

ЭОЖ 81` 33: 004.8

ГТАХР 16.31.51; 16.21.47

<https://doi.org/10.48371/PHILS.2026.2.81.026>

## ҚАЗАҚ ТІЛІНДЕГІ ГЕНЕРАТИВТІ ЖАСАНДЫ ИНТЕЛЛЕКТ МӘТІНДЕРІНІҢ ТІЛДІК-СТИЛЬДІК ЕРЕКШЕЛІКТЕРІ (OPENAI ЖӘНЕ ОТАНДЫҚ ПЛАТФОРМА НЕГІЗІНДЕ)

\*Тәшпеннова Қ.Е.<sup>1</sup>, Садирова К.К.<sup>2</sup>

\*<sup>1,2</sup>Қ.Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе, Қазақстан

**Аңдатпа.** Бұл мақалада генеративті жасанды интеллект (ЖИ) платформаларының қазақ тіліндегі мәтіндерге тән тілдік және стильдік ерекшеліктері салыстырмалы талдау арқылы зерттеледі. Генеративті модельдердің тілдік сапасын бағалау – қазіргі лингвистика мен цифрлық технологиялардың тоғысындағы өзекті мәселе. *Зерттеу нысаны* ретінде OpenAI ұсынған ChatGPT моделі мен отандық ЖИ платформасы таңдалып, олардың қазақ тіліндегі мәтіндерді генерациялау қабілеті бес негізгі критерий бойынша бағаланды: промт талабын орындау, әдеби тіл нормаларына сәйкестік, калька мен кірме құрылымдардың ықпалы, стильдік регистрдің сақталуы және лексикалық байлық. Бағалау процесінде 0–2 балдық шкала қолданылып, әр жауаптың сапалық көрсеткіштері нақты сандармен сипатталды. Талдау нәтижелері OpenAI платформасының мәтіндері мазмұн тереңдігі, стильдік байлық, аргументациялық логика мен бейнелілік тұрғысынан анағұрлым жоғары деңгейде екенін көрсетті. Сонымен қатар, оның көркем-публицистикалық стиль талаптарын орынды қолданатыны байқалды. Ал отандық платформа мәтіндері қысқа, құрылымы жиі қайталанатын, кейде калька үлгілері мен грамматикалық қателерге толы болып шықты. Мұндай айырмашылықтар көбіне платформаларда қолданылған тілдік корпустардың құрамымен, ресми-іскерлік стильге бағдарланған материалдардың басымдығымен байланысты. Мәтіндердегі мәдени-ұлттық мазмұнның көрініс табуы да *зерттеудің маңызды бағыты* болды. *Нәтижелер* қазақ тіліндегі образдылық, мақал-мәтел мен теңеу секілді көркемдік құралдардың генеративті жүйелер арқылы жеткізілуінде шектеулер бар екенін көрсетті. Бұл зерттеу қазақ тілінде сапалы генеративті ЖИ моделін дамыту үшін тілдік нормалар мен стиль үйлесімін ескеру қажеттігін дәлелдейді. Сонымен қатар, зерттеу барысында ЖИ платформаларының мәтін құраудағы прагматикалық сәйкестік пен контексті түсіну деңгейі де жанама түрде қарастырылды. Бұл аспектілер қазақ тілінің семантикалық және дискурстық ерекшеліктерін меңгерудің генеративті модельдер үшін күрделі екендігін аңғартады. Алынған нәтижелер болашақта қазақ тіліне бейімделген ЖИ жүйелерін жасау мен тілдік деректер қорын сапалы қалыптастыруда әдістемелік негіз бола алады.

**Тірек сөздер:** қазақ тілі, генеративті жасанды интеллект, мәтін стилистикасы, калька, стильдік регистр, тілдік корпус, отандық платформа, OpenAI

### **Кіріспе**

Қазақ тіліндегі генеративті жасанды интеллект құралдары білім беру, ғылыми жазылым, медиа және ресми коммуникацияда күнделікті қолданылатын технологияға айналып келеді. Соған қарамастан, қазақ тіліндегі жасанды интеллект (әрі қарай ЖИ) мәтіндерінің сапасы көбіне «мазмұн дұрыстығымен» ғана өлшеніп, әдеби тіл нормасына сәйкестік, төл сөздер мен бейнелі-әдеби лексиканы табиғи қолдану, калька мен код-ауысымның деңгейі сияқты тілдік-стилистикалық өлшемдер жеткілікті деңгейде жүйелі талдана бермейді. Әсіресе отандық платформалар мен халықаралық платформалардың қазақша мәтін тудыру қабілеті салыстырылғанда, «қайсысы жақсы?» деген жалпылама бағадан гөрі, нақты тілдік индикаторлар арқылы дәлелді қорытынды жасау қажет. Бұл мәселе қазақ тілінің цифрлық экологиясын сақтау, ресми-іскери стильдің кодификациясын күшейту және көркем-публицистикалық дискурста табиғи сөзқолданысты қолдау тұрғысынан да өзекті. Біздің ұсынып отырған зерттеуіміздің мақсаты – қазақстандық платформа мен OpenAI негізіндегі чат-ассистенттің қазақ тіліндегі мәтіндерінде көрінетін тілдік ерекшеліктерді салыстырмалы түрде сипаттау және әдеби нормаға сәйкестік пен төл тіліміздің әдеби лексикасын қолдану деңгейін нақты тапсырмалар арқылы айқындау. Зерттеу өзегіне «қазақша тапсырма шартын дәл орындау», регистрді ұстану (ғылыми, ресми, публицистикалық, көркем), калька мен орынсыз кірме элементтердің көрінуі, мақал-мәтелді мәтінге табиғи кіріктіру және редакторлық қайта жазудағы нормалау сапасы алынды. Эмпирикалық материал бірдей 5 промт негізінде жиналып, екі платформаның жауаптары мәтін көлемі, құрылымдық толықтық және лексика-стилистикалық сапа тұрғысынан салыстырылды. Бұл тәсіл шағын корпусқа сүйенгенімен, нақты жанрлық тапсырмалар арқылы платформалардың қазақ тіліндегі «тілдік мінезін» айқын ажыратуға мүмкіндік береді. Зерттеудің ғылыми жаңалығы – қазақ тіліндегі генеративті жүйелерді «әдеби тіл» өлшемімен бағалауда мазмұндық дұрыстықты, тілдік норма, регистр және лексикалық табиғилық сияқты көрсеткіштерді біртұтастықта алып кешенді талдауы. Практикалық маңыздылығы нәтижелер қазақша ресми хат жазу, ғылыми түсіндірме құрастыру, публицистикалық мәтін жасау және көркем миниатюра жазу сияқты қолданбалы сценарийлерде қай платформаның қай параметр бойынша сенімді екенін көрсетуге бағытталуымен анықталады.

### **Материалдар мен әдістердің сипаттамасы**

Зерттеудің тәжірибелік материалы үшін отандық ISSAI платформасы мен OpenAI ChatGPT (GPT-5.2 Thinking) жүйесінің қазақ тілінде

генерациялаған мәтіндері алынды. Салыстырмалы корпус 5 бірдей промттан құралды:

- 1) «когнитивті диссонанс» ұғымын қазақ тілінде орыс, ағылшын сөздерін қоспай түсіндіру, 120-150 сөз;
- 2) деканға ресми стильде өтініш хат;
- 3) «Еңбек түбі – береке» мәтелін публицистикалық абзацқа кіріктіру;
- 4) теңеу, эпитет қолданылған көркем миниатюра;
- 5) калька мен кірме сөздерді алып тастап, мәтінді әдеби нормаға келтіру.

Әр промт бойынша екі платформаның жауаптары жинақталып, бір құжатқа енгізілді.

Әдістерге келер болсақ, бірдей тапсырмалар арқылы жұптық салыстыру қолданылды. Екі жүйеге бірдей промт мәтіндері бірдей шарттармен беріліп, алынған жауаптар лексика-стистикалық және нормалық көрсеткіштер бойынша салыстырылды. Әрбір мәтін 5 негізгі критерий бойынша сарапталып, 0–2 балл аралығында ұпаймен бағаланды:

1. Промт талабын орындау (көлем, мазмұнның толықтығы, нақты жауап беру),

2. Әдеби нормаға сәйкестік (грамматикалық және синтаксистік тазалық),

3. Калька мен кірме құрылымдардың ықпалы (орыс тілінен сөзбе-сөз аударылған тіркестердің кездесуі),

4. Стильдік регистрдің сақталуы (ресми, публицистикалық, көркем стиль талаптары),

5. Лексикалық байлық пен образдылық (әсіресе көркем мәтінде тіл өрнегі мен сөзқолданыс әртүрлілігі).

Жалпы балл әрбір критерийден алынған ұпайлардың қосындысы арқылы есептелді (максималды балл – 10). Алынған нәтижелер төмендегідей сапалық деңгейлер бойынша жіктелді:

- 9-10 балл – *өте жоғары сапа*: промт шарты толық орындалған, стильдік және грамматикалық норма сақталған;

- 7-8 балл – *жоғары сапа*: негізгі талаптар орындалған, жекелеген ауытқулар болса да, жалпы мазмұн мен тілдік құрылым орнықты;

- 5-6 балл – *орташа сапа*: мәтін талапқа сай, бірақ тілдік және стильдік кемшіліктер кездеседі;

- 3-4 балл – *төмен сапа*: жауап толық емес, стильдік сәйкессіздік, калька және қателер жиі байқалады;

- 0-2 балл – *қанағаттанарлықсыз деңгей*: промтқа жауап берілмеген немесе мазмұн мен құрылым айқын бұзылған.

Мұндай шкала генеративті платформалар тудырған мәтіндерді объективті салыстыруға, олардың қазақ тілінің нормалары мен стильдік ерекшеліктеріне сәйкестігін жүйелі түрде бағалауға мүмкіндік береді.

## Нәтижелер

Нәтижелер бөлімінде екі платформадан алынған жауаптар 5 промт бойынша топтастырылып, салыстырмалы түрде ұсынылды. Әр промтқа берілген мәтіндерден платформалардың қазақ тіліндегі генерациясының негізгі айырмашылықтары – тапсырма шартын сақтау деңгейі, жауаптың құрылымдық толықтығы, стильдік регистрдің орнықтылығы, әдеби тіл нормасына сәйкестік, сондай-ақ төл және бейнелі лексиканы қолдану сипаты – нақты мысалдар арқылы көрсетіледі. Төменде әр тапсырма бойынша екі жүйенің жауаптары қатар беріліп, байқалған айырмашылықтар қысқа түсіндірмемен белгіленді.

1-промт: «Когнитивті диссонанс» ұғымын қазақ тілінде 120 - 150 сөзбен түсіндір. Шарт: орысша, ағылшынша сөз қоспа, халықаралық терминді қолдансаң да, оны қазақша баламамен қатар бер, сөйлемдер әдеби нормада болсын.

*Когнитивті диссонанс (танымдық үйлеспеушілік) – адамның санасында бір-біріне сыйыспайтын ой, сенім, бағалау немесе әрекет қатар тұрған кезде пайда болатын ішкі жайсыздық күйі. Мұндайда адам өзін дұрыс сезінбей, көңілі алаңдап, шешіміне күмәндануы мүмкін. Үйлеспеушілік көбіне біз ұстанатын қағида мен нақты іс-әрекет қабыспағанда, не жаңа ақпарат бұрынғы түсінігімізге қайшы келгенде туындайды. Ішкі жайсыздықтан құтылу үшін адам әдетте үш жолдың бірін таңдайды: көзқарасын өзгерту, әрекетін түзету немесе қайшылықты әлсірететін уәж табу. Мысалы, денсаулықты бағалаймын дей тұра зиянды әдеттен бас тартпаған адам өзін ақтау үшін оның зиянын кемітіп көрсетуі мүмкін. Осы құбылыс мінез-құлық пен пікірдің өзгеруіне әсер етеді. <https://chatgpt.com/c/69641c20-b734-8320-97d0-ab00f0d8a51b>*

*Когнитивті диссонанс — психологиядағы негізгі ұғым, адамның ішкі қайшылықтарынан туындайтын әлеуметтік-психологиялық құбылыс. <https://oylan.nu.edu.kz/oylan/12005>*

2-промт. Қазақ тілінде ресми стильде деканға арналған өтініш хат жаз: студент ғылыми конференцияға қатысу үшін 2 күн сабақтан босатуды сұрайды. 120–150 сөз. Шарт: эмоциялық сөздер мен ауызекі тіркестер қолданба.

*Құрметті факультет деканы \_\_\_\_\_ мырза/ханым!*  
*Мен, \_\_\_\_\_ факультетінің \_\_\_\_\_*  
*мамандығы бойынша білім алушы \_\_\_\_\_ курс студенті*  
*\_\_\_\_\_ (аты-жөні), Сізден \_\_\_\_\_ (күні)*  
*және \_\_\_\_\_ (күні) күндері оқу сабақтарынан босатуыңызды*  
*сұраймын. Аталған күндері мен \_\_\_\_\_*  
*университетінде/қаласында өтетін \_\_\_\_\_ атты*  
*ғылыми конференцияға қатысамын және баяндама жасаймын (бағыты: \_\_\_\_\_).*

Конференцияға қатысуым оқу бағдарламасы аясындағы ғылыми-зерттеу жұмысымның нәтижелерін ұсынуға, кәсіби құзыреттерімді жетілдіруге және академиялық тәжірибе алмасуға бағытталған. Сабақтан босатылған күндер бойынша оқу материалдарын өз бетімше меңгеріп, өткізіп алған сабақтарды белгіленген тәртіпке сәйкес толықтыруға міндеттенемін.

Қосымша: конференцияға шақыру хаты/бағдарлама көшірмесі.

Күні: \_\_\_\_\_

Қолы: \_\_\_\_\_

Байланыс: \_\_\_\_\_

<https://chatgpt.com/c/69641c20-b734-8320-97d0-ab00f0d8a51b>

[Студенттің аты-жөні]

[Топ нөмірі]

[Электрондық пошта мекенжайы / телефон нөмірі]

[Күні], [Айы], [Жылы]

Декан [Деканның аты-жөні],

**Тақырыбы:** Ғылыми конференцияға қатысу үшін сабақтан босату туралы өтініш

Құрметті дегенмен,

Мен [Конференция атауы] атты ғылыми конференцияға қатысу үшін [Басталу күні] мен [Аяқталу күні] күндері арасында сабақтардан босатуды сұраймын. Бұл конференция [нақты пәндерге немесе зерттеулерге байланысты] бағытталған. Сабақтарды өзімнің жұмысым аяқтағаннан кейін қайта оқимын. Жоғалған материалдарды қайта оқып шығу және тапсырмаларды уақтылы орындау үшін қажетті уақытты бөлемін. Конференцияның ресми хаттамасында көрсетілгендей, бұл іс-шара студенттердің академиялық дамуына қатысты.

Өтінішімді қанағаттандыруыңызды сұраймын.

Сіздің құрметпен,

[Студенттің аты-жөні] <https://oylan.nu.edu.kz/oylan/12005>

3-прокт. «Еңбек түбі – береке» мақалын қолдана отырып, жас маман туралы шағын публицистикалық абзац жаз (120–150 сөз). Мақал мәтіннің ішінде табиғи кіріксіз, жасанды болып тұрмасын.

Жас мамандар өз кәсіби жолдарында еңбекке негізделген ұстаныммен алға басады. «Еңбек түбі — береке» деген мақал олардың әрбір қадамындағы сабыр мен ұмтылыстың нәтижесін көрсетеді. Технологиялық өзгерістерге ұшыраған заманда жас мамандар үздіксіз білім алу, инновацияларды іске асыру және өзін-өзі дамыту арқылы өзінің мүмкіндіктерін кеңейтеді. Олардың әрбір жұмысы қоғамға қосатын үлесін арттырып, болашақтағы

*жетістіктерге жол ашады. Еңбек пен ұмтылыс – бұл тек жеке өсу ғана емес, сонымен қатар ұрпақтың өркендеуіне үлес қосу. Жас мамандардың әрбір қадамындағы еңбегі қоғамдық прогреске бағытталған, сондықтан олардың әрекеттерінің маңызы зор. <https://oylan.nu.edu.kz/oylan/12005>*

*Жас маманның ең үлкен капиталы – диплом ғана емес, күн сайынғы табанды еңбегі. Бүгінгі еңбек нарығында тәжірибе талап етілетіні рас, алайда тәжірибе күтумен емес, жауапкершілік алып, ұсақ міндеттен бастап, нәтижеге ұмтылумен жиналады. Қызметке жаңа кірген маман әр тапсырманы “менікі емес” демей, уақытында орындап, қателігін талдап, білігін толықтырса, оның кәсіби беделі де қалыптасады. Ұжыммен тіл табысу, талапқа сай құжат жүргізу, уақытты дұрыс жоспарлау сияқты дағдылар да осы кезеңде бекіп шығады. Тәртіп пен ізденіс қатар жүрген жерде өсім болады: бір күннің еңбегі бір күндік нәтиже бермесе де, тұрақты әрекет алға жетелейді. Сондықтан “Еңбек түбі — береке” деген нақыл жас маманға бағыт-бағдар болып, ертеңгі жетістікке бастайтын қағидаға айналады. <https://chatgpt.com/c/69641c20-b734-8320-97d0-ab00f0d8a51b>*

4-промпт. Көркем стильде қысқа миниатюра жаз: қысқы кеш, қардың дыбысы, алыстағы шам, үй ішіндегі жылулық. 120–150 сөз. Шарт: кемінде 2 теңеу, 2 эпитет болсын; сөйлемдер әуезді әрі әдеби нормаға сай болсын.

*Күн батқан қысқы кеште ақ қар қабырғаларды қаптап, аспандағы жұлдыздар сияқты әлсіз жарқырайды. Қардың сыбдыры құлағымызға ұйқыны ұялататын әуен сияқты естіледі. Алыстағы шам алтын сәулесімен үйдің терезесіне қарап, оның жылулығын өзімізге тартады. Үй ішінде шай қайнап, кітаптардың беттерінің шуы қардың сыбдырына қарама-қарсы келеді. Жылы үйдің ішінде уақыт тоқтап, әлем сыртта ұйқыға батқан сияқты болады. Шамның жарығы қардың ақ қабырғасын қызыл-сары түстерге бояйды, ал үйдің ішінде қоңырау соғып, кітаптардың беттерінің шуы қардың сыбдырына қарама-қарсы келеді. Сырттағы қардың сыбдыры мен үй ішіндегі жылулықтың қарсылығы адамды өзінің ішкі әлеміне қайта оралуға итермелейді. <https://oylan.nu.edu.kz/oylan/12005>*

*Қысқы кеш аспанның қалың жамылғысындай тұнжырап, даланы үнсіздікке бөлеп тұр. Қардың дыбысы – жібек шапанның сыбдыры секілді, әр қадамды жұмсақ құшақтайды. Аязды ауа мөлдір айнадай жарқырап, тынысты қарып өтеді. Алыстағы жалғыз шам боз тұманның арасынан жұлдыздай жылтырап, жолға үміт шырағын төгеді. Терезе сыртында ақ қар ұлта мақтадай бұрқырап, қараңғылыққа нәзік сәуле қосқандай. Үй ішіне кіргенде, жылы пештің мейірімді лебі бетке тиіп, көңілдің мұзын ерітеді. Бөлмені жұмсақ жарық тербетіп, шайдың қоңыр иісі тыныштыққа сән береді. Қабырғадағы тынық көлеңкелер баяу сырғып, ошақтың қызыл шоғы жүрекке жылу құяды. Сырттағы сұсты табиғатқа қарсы, іште*

жайлы мекен – қорғандай берік. <https://chatgpt.com/c/69641c20-b734-8320-97d0-ab00f0d8a51b>

5-промпт. Төмендегі мәтінді әдеби қазақ тіліне келтіріп, калька мен орынсыз кірме сөздерді алып таста. Мағына сақталсын. Соңында 3 қысқа ереже жаз: қандай түзету принциптерін ұстандың. Мәтін: «Біз бұл зерттеуде өте маңызды проблеманы қарастырамыз, себебі қазіргі таңда бұл сұрақ актуалды болып табылады. Авторлар әртүрлі методтарды қолдана отырып, нәтиженің эффективтілігін көтеруге тырысады. Осыған байланысты қорытынды жасауға болады.»

Біз бұл зерттеуде өте маңызды мәселені қарастырамыз, себебі қазіргі таңда бұл сұрақ өзекті болып табылады. Авторлар әртүрлі әдістерді қолдана отырып, нәтиженің тиімділігін арттыруға тырысады. Осыған байланысты қорытынды жасауға болады. <https://oylan.nu.edu.kz/oylan/12005>

Біз бұл зерттеуде қазіргі кезең үшін өзекті мәселені қарастырамыз. Зерттеу барысында авторлар түрлі тәсілдерді қолданып, алынған нәтижелердің нәтижелілігін арттыруға ұмтылады. Осы негізде тиісті қорытынды жасауға болады. <https://chatgpt.com/c/69641c20-b734-8320-97d0-ab00f0d8a51b>

Бес түрлі промпт негізінде алынған жауаптар екі платформа бойынша жүйеленіп, салыстырмалы түрде 1-кестеде берілді.

Кесте 1. 5 промпт бойынша екі платформадан алынған жауаптардың салыстырмалы кестесі

Көрсеткіш (0–2)	OpenAI (ChatGPT)	Қазақстандық платформа
1) Промт талабын орындау (көлем, шарт, толықтық)	1 — 1-промптта талап орындалған, бірақ 5-промптта “3 ереже” бөлігі берілмеген	0 — 1-промптта жауап бір сөйлеммен шектелген, 5-промптта “3 ереже” берілмеген
2) Әдеби норма (грамматика/синтаксис тазалығы)	2 — сөйлем құрылысы орнықты, түсіндірме логикасы дұрыс	1 — ресми мәтінде синтаксистік дәлсіздік бар (“өзімнің жұмысын...”)
3) Калька/кірме ықпалы (болып табылады, қорытынды жасауға болады т.б.)	1 — калькаға жақын тіркес сақталады (“тиісті қорытынды жасауға болады”)	— калька үлгілері сақталған (“өзекті болып табылады”, “қорытынды жасауға болады”)
4) Стильдік регистрді ұстау (ресми/публицист./көркем)	2 — публицистикада нақты тірек ұғымдар, аргументация жинақы; мақал табиғи кіріккен	1 — публицистикада жалпылама клише көп (“инновация”, “қоғамдық прогрес”); ресми стильде қаратпа/құрылым ауытқуы бар
5) Лексикалық байлық, образдылық (көркем мәтін)	2 — бейнелі теңеу-эпитеттер табиғи (“жібек шапанның сыбдыры”, “ұлпа мактадай”)	1 — көркем мәтінде қайталау мен композициялық ауырлық байқалады

Бес тапсырма бойынша екі платформадан екі түрлі нәтиже алдық. Бағалау нәтижесінде OpenAI жүйесі 8 ұпай алып, жоғары сапа деңгейін көрсетті. Оның мәтіндерінде стильдік регистр сақталып, образдылық пен аргументация көрініс тапты. Ал Қазақстандық платформа 3 ұпай алып, төмен сапа деңгейінде қалып отыр. Тапсырма шартын орындаудан бастап, қазақ тіліне тән әуезділікті, бейнелілікті бере алмады. Бейнелі сөздерді орынсыз қолданғаны мәтіннің мазмұнына нұқсан келтірді. Дегенмен OpenAI платформасында да метатілдік компоненттің әлсіз екенін байқадық. Калькаға қатысты тапсырмада екеуінде де мәтін толықтай таза табиғи түрге ауыспады, сөзбе-сөз аударма үлгілер сақталды.

### **Талқылау**

Қазақ тіліндегі генеративті жасанды интеллект (ЖИ) мәтіндерінің тілдік және стильдік ерекшеліктерін талдау қазіргі таңда маңызды зерттеу бағыттарының біріне айналып отыр. Бұл салада Dong және т.б. ұсынған табиғи тіл генерациясына (NLG) арналған жан-жақты шолу жұмысында қазіргі модельдердің мәтін түзу қабілеті, стильдік вариативтілігі мен бағалау тәсілдері кеңінен қарастырылады [1,15-б.]. Осы теориялық негізге сүйене отырып, біз OpenAI мен отандық ISSAI жүйелерінің қазақ тілінде генерациялаған мәтіндеріндегі лексикалық талдау, синтаксистік құрылым және стильдік үйлесімділікті салыстырмалы түрде талдаймыз. Бұл талдау ЖИ мәтіндерінің тілдік сапасына, прагматикалық сәйкестігіне және оқырман қабылдауына әсер ететін негізгі факторларды ашуға бағытталған.

Зерттеудің нәтижелері екі платформаның қазақ тіліндегі генерацияланған мәтіндерінде әдеби нормаға бағыну, тапсырма шартын орындау және лексикалық-стильдік жағы біркелкі емес екенін көрсетті. Р.Боммазани зерттеуінде ЖИ мәтінді бағалау кезінде HELM тәсілін қолдануды ұсынады [2]. Бұл мәтінді бағалау кезінде жай ғана өз көзқарас тұрғысынан дұрыс не бұрыс демей, нақты регистр, норма, орындаушылық, фразеология сияқты өлшемдермен бағалауды айтамыз. Біздің ойымызша, осындай бағалау ЖИ мәтіндердің дұрыстығын өлшеуге, оны одан әрі жетілдіруге дұрыс бағыт береді деп санаймыз. Жалпы ЖИ-нің нормадан ауытқуы қалыпты нәрсе. Норма контекстке байланысты уәжді, уәжсіз ауытқу жасайды. Кейде ЖИ мәтіндері сенімді көрінгенмен, неге нормадан ауытқиды деген сұраққа Э.Бендер бастаған зерттеушілер былай дейді [3, 612-б.]: LLM-дер мәтін шығару үшін өте үлкен деректерде оқытылатынын, ал ол деректер кітаптар мен сайттардағы мәтіндерді қамтитынын айтады. Мұндай деректердің барлығы бірдей тілдік нормаға сай болмайды. Сол себепті ЖИ-де соған ұқсас мәтіндер тудыруы заңды құбылыс деп санайды. Ал нормадан ауытқуды азайту үшін алдын ала үйретілген модельді нөлден қайта оқытпай-ақ, алдын-ала белгіленген нормаларға сезімтал ету ұсынылады.

ЖИ мәтіндері адам жазған мәтіндерден ерекшеленіп тұратыны белгілі. Ал екі ЖИ платформасындағы айырмашылық неден деген сұрақ туады? «Студент жұмыстарындағы чат GPT генерациялаған мәтінді қалай танимыз» деген А.Мидзумото мынадай қызықты дерек келтіреді [4]. Ол ЖИ мәтіндерінің оңай танылатындығын оның лингвистикалық профильдерінің жүйелілігімен байланыстырады: яғни ЖИ мәтіні бірқалыпты, құрылымы жинақы болуы мүмкін. Соның нәтижесінде ол көп жағдайда табиғи тілдегі жақсы жазылған мәтін әсерін береді. Біздің зерттеуде OpenAI бірдей промтқа жауап бергенде жанрлық талаптарды толық орындап, құрылымды сақтаған. Бұл оның нұсқауды орындау, жанрлық имитациялау қабілеті жоғары екенін көрсетеді.

Осы орайда отандық платформа мәтін генерациялау немесе жауап алудың әлсіздігіне мына факторлар әсер етуі мүмкін:

- қазақ тіліндегі сапалы әдеби, ресми корпустың салыстырмалы тапшылығы;
- жанрлық нормаға сай ресми стильге сай құжаттар үлгілерінің аздығы;
- жауап ұзындығын, құрылымын тұрақты ұстайтын компоненттердің шектеулігі;
- ең соңғысы техникалық себептерге тікелей байланысты дейміз.

М.Маккарти зерттеуінде ЖИ мәтіндеріндегі лексикалық әртүрлілікті өлшеудің тәсілін ұсынады [5, 390-б.]. Ол лексикалық әртүрлілікті (lexical diversity) өлшеудің сенімді тәсілдері ретінде MTLД, vocd-D және HD-D көрсеткіштерін валидтеп, MTLД-тің мәтіндегі сөздік әркелкілікті жолақ-жолақ (string) бойынша есептеп, тек TTR-ға (type-token ratio) сүйенгендегі ұзындыққа тәуелді қателікті азайтатынын түсіндіреді. Мәтінде бейнелі сөздер көбейген сайын мәтінде қайталанбайтын сөздер мен сирек бірліктердің үлесі өседі, демек лексикалық әртүрлілік жоғарылауы ықтимал. Көркем регистрдегі сөздің байлығы мәтіннің көркемдігіне, образдылығына жауап береді деген сөз. Осы автордың басқа мақаласында лексикалық әртүрлілік болса да, қысқа мәтіндерді генерациялау кезінде абай болу керегін ескертеді. Өйткені кейбір индекстер мәтін ұзындығына тым сезімтал. Мәтін қысқарған сайын тип-токен қатынасы (TTR) секілді көрсеткіштер жасанды түрде өсіп, лексикалық байлықты асыра көрсетеді, ал бұл қысқа генерацияланған мәтіндерді көркем деп қате бағалауға әкеледі.

Лексикалық әртүрлілік индексі мәтін ұзындығына тәуелсіз деген кең тараған сенім толық дұрыс еместігін айтады. Яғни мәтін көлеміне қарай лексикалық әртүрлілік те өзгеріп отырады деген мәселені қарастырады. Лексикалық әртүрлілік индексі (vocd) идеясы мәтіннен белгілі бір көлемде сөздерді қайта-қайта іріктеп алып, тип-токен (type-token) қатынасының қисығын модельдеу арқылы лексикалық әртүрліліктің бір санын шығару. Бірақ McCarthy мен Jarvis бұл процедурадағы бір түйінді мәселені ашады:

- D, шын мәнінде, мәтіндегі сөздердің кездесу ықтималдықтарының агрегатталған бейнесін береді,

- ал ол ықтималдықтардың өзі (әсіресе сирек сөздер немесе қайталаным үлесі) мәтін ұзындығымен өзгеріп отырады, сондықтан D де өзгеруі мүмкін [6].

Осы ойды біздің зерттеуге қатысты айтатын болсақ, мәтіндердегі әдеби, бейнелі сөздердің орнымен қолданылуы, бар болуы мәтін ұзындығына байланысты. Біздің ойымызша, промт бойынша 100-120 сөзден тұратын мәтін үшін тым көп бейнелі сөздер қолданылған. Ол мәтінге жасандылық беріп тұр деген ойдамыз. Сондықтан осы критерийге қатысты мәтінді көлемін теңестіріп, бірнеше рет генерация жасаған дұрыс деп санаймыз.

Отандық зерттеушілер М.Қ.Шенгелбай, Д.А.Әлкебаева, Ж.С.Таласпаева жазған мақалада авторлар ЖИ өнімдері тілдің сыртқы формасын жүйелі берсе де, ұлттық мәдениет пен эмоционалдық мазмұнды жеткізуде әлсіздік танытады деп нақты айтады [7, 140-б.]. Мақалада табиғи тіл мен ЖИ тілі салыстырылғандықтан авторлар табиғи тілдегі мәтіндердің лексикалық әртүрлілігі мол, синтаксисі күрделірек және эмоционалдық-идеологиялық контексті бай деп тұжырымдайды. Біздің зерттеуде екі ЖИ платформасын салыстыру болғандықтан, табиғи тілге қатысты ештеңе айта алмаймыз. Дегенмен, табиғи тілге жақындық OpenAI -да байқалып тұр. Авторлардың ойы біз жоғарыда айтқан тұжырыммен сәйкес келеді. Отандық платформада жауаптардың шаблондануы мен бейнелі сөзқолданыстың сиреуі ЖИ-ге тән экспрессиялық шектеумен, сондай-ақ, қазақ тіліндегі жанрлық-стильдік ресурстардың жеткіліктілігі мәселесімен байланысты, ал OpenAI жүйесінің салыстырмалы түрде жақсы көрінуі көбіне құрылымдық толықтық пен тапсырма талабын орындаудың тұрақтылығынан байқалады. ЖИ мәтіндері көбінесе тілді формалды түрде игергенімен, олардың мазмұнында ұлттық-мәдени мәндер мен тілдік символикалар толық көрініс таба бермейді. Бұл мәселе Г.К.Ахметова мен Ж.М.Байғожина еңбегінде тілдің мәдени кодтарды жеткізуші негізі құрал екендігімен түсіндіріледі [8].

Д.Ахмед-Заки бастаған отандық зерттеушілер қазақ тіліндегі ЖИ мәтін өңдеудің негізгі мәселесін айқындап, «қазақ тілі – ресурсы шектеулі тіл және агглюнативті тіл, сондықтан сапалы генерация үшін корпус, морфологиялық құралдар, дизамбигуация сияқты базалық инфрақұрылым керек», - дейді [9]. ISSAI платформасындағы кемшілікті ресурсы аз тілде модель жанрлық алуандықты толық игеріп үлгермейді, соның нәтижесінде шаблондану, қысқа жауап, калька жиілеуі мүмкін деп санаймыз. Ал OpenAI секілді жүйелер көптілді, үлкен көлемді дерекпен оқытылатындықтан, жанрлық үлгісі кең болады. Бұл шетелдік платформаның тым «ақылды» болуынан емес, дерек көлеміне тікелей байланысты деп тұжырым жасаймыз. Д.Ахмед-Заки және басқа да зерттеуші ғалымдар отандық платформа үшін негізгі ресурс қазақ тілінің медиа корпусы ұсынылатынын айтады.

Қорытындылай айтқанда, егер отандық жүйелердің негізгі дерекқоры көбіне жаңалықтар мен ресми стильдегі мәтіндерден тұрса, модельдің тілдік мінезі заңды түрде ақпараттық-нейтрал реңкке бейімделеді. Соның салдарынан мәтінде көркемдік, бейнелеуіш тәсілдер мен экспрессивті құралдар сирек ұшырасуы мүмкін.

Осы ойды В.Карюкин, Д.Рахимова, А.Карибаева, А.Турганбаева, А.Турарбек те өз зерттеулерінде дамытып, осыған ұқсас тұжырым жасаған. Олар моно-қазақ мәтінін ғылыми журналдардағы редакцияланған сапалы мәтіндерден алғанын арнайы атап өтеді [10]. Бұл тағы да корпустың тым ресми деректерге жүгінетінін көрсетеді. Сонымен қатар авторлар корпусты мақсатты түрде жинап, нормалап, редакциялап, ресми мемлекеттік сайттардан алынған параллель мәтіндер және аударма арқылы жинаған синтетикалық мәтіндерді талқылайды. Бұл өз кезегінде отандық платформа жауаптарының тым қысқалығы, шаблондануы немесе калькаға бейімділігінің себебін көрсетеді. Бұл тұжырым А.Алдаш пен Ә.Сейдадат зерттеуінде сипатталған қазақ ресми-іскерлік стилінің ерекшеліктерімен үндеседі [11]. Авторлар ресми стильдің прагматикалық табиғаты мен тілдік бірліктерінің стандартты құрылымын сипаттай отырып, ондағы формализм мен бірізділікке негізделген тілдік нұсқалар мен синтаксистік үлгілердің жиілігін атап көрсетеді. Демек, отандық платформалар негізге алған корпус табиғаты – генерацияланған мәтіннің стильдік біртектілігі мен табиғи емес құрылымдардың басты факторларының бірі.

Қазақ тіліне арналған тілдік платформа сапасының төмендеуі көбіне интеллект әлсіздігі емес, сапалы деректің тапшылығынан және синтетикалық деректің әсерінен болуы мүмкін деген тұжырымды Н.Кадырбек, Ж.Түймебаев, М.Мансурова, В.Виегас келтірген [12]. Зерттеушілер корпус үшін сапалы мәтіндерді жинап, оны аударып, кейін қайта базаға қосу керегін айтқан. Олардың ойынша, аударманы негізінен ресми әлеуметтік, саяси және ғылыми мәтіндер қорында сапалы етуге болады.

### **Қорытынды**

Зерттеу нәтижелері қазақстандық платформа мен OpenAI жүйесінің қазақ тілінде мәтін тудыру сапасы бірдей тапсырмалар жағдайында тең деңгейде емес екенін көрсетті. Бес промт бойынша OpenAI жауаптары көп жағдайда тапсырма шартын сақтап, мәтінді құрылымдық тұрғыдан толық құрастырып, стильдік регистрді (ғылыми, ресми, публицистикалық, көркем) бірізді ұстауға бейім болды. Отандық платформада кей жауаптардың қысқалығы, шаблонды формулаларға сүйенуі және ресми-іскери мәтінде синтаксистік дәлсіздіктердің көрінуі байқалып, көркем стильде бейнелі-әдеби сөзқолданыстың ауқымы шектеулі екенін аңғартты. Сонымен қатар редакторлық қайта жазу тапсырмасы екі жүйеде де калька құрылымдарының толық жойылмайтынын және метатапсырмалар (мысалы, түзету принциптерін бөлек көрсету) кейде орындалмай қалатынын көрсетті.

Алынған айырмашылықтарды түсіндіруде қазақ тілінің цифрлық ресурс экожүйесі шешуші фактор ретінде көрінеді: қазақ тіліндегі жанрлық тұрғыдан әртарап, редакцияланған корпус пен морфологиялық өңдеу құралдарының жеткіліктілігі генеративті жүйелердің әдеби нормаға жақындауына тікелей ықпал етеді. Сондықтан отандық платформалардың сапасын арттыру үшін ресми үлгілер, публицистикалық мәтіндер, көркем әдебиет және нормаланған редакторлық жұптар (қате→дұрыс) сияқты деректер қорын кеңейту, стильдік бейімдеу критерийлерін күшейту маңызды. Зерттеу шағын промт-жиынтыққа негізделгендіктен, болашақта тапсырмаларды көбейтіп, бір промтты бірнеше рет қайталау арқылы вариативтілікті есепке алу және лексикалық әртүрлілікті ұзындық әсерін бақылай отырып өлшейтін көрсеткіштермен толықтыру көзделіп отыр. Жалпы алғанда, біздің зерттеу нәтижеміз қазақ тіліндегі генеративті платформаларды бағалауда мазмұннан бөлек, әдеби норма мен стильдік табиғилықты өлшем ретінде енгізудің қажеттігін және отандық жүйелерді дамытуда тілдік ресурстарды жүйелі толықтырудың стратегиялық маңызын айқындады.

### ӘДЕБИЕТ

[1] Dong C., Wang J., Zheng Y., Chen Y., Feng Y. A survey of natural language generation // *ACM Computing Surveys*. – 2022. – Vol. 55, №. 8. – Pp. 1-38. DOI:10.1145/3554727\_

[2] Bommasani R. et al. On the Opportunities and Risks of Foundation Models // *arXiv*, 2022. <https://arxiv.org/abs/2211.09110> 22.01.2026

[3] Bender E. M., Gebru T. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. – 2021. – Pp. 610-623. DOI: 10.1145/3442188.3445922

[4] Mizumoto T., Fujita A., Komachi M. Exploring Evaluation Methods for Large Language Models in Academic Writing Assistance // *Artificial Intelligence in Corporate Communications*, 2024. – Article 100106. DOI: 10.1016/j.acorp.2024.100106

[5] McCarthy P. M., Jarvis S. MTL, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment // *Behavior Research Methods*. – 2010. – Vol. 42, № 2. – Pp. 381-392. DOI: 10.3758/BRM.42.2.381

[6] Koizumi R., In'nami Y. Vocabulary Knowledge and Speaking Proficiency among Second Language Learners of English // *Language Testing*. – 2007. – Vol. 24(3). – P. 289–312. – DOI: 10.1177/0265532207080767

[7] Шенгелбай М., Алкебаева Д., Таласпаева Ж. Жасанды интеллект тілі: лингвистиканың жаңа көкжиектері // *Bulletin of Shokan Ualikhanov Kokshetau University. Philological Series*. – 2025. – № 3. – С. 138–149. DOI: 10.59102/kufil/2025/iss3pp138-149

[8] Ахметова Г. К., Байгожина Ж. М. Тіл – мәдениетті танытушы құрал // *Известия. Серия: Педагогические науки*. – 2021. – Т. 60. – №1. – С. 34-44. <https://doi.org/10.48371/PEDS.2021.60.1.003>

[9] Ahmed-Zaki, D., Mansurova, M., Madiyeva, G., Kadyrbek, N., & Kyzrbaeva, M. Development of an information system for preliminary processing of texts in the Kazakh language. // Cogent Engineering. – 2021. – Vol. 8(1). DOI: 10.1080/23311916.2021.1896418

[10] Karyukin V., Zhanabergenova D., Kalekeyeva A., Madiyeva G., Sarsenbayev K. Efficient and interpretable text style transfer with mutual information maximization // PeerJ Computer Science. – 2023. – Vol. 9. – Article e1224. DOI: 10.7717/peerj-cs.1224

[11] Алдаш А., Сейдапат Ә. Қазақ ресми-іскерлік стилі: сипаттамасы, стандартты тілдік бірліктердің құрылымы мен мазмұны // Абылай хан атындағы ҚазХҚжәнеӘТУ Хабаршысы. Филология ғылымдары сериясы. – 2024. – Т. 74. – № 3. – С. 12-31. <https://doi.org/10.48371/PHILS.2024.3.74.001>

[12] Kadyrbek N., Tuimebayev Zh., Mansurova M., Viegas V. Automatic Detection of Language Style and Emotional Tone in Multilingual Texts // Big Data and Cognitive Computing. – 2023. – Vol. 9(5). – Article 137. DOI: 10.3390/bdcc9050137

## REFERENCES

[1] Dong C., Wang J., Zheng Y., Chen Y., Feng Y. A survey of natural language generation // ACM Computing Surveys. – 2022. – Vol. 55, No. 8. – Pp. 1-38. DOI:10.1145/3554727\_

[2] Bommasani R. et al. On the Opportunities and Risks of Foundation Models // arXiv, 2022. <https://arxiv.org/abs/2211.09110> 22.01.2026

[3] Bender E. M., Gebru T. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). – 2021. – Pp. 610–623. DOI: 10.1145/3442188.3445922

[4] Mizumoto T., Fujita A., Komachi M. Exploring Evaluation Methods for Large Language Models in Academic Writing Assistance // Artificial Intelligence in Corporate Communications, 2024. – Article 100106. DOI: 10.1016/j.acorp.2024.100106

[5] McCarthy P. M., Jarvis S. MTL, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment // Behavior Research Methods. – 2010. – Vol. 42, No. 2. – Pp. 381-392. DOI: 10.3758/BRM.42.2.381

[6] Koizumi R., In'nami Y. Vocabulary Knowledge and Speaking Proficiency among Second Language Learners of English // Language Testing. – 2007. – Vol. 24(3). – P. 289–312. DOI: 10.1177/0265532207080767

[7] Shengelbay M., Alkebaeva D., Talaspayeva Zh. Zhasandy intellekt tili: lingvistikaning zhana kokzhiakteri [Language of Artificial Intelligence: New Horizons in Linguistics] // Bulletin of Shokan Ualikhanov Kokshetau University. Philological Series. – 2025. – № 3. – Pp. 138-149. – DOI: 10.59102/kufil/2025/iss3pp138-149\_[in Kaz].

[8] Akhmetova G. K., Baigozhina Zh. M. Til – madeniatti tanytushy

kural [Language as a Tool for Representing Culture] // Izvestiya. Seriya: Pedagogicheskie nauki. – 2021. – Vol. 60. – № 1. – Pp. 34-44. <https://doi.org/10.48371/PEDS.2021.60.1.003> [in Kaz.]

[9] Ahmed-Zaki D., Mansurova M., Madiyeva G., Kadyrbek N., & Kyzrbaeva M. Development of an information system for preliminary processing of texts in the Kazakh language // Cogent Engineering. – 2021. – Vol. 8(1). DOI: 10.1080/23311916.2021.1896418

[10] Karyukin V., Zhanabergenova D., Kalekeyeva A., Madiyeva G., Sarsenbayev K. Efficient and interpretable text style transfer with mutual information maximization // PeerJ Computer Science. – 2023. – Vol. 9. – Article e1224. DOI: 10.7717/peerj-cs.1224

[11] Aldash A., Seydamat A. Qazaq resmi-iskerlik stili: sipattamasy, standarty tildik birlikterdin qurylymy men mazmuny [Kazakh Official-Business Style: Description, Structure and Content of Standard Language Units] // Ablai Khan atyndagy QazHKhZhATU Khabarshysy. Filologiya gylymdary seriasy. – 2024. – T. 74. – № 3. – B. 12-31. <https://doi.org/10.48371/PHILS.2024.3.74.001> [in Kaz.]

[12] Kadyrbek N., Tuimebayev Zh., Mansurova M., Viegas V. Automatic Detection of Language Style and Emotional Tone in Multilingual Texts // Big Data and Cognitive Computing. – 2023. – Vol. 9(5). – Article 137. – DOI: 10.3390/bdcc9050137

## **ЯЗЫКОВЫЕ И СТИЛЕВЫЕ ОСОБЕННОСТИ КАЗАХСКОЯЗЫЧНЫХ ТЕКСТОВ, СГЕНЕРИРОВАННЫХ ГЕНЕРАТИВНЫМ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ (НА ПРИМЕРЕ OPENAI И ОТЕЧЕСТВЕННОЙ ПЛАТФОРМЫ)**

\*Ташпеннова К.Е.<sup>1</sup>, Садирова К.К.<sup>2</sup>

\*<sup>1,2</sup> Актюбинский региональный университет имени К. Жубанова  
Актобе, Казахстан

**Аннотация.** В данной статье проводится сравнительный анализ языковых и стилевых особенностей текстов на казахском языке, сгенерированных генеративными платформами искусственного интеллекта (ИИ). Оценка лингвистического качества генеративных моделей представляет собой актуальную проблему на пересечении современной лингвистики и цифровых технологий. В качестве объекта исследования были выбраны модель ChatGPT от OpenAI и отечественная ИИ-платформа. Их способность генерировать тексты на казахском языке оценивалась по пяти основным критериям: соответствие требованиям промпта, соблюдение литературных норм, влияние кальки и заимствованных конструкций, стилистическая регистровая точность и лексическое разнообразие. Оценка проводилась по шкале от 0 до 2 баллов, с количественным описанием каждого ответа. Результаты анализа показали, что тексты, созданные платформой

OpenAI, отличаются большей содержательной глубиной, стилистическим богатством, логичностью аргументации и выразительностью. Также отмечается адекватное применение норм публицистического и художественного стиля. Тексты отечественной платформы, напротив, были более краткими, шаблонными, в них встречались элементы кальки и грамматические ошибки. Эти различия в первую очередь объясняются составом используемых языковых корпусов и преобладанием материалов в официально-деловом стиле. Особое внимание в исследовании уделено отражению культурно-национального содержания, включая пословицы, метафоры и фольклорные элементы, которые генеративные системы передают с ограничениями. Работа подчеркивает необходимость учета языковых норм и стилевой согласованности при разработке ИИ-моделей для казахского языка. Также косвенно рассматриваются аспекты прагматической уместности и понимания контекста при генерации текстов. Итоговые выводы могут служить методологической основой для создания более качественных ИИ-систем и языковых корпусов, адаптированных под казахский язык.

**Ключевые слова:** казахский язык, генеративный искусственный интеллект, стилистика текста, калька, стилевой регистр, языковой корпус, отечественная платформа, OpenAI

## LINGUISTIC AND STYLISTIC FEATURES OF GENERATIVE AI TEXTS IN THE KAZAKH LANGUAGE (BASED ON OPENAI AND A DOMESTIC PLATFORM)

\*Tashpenova K.E.<sup>1</sup>, Sadyrova K.K.<sup>2</sup>

\*<sup>1,2</sup> K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan

**Abstract.** This article presents a comparative analysis of the linguistic and stylistic features of texts generated in the Kazakh language by generative artificial intelligence (AI) platforms. Evaluating the linguistic quality of generative models is a pressing issue at the intersection of modern linguistics and digital technologies. The study focuses on OpenAI's ChatGPT model and a domestic Kazakh AI platform, assessing their ability to generate Kazakh-language texts based on five core criteria: prompt fulfillment, compliance with literary language norms, influence of calques and borrowings, maintenance of stylistic register, and lexical richness. A 0–2 point scale was used for evaluation, with quantitative indicators describing the quality of each response. The results indicate that OpenAI's texts demonstrate greater depth, stylistic richness, coherent argumentative logic, and expressive language. Additionally, its ability to adhere to the norms of publicistic and artistic style was evident. In contrast, the domestic platform's texts tended to be shorter, more templated, and often included calqued structures and grammatical inaccuracies. These differences are primarily linked to the linguistic corpora

used and the predominance of official-business style materials. The study also examines the presence of culturally and nationally specific content, revealing that generative systems have limitations in conveying Kazakh-specific figurative language, such as proverbs and metaphors. The findings emphasize the importance of aligning linguistic norms and stylistic consistency when developing Kazakh-language AI models. The analysis also touches upon pragmatic appropriateness and contextual understanding, highlighting the challenges generative models face in mastering the semantic and discursive nuances of the Kazakh language. These results may serve as a methodological foundation for future improvements in Kazakh-language AI systems and linguistic data resources.

**Keywords:** Kazakh language, generative artificial intelligence, text stylistics, calque, stylistic register, language corpus, OpenAI, domestic platform

*Мақала түсті / Статья поступила / Received: 25.02.2026.*

*Жариялауға қабылданды / Принята к публикации / Accepted: 25.06.2026.*

***Авторлар туралы мәлімет:***

Тәшпеннова Құралай Елібайқызы – докторант, Қ.Жұбанов атындағы Ақтөбе өңірлік университеті Ақтөбе, Қазақстан, e-mail: kuralai.tashpenova@mail.ru, <http://orcid.org/0009-0001-4732-2598>

Садирова Қулзат Каниевна – филология ғылымдарының докторы, профессор. Қ.Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе, Қазақстан, e-mail: kulzat.sadirova.70@mail.ru, <http://orcid.org/0000-0001-6092-8191>

***Информация об авторах:***

Ташпеннова Куралай Елибайқызы – докторант, Актюбинский региональный университет имени К. Жубанова, Актөбе, Казахстан, e-mail: kuralai.tashpenova@mail.ru, <http://orcid.org/0009-0001-4732-2598>

Садирова Қулзат Каниевна – доктор филологических наук, профессор, Актюбинский региональный университет имени К. Жубанова, Актөбе, Казахстан. e-mail: kulzat.sadirova.70@mail.ru, <http://orcid.org/0000-0001-6092-8191>

***Information about the author:***

Kuralai Yelibaykyzy Tashpenova – PhD student, K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan, e-mail: kuralai.tashpenova@mail.ru, <http://orcid.org/0009-0001-4732-2598>

Kulzat Kanievna Sadyrova – Doctor of Philology, Professor, K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan, e-mail: kulzat.sadirova.70@mail.ru, <http://orcid.org/0000-0001-6092-8191>