



ХҒТАР 19.01.11
Ғылыми мақала

<https://doi.org/10.32523/2616-7174-2026-156-3-64-82>

Жасанды интеллект жүйелеріндегі қазақ тілі: мәдени контекст және цифрлық даму мәселелері

Ұ.М. Есенбекова*¹, Н.Д. Стамбекова², Ы.Д. Егембердиев³

^{1,2}А.Н.Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Астана

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

(E-mail: ¹*ultmudde@mail.ru, ²nazymstambekova@gmail.com, ³iskak_khabar@mail.ru)

Аңдатпа. Зерттеу мақаласында жасанды интеллект жүйелерінің қазақ тілі мен ұлт мәдениетін дамытуға ықпалы талданды. Қазақ тіліндегі ЖИ жүйелерінің грамматикалық, семантикалық және мәдени сапасына баға берілді. Сауалнама, салыстырмалы аударма және мәтіндік талдау нәтижелері ЖИ жүйелерінде қазақ тіліндегі деректер қорының жеткіліксіздігін көрсетті. Қарастырылған аудармаларда тілдік сәйкессіздіктер мен мәдени контекстің ішінара жоғалуы байқалды.

Методологиялық тұрғыда зерттеу жұмысы Стиг Хьярвард (Stig Hjarvard) пен Андреас Хепптің (Andreas Hepp) медиатизация, Элиху Кац (Elihu Katz) және Джей Блумлердің (Jay G. Blumler) пайдалану, қанағаттану теорияларына сүйенеді. Аталған теориялар ЖИ жүйелерін қазіргі медиамәдени кеңістіктің маңызды бөлігі ретінде қарастырып, тілдің цифрлық ортадағы рөлін түсіндіруге мүмкіндік береді. Қанағаттану теориясы ЖИ технологияларының қоғамдық ақпараттық қажеттіліктерін қанағаттандырудағы орнын көрсетеді.

Зерттеу нәтижелері негізінде авторлар қазақ тілін цифрлық ортада дамыту үшін ұлттық тіл корпустарын кеңейтуді, IT және лингвистика салаларының өзара байланысын нығайтуды, білім беру жүйесінде цифрлық тіл мәдениеті пәндерін енгізуді ұсынады. Бұл бағыттар жүзеге асқан жағдайда қазақ тілі тек ұлттық мұра емес, сонымен қатар заманауи технологияларда дамиды, бәсекеге қабілетті тілге айналатыны туралы тұжырым жасалады.

Түйін сөздер: жасанды интеллект, машина аудармасы, тіл корпусы, мәдени контекст, медиатизация.

Кіріспе

Цифрлық орта жасанды интеллект жүйелерінің қоғамдық қарым-қатынастарға еркін енуіне ыңғайлы ережелер жасауда. Адамзаттың жаңа технологияны ойлап тауып, игеріп, игілігін көруі қуантарлық жағдай. Бірақ цифрлық технология мен жасанды интеллекттің

Түсті: 4.09.2026; Мақұлданды: 26.09.2026; Онлайн қолжетімді: 30. 09. 2026

қарқынды дамуына бейімделу де үлкен проблемалар туғызуда. Жаһанды өзгертіп жатқан технологиядан Қазақстанның қалыс қалмайтыны анық. Ел алдында тұрған мақсат – тілімізді цифрлық кеңістікте өркендету болып отыр.

Мемлекет басшысы да озық технологияларды қазақ тілінде сөйлетуді стратегиялық міндет етіп қойды. Өйткені жасанды интеллект жүйелері мемлекеттік басқару мен экономикалық өсімге, ғылым, білім мен мәдени дамуға ықпал етуде. Тіл мен мәдениеттің дамуына ЖИ құралдары үлкен серпін беріп отыр. Аударма жасауға, шығармашылыққа, мәтін жазуға, тіл үйренуге жаңа технология зор мүмкіндік ұсынуда. Бұл құбылысты медиатизация теориясының аясында да қарастырған жөн. Себебі медиа мен технология тілдің қолданысын, формасын өзгеріске ұшыратады. Ал қазақ тілінің болмысы мен қалпын бұзбай, сақтап қалуына осы технология көмектесе ала ма? Бұл жерде ғылыми, тәжірибелік түйткілдердің туындайтыны сөзсіз.

Жасанды интеллект ұғымын да, бағдарламалық құралдарын да 1950 жылдары алғаш рет америкалық ғалымдар ойлап тауып, жасап шығарған. Бастапқыда бұл технология батыс елдерінде дамып, ағылшын тілі негізінде жасалды. Сондықтан аймақтық тілдерді бағдарламаға енгізу бірқатар мәселелер туындатып келеді. Оның үстіне жасанды интеллект ең алдымен тілдік деректерге сүйенеді. Белгілі бір тілде жеткілікті дерек болмаса, ол тіл жасанды интеллект жүйелерінде толық бейнеленбейді. Аз қолданылатын тілдерде дерек шағын болғандықтан, бұл тілдердегі аударма да, мәтін де сапасыз болады.

Бүгінгі цифрлық технологиядағы ілгерілеушіліктің ана тіліміз бен төл мәдениетіміздің дамуындағы іркілістерге әкелетін тұстары да бар. ChatGPT, Google Translate, DeepL сияқты тілдік модельдер мен аударма жүйелері адамдардың қарым-қатынасын жеңілдетері анық. Бірақ қазақ тілінің терең мағынасын ұғыну, түсіндіру, аудару ЖИ-дің қолынан келмеуі мүмкін. Бұндай кемшіліктер Google Translate аудармаларында жиі байқалады. Егер бұндай үдерістерді бақылаусыз жіберсек, бай, көркем әдеби тілімізді жұтаңдатып, технология тіліне айналдырып жіберу қаупі туындайды. Оның салдары онсыз да мемлекеттік тілді дамытып, қолданысын кеңейте алмай отырған ұлт үшін үлкен соққы екені рас (Qazconpro, 2025).

Ян Лекун өзінің *How the Machine Learns* атты еңбегінде машиналық оқыту жүйелерінің адам миының жұмыс принципіне ұқсастығын айтады. Оның пікірінше, жасанды интеллект тіл мен ақпаратты талдауда үлкен жетістікке жеткенімен, мағыналық және мәдени контексті түсінуде әлі де шектеулі. Сондықтан адам жаңа тілді қалай меңгерсе, ЖИ жүйелеріне де қазақ тіліндегі деректер базасын ауқымды түрде енгізіп, тілдің грамматикалық және семантикалық ерекшеліктерін қамтып, ереже-заңдылықтарға бейімдеу қажет. Бұл бағытта отандық IT мамандары мен тіл мамандарының бірлескен жұмысы қажет (LeCun, 2023).

Осы зерттеу жұмысында жасанды интеллекттің қазақ тілі мен мәдениетке ықпалы тек теориялық тұрғыда емес, тәжірибелік әдістер арқылы да талданды. Зерттеуде ЖИ жүйелеріндегі қазақ тілінің қолданылу сапасын анықтау үшін бірқатар тәжірибелер жүргізілді. Атап айтқанда:

- әртүрлі ЖИ құралдарының аударма сапасы салыстырылып, дәлдігі талданды (қазақша – орысша – қазақша мәтін аудармаларын сарапқа салу);
- морфологиялық талдау жүргізілді (ЖИ жүйелерінің қазақ тілінің грамматикасын дұрыс түсінуі тексеріледі);
- көркем әдебиеттің мағынасын тексеру (көркем әдебиеттен үзінді жазылып, ЖИ-дің түсіну деңгейі бағаланады).

Осындай тәжірибелер зерттеудің ғылыми негізін нақты мысалдармен толықтырып, негізгі проблемаларды айқын көрсетіп берді.

Жасанды интеллекттің қазақ тілі мен ұлттық мәдениетке ықпалы сауалнама жүргізу арқылы сарапталды. Сауалнама көмегімен қоғамның жасанды интеллектке деген көзқарасын, қазақ тіліндегі ЖИ жүйелерін пайдалану деңгейі мен олардың сапасы талданды. Сауалнама нәтижелері қазақ тілін цифрлық ортада дамытуға кедергі болып тұрған негізгі мәселелерді айқындап, оларды шешу жолдарын ұсынуға мүмкіндік берді. Сауалнамаға студенттер, оқытушылар, IT және филология мамандары, сондай-ақ ЖИ құралдарын күнделікті қолданатын пайдаланушылар қатысты. Сұрақтар қазақ тіліндегі жасанды интеллект жүйелерінің қолданысын, аударма сапасы мен мәдени ұғымдарды жеткізу дәлдігін қамтыды. Әлеуметтік сауалнама тәсілі зерттеудің ғылыми өзегін нақты деректермен толықтырып, теориялық тұжырымдарды тәжірибемен ұштастыруға мүмкіндік береді.

Жасанды интеллект жүйелеріндегі қазақ тілінің сапасын арттыру тек технологиялық емес, ұлттық деңгейдегі маңызды міндет. Тілдің цифрлық ортада белсенді, ауқымды және дұрыс қолданылуы оның өміршеңдігін арттырады. Қазақ тілі заманауи жүйелерге толық бейімделмесе, болашақ ұрпақтың ана тілінде ақпарат іздеу, оқу немесе жазу мүмкіндігі шектеледі. Қазақ тілін цифрлық кеңістікке дұрыс енгізу – елдің мәдени тұтастығы мен тілдік тәуелсіздігін қамтамасыз етудің бірден бір кепілі. Бұл бағытта мемлекеттік қолдаумен арнайы тілдік деректер базасын құру, қазақ тіліндегі қолданбаларды көбейту және білім беру саласында цифрлық сауаттылықты арттыру маңызды.

Зерттеудің мақсаты – жасанды интеллект жүйелерін тіл мен мәдениетті дамыту саласында қолдану мүмкіндіктері мен проблемаларын талдау және оларды шешу жолдарын ұсыну. Зерттеу нысаны – ЖИ-дің тіл мен мәдениетке ықпалы, ЖИ технологияларын қолдану барысында туындайтын, тілдік, мәдени, этикалық мәселелер.

ЖИ технологиялары бұқаралық коммуникация құралдарын түбегейлі өзгерте бастады. Бүгінде ақпараттық порталдар, жаңалық тарататын агенттіктер мәтіндерін, тақырыптарын, тіпті қысқаша шолуларды ЖИ құралдарымен дайындайды. Бұндай жағдайда тілдің дұрыс қолданылуы, стильдің бірізділігі мен мағыналық нақтылық ерекше маңызға ие. Себебі жасанды интеллекттің тілдік қателері мен сапасыз аудармалары ақпараттың бұрмалануына, ана тілдің дамуының тежелуіне әкелері анық. Бұл оқырмандардың білімін көтеріп, сауатын арттыруы тиіс медиа құралдардың халыққа кері әсерін күшейтеді.

Әдебиетке шолу

Жасанды интеллект жүйелерінің тіл мен мәдениетке ықпалын зерттеу қазіргі медиакоммуникация, цифрлық лингвистика және машина этикасы салаларындағы маңызды ғылыми бағыттардың біріне айналды. Бұл, әсіресе, цифрлық ортада дерек қоры шектеулі, морфологиялық құрылымы күрделі және мәдени-тілдік контексті терең тілдер үшін өзекті. Қазақ тілі дәл осындай тілдердің қатарына жатады. ЖИ жүйелеріндегі қазақ тілінің сапасын талдауды тек технологиялық мәселе емес, ұлттық тілдің цифрлық өміршеңдігі мен мәдени тұтастығын сақтауға байланысты ғылыми проблема ретінде қарастыру керек.

Бұл зерттеудің теориялық негізін медиатизация теориясы, пайдалану мен қанағаттану теориясы құрайды. С. Хьярвард медиатизацияны қоғам мен мәдениеттің медиа логикасына тәуелді болу үдерісі түрінде сипаттайды. Оның тұжырымы

бойынша, медиа тек ақпарат жеткізуші құрал емес, әлеуметтік институттардың, мәдени тәжірибелердің және күнделікті коммуникацияның құрылымын өзгертетін күшке айналуға. А. Хепп медиатизацияны мәдени өзгерістермен байланыстырып, қазіргі қоғамдағы коммуникациялық тәжірибелердің технологиялық жүйелер арқылы қайта қалыптасатынын көрсетеді. Осы тұрғыдан алғанда, ЖИ жүйелері қазіргі медиатизацияның жаңа кезеңін айқындайды. Өйткені олар мәтінді таратып қана қоймай, оны автоматты түрде өңдейді, ықшамдайды, аударарды және қайта құрастырады (Hjarvard, 2013; Hepp, 2013; Naveen, 2024).

Пайдалану мен қанағаттану теориясы да жасанды интеллект құралдарын түсіндіруде маңызды. Э. Кац, Дж. Блумлер және М. Гуревич аудиторияның медиа құралдарды белгілі бір ақпараттық, танымдық және практикалық қажеттіліктерді қанағаттандыру үшін пайдаланатынын дәлелдеген. Қазіргі жағдайда ChatGPT, Google Translate, DeepL және басқа да ЖИ жүйелерін пайдаланушылар жылдам аударма жасау, мәтін жазу, тіл үйрену және шығармашылық контент дайындау үшін қолдануда. Бұл жерде басты мәселе – қолжетімділік пен сапаның арақатынасына келіп тіреледі. ЖИ құралы мәтінді жылдам ұсынғанымен, оның грамматикалық дұрыстығы, мағыналық дәлдігі және мәдени контексті жеткізу қабілеті әрдайым жоғары бола бермейді (Katz et al., 1973).

Халықаралық зерттеулерде жасанды интеллект жүйелерінің көптілді ортадағы мүмкіндіктері мен шектеулері кеңінен талқыланды. ЮНЕСКО генеративті жасанды интеллектті білім мен зерттеу саласында қолдану кезінде адамға бағытталған қағидаттарды, мәдени әртүрлілікті және тілдік инклюзияны сақтаудың маңызын атап өтеді. Бұл тұжырым қазақ тілінің цифрлық ортадағы болашағына да тікелей қатысты. Өйткені тілдік әртүрлілік ескерілмеген жағдайда ЖИ жүйелері ірі тілдердің құрылымын басым көрсетіп, шағын әрі дерек қоры шектеулі тілдердің ерекшеліктерін толық қамтымауы мүмкін (Miao & Holmes, 2023).

Машина аудармасы мен көптілді модельдерге арналған соңғы зерттеулер де осы мәселені растайды. W. Zhu және әріптестері ірі тілдік модельдердің көптілді аудармада үлкен әлеуеті бар екенін көрсеткенімен, төмен ресурсты тілдерде олардың нәтижесі тұрақты емес екенін атап өтеді. Машина аудармасындағы басты қиындықтардың бірі – контексті, көпмағыналы сөздерді, идиомаларды, мәдени реңктерді және салалық терминдерді дәл жеткізу. Бұл қазақ тілі үшін ерекше маңызды. Себебі қазақ тіліндегі мақал-мәтелдер, тұрақты тіркестер, көркем мәтіндер мен ұлттық ұғымдар сөзбе-сөз аударма арқылы толық ашылмайды. Ұсынылған зерттеу жұмысы да осы проблеманы нақты мысалдармен ашып берді (Zhu et al., 2024; Russell, & Norvig, 2021).

Қазақ тілін табиғи тілді өңдеу жүйелеріне енгізу бағытында соңғы жылдары бірқатар маңызды зерттеулер жүргізілді. Қазақ тілінің ұлттық корпусы мен оның субкорпустары қазақ тіліндегі мәтіндерді жинақтау, жүйелеу, аннотациялау және ғылыми талдау үшін маңызды база қалыптастыра бастады. Бұндай корпустар тілдің сөздік қорын, грамматикалық құрылымын, стильдік ерекшеліктерін және нақты қолданыстағы мағыналық реңктерін зерттеуге көмектеседі. ЖИ жүйелерінің сапасы оқытуға берілетін деректердің көлемі мен сапасына тәуелді болғандықтан, ұлттық тіл корпустарын кеңейту қазақ тіліндегі жасанды интеллект жүйелерін дамытудағы негізгі шарттардың біріне айналды.

Қазақ тіліндегі арнайы ЖИ модельдері мен деректер жиынтықтары да қалыптасып келеді. Мысалы, Small Kazakh Language Corpus жобасы қазақ тіліндегі мәтіндерді жинақтап, көптілді NLP жүйелерінің тиімділігін арттыруды мақсат етеді. Astana IT

University мен Назарбаев университетінің ғалымдарының бірлескен зерттеуі тілге бейімделген корпус арқылы XLM-RoBERTa сияқты көптілді модельдердің нәтижесін жақсартуға болатынын көрсетті. Ал KazQAD деректер жиынтығы қазақ тіліндегі ашық доменді сұрақ-жауап жүйелерін дамытуға бағытталған. KazMMLU деректер жиынтығы қазақ, орыс және Қазақстанға қатысты аймақтық білім бойынша тілдік модельдерді бағалауға арналған. Бұл жұмыстар қазақ тіліндегі ЖИ жүйелерін тек аударма құралы ретінде емес, білім, мәдениет, мемлекеттік басқару және қоғамдық коммуникация салаларында қолдануға болатын кешенді технология ретінде дамытуға жол ашады (Togmanov et al., 2025; Yessenbekova et al., 2024).

Назарбаев университетінде іске асырылған ISSAI KAZ-LLM жобасы Қазақстанның ұлттық тілдік моделін қалыптастыру бағытындағы маңызды қадам ретінде бағаланады. Бұл модель қазақ, орыс, ағылшын және түрік тілдерінде контент генерациялауға бейімделген. Және Қазақстанның көптілді, көпмәдениетті жағдайын ескереді. Бұндай ұлттық модельдердің пайда болуы қазақ тілінің ЖИ жүйелеріндегі қолданысын күшейткенімен, оларды тұрақты дамыту, деректерді жаңарту, мәдени контексті тереңдету және тіл мамандарының сараптамасын енгізу әлі де өзекті болып қалуда (Tleubayeva et al., 2025; Borisov et al., 2025).

Қазақ тілінің морфологиялық ерекшеліктері де ЖИ жүйелерін дамытуда маңызды фактор саналады. Қазақ тілі жалғамалы тіл болғандықтан, бір сөздің құрамында бірнеше морфема, септік, көптік, тәуелдік, шақ және жақ көрсеткіштері қатар беріледі. Сондықтан автоматты морфологиялық талдау, сөздерді дұрыс бөлу, лемматизация және грамматикалық мағынаны анықтау күрделі міндет болып қалуда. Қазақ тіліндегі мәтінді автоматты түзетуге арналған гибриді ЖИ архитектуралары туралы зерттеулер ережеге негізделген әдістер мен нейрондық модельдерді ұштастырудың тиімді екенін көрсетеді. Бұл бағыт қазақ тіліндегі грамматикалық қателерді азайту, машина аудармасын жақсарту және мәтін сапасын арттыру үшін маңызды (Yeshpanov et al., 2023).

Осы әдебиеттерді талдау негізінде бірнеше ғылыми қорытынды жасауға болады. Біріншіден, жасанды интеллект жүйелері тілдік коммуникацияны жылдамдатып, мәтін жасау мен аударманы жеңілдетеді. Екіншіден, қазақ тілі сияқты қолданушысы аз, дерек қоры шектеулі тілдерде ЖИ жүйелерінің сапасы ұлттық корпустардың көлеміне, аннотация сапасына және мәдени деректердің қамтылуына тікелей байланысты. Үшіншіден, қазіргі зерттеулердің басым бөлігі техникалық модельдер мен деректер жиынтықтарына бағытталған. Әдебиетке шолуда қарастырылған еңбектер бойынша қазақ тіліндегі автоматты аударманың редакциялық сапасы мен мәдени ерекшеліктердің сақталуын өзара байланыстыра талдау қажеттігі байқалады. Авторлардың мақсаты – пайдаланушылар пікірлерін нақты аударма мысалдарымен ұштастырып, редакторлық тексеруді қажет ететін тұстарды көрсету болды.

Зерттеу әдістемесі

Зерттеу әдістемесі сапалық және сандық тәсілдерді ұштастырған аралас әдіске негізделді. Бұндай әдіс жасанды интеллект жүйелеріндегі қазақ тілінің қолданылу сапасын тек теориялық тұрғыдан емес, нақты тәжірибелік материалдар мен пайдаланушылардың пікірлері арқылы кешенді бағалауға мүмкіндік берді.

Зерттеу дизайны үш негізгі бағыттан тұрады. Біріншіден, қазақ тіліндегі жасанды интеллект жүйелерін қолдану тәжірибесін анықтау мақсатында сауалнама жүргізілді; екіншіден, машина аудармасының сапасын бағалау үшін салыстырмалы аударма талдауы

жасалды; үшіншіден, қазақ тіліндегі көркем және ақпараттық мәтіндердің жасанды интеллект жүйелері арқылы өңделу сапасы мәтіндік, семантикалық және мәдени талдау әдістері арқылы сарапталды.

Зерттеу барысында дерек жинаудың негізгі құралдарының бірі ретінде онлайн сауалнама қолданылды. Сауалнама 2025 жылдың қазан айында Google Forms платформасы арқылы жүргізілді. Респонденттер мақсатты іріктеу әдісімен таңдалды. Сауалнамаға тіл, медиа, білім беру және ақпараттық технология саласына қатысы бар студенттер, оқытушылар, журналистер, редакторлар мен IT саласының мамандары, сондай-ақ жасанды интеллект құралдарын күнделікті қолданатын пайдаланушылар тартылды.

Сауалнама сұрақтары бірнеше тақырыптық блокқа бөлінді. Бірінші блокта респонденттердің әлеуметтік-демографиялық сипаттамалары мен тілдік профилі анықталды. Екінші блокта жасанды интеллект жүйелерін қолдану жиілігі мен қолдану мақсаты қарастырылды. Үшінші блок қазақ тіліндегі жасанды интеллект құралдарының сапасын бағалауға арналды. Бұл бөлімде респонденттер аударма сапасы, грамматикалық дұрыстық, мағыналық дәлдік және мәдени контекстің сақталуы сияқты өлшемдер бойынша өз тәжірибесін көрсетті. Төртінші блокта қазақ тілін цифрлық ортада дамыту үшін қажетті ұсыныстар мен шешу жолдары жинақталды.

Сандық деректер сипаттамалық статистика әдісі арқылы өңделді. Респонденттердің жауаптары пайыздық көрсеткіштер арқылы жүйеленіп, жасанды интеллект құралдарын қолдану жиілігі мен қолдану мақсаты, қазақ тілінің сапасын арттыруға қатысты ұсыныстары талданды. Бұл әдіс қазақ тіліндегі ЖИ жүйелеріне қатысты негізгі мәселелерді анықтауға мүмкіндік берді.

Зерттеудің екінші кезеңінде салыстырмалы аударма талдауы жүргізілді. Бұл әдіс жасанды интеллект жүйелерінің қазақ тіліндегі грамматикалық, семантикалық және стильдік сапасын бағалау үшін қолданылды. Талдау барысында орыс тілінен қазақ тіліне және қазақ тілінен орыс тіліне аударылған мәтіндер салыстырылды. Аударма сапасын бағалау үшін ChatGPT 4 және Google Translate жүйелерінің нәтижелері қарастырылды. Бұл құралдар таңдалуының себебі – олардың көпшілік арасында кең таралуы және күнделікті аударма, мәтін жазу, ақпарат іздеу тәжірибесінде жиі қолданылуы болды.

Бұл жерде ChatGPT – үлкен тілдік модельге негізделген генеративті жүйе, ал Google Translate – машина аудармасына арналған сервис. Бұл зерттеуде олардың технологиялық құрылымы емес, бірдей мәтіндерді аудару нәтижелері салыстырылды. Аударма тәжірибесі 2025 жылдың 10-шы желтоқсанында жүргізілді. OpenAI әзірлеген ChatGPT 4 және Google Translate веб-сервисі пайдаланылды; қолжетімділік күні 2025 жылдың 10-шы желтоқсаны. ChatGPT-ге берілген нұсқаулық: «Мына мәтінді қазақшаға аудар». Әр мәтін бойынша алынған 2 жауап талданды. Талдауға бір жаңалық сөйлемі мен Абайдың он сегізінші қара сөзінен бір үзінді алынды. Екі жүйеден барлығы төрт аударма қарастырылды. Бұл материалдар жекелеген қателерді сипаттауға мүмкіндік береді, бірақ жүйелердің жалпы сапасын бағалауға жеткіліксіз.

Аударма мәтіндері бірнеше өлшем бойынша талданды. Біріншіден, грамматикалық дұрыстық бағаланды, яғни септік жалғауларының дұрыс қолданылуы, сөйлем құрылымы, сөздердің байланысуы және синтаксистік үйлесімділігі қарастырылды. Екіншіден, семантикалық дәлдік тексерілді, түпнұсқа мәтіндегі негізгі мағына аудармада қаншалықты сақталғаны тексерілді. Үшіншіден, стильдік сәйкестік бағаланды. Бұл өлшем аударманың қазақ тілінің табиғи қолданысына, журналистік немесе әдеби стиль

нормаларына қаншалықты сәйкес келетінін анықтауға көмектесті. Төртіншіден, мәдени контекстің берілуі қарастырылды. Бұнда ұлттық ұғымдар, тұрақты тіркестер, астарлы мағына, ирония және көркемдік реңктердің сақталу деңгейі сарапталды.

Бағалауды мақала авторлары сапалық талдау арқылы жүргізді. Балдық шкала қолданылмады. Түпнұсқа мен аударма салыстырылып, грамматикалық, мағыналық, стильдік және мәдени сәйкессіздіктер нақты үзінділермен сипатталды. Тәуелсіз бағалаушылар қатыспағандықтан, авторлық пайымның субъективтілігі зерттеудің шектеуі болып саналады.

Зерттеудің үшінші кезеңінде мәтіндік және семантикалық талдау әдісі қолданылды. Бұл әдіс жасанды интеллект жүйелерінің қазақ тіліндегі күрделі мәтіндерді, әсіресе көркем әдебиет үзінділерін түсіну және аудару қабілетін бағалауға бағытталды. Талдау материалы ретінде Абай Құнанбайұлының қара сөзінен алынған үзінді пайдаланылды. Бұл мәтіннің таңдалуы да кездейсоқ емес. Себебі Абайдың қара сөздерінде философиялық ой, ұлттық дүниетаным, ирония, астарлы мағына және бейнелі тіл кеңінен қолданылады. Сондықтан бұндай мәтін жасанды интеллект жүйелерінің тек сөздік деңгейдегі аударма жасауын емес, мәдени және мағыналық қабатты түсіну қабілетін бағалау үшін таңдалды.

Мәтіндік талдау барысында түпнұсқа мен жасанды интеллект жүйелері ұсынған аудармалар салыстырылып, мағыналық жоғалтулар, сөзбе-сөз аударма, калька, стильдік жұтаңдау және мәдени реңктің бұрмалануы анықталды. Бұл талдау қазақ тіліндегі көркем мәтіндердің жасанды интеллект жүйелерінде толық әрі дәл берілуі үшін тек тілдік дерек қоры ғана емес, мәдени және әдеби контексті қамтитын сапалы корпус қажет екенін көрсетіп берді.

Зерттеудің теориялық интерпретациясы медиатизация теориясы және пайдалану мен қанағаттану теориясы негізінде жасалды. Медиатизация теориясы арқылы жасанды интеллект жүйелерінің тілдік коммуникацияны қалай өзгертетіні, мәтін өндіру мен тілдік нормаларды цифрлық медиа логикасына қалай бейімдейтіні түсіндірілді. Ал пайдалану мен қанағаттану теориясы жасанды интеллект құралдарының пайдаланушылардың ақпараттық, танымдық және практикалық қажеттіліктерін қаншалықты қанағаттандыратынын бағалауға қолданылды. Осы екі теориялық бағыт зерттеу нәтижелерін тек техникалық проблема ретінде емес, медиамәдени және әлеуметтік құбылыс ретінде қарастыруға негіз болды.

Бұл мақалада зерттеу этикасы да толық ескерілді. Сауалнамаға қатысу ерікті түрде жүргізілді, респонденттердің жеке деректері жарияланбады және жауаптар тек ғылыми талдау мақсатында пайдаланылды. Осылайша зерттеу деректерінің құпиялылығы сақталып, қатысушылардың құқықтары қорғалды.

Зерттеудің шектеулері де болды. Зерттеуге 79 респондент қатысқандықтан, сауалнама нәтижелері тұтастай қоғамның пікірін білдірмейді. Аударма мысалдарының шектеулі саны да жүйелердің жалпы сапасын немесе бірінің екіншісінен тұрақты артықшылығын анықтауға жеткіліксіз. Нәтижелер тек зерттелген мәтіндерге және тәжірибе жүргізілген кездегі жүйе нұсқаларына ғана қатысты.

Осылайша зерттеу әдіснамасы қазақ тіліндегі ЖИ жүйелерінің сапасын жан-жақты бағалауға ұшталды. Сауалнама пайдаланушылардың тәжірибесі мен қабылдауын анықтаса, салыстырмалы аударма талдауы ЖИ жүйелерінің нақты тілдік қателерін көрсетті. Ал мәтіндік және мәдени талдау қазақ тілінің семантикалық, стильдік және ұлттық ерекшеліктерінің жасанды интеллект жүйелерінде қаншалықты сақталатынын

айқындады. Бұл әдістердің үйлесімі зерттеу мақсатына сәйкес келіп, қазақ тілін цифрлық ортада дамытуға қатысты ғылыми негізделген қорытындылар жасауға жол ашты.

Талдау мен басты нәтижелер

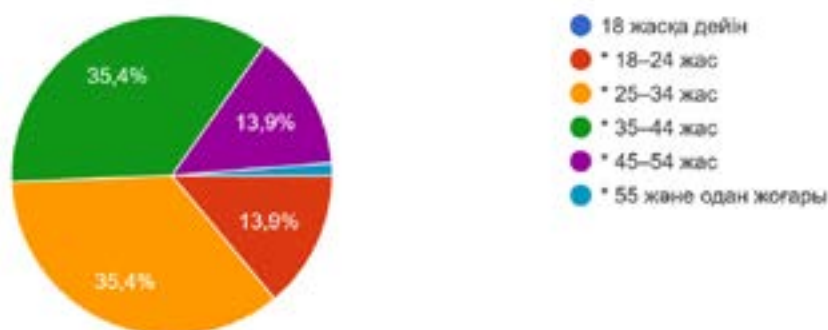
Жасанды интеллект жүйелері әр салаға, соның ішінде медиакеңістікке енгелі ақпарат өндірісі түбегейлі өзгерді. Бұрын тек адамға тән шығармашылық қызметті енді интеллектуалды алгоритм атқара бастады. Бұл құбылыс қазіргі медиатизация теориясының негізгі қағидаларын дәлелдейді. Аталған тұжырымға сәйкес, қоғамның барлық салалары медианың логикасына бейімделеді. Тілдік нормалар да сол логиканың ыңғайына қарай өзгереді. Маршал Маклюеннің тұжырымдамасы да ЖИ жүйелеріне қатысты. Ғалымның теориясы бойынша, ақпараттың мәні оны жеткізуші құралмен бірге өзгереді. Сол сияқты қазіргі жасанды интеллект жүйелері ақпаратты жай жеткізбейді, оны өзгертеді, бейімдейді. Салдарынан мағына бұрмаланып, контекст өзгереді және тілдің өзіндік ерекшеліктері трансформацияға ұшырайды.

Қанағаттану теориясы тұрғысынан қарағанда (Katz, Blumler, 1973) жасанды интеллект жүйелері пайдаланушылардың ақпаратқа деген қолжетімділігін жылдам әрі ыңғайлы етеді. Алайда ақпараттың қолжетімділік пен сапа арасындағы теңгерім бұзылған жағдайда тіл байлығына қауіп төнеді. Осылайша жасанды интеллекттің тіл мен коммуникация саласына енуі медиатизацияның жаңа кезеңінің басталғанын көрсетеді. Зерттеу нәтижелері бойынша, жасанды интеллект қолданылуының екі жақты сипаттамасы анықталды. Бір жағынан, тиімділік, жеделдік, факт-чек пен іздеу-талдау мүмкіндіктері. Екінші жағынан, тілдік, мәдени, мағыналық бұрмаланулар. Бұндай қорытындыға жүргізілген сауалнама нәтижелері де сәйкес келеді (Katz et al., 1973).

Сауалнама 2025 жылдың қазан айында онлайн форматта (Google.Forms) жүргізілді. Бұл зерттеу қазақ тілінің қолданысына қатысты болғандықтан, респонденттер арнайы іріктеліп, сауалнама журналистер, филолог мамандар арасында жүргізілді. Респонденттердің әлеуметтік профильдері: студенттер, оқытушылар, редакторлар, журналистер. Сауалнамаға 79 адам қатысты. Олардың 35,4% – 35-44 жас аралығында болса, 35,42% – 25-34 жас аралығындағы жастар. Қалған 13,9% – 45-54 жас аралығындағы орта буын өкілдері. Тағы 13,9% – 18-24 жас аралығындағы студенттер болды.

Жасыңыз нешеде?

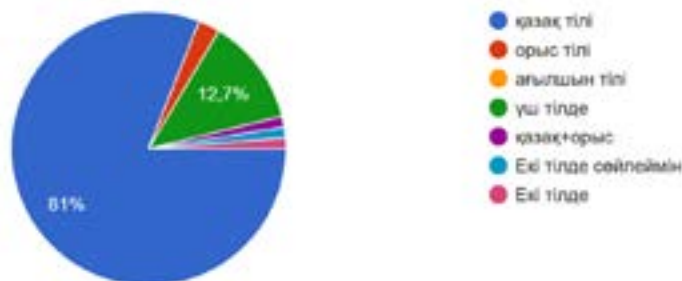
79 ответов



Сауалнамаға қатысқандардың басым бөлігі, яғни 81%-ы қазақтілді. 12,7% үш тілді меңгерген. Бұл тілдік бағыттағы зерттеудің өзектілігін арттыра түседі.

Қай тілде сөйлесесіз?

79 ответов



ЖИ жүйесін қолдану жиілігі:

- 54,4% - анда-санда қолданады
- 40,5% - үнемі пайдаланады
- 5,1% - мүлдем қолданбайды.

ЖИ құралдарынан бойын аулақ ұстайтындардың қатары сирек. Бұл дерек жасанды интеллект жүйелерінің күнделікті өмір мен әртүрлі салаға қарқынды түрде ене бастағанын дәлелдейді.

Жасанды интеллект жүйелерін (мысалы ChatGPT, Google Translate және т.б.) қолданасыз ба?

79 ответов



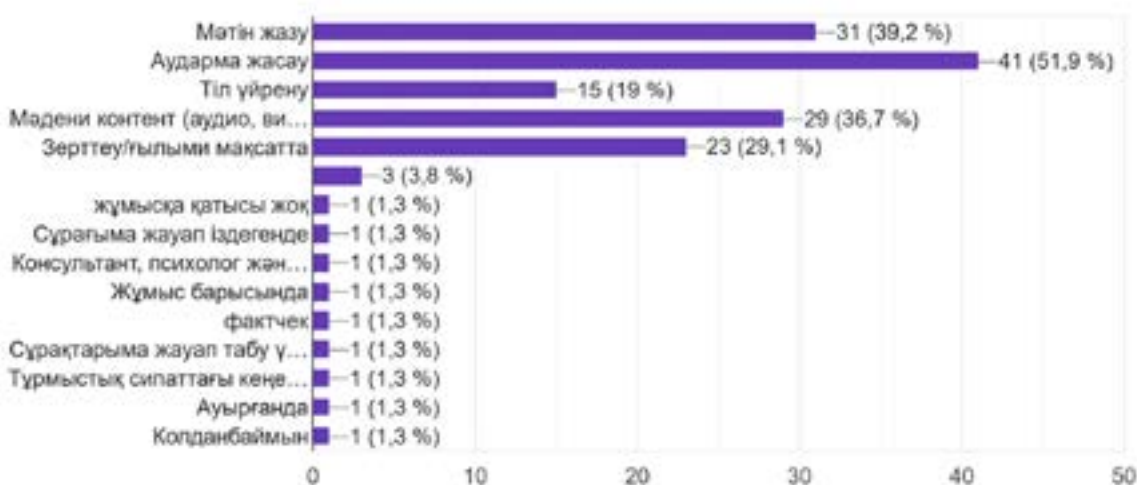
Жасанды интеллектіні қолдану мақсаты бойынша респонденттердің жауабы төмендегідей болды:

- 51,9% - аударма жасау
- 39,2% - мәтін жазу
- 36,7% - аудио, видео дайындау
- 29,1 % - ғылыми зерттеу
- 19% - тіл үйрену.

Бұл ЖИ жүйелерінің тілге қатысты еңбектерде және шығармашылық бағыттарда белсенді қолданылатынын көрсетеді.

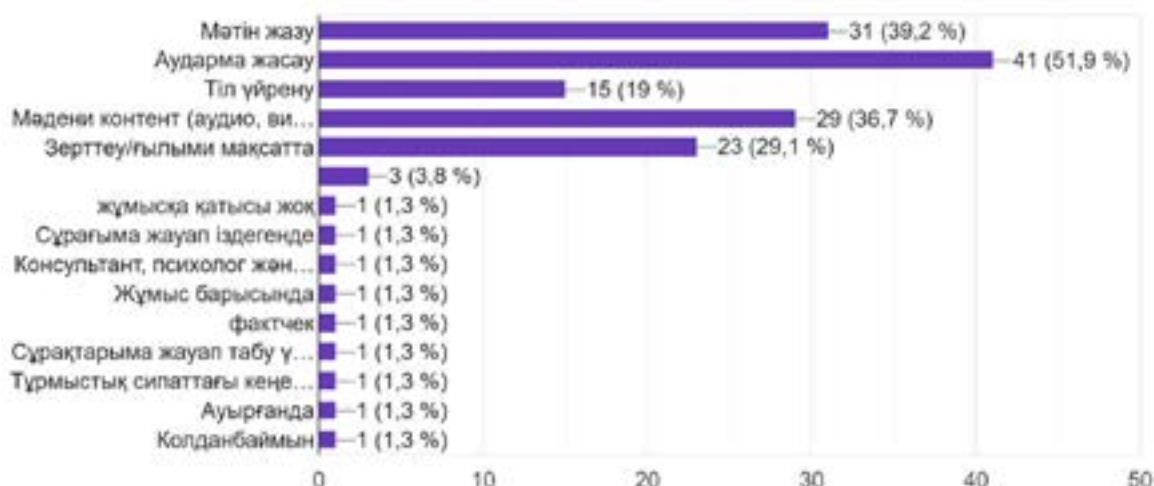
Жасанды интеллектті қандай мақсатта қолданасыз?

79 ответов



Жасанды интеллектті қандай мақсатта қолданасыз?

79 ответов



ЖИ кемшіліктері туралы пікірлер:

- 57% - мағынаның бұрмалануы
- 36,7% - грамматикалық қате
- 32,9% - жалған ақпарат
- 31,6% - жеке деректердің қауіпсіздігі.

Респонденттердің жауаптары ЖИ ұсынған ақпараттың тілдік сапасын және фактілердің дұрыстығын тексеру қажеттігін көрсетеді. Жасанды интеллекттің жалған ақпаратты анықтау және әшкерелеу қабілеті бар деген тұжырым біздің сауалнамаға қатысқан респонденттердің жауаптарында расталмады.

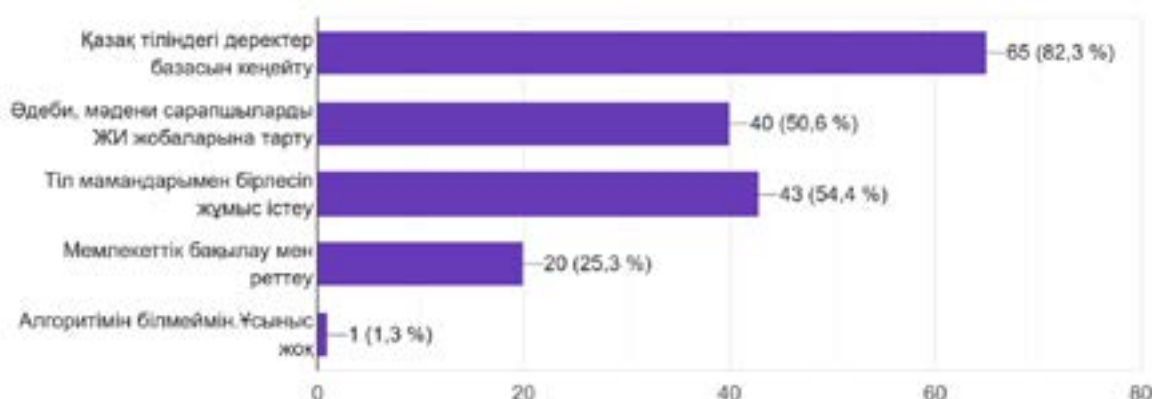
Қазақ тілі бұрмаланбас үшін не істеу керек?

- 82,3% - қазақ тіліндегі деректер базасын кеңейту
- 54,4% - тіл мамандарымен бірлесіп жұмыс істеу
- 50,6% - әдеби, мәдени сарапшыларды ЖИ жобаларына тарту.

Нәтиже ЖИ жүйелеріндегі тілдің сапасын арттыру үшін ана тілдегі мәтіндердің сандық көлемін көбейту қажеттігін анықтап берді.

ЖИ-де қазақ тілі бұрмаланбас үшін не істеу керек деп ойлайсыз?

79 ответов



Жалпы сауалнама нәтижелері жасанды интеллект жүйелерінің қазақ қоғамына белсенді енгенін, бірақ қазақ тілінің сапасы мен контенттің сенімділігі өзекті мәселе екенін көрсетті. Қатысушылардың көпшілігі ЖИ жүйелерін пайдалы құрал ретінде қабылдағанымен, тілдің бұрмалануы мен мағыналық, грамматикалық қателерді назар аударарлық түйткіл деп бағалаған.

Цифрлық дәуірде тіл өзгерісін зерттейтін сандық лингвистикаға сәйкес, тілдік модельдер тілдің грамматикасын, семантикасын, басқа ерекшеліктерін толық меңгермейінше, сапалы өнім шығара алмайды. Қазақ тілінің ЖИ жүйелерінде жиі бұрмалануы – осы теорияның дәлелі. Тілдің автоматты өңделуі, мәтінді машиналық талдау сияқты процестер цифрлық лингвистиканың өзегі саналады. Бұл тілдің цифрлық ортада мүлде өзгеше дамуына алып келеді.

Қазақ тілін ЖИ жүйелерінде дұрыс бағытта дамыту үшін ең алдымен ұлттық тілдік деректер қорын құру, сапалы контент жинау және машиналық оқыту процестеріне лингвистикалық бақылау енгізу қажет. Дәл осы жұмысты қазір Әл-Фараби атындағы Қазақ Ұлттық университеті Жасанды интеллект және Big Data кафедрасының аға оқытушылары Ахмет Байтұрсынов атындағы тіл білімі институты мен «Тіл қазына» мемлекеттік орталығының мамандарымен бірлесіп атқаруда. Кітапханалардан, газет-журналдардан PDF форматтағы материалдар OCR жүйесі арқылы мәтіндік форматқа ауыстырылуда.

Машиналық оқытуда басты нәрсе – оған берілетін «шикізат», яғни деректердің сапасы. Жасанды интеллекттің беретін жауабы сол деректердің сапасына тікелей байланысты. Ғаламтордағы қазақ тіліндегі контенттің көлемі басқа тілдерге қарағанда әлдеқайда аз. Сол себепті жасанды интеллект ағылшын, орыс және басқа тілдерде үйретілген ақпаратты қазақшаға аударып береді. Кейде ол терминдерді аудармай, араластырып қолданады. Мысалы, ағылшын және қазақ тіліндегі сөздерді қатар қолдануы мүмкін. Бұл да мағынаның бұрмалануына себеп (Yeshpanov et al., 2024).

Дегенмен қазақ тілі цифрлық кеңістікке толық бейімделгенше жасанды интеллект жүйелерінде қателіктер мен елестердің (галлюцинация) жиі кездесуі заңдылық. Бұл тұжырымға дәлел ретінде ЖИ құралдарымен аударылған мәтіндердің сапасын салыстыруға болады. Мәселен, «Казахстанка взыскала с врача через суд 3,2 миллиона тенге за неправильные диагноз и лечение» деген сөйлемді ChatGPT былай аударып

берді: «Қазақстандық әйел дәрігерден қате диагноз бен ем тағайындағаны үшін сот арқылы 3,2 миллион теңге өндіріп алды». Дәл сол сөйлемнің Google Translate арқылы аударылған нұсқасы мынадай: «Қазақстандық әйел дәрігерді дұрыс диагноз қоймағаны және емі дұрыс емес деп 3,2 миллион теңге өндіріп алды».

Екі аударманы салыстырғанда Google Translate ұсынған аудармада грамматикалық қате бірден көзге түседі. «Дәрігерді... өндіріп алды», «ді» деген табыс септігінің жалғауы жалғанған, ал өндіріп алу етістігі шығыс септігін қажет етеді. Кімнен? Дұрыс нұсқасы: «Дәрігерден...» болуы керек еді. «Емі дұрыс емес» деген сөз тіркесінің орнына «дұрыс емдемегені үшін» деген қарапайым сөз тіркесін де қолдануға болатын еді. Үшіншіден, бұл сөйлемде қажетсіз қайталамаға (тавтология) жол берілген. Бір сөйлемде «дұрыс» сөзі екі рет қайталанады.

Осы мысалда ChatGPT аудармасы бастапқы сөйлемнің негізгі деректерін сақтағанын байқаймыз. Редакциялық нұсқа ретінде: «Қазақстандық әйел дәрігерден сот арқылы 3,2 миллион теңге өндіріп алды. Бұған қате диагноз бен дұрыс жүргізілмеген ем себеп болған» деп ұсынуға болады. Редакторлық өңдеу кезінде бастапқы мәтінде жоқ өңір атауын қосуға немесе «сот арқылы» деген маңызды деректі алып тастауға болмайды. Адам аударған бұл нұсқа әлдеқайда құлаққа жағымды, грамматикалық ережелерге сай, лексикалық калькасы жоқ, таза қазақ тілінде құрылған. «Сот арқылы» деген тіркесті қолданбаса да, өтемақының сот жүйесі арқылы өндірілетіні белгілі.

Енді бұдан күрделірек мәтіннің қазақ тілінен орысшаға аудармасын салыстырайық. Мысалға, Абай Құнанбаевтың «Қара сөзінен» үзінді алайық, бұл түпнұсқа:

«Кербездің екі түрлі қылығы болады: бірі бет-пішінін, мұртын, мүшесін, жүрісін, қас-қабағын қолдан түзетіп, шынтағын көтеріп, қолын тарақтап әуре болмақ. Біреуі атын, киімін «айран ішерім» деп, солардың арқасында сыпайы, жұғымды жігіт атанбаққа, өзінен ілгерілерге елеулі болып, өзі қатардағының ішін күйдіріп, өзінен кейінгілерге «әттең, дүние-ай, осылардың атындай ат мініп, киіміндей киім кигеннің не арманы бар екен?!», – дейтұғын болмаққа ойланбақ» (Абай, 18-қара сөз).

ChatGPT аудармасы:

«У щеголя два вида поведения: один старается исправить свое лицо, усы, фигуру, походку, брови, поднимает локоть, расчёсывает руки, тратит время на внешность. Другой – старается стать заметным благодаря своей лошади и одежде, чтобы его считали вежливым, приятным человеком, чтобы старшие обращали на него внимание, сверстники завидовали, а младшие думали: «Ах, вот бы иметь такую же лошадь и такую одежду!»».

Google Translate аудармасы:

«Красивый мужчина имеет два типа поведения: один — выпрямлять лицо, усы, конечности, походку и брови, поднимать локти и расчесывать руки. Второй — притворяться одновременно лошадью и мужчиной, и благодаря этому вежливый, обаятельный молодой человек становится знаменитым, важным для тех, кто впереди, и заставляет тех, кто позади, восклицать: «Боже мой, какой смысл ездить на лошади, как эти парни, и носить такую же одежду?»».

Бірінші нұсқада «щеголь» «кербез» сөзіне ең жақын аударма. Бірақ мағыналық өрісі тар. Абайдағы «кербез» тек сәнқой емес, сыртқы әсемдікке алданып, рухани мәннен айырылған адам. Демек, аударма лексикалық тұрғыдан дәл, бірақ философиялық мағынасын толық бере алмаған. Қазақ мәтінінде «айран ішерім» тіркесі ұлттық идиома,

ол ерке, сәнқой, өмірі жеңіл деген мағынада беріліп тұр. Аудармада бұл тіркес мүлдем түсіп қалған. «Қолын тарақтап әуре болмақ», қазақ тілінде ирония мен бейнелілік бар, ал аудармасы «тратит время на внешность» ақпараттық сипатта ғана берілген.

Google Translate аудармасында бірнеше семантикалық сәйкессіздік байқалады. Мысалы, «кербез» сөзінің «красивый мужчина» деп берілуі түпнұсқадағы сыни реңкті жеткізе алмаған. Абайда бұл сөз сыни тұрғыдан пайдаланса, аудармада позитивті мағынада берілген. «Притворяются одновременно лошадей и мужчиной» деген аударма қара сөздің мағынасын мүлдем басқа арнаға бұрып жіберген. Google Translate қазақ сөйлеміндегі «атпен, киіммен мақтану» тіркесін түсінбеген. Салдарынан «жылқыға ұқсауға тырысып жатыр» деген мағыналық тұрғыдан сәйкес емес тіркес жазылып, Абайдың философиялық тұжырымы мағынасын жоғалтады. Аударманың екі нұсқасында да Абайдың айтқан ойы мен шешендік стилі берілмеген. Түпнұсқада ирония, сын, астар, образ бар. Аудармада бұл философиялық салмақ логикалық тізбектен тұратын құрғақ сөйлемге айналған. Бұл мысал машина аудармасы жүйелерінің шектеулерін айқын көрсетіп, дәлелдейді.

Бұл мысалдар мәдени контекст пен идиоматика сақталмаса, аудармада мәтіннің мәні жоғалатынын көрсетеді. Қорыта айтқанда, жасанды интеллект жүйелері сөзбе-сөз аударма жасап, мәтіннің мәдени астарын, иронияны, әлеуметтік реңкті дәл жеткізе алмайды. Иляхина мен Деева (2024) өз зерттеуінде жасанды интеллект жүйелері мәтінді өңдеу, генерациялау және таратуды автоматтандыру арқылы медиа саласында біртұтас стандарт қалыптастырып, шығармашылық бірегейлікті азайтатынын болжаған. Бұл құбылысты зерттеушілер «журналистиканың цифрлық конвергенциясы» деп атайды (Ilyakhina & Deeva, 2024).

Қарастырылған аудармаларда бейнелілік пен мәдени реңктің ішінара жоғалуы байқалды. Бұл нәтижелер мәдени мазмұны күрделі мәтіндерді автоматты аударудан кейін редакторлық сараптамадан өткізу қажеттігін дәлелдеп берді.

Қорытынды

Медиатизация теориясы зерттелген аудармаларды цифрлық коммуникация аясында түсіндіруге мүмкіндік береді. Тіл де медиатизацияға ұшырайтынын уақыт көрсетуде. Оның қолданыс формалары, нормалары мен мағыналық құрылымдары цифрлық медиа логикасына бейімделе бастады. Осы үрдіс аясында қазақ тілі де, өкінішке орай, жаңа медианың тіліне айналып барады. Әлеуметтік желілерде, ЖИ жүйелерінде, аударма құралдарында қазақ тілінің синтаксисі мен сөздік қоры қысқарып, «технологиялық стильге» ауысуда. Бұл медиатизация үдерісінің нақты дәлелі.

Сауалнамада респонденттердің 57%-ы мағынаның бұрмалануын, 36,7%-ы грамматикалық қателерді, 32,9%-ы жалған ақпаратты ЖИ құралдарының кемшіліктері деп атаған. Қарастырылған аудармаларда грамматикалық және мәдени-мағыналық сәйкессіздіктер байқалды. Бұл нәтижелер автоматты аудармаларды жариялау алдында тексеру қажеттігін көрсетіп, ЖИ жүйелеріндегі қазақ тілінің қолдану проблемаларын да айшықтайды.

Зерттеу нәтижелері негізінде мынандай мәселелер анықталды:

- ЖИ жүйелерінде қазақ тіліндегі грамматикалық, семантикалық және мәдени қателер жиі кездеседі;
- сауалнама нәтижесінде аударма мен мәтін жазуда ЖИ құралдары жиі қате жібереді;

- деректер қорының жеткіліксіздігі тілдің қолданыс сапасына әсер етуші ықтимал фактор болуы мүмкін, бірақ оның әсері бұл зерттеуде тексерілген жоқ.

Машиналық оқыту модельдері тілдік деректердің көлемі мен сапасына тәуелді. Қазақ тілінде мәтіндік мәліметтер қоры шектеулі болғандықтан, ЖИ модельдері орыс және ағылшын тіліндегі құрылымдарға сүйенеді. Бұл көшірме (калька), грамматикалық қате және мағыналық бұрмалану түрінде байқалады. Иляхина мен Деева (2024) атап өткендей, автоматтандырылған контент пен алгоритмдік мәтін генерациясы «шығармашылық бірегейлікті азайтып, стандартталған медиа тіл қалыптастырады» (Ilyakhina & Deeva, 2024).

Бұл құбылыс тек медиада ғана емес, әдебиет пен ғылым саласында да көрініс табуда. Қазақ тілінің бай бейнелілігі мен поэтикалық құрылымы біртіндеп технологиялық стильге бейімделуін заңдылық деп қабылдаған жөн. Дегенмен тілдің мәдени ерекшеліктерін сақтау үшін ЖИ жүйелерінде ұлттық мазмұнды, фольклорлық және әдеби деректерді пайдалану маңызды. Бұл бағытта «Тіл қазына» орталығы мен Ахмет Байтұрсынов атындағы Тіл білімі институты жасаған Тілдік корпустар мен деректер базасы – үлкен мүмкіндік ұсынады. Осы мәліметтердің барлығы цифрландырылып, жасанды интеллект жүйелеріне енгізілсе, жаңа технологиядағы қазақ тілінің дамуына оң әсерін тигізеді. Бұл айтылған сын-ескертулер жасанды интеллект тек кемшіліктерден тұрады дегенді білдірмейді. Жаңа технологияның артықшылықтары мол, тәуекелдері де жеткілікті.

ЖИ ақпаратқа қолжетімділікті арттырады, мәтінді талдау мен жүйелеуді жеңілдетеді. Білім беру мен зерттеу жұмысын тиімді жасайды, тіл үйренуге жаңаша мүмкіндіктер туғызады. Ал оның ақпарат саласындағы кемшіліктері мынадай:

- ақпараттарды тексермей, факт-чек жасамай ұсыну;
- мәдени және мағыналық бұрмалану, грамматикалық қателер, идиомалар мен мақал-мәтелдердің жоғалуы, ұлттық стильдің жойылуы және көркем тілдің технологиялық тілге айналуы.

Авторлар зерттеуде қарастырылған мысалдардан автоматты аудармада тілдік және мәдени сәйкессіздіктер туындауы мүмкін екенін анықтады. Жасанды интеллект технологиялары тілдің дамуына оңтайлы жағдай жасай отырып, оның табиғи құрылымына қандай зиян келтіретінін де талдау қажет деп ойлаймыз. Бұл парадокс қазіргі заманның басты мәдени мәселелерінің біріне айналды. Біздің ойымызша, проблеманы шешудің дұрыс жолы – ЖИ жүйелерін жетілдіру. Сондықтан зерттеу нәтижелері бойынша төмендегі ұсыныстар енгізіледі:

- «Жасанды интеллект туралы» заң жобасына тіл нормалары мен ұлттық контентті қорғау нормаларын енгізу;
- ұлттық тілдік деректер қорын құру және оны ұдайы кеңейтіп отыру;
- OCR арқылы архивтік деректерді толық цифрландыру;
- ЖИ генерациялаған мәтіндерді міндетті редакторлық тексеруден өткізу;
- университеттер мен орта білім жүйесінің білім бағдарламаларына «Цифрлық тіл мәдениеті», «Машина аудармасының этикасы» атты пәндерді енгізу;
- қазақ тілінің синтаксисін ЖИ модельдеріне бейімдеуге арналған арнайы алгоритмдер әзірлеу.

Зерттеу қазақ тіліндегі автоматты аударманың тілдік және мәдени ерекшеліктерін сауалнама мен жекелеген мәтіндік мысалдар негізінде қарастырады. Болашақта бұндай

зерттеулер тереңдетіліп, бірнеше бағытта жүргізілуі қажет деп санаймыз. Мысалы, қазақ тілінің фонетикалық және интонациялық ерекшеліктерін ЖИ жүйелеріне енгізу, ЖИ құралдарының көмегімен ұлттық әдебиет пен фольклорды автоматты талдау әдістерін жетілдірген дұрыс болар еді. Осындай жүйелі зерттеулер жаңа дәуірдегі қазақ тілінің сапасын өсіріп, қолданыс аясын кеңейтетінін зор сеніммен айта аламыз.

Қаржыландыру

Бұл зерттеу мақаласы ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің ғылыми гранты есебінен қаржыландырылды (грант №AP32723927).

Авторлардың қосқан үлесі.

Есенбекова Ұ. М. – хат-хабар авторы, бақылау және ұсыныс беру, мақсат-міндетті анықтау, жалпы ғылыми аппаратты басқару, деректерді сұрыптау, редакциялау.

Стамбекова Н. Д. – хат-хабар авторы, әдебиетпен жұмыс, материалды жинау, негізгі бөлімді жазу.

Егембердиев Ы. Д. – материалды жинау, мақаланы техникалық талапқа сай рәсімдеу.

Әдебиеттер тізімі

Borisov, M., Kozhirbayev, Z. and Malykh, V. (2025) 'Low-resource machine translation for code-switched Kazakh-Russian language pair', NAACL Student Research Workshop. doi: <https://doi.org/10.18653/v1/2025.naacl-srw.7>

Hepp, A. (2013) Cultures of mediatization. Polity Press. ISBN: 978-0-745-66226-8.

Hjarvard, S. (2013) The mediatization of culture and society. Routledge. doi: <https://doi.org/10.4324/9780203155363>

Ilyakhina, A.A. and Deeva, I.V. (2024) 'Perspectives on the application of artificial intelligence technologies in journalism', Vestnik nauki, 2, pp. 580–586. ISSN: 2712-8849. Available at: <https://www.vestnik-nauki.com/article/12508> (қаралған күні: 15 қазан 2026).

Katz, E., Blumler, J.G. and Gurevitch, M. (1973) 'Uses and gratifications research', Public Opinion Quarterly, 37(4), pp. 509–523. doi: <https://doi.org/10.1086/268109>

LeCun, Y. (2023) How the machine learns: The revolution in neural networks and deep learning. Moscow: Al'pina Publisher, 335 p. ISBN: 9785907394926.

Miao, F. and Holmes, W. (2023) Guidance for generative AI in education and research. UNESCO. doi: <https://doi.org/10.54675/EWZM9535>

Naveen, P. et al. (2024) 'Overview and challenges of machine translation for contextually appropriate translations', iScience. doi: <https://doi.org/10.1016/j.isci.2024.110878>

Qazcorp.kz (2025) Sub-corpus of the National Corpus of the Kazakh Language. Available at: <https://qazcorp.kz/> (қаралған күні: 15 қазан 2026).

Russell, S. and Norvig, P. (2021) Artificial intelligence: A modern approach. 4th edn. Pearson. Available at: <https://aima.cs.berkeley.edu/> (қаралған күні: 15 қазан 2026).

Tleubayeva, A., Aubakirov, S., Tabuldin, A. and Shomanov, A. (2025) 'Development and evaluation of a small Kazakh language corpus to improve the efficiency of multilingual NLP systems in low-resource environments', IEEE SIST 2025. doi: <https://doi.org/10.1109/SIST61657.2025.11139363>

Togmanov, M., Mukhituly, N., Turmakhan, D. et al. (2025) 'KazMMLU: Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan'. doi: <https://doi.org/10.18653/v1/2025.acl-long.701>

Yeshpanov, R., Efimov, P., Boytsov, L., Shalkarbayuli, A. and Braslavski, P. (2023) 'KazQAD: Kazakh open-domain question answering dataset'. doi: <https://doi.org/10.48550/arXiv.2404.04487>

Yeshpanov, R., Polonskaya, A. and Varol, H.A. (2024) 'KazParC: Kazakh parallel corpus for machine translation', arXiv preprint. doi: <https://doi.org/10.48550/arXiv.2403.19399>

Yessenbekova, U., Syzdykova, A. and Kutym, B. (2024) 'Demand for TV content in a national language: The case of Kazakhstan', International Journal of Innovative Research and Scientific Studies, 7(3), pp. 951–964. doi: <https://doi.org/10.53894/ijirss.v7i3.2984>

Zhu, W., Liu, H., Dong, Q., Xu, J., Kong, L., Chen, J., Li, L. and Huang, S. (2024) 'Multilingual machine translation with large language models: Empirical results and analysis', Findings of the Association for Computational Linguistics: NAACL 2024, pp. 2765–2781. doi: <https://doi.org/10.18653/v1/2024.findings-naacl.176>

У. М. Есенбекова¹, Н.Д. Стамбекова², И.Д. Егембердиев³

^{1,2}*Евразийский национальный университет им. Л.Н.Гумилева, Астана, Казахстан*

³*Казахский национальный университет им. аль-Фараби, Алматы, Казахстан*

Казахский язык в системах искусственного интеллекта: языковое качество, культурный контекст и проблемы цифрового развития

Аннотация. В исследовательской статье проанализировано влияние систем искусственного интеллекта на развитие казахского языка и национальной культуры. Дана оценка грамматическому, семантическому и культурному качеству ИИ-систем на казахском языке. Результаты опроса, сравнительного анализа переводов и текстового анализа выявили недостаточность данных на казахском языке в системах искусственного интеллекта. В рассмотренных переводах наблюдались языковые несоответствия и частичная утрата культурного контекста.

В методологическом плане исследование опирается на теории медиатизации Стига Хьярварда (Stig Hjarvard) и Андреаса Хеппа (Andreas Hepp), а также на теории удовлетворения потребностей Элиху Каца (Elihu Katz) и Джея Блумлера (Jay G. Blumler). Эти теории рассматривают ИИ-системы как важную часть современного медиакультурного пространства и позволяют объяснить роль языка в цифровой среде. Теория удовлетворения потребностей демонстрирует место ИИ-технологий в удовлетворении общественных информационных потребностей.

На основе результатов исследования авторы предлагают расширять национальные языковые корпуса для развития казахского языка в цифровой среде, укреплять взаимосвязь ИТ и лингвистики, а также внедрять в систему образования дисциплины по цифровой языковой культуре. Делается вывод о том, что при реализации этих направлений казахский язык станет не только национальным наследием, но и конкурентоспособным языком, развивающимся в современных технологиях.

Ключевые слова: искусственный интеллект, машинный перевод, языковой корпус, культурный контекст, медиатизация.

U.M. Yessenbekova¹, N. D. Stambekova², I. D. Yegemberdiyev³
^{1,2,3} *L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*

The Kazakh Language in Artificial Intelligence Systems: Linguistic Quality, Cultural Context, and Challenges of Digital Development

Abstract. This research article analyzes the impact of artificial intelligence (AI) systems on the development of the Kazakh language and national culture. The study evaluates the grammatical, semantic, and cultural quality of Kazakh-language AI systems. The findings from the survey, comparative translation analysis, and textual analysis revealed insufficient Kazakh-language data in artificial intelligence systems. Linguistic inconsistencies and a partial loss of cultural context were observed in the translations examined.

Methodologically, the research draws on the mediatization theories of Stig Hjarvard and Andreas Hepp, as well as the uses and gratifications approach associated with Elihu Katz and Jay G. Blumler. These frameworks conceptualize AI systems as an important component of today's media-cultural environment and help explain the role of language in digital settings. The uses and gratifications perspective highlights how AI technologies contribute to meeting society's information needs.

Based on the results, the authors recommend expanding national language corpora to support the development of Kazakh in the digital environment, strengthening collaboration between information technology and linguistics, and introducing courses on digital language culture within the education system. The article concludes that, if these measures are implemented, Kazakh will become not only a national heritage language but also a competitive language that can evolve alongside modern technologies.

Keywords: artificial intelligence, machine translation, language corpus, cultural context, mediatization.

References

- Borisov, M., Kozhirbayev, Z. and Malykh, V. (2025) 'Low-resource machine translation for code-switched Kazakh-Russian language pair', NAACL Student Research Workshop. doi: <https://doi.org/10.18653/v1/2025.naacl-srw.7>
- Hepp, A. (2013) *Cultures of mediatization*. Polity Press. ISBN: 978-0-745-66226-8.
- Hjarvard, S. (2013) *The mediatization of culture and society*. Routledge. doi: <https://doi.org/10.4324/9780203155363>
- Ilyakhina, A.A. and Deeva, I.V. (2024) 'Perspectives on the application of artificial intelligence technologies in journalism', *Vestnik nauki*, 2, pp. 580–586. ISSN: 2712-8849. Available at: <https://www.vestnik-nauki.com/article/12508> (Accessed: 15 October 2026). [in Russian].
- Katz, E., Blumler, J.G. and Gurevitch, M. (1973) 'Uses and gratifications research', *Public Opinion Quarterly*, 37(4), pp. 509–523. doi: <https://doi.org/10.1086/268109>
- LeCun, Y. (2023) *How the machine learns: The revolution in neural networks and deep learning*. Moscow: Al'pina Publisher, 335 p. ISBN: 9785907394926.
- Miao, F. and Holmes, W. (2023) *Guidance for generative AI in education and research*. UNESCO. doi: <https://doi.org/10.54675/EWZM9535>
- Naveen, P. et al. (2024) 'Overview and challenges of machine translation for contextually appropriate translations', *iScience*. doi: <https://doi.org/10.1016/j.isci.2024.110878>

Qazcorporaz.kz (2025) Sub-corpus of the National Corpus of the Kazakh Language. Available at: <https://qazcorporaz.kz/> (Accessed: 15 October 2026).

Russell, S. and Norvig, P. (2021) Artificial intelligence: A modern approach. 4th edn. Pearson. Available at: <https://aima.cs.berkeley.edu/> (Accessed: 15 October 2026).

Tleubayeva, A., Aubakirov, S., Tabuldin, A. and Shomanov, A. (2025) 'Development and evaluation of a small Kazakh language corpus to improve the efficiency of multilingual NLP systems in low-resource environments', IEEE SIST 2025. doi: <https://doi.org/10.1109/SIST61657.2025.11139363>

Togmanov, M., Mukhituly, N., Turmakhan, D. et al. (2025) 'KazMMLU: Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan'. doi: <https://doi.org/10.18653/v1/2025.acl-long.701>

Yeshpanov, R., Efimov, P., Boytsov, L., Shalkarbayuli, A. and Braslavski, P. (2023) 'KazQAD: Kazakh open-domain question answering dataset'. doi: <https://doi.org/10.48550/arXiv.2404.04487>

Yeshpanov, R., Polonskaya, A. and Varol, H.A. (2024) 'KazParC: Kazakh parallel corpus for machine translation', arXiv preprint. doi: <https://doi.org/10.48550/arXiv.2403.19399>

Yessenbekova, U., Syzdykova, A. and Kutym, B. (2024) 'Demand for TV content in a national language: The case of Kazakhstan', International Journal of Innovative Research and Scientific Studies, 7(3), pp. 951–964. doi: <https://doi.org/10.53894/ijirss.v7i3.2984>

Zhu, W., Liu, H., Dong, Q., Xu, J., Kong, L., Chen, J., Li, L. and Huang, S. (2024) 'Multilingual machine translation with large language models: Empirical results and analysis', Findings of the Association for Computational Linguistics: NAACL 2024, pp. 2765–2781. doi: <https://doi.org/10.18653/v1/2024.findings-naacl.176>

Авторлар туралы мәлімет:

Есенбекова Ұ. М. – хат-хабар авторы, филология ғылымдарының кандидаты, қауымдастырылған профессор, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, медикоммуникация және қоғаммен байланыс кафедрасының профессоры м.а., 010000, Астана, Қазақстан.

Стамбекова Н. Д. – әлеуметтік білім магистрі, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, журналистика және әлеуметтік ғылымдар жоғары мектебі, 010000, Астана, Қазақстан.

Егембердиев Ы. Д. – PhD докторы, Әл-Фараби атындағы Қазақ ұлттық университеті Журналистика факультетінің деканы, 0500040, Алматы, Қазақстан.

Сведения об авторах:

Есенбекова У. М. – автор для корреспонденции, кандидат филологических наук, ассоциированный профессор, и.о. профессора кафедры Медиакommunikации и связи с общественностью Евразийского национального университета имени Л.Н. Гумилева, 010000, Астана, Казахстан.

Стамбекова Н. Д. – магистр социального образования, Высшая школа журналистики и социальных наук Евразийского национального университета им. Л.Н. Гумилева, 010000, Астана, Казахстан.

Егембердиев И. Д. – доктор PhD, декан Факультета Журналистики Казахского национального университета им. аль-Фараби, 050040, Алматы, Казахстан.

Author information:

Yessenbekova U. M. – corresponding author, Candidate of Philological Sciences, Associate Professor, Acting Professor of the Department of Media Communications and Public Relations, L.N. Gumilyov Eurasian National University, 010000, Astana, Kazakhstan

Stambekova N. D. – master of Social Education, Higher School of Journalism and Social Sciences, L.N. Gumilyov Eurasian National University, 010000, Astana, Kazakhstan

Yegemberdiyev I. D. – PhD, Dean of the Faculty of Journalism at Al-Farabi Kazakh National University, 050040, Almaty, Kazakhstan