-0771. Хирш индексі: 1. 6 жылдан астам тәжірибесі бар, 8 жарияланым (оның ішінде 2 – Scopus-та). CUDA, визуализация, OpenGL/Vulkan саласының маманы. Мұханбет Ақсұлтан Айтуарұлы – орындаушы, әл-Фараби атындағы ҚазҰУ-дың докторанты. LLM міндеттерінің ерекшеліктерін ескере отырып, ресурстарды басқару жүйесін әзірлейді. Scopus ID: 57214259277, ORCID ID: 0000-0003-4699-0436. Хирш индексі: 2. 5 жылдан астам тәжірибесі бар, кванттық алгоритмдер мен машиналық оқыту бағытында жұмыс істейді. Махмұт Ерлан – орындаушы, әл-Фараби атындағы ҚазҰУ-дың оқытушысы. Таратылған оқыту әдістерін зерттеп, LLM үшін тиісті инфрақұрылымды баптаумен айналысады. Scopus ID: 58042934300, ORCID ID: 0009-0002-3451-415X. Хирш индексі: 0. 5 жылдан астам тәжірибесі бар. CUDA және деректер визуализациясы саласының маманы. Аетова Багдат Алтынбекқызы – жобаның бухгалтері, әл-Фараби атындағы ҚазҰУ. Қаржылық есептілікті жүргізуді қамтамасыз етеді. Ғылыми идентификаторлары жоқ.

	Adalı Eşref – профессор, PhD, Стамбұл техникалық университеті (Түркия). Мәтіндердің синтаксистік және морфологиялық талдау әдістерін әзірлеуге қатысады. Агглютинативті тілдерге арналған сөз таптарын белгілеу (POS-tagging) және стемминг әдістерінің негізін қалаушылардың бірі. Scopus ID: 14834024800, ORCID ID: 0000-0002-1561-8255. Хирш индексі: 9. 18 жылдық тәжірибесі бар, 49 мақала (WoS, Scopus). Машиналық аударма жобаларына қатысқан (түрік–ағылшын, түрікмен–түрік). Барахнин Владимир Борисович – техникалық ғылымдар докторы, доцент, ФИЦ ИВТ (Ресей, Новосібір). Корпус қалыптастыруға, LLM баптауға, тест әзірлеуге қатысады. Scopus ID: 6508258628, ORCID ID: 0000-0003-3299-0507, WoS ID: A-5856-2014. Хирш индексі: Scopus – 7, WoS – 4. 200-ден астам жарияланымның авторы, бірнеше монография мен оқулық жазған. РФФИ және РНФ жобаларын басқарған. NLP және математикалық модельдеу саласының сарапшысы. Амирғалиев Еділхан Несіпханұлы – орындаушы, ҚР Ұлттық инженерлік академиясының академигі, ҚР ҰҒА-ның корреспондент-мүшесі, техникалық ғылымдар докторы, профессор, ҚР ҒЖБМ ҒК Ақпараттық және есептеуіш технологиялар институтының (АЕТИ) жасанды интеллект және робототехника зертханасының меңгерушісі. Ол заманауи үлкен тілдік модельдерді (GPT-3, BERT, LaMDA, BLOOM және т.б.) салыстырмалы талдаудан өткізіп, қазақ тіліне бейімдеудің артықшылықтары мен кемшіліктерін айқындайды. Scopus ID: 56167524400, ORCID: 0000-0002-6528-0619, Researcher ID: C-6963-2015. Хирш индексі: 15. 8 монография мен 350-ден астам ғылыми жарияланымның, оның ішінде Scopus-тағы 95-тен астам мақаланың авторы. ҚР БҒМ-нің ГФ және ПЦФ жобаларына жетекшілік еткен. Мамырбаев Өркен Жұмажанұлы – орындаушы, PhD, профессор, ҚР ҒЖБМ ҒК АЕТИ ғылым жөніндегі бас директорының орынбасары. Ол үлкен тілдік модельді (LLM) оқытудың негізгі параметрлерін (корпус көлемі, токен саны, GPU және т.б.) анықтап, оны білім беру, мемлекеттік қызмет, заңнама, қазақ тілін

	қолдау сияқты салаларда қолдану аясын белгілейді. Scopus ID: 55967630400, ORCID: 0000-0001-8318-3794, Researcher ID: 0-1265-2017. Хирш индексі: 11. Сөйлеу және бейне тану саласының сарапшысы. Қалиханова Әлия Уәлиқызы – орындаушы, физика-математика ғылымдарының кандидаты, қауымдастырылған профессор, ҚР ҒЖБМ ҒК АЕТИ бас ғылыми қызметкері. Ол оқыту деректеріндегі зиянды контентті сүзгілеу ережелерін әзірлеп, LLM құрудың этикалық аспектілеріне жауап береді. Scopus ID: 57034267000, ORCID: 0000-0002-5979-9756, Researcher ID: AAR-7411-2020. Хирш индексі: 7. AP09259547, AP05132778 және басқа жобалардың жетекшісі. Нарынов Серғазы Сәкенұлы – орындаушы, техникалық ғылымдар кандидаты, ҚР ҒЖБМ ҒК АЕТИ жетекші ғылыми қызметкері. Қазақ тіліндегі мәліметтерді жинау, корпус қалыптастыру және GAN арқылы синтетикалық мәтіндер генерациялауға жауапты. Scopus ID: 56582431600, ORCID: 0000-0002-9611-9772. Хирш индексі: 5. AP09259140, AP08855520 жобаларының жетекшісі, BR05236699-OT-19 ПЦФ жобасының қатысушысы. Қозбақова Айнұр Қолдасқызы – орындаушы, PhD, қауымдастырылған профессор, ҚР ҒЖБМ ҒК АЕТИ жетекші ғылыми қызметкері. Орыс тіліндегі деректер корпусын қалыптастырып, GAN көмегімен синтетикалық мәтіндер генерациялайды. Scopus ID: 57195683902, ORCID: 0000-0002-5213-4882, Researcher ID: K-5077-2018. Хирш индексі: 6. BR05236693, AP05132648, AP08855903 және басқа ПЦФ/ГФ жобаларына қатысқан. Оралбекова Дина Орымбайқызы – орындаушы, PhD, ҚР ҒЖБМ ҒК АЕТИ аға ғылыми қызметкері. Ағылшын тіліндегі корпус құрып, оны GAN арқылы кеңейтумен айналысады. Scopus ID: 57226648854, Web of Science: профилі. Хирш индексі: Scopus – 5, Web of Science – 2. Бейне және сөйлеу тану саласының сарапшысы. Мерембаев Тимур Жұмақанұлы – орындаушы, математика магистрі, ғылыми қызметкер (Research Assistant), Халықаралық ақпараттық технологиялар университеті.

	Қазақ тіліне бейімделген трансформер негізіндегі модельдерді әзірлеп, лингвистикалық және мәдени ерекшеліктерді ескере отырып оқытумен айналысады. Scopus ID: 57207766879, ORCID: 0000-0001-8185-235X. Хирш индексі: 4. Қашықтықтан зондтау, ГАЖ және деректер базасы саласында 10 жылдан астам тәжірибесі бар. Черикбаева Ляйля Шариповна – орындаушы, PhD, информатика магистрі, әл-Фараби атындағы ҚазҰУ-дың оқытушысы (доцент м.а.). LLM арқылы зиянды контент генерациясының негізгі тәуекелдерін анықтайды. Модельді оқытуда қолданылатын деректер сапасын бақылап, тәуекелдерді азайтуға жауапты. Scopus ID: 57200073690. Хирш индексі: 1. Бейне тану және компьютерлік көру салаларында тәжірибесі бар. BR05236693 жобасының қатысушысы. Қуанышбай Дархан Нұрғазыұлы – орындаушы, Сүлеймен Демирел университетінің өкілі. ҚР-ның IT қауіпсіздік талаптарына сәйкес жеке деректерді және құпиялықты қорғауды қамтамасыз етеді. LLM-ді қазақ тілінің мәдени және тілдік ерекшеліктеріне бейімдеу әдістерін әзірлейді. Scopus ID: 57215602501. Хирш индексі: 1. Робототехника, компьютерлік көру және машиналық оқыту салаларында тәжірибесі бар. Ыбытаева Ғалия Сейтқалиевна – орындаушы, «Ақпараттық және есептеуіш технологиялар институты» РМК кіші ғылыми қызметкері. Қазақ тілінің лингвистикалық талдауын (морфология, синтаксис, семантика) жүргізеді, модельді оқытуға арналған жергілікті тіркестер мен идиомалар базасын жасайды. Scopus ID: 57221333455. Хирш индексі: 2. Робототехника, компьютерлік көру және машиналық оқыту салаларында тәжірибесі бар. Өтегенова Анар Урантаевна – орындаушы, PhD, «Ақпараттық және есептеуіш технологиялар институты» РМК жетекші ғылыми қызметкері. LLM-ді білім беру, мемлекеттік қызметтер, заңнамалық процестер және т.б. салаларға енгізу бағыттарын анықтайды. API-ге техникалық сипаттама әзірлейді. Scopus ID: 56826094200.

	Хирш индексі: 5. Ғылыми қызығушылықтары: ақпараттық және ғарыштық технологиялар, ГАЖ, телекоммуникация, машиналық оқыту. 50-ден астам жарияланымның авторы, 20 жылдан астам ғылыми тәжірибесі бар. Сүндетов Талғат Рысбекұлы – орындаушы, PhD-докторант, «Ақпараттық және есептеуіш технологиялар институты» РМК кіші ғылыми қызметкері. API сервистерін, құжаттаманы әзірлейді және оларды таңдалған платформаларға бейімдейді. Scopus ID: 57211753744, ORCID: 0000-0003-1193-5517. Хирш индексі: 1. ГФ №AP05132648 жобасының қатысушысы. Робототехника және компьютерлік көру салаларында тәжірибесі бар. Атаниязова Айсұлу Саламатовна – орындаушы, PhD-докторант, «Ақпараттық және есептеуіш технологиялар институты» РМК. API-сервистер мен тиісті құжаттаманы әзірлеуге және бейімдеуге қатысады. Scopus ID: 57298357700. Хирш индексі: 0. Робототехника, компьютерлік көру және машиналық оқыту салаларында тәжірибесі бар. Хусейн Атакан Варол – орындаушы, ISSAI-дың бас директоры, Назарбаев Университеті ДББҰ. Робототехника, машиналық оқыту, зияткерлік жүйелер мен жасанды интеллект салаларындағы жоғары білікті маман. 100-ден астам ғылыми жарияланымның авторы, халықаралық патенттердің иегері. ҚР БҒМ мен Назарбаев Университеті жобаларының жетекшісі. Scopus ID: 24484134500, ORCID ID: 0000-0002-4042-425X. Хирш индексі: 28. Аскат Күздеуов – орындаушы, ISSAI-дың аға деректер талдаушысы, Назарбаев Университеті ДББҰ. Компьютерлік көру, сөйлеуді тану, табиғи тілді өңдеу (NLP) және робототехника салаларының маманы. 17 ғылыми мақаланың авторы, 30-дан астам жобаның қатысушысы. Scopus ID: 57208734478. Хирш индексі: 9. Абдрахманова Мадина – орындаушы, ISSAI-дың деректер талдаушысы, Назарбаев Университеті ДББҰ. Білімі: Иллинойс университеті (Урбана-Шампейн) – бакалавр, Калифорния университеті (Ирвайн) – магистр. Мультимодальды оқыту, деректерді

	аугментациялау және кластерлік жүйелерді басқару саласында тәжірибесі бар. Scopus ID: 57191414950. Хирш индексі: 3. Полонская Алина – орындаушы, ISSAI-дың деректер талдаушысы, Назарбаев Университеті ДББҰ. Санкт-Петербург мемлекеттік университетінің магистратура түлегі. Ғылыми қызығушылықтары: табиғи тілді өңдеу (NLP), мәтін генерациясы, сөйлеуді тану. Көптілді жасанды интеллект модельдерімен жұмыс істейді. Scopus ID: 000000000, ORCID ID: 0000-0000-0000-0000. Хирш индексі: 0. Мусаходжаева Саида – орындаушы, ISSAI-дың ғылыми қызметкері, Назарбаев Университеті ДББҰ. KAIST (Оңтүстік Корея) түлегі, бакалавриат – Назарбаев Университеті. 12 ғылыми мақаланың авторы, қазақ тіліне бағытталған жасанды интеллект жобаларына қатысады. Scopus ID: 57189046541. Хирш индексі: 6. Ешпанов Рүстем – орындаушы, ISSAI-дың деректер талдаушысы, Назарбаев Университеті ДББҰ. Оксфорд университетінің лингвистика магистрі. Табиғи тілді өңдеу және жасанды интеллект салаларының маманы, ғылыми жарияланымдардың авторы, қазақ тіліне қатысты жобаларға қатысқан. Ешпанов Рүстем – орындаушы, деректер талдаушысы, ISSAI, «Назарбаев Университеті» ДББҰ. Оксфорд университетінің лингвистика магистрі. NLP және жасанды интеллект саласының маманы, қазақ тілі бойынша жобаларға қатысқан, ғылыми жарияланымдардың авторы. Scopus ID: 57364291200. Хирш индексі: 2. Аспандияр Нұріманов – орындаушы, деректер талдаушысы, ISSAI, «Назарбаев Университеті» ДББҰ. Data Science магистрі, Frankfurt School of Finance & Management. Бұрын стартаптар мен консалтинг салаларында жұмыс істеген. LLM және қазақ тіліне қатысты зерттеулермен айналысады. Scopus ID: 000000000, ORCID ID: 0000-0000-0000-0000. Хирш индексі: 0.

	Серікбай Бауыржан – AI-маманы, магистр, Шайсұлтан Шаяхметов атындағы «Тіл – Қазына» ұлттық ғылыми-практикалық орталығы» КЕАҚ ғылыми қызметкері. Жоба аясындағы жұмыс тәжірибесі – 4 жылдан астам. Қазақ ұлттық корпусындағы бес ішкі корпустың (2021–2024 жж.) әзірлеуіне, қазақ тіліне арналған TTS және STT жобаларына, «Бала тілі» балалар порталы мен зағип жандарға арналған «Экран дикторы» жобасына қатысқан. Мәліметтерді парсингтеу, Python арқылы өңдеу үдерістерін автоматтандыру, скрипттер әзірлеу, деректерді нормализациялау мен форматтау салаларында маманданған. Scopus ID: жоқ, ORCID ID: 0000-0000-0000-0000, Хирш индексі: 0. Алшынбекова Мәдина Серікболсынқызы – жобаның орындаушысы, ҚХР Сиань шет тілдер университетінің гуманитарлық ғылымдар магистрі, ЕҰУ докторанты. «Тіл – Қазына» ұлттық ғылыми-практикалық орталығы» КЕАҚ қолданбалы лингвистика бөлімінің ғылыми қызметкері. Ғылыми қызығушылықтары: салыстырмалы фразеология. Жұмыс тәжірибесі – 10 жылдан астам. 10-нан астам ғылыми мақаланың авторы, олардың бірқатары Scopus базасында индекстелген. Күнтізбелік жоспар бойынша жұмыстарды үйлестіру, контентті тексеру мен талдау, мақалаларды жариялауға дайындау жұмыстарын атқарады. Scopus ID: жоқ, ORCID ID: 0009-0002-9739-1775, Хирш индексі: 0. Серікбай Бауыржан – магистр, AI-специалист, Шайсұлтан Шаяхметов атындағы «Тіл-Қазына» ұлттық ғылыми-практикалық орталығының ғылыми қызметкері. Жоба аясындағы еңбек өтілі – 4 жылдан астам. 2021–2024 жылдар аралығында Қазақ ұлттық корпусының бес подкорпусы, «Бала тілі» балалар порталы, қазақ тіліндегі TTS жобасы, сөйлеу синтезіне арналған чат-боттар, көру қабілеті бұзылған жандарға арналған «Экран дикторы» жобасы сияқты жұмыстарға қатысқан. Негізгі міндеттері: тапсырмаларды үлестіру, деректерді парсингтеу, скрипттерді әзірлеу, деректерді жинау және өңдеуді автоматтандыру, нормализациялау. Scopus ID: — ORCID ID: 0000-0000-0000-0000 Хирш индексі: 0 Алшынбекова Мәдина Серікболсынқызы – магистр, ЕҰУ-дің PhD докторанты, «Тіл-Қазына» ұлттық орталығының қолданбалы лингвистика бөлімінің ғылыми қызметкері.

	Жобаның негізгі орындаушысы, үйлестіру, редакциялау, контентті талдау және жарияланымдарды дайындау процесін бақылайды. Қытай Халық Республикасындағы Сиань шет тілдер университетінің гуманитарлық ғылымдар магистрі. Ғылыми қызығушылығы – салыстырмалы фразеология. 10-нан астам ғылыми жарияланымның авторы, олардың кейбірі Scopus базасына енгізілген. Scopus ID: — ORCID ID: 0009-0002-9739-1775 Хирш индексі: 0 Демесінова Лэйлә Мұратқызы – PhD, ғылыми қызметкер. Педагогикалық өтілі – 17 жылдан астам. Жоба аясындағы негізгі міндеттері: үйлестіру, редакциялау, грамматикалық және құрылымдық тексеру, мета-белгілеу, талдау және ғылыми мақалаларды дайындау. 40-тан астам ғылыми жарияланымның, 2 монография мен 2 оқулықтың авторы, олардың бір бөлігі Scopus және КОКСОН базаларына енген. Scopus ID: 58242361100 ORCID ID: 0000-0002-2414-7360 Хирш индексі: 0 AP08855981 (2020–2022), AP19677903 (2023–2026) жобаларына қатысқан. «Тіл-Қазына» орталығының «Тіл-Құрал» мобильді қосымшасы, сөйлеу синтезі, лексика-грамматикалық минимум жобаларына атсалысқан. Гүлназ Төкенқызы – филология ғылымдарының кандидаты, доцент, «Тіл-Қазына» орталығының қолданбалы лингвистика бөлімінің жетекші ғылыми қызметкері. Ассоциирленген профессор, ғылыми өтілі – 30 жыл. 100-ден астам ғылыми мақала, оның ішінде КОКСОН базасында 30-дан астам және Scopus-та жарияланған еңбектері бар. 17 оқулық пен электронды оқу құралдарының авторы және теңавторы. ORCID ID: 0000-0003-4993-9927 Хирш индексі: 3 2011 жылғы «ЖОО-ның үздік оқытушысы» мемлекеттік сыйлығының иегері, KazTest сарапшысы (2011–2023), ЕҰУ жанындағы диссертациялық кеңестің мүшесі. Эпостық мұра, жастар поэзиясы, Ахмет Байтұрсынұлының тұлғасы бойынша ғылыми жобалардың жетекшісі және орындаушысы. Қазақ ұлттық корпусы жобасы мен «Qazgramma» жобасының контент редакторы. Дана Оспанова Жаңабекқызы – магистр («Қазақ тілі мен әдебиеті», 2017), Л.Н. Гумилев атындағы Еуразия ұлттық университетінің PhD

	докторанты. Шайсұлтан Шаяхметов атындағы «Тіл–Қазына» ұлттық ғылыми-практикалық орталығының қолданбалы лингвистика бөлімінің ғылыми қызметкері, жоба жетекшісі. 8 жылдан астам ғылыми тәжірибесі бар. 2021–2024 жылдары Қазақ ұлттық корпусының бес ішкі корпусы жобасының жетекшісі болды. Жобалық жұмыстарды үйлестіру, контентті өңдеу, тексеру, талдау және ғылыми жарияланымдарды дайындауды бақылау бойынша жұмыстарды атқарады. 15-тен астам ғылыми жарияланымның авторы, соның ішінде Scopus базасындағы мақалалары бар. «Qazgramma» жобасында контент-корректор ретінде, «Тіл және қоғам» журналы мен qazcorpora.kz порталына материал жинау және дайындау бойынша орындаушы ретінде жұмыс істеген. Scopus ID: — ORCID ID: 0000-0001-8901-8060 Хирш индексі – 0. София Кожиакпарқызы Омарова – филология ғылымдарының кандидаты, Шайсұлтан Шаяхметов атындағы «Тіл–Қазына» ұлттық ғылыми-практикалық орталығының қолданбалы лингвистика бөлімінің меңгерушісі. Қазақ тілін оқыту әдістемесі бойынша 45-тен астам ғылыми мақала, 2 монография және 20 патенттің авторы. Қазақ тілін латын графикасына көшіру, аударма ісі (2013 жылы «Түркі альманағы» мен «Түркілік тәрбие» кітабының аудармасы), туыстас тіл ретінде оқыту әдістемесі, оқу сөздіктер мен бейнесабақтар жасау бойынша ғылыми жобаларға кеңінен қатысқан. Scopus ID: — ORCID ID: 0009-0009-0819-0543 Хирш индексі – 0. Мемлекеттік марапаттар мен құрмет белгілерінің иегері. Меруерт Ермахан – магистр, Шайсұлтан Шаяхметов атындағы «Тіл–Қазына» ұлттық ғылыми-практикалық орталығының ғылыми қызметкері. Білімі – халықаралық және еуропалық құқық магистрі, Римдегі Ла Сапиенца университеті (Италия), LazioDisco стипендиаты. Деректерді жинау, талдау, өңдеу және ағылшын тілінен қазақ тіліне аудару бойынша біліктілігі бар. «Тіл–Қазына» орталығының жобаларына қатысқан. Контентті редакциялау, тексеру және пішімдеуге маманданған. Scopus ID: — ORCID ID: 0000-0000-0000-0000 Хирш индексі – 0.

	Молдир Бақытқызы – Шайсұлтан Шаяхметов атындағы «Тіл–Қазына» ұлттық ғылыми-практикалық орталығы Til-Hub бөлімінің жетекшісі, магистр. Рөлі: Орындаушы. Міндеттері: күнтізбелік жоспарға сәйкес жобалық жұмыстарды үйлестіру, контентті өңдеу, орфографиялық және грамматикалық түзету, мәтіннің құрылымдық талдауы, метабелгілеу және мақалаларды жариялауға дайындау. Лиза Есбосын Қожаққызы – «Тіл – Қазына» ұлттық ғылыми-практикалық орталығының атқарушы директоры, магистр. Рөлі: Орындаушы. Міндеттері: жобалық жұмысты үйлестіру, ғылыми контентті тексеру және редакциялау, жарияланым қызметін қадағалау. Нұрсұлтан Еламанов Әлібекұлы – Ш. Шаяхметов атындағы «Тіл – Қазына» ұлттық ғылыми-практикалық орталығының IT-маманы, магистр. Міндеттері: үлкен көлемдегі деректерді сақтау және басқаруға арналған дерекқорды әзірлеу және қолдау, деректерді сақтық көшіру және қалпына келтіру жүйесін баптау. Веб-порталдарды әзірлеу саласында 7 жылдық тәжірибесі бар. Сандуғаш Кенжебаева Баймаханқызы – Ш. Шаяхметов атындағы «Тіл – Қазына» ұлттық ғылыми-практикалық орталығы жобаларды басқару орталығының жетекшісі, магистр. Міндеттері: тапсырмалар мен күнтізбелік жоспарға сәйкес жұмыстарды үйлестіру, жиналған контентті тексеру, өңдеу және талдау. Медиа саласының сарапшысы, сертификатталған медиа-тренер. Әртүрлі халықаралық жобаларда үлкен тәжірибесі бар. АҚШ Мемлекеттік департаменті ұйымдастырған Digital Communication Network грантының иегері. Әртүрлі дереккөздермен (аудио, видео, мәтін) жұмыс істеу, түрлендіру, метабелгілеу бойынша білімі мен тәжірибесі бар. Scopus ID: жоқ, ORCID ID: 0000-0000-0000-0000, Хирш индексі: 0. Олжас Ибраев Пазылбекұлы – магистр, «Тіл – Қазына» ұлттық ғылыми-практикалық орталығының бас директорының кеңесшісі. Рөлі: Орындаушы. Міндеттері: контентті үйлестіру, тексеру және редакциялау. Басқарушылық және кәсіпкерлік тәжірибесі бар: DRIFT компаниясының

	директоры (қазіргі уақытта), Green Innovation ЖШС-нің бас директоры (2018–2024), Market Connect ЖШС-нің атқарушы директоры (2017–2018), ННТЭО-да бас маман (2014–2016). Әсел Исмайлова Амангелдіқызы – экономика ғылымдарының магистрі, «Тіл – Қазына» ұлттық ғылыми-практикалық орталығының IT бөлімінің маманы. Рөлі: Орындаушы. Міндеттері: контентті үйлестіру және редакциялауға қатысу, оның ішінде қазақ ауызша сөйлеуін синтездеу жобасы аясында. Ұлттық LLM моделін әзірлеу жобасының үйлестірушісі. Байқаулар, семинарлар мен конференциялар ұйымдастырады, ғылыми-әдістемелік және ұйымдастырушылық мәселелер бойынша мемлекеттік органдармен байланыс жүргізеді. Латын графикасы бойынша конференция материалдар жинағының корректор-редакторы болған. Scopus ID: жоқ, ORCID ID: 0000-0000-0000-0000, Хирш индексі: 0. Жазира Искақова Мақсұтқызы – филология ғылымдарының кандидаты, Шайсұлтан Шаяхметов атындағы «Тіл – Қазына» ұлттық ғылыми-практикалық орталығының әдістемелік бөлімінің меңгерушісі. Жобаның орындаушысы. Жоба аясында тапсырмалар мен күнтізбелік жоспар бойынша жұмысты үйлестірумен, жиналған контентті тексеріп, редакциялаумен, ғылыми мақалаларды жариялауға дайындау барысын бақылаумен айналысады. Ғылыми-әдістемелік салада 20 жылдан астам тәжірибесі бар. Қазақ тілін екінші тіл ретінде оқыту және қазақ әдебиеті бойынша 60-тан астам ғылыми мақала мен 5 оқу құралының авторы. ORCID ID: 0000-0002-0940-615X Жамал Айтқалиевна Манкеева – Ахмет Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері, филология ғылымдарының докторы, лексикология, лексико-семантика және корпус тіл білімі саласындағы жетекші мамандардың бірі. 200-ден астам ғылыми жарияланымдардың, оның ішінде монографиялар мен халықаралық басылымдардағы мақалалардың авторы. Жобада әртүрлі тілдік құрылымдардан алынатын дереккөздерді таңдау және негіздеу жұмыстарын атқарады. Scopus ID: 57204363069 ORCID: 0000-0002-7390-4778

	h-индексі: 2 Рүстембек Нұсқабекұлы Шойбеков – Ахмет Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері, лексикология, этнолингвистика және лингвомәдениеттану салаларында жетекші маман. Энциклопедиялық сипаттағы іргелі еңбектердің авторы. Жоба аясында әртүрлі стильдегі тілдік құрылымдардан алынатын дереккөздерді іріктеу және негіздеу жұмыстарымен айналысады. Scopus ID: жоқ ORCID: 0000-0002-8979-3566 h-индексі: 0 Мырзаберген Малбақұлы Малбақов – Ахмет Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері, түркітану, тіл тарихы және лексикография салаларының маманы. Бұл бағыттағы отандық жетекші зерттеушілердің бірі. Жоба аясында лексикографиялық ресурстардан алынатын дереккөздерді таңдау және негіздеуге жауапты. Scopus ID: 56195379100 ORCID: 0000-0001-8485-0544 h-индексі: 0 Бағдан Қатайқызы Мамынова – Ахмет Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері. Грамматика, мәтінтану, морфология және прагмалингвистика саласындағы белгілі ғалым. 250-ден астам ғылыми жарияланымның, оның ішінде 15 монография мен Scopus базасында индекстелетін 3 іргелі мақаланың авторы. Жоба аясында тарихи және әлеуметтік контекстердегі тілдік дереккөздердің негіздемесімен айналысады. Scopus ID: 56176636300 ORCID: жоқ h-индексі: 1 Сарсенбай Қуантайұлы Құлманов – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері, Терминология орталығының жетекшісі. Грамматика, терминология және аудармашылық саласының маманы. 100-ден астам ғылыми еңбектің авторы. Жоба аясында оқыту модельдеріндегі әртүрлілікті қамтамасыз ету үшін терминологиялық деректерді негіздеу және жүйелеумен айналысады. Scopus ID: жоқ ORCID: 0000-0001-8289-9299 h-индексі: 0

	Әлфия Әбдікенқызы Сұлтанбекова – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері, морфология, синтаксис, функционалды грамматика және лексикография саласының маманы. 100-ге жуық ғылыми еңбектің авторы. Жоба аясында түрлі типтер мен контекстерде ұсынылған тілдік құрылымдар негізінде дереккөздерді іріктеу және негіздеумен айналысады. Scopus ID: жоқ ORCID: 0000-0002-3513-4012 h-индексі: 0 Айнұр Аташбекқызы Сейтбекова – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері, тіл тарихы, түркітану, араб және парсы тілдері саласының маманы. Қазақ тілінің Ұлттық корпусының тарихи ішкікорпусын әзірлеуге қатысады. Жоба аясында тарихи дереккөздерден тілдік деректерді іріктеумен айналысады. Scopus ID: 572055459500 ORCID: 0000-0002-3513-4012 h-индексі: 0 Эльмира Әбдуғалиқызы Өтебаева – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері. Этнолингвистика, концептология, лексикография және неография саласының маманы. 60-тан астам ғылыми мақаланың авторы. Қазақ тілінің мәдени-репрезентативті ішкікорпусының бірлескен әзірлеушісі. Жоба аясында этномәдени контексті бар тілдік деректерді талдаумен айналысады. Scopus ID: 57200244748 ORCID: жоқ h-индексі: 1 Нұрсәуле Мақсұтқызы Рсалиева – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері, терминология, аудармотану, ономастика және лексикография саласының маманы. Oxford қазақша-ағылшынша және ағылшынша-қазақша сөздігінің бас редакторы және құрастырушысы. Ұлттық корпустың параллель және жазушы ішкікорпустарының әзірлеушілерінің бірі. Жоба аясында түрлі стиль, жанр және типтегі дереккөздерді тексеру және талдауға жауапты. Scopus ID: 410650707 ORCID: 0000-0003-2065-0505 h-индексі: 0

	Гүлжихан Байқазықызы Кубденова – А. Байтұрсынұлы атындағы Тіл білімі институтының жетекші ғылыми қызметкері, түркі тілдерін салыстырмалы-тарихи тұрғыда зерттеу саласының маманы. 50-ге жуық ғылыми мақаланың авторы. Қазақ тілінің Ұлттық корпусының тарихи ішкі корпусының әзірлеушілерінің бірі. Жоба аясында әртүрлі тарихи мәтіндерден мәліметтерді іріктеумен айналысады. Scopus ID: жоқ ORCID: 0000-0001-7514-8710 h-индексі: 0 Арайлым Ботанқызы Шормақова – А. Байтұрсынұлы атындағы Тіл білімі институтының аға ғылыми қызметкері, этнолингвистика, лингвомәдениеттану, лексикология және лексикография саласының маманы. 30-дан астам ғылыми жарияланымның авторы. Жоба аясында іріктелген тілдік деректерді топтастырумен айналысады. Scopus ID: 57211905258 ORCID: 0000-0001-9551-3818 h-индексі: 1 Ақмарал Зинол-Ғабденқызы Бисенғали – А. Байтұрсынұлы атындағы Тіл білімі институтының аға ғылыми қызметкері. Түркология, терминология, лексикология және этнолингвистика саласының маманы. 40-қа жуық ғылыми мақаланың авторы. Жоба аясында қоғамдық өмірдегі нақты тілдік қолданыстарды бейнелейтін терминологиялық дереккөздерден алынған мәліметтерді сәйкестендіру және салыстырумен айналысады. Scopus ID: 57200075552 ORCID: 0000-0002-1964-9948 h-индексі: 1 Нұрлан Құнпияұлы Шүленбаев – А. Байтұрсынұлы атындағы Тіл білімі институтының аға ғылыми қызметкері, аудармատану және терминология саласының маманы. 40-тан астам ғылыми мақаланың авторы. Жоба аясында электронды ресурстардан дереккөздерді таңдаумен айналысады. Scopus ID: жоқ ORCID: 0000-0003-2203-9822 h-индексі: 0 Зубайра Боранбекқызы Садырбаева – А.Байтұрсынұлы атындағы Тіл білімі институтының ғылыми қызметкері, ономастика және этнолингвистика саласының маманы. Жоба аясында ономастикалық есімдерге

	лингвистикалық белгілеу жұмыстарын жүргізеді. Scopus ID: жоқ ORCID: 0000-0002-2729-4198 h-индекс: 0 Аида Серікханқызы Барменқұлова – А.Байтұрсынұлы атындағы Тіл білімі институтының ғылыми қызметкері, поэзиялық мәтінге стилистикалық талдау жасау бойынша маман. Жоба аясында үлкен тілдік модельдерді оқытуға арналған деректердің лингвобелгілеуін әзірлеумен айналысады. Scopus ID: жоқ ORCID: 0000-0002-1471-5846 h-индекс: 0 Талғат Бекболатұлы Рамазанов – А.Байтұрсынұлы атындағы Тіл білімі институтының ғылыми қызметкері, грамматика, тіл тарихы және салыстырмалы тіл білімі саласының маманы. Жоба аясында үлкен тілдік модельдерді оқытуға арналған мета және лингвобелгілеу жұмыстарын жүргізеді. Scopus ID: жоқ ORCID: 0000-0002-4405-5914 h-индекс: 0 Гүлден Бақытқазықызы Тіленова – А.Байтұрсынұлы атындағы Тіл білімі институтының ғылыми қызметкері, қолданбалы лингвистика саласының жас маманы. Жоба аясында үлкен тілдік модельдерді оқытуға арналған тілдік деректердің метабелгілеуін жасаумен айналысады. Scopus ID: жоқ ORCID: 0000-0001-9842-3324 h-индекс: 0 Әсел Қазбекқызы Сейдамат – А.Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, PhD докторанты, Қазақ ұлттық корпусының тарихи ішкі корпусының негізгі әзірлеушілерінің бірі. Жоба аясында әртүрлі стильдегі мәтіндердің лингвобелгілеуін жүргізеді. Scopus ID: жоқ ORCID: 0000-0003-1893-4920 h-индекс: 0 Меруерт Ануарқызы Имангазина – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, грамматика және салыстырмалы тіл білімі саласының маманы. Ағылшын тілін еркін меңгерген (IELTS – 7.0, C1 деңгейі), 10-ға жуық

	ғылыми жарияланымның авторы. Жоба аясында тілдік модельдерді үйретуге арналған деректерге лингвистикалық таңбалау жүргізеді. Scopus ID: жоқ ORCID: 0000-0001-8473-2812 h-индексі: 0 Айкерім Мурсал – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, PhD-докторант, Қазақ ұлттық корпусының тарихи қосалқы корпусы бойынша жетекші мамандардың бірі. Жоба аясында тарихи мәтіндердің деректеріне лингвистикалық таңбалау жүргізеді. Scopus ID: жоқ ORCID: 0000-0001-8035-4236 h-индексі: 0 Қымбат Берікұлы Слямбеков – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, грамматика және стилистика саласының маманы. 10-ға жуық ғылыми мақаланың авторы. Жобада ірі тілдік модельдерді оқытуға арналған деректерге мета және лингвистикалық таңбалау жүргізумен айналысады. Scopus ID: жоқ ORCID: 0000-0002-9731-0448 h-индексі: 0 Маржан Серікқызы – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, PhD-докторант, этнолингвистика саласының маманы. 10-ға жуық ғылыми жарияланымның авторы. Жобада этнолингвистикалық деректерге мета және лингвистикалық таңбалау жүргізумен айналысады. Scopus ID: жоқ ORCID: 0000-0002-4547-1928 h-индексі: 0 Айдана Мұхымбайқызы Махамбетова – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, социолингвистика саласының маманы. Жоба аясында социолингвистикалық материалдарды жинақтау және жүйелеумен айналысады. Scopus ID: жоқ ORCID: 0009-0005-5846-7433 h-индексі: 0 Аягүл Даниярқызы Омарова – А. Байтұрсынұлы атындағы Тіл білімі институтының кіші ғылыми қызметкері, PhD-докторант, терминология және терминотворчество саласының маманы. 10-ға

	жуық ғылыми жарияланымның авторы. Жоба аясында терминологиялық деректерге мета және лингвистикалық таңбалау жүргізеді. Scopus ID: жоқ ORCID: 0000-0002-4948-6764 h-индексі: 0
Басылымдар тізімі және олардың сілтемелері	1. Sofya Omarova. A Corpus Approach in Language Discovery: A Word Frequency Analysis Based on the Corpus Outcomes in Kazakh. Forum for Linguistic Studies (FLS). View Vol. 7 , Iss. 4 (April 2025): In Progress ISSN: 2705-0602 (Online) 2. Akbayan Bekarystankyzy, Abdul Razaque, Orken Mamyrbayev Integrated end-to-end multilingual method for low-resource agglutinative languages using Cyrillic scripts // Journal of Industrial Information Integration, Volume 43, January 2025, 100750, https://doi.org/10.1016/j.jii.2024.100750 3. Shaikhiyeva, Z., Mansurova, M., & Amirkhanova, G. (2024). Text to SQL Transformation Using LLM: A Comparative Research of T5, Seq2Seq, and SQLNet Models. Proceedings of the IX International Conference on Computer Science and Engineering (UBMK-2024) (pp. 967–972). IEEE. ISBN: 979-8-3503-6587-0, p.967-972. 4. Kairatuly, B., & Mansurova, M. (2024). Named Entity Recognition from Kazakh Speech. Proceedings of the IX International Conference on Computer Science and Engineering (UBMK-2024). IEEE. ISBN: 979-8-3503-6587-0, p.153-156. 5. Шайхиева, Ж., Мансурова, М., Амирханова, Г., & Оспан, А. (2024). Кластеризация комментариев с использованием большой языковой модели в финансовых технологиях. Информатика и прикладная математика: Материалы IX Международной научно-практической конференции (31 октября – 1 ноября 2024 г.) (сс. 444–451). Алматы. 6. Mansurova, M. E., Rakhimova, D. R., & Duisenbekkyzy, Zh. (2024). Morphological Parsing of Kazakh Texts with Deep Learning Approaches. Journal of Mathematics, Mechanics and Computer Science (JMMCS). ISSN: 1563–0277, eISSN: 2617-4871. Retrieved from https://bm.kaznu.kz. DOI: https://doi.org/10.26577/JMMCS.2022.v113.i1.13. 7. Rakhimova, D., Mansurova, M., Matanov, N., & Yerik, A. (2024). Machine Learning-Based Synonymous Word Detection in Kazakh. Proceedings of the IX International Conference on Computer Science and Engineering (UBMK-2024). IEEE. ISBN: 979-8-3503-6587-0.