

Т. Рамазанов<sup>1\*</sup>, А. Жаңабекова<sup>1</sup>, А. Садық<sup>2</sup>

<sup>1</sup>А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

<sup>2</sup>К. Сағадиев атындағы Халықаралық бизнес университеті, Алматы, Қазақстан

\*e-mail: tphd-92@mail.ru

## QAZLLM VS. CHATGPT: ҚАЗАҚ ЭТНОМӘДЕНИ КОДЫН ЖЕТКІЗУДЕГІ ЛИНГВИСТИКАЛЫҚ ҰҚСАСТЫҚТАР МЕН БАСЫМДЫҚТАР

Үлкен тілдік моделдер қазақ тіліндегі цифрлық қызметтерді кеңейтті, алайда көптеген жүйелердің ағылшын тілді дерекке көбірек сүйенуі ұлттық мазмұнды, әсіресе қазақтың этномәдени кодын дәл жеткізуде қиындық туғызады. Осы жұмыста QazLLM мен ChatGPT жауаптары қазақ тіліндегі мәдени-тарихи, құқықтық және идиомалық контексті түсіндіру қабілеті тұрғысынан салыстырмалы лингвистикалық талдаудан өтті.

Зерттеудің өзектілігі қазақ қоғамындағы мемлекеттік сервистер, білім беру және медиакөністікте мәдени тұрғыдан дұрыс, прагматикалық тұрғыдан орнықты қазақша генерацияға сұраныстың артуымен айқындалады. Эмпирикалық негіз ретінде сарапшылар белгілеген 140 000 мысал қолданылып, нәтижелер қосымша түрде OpenAI жүйелері және Назарбаев Университеті жанындағы ISSAI әзірлемелерімен салыстырылды.

Зерттеу барысында ұлттық-мәдени реалияларды тану, фразеологизмдер мен тұрақты тіркестерді уәжді түсіндіру, бағалауыштық реңкті дәл беру, сондай-ақ диалогтік нормаларды сақтау сияқты көрсеткіштер бойынша моделдердің жауаптары бағаланды. Сондай-ақ қателердің типологиясы жасалып, мәдени жалпылау, ұғымды алмастыру, контексті үстірт түсіндіру тәрізді ауытқулар жіктелді.

Талдау QazLLM-нің ұлттық болмысқа тән ұғымдарды, фразеологизмдер мен бағалау реңктерін контекспен қабыстыра беруі жиірек екенін, ал ChatGPT-дің әмбебап ақпараттық сұрақтарда тұрақтырақ екенін көрсетті. Зерттеудің ғылыми құндылығы мәдени релеванттылықты өлшейтін лингвистикалық критерийлерді нақтылауымен және локалданған корпустың рөлін дәлелдеуімен маңызды. Практикалық тұрғыдан нәтиже қазақша чатботтар, оқу-әдістемелік материал, аудармаға көмекші құралдар және мәдени анықтамалық сервистерді жетілдіруге негіз бола алады.

**Түйін сөздер:** үлкен тілдік модель, жасанды интеллект, қазақ тілі, мәдени контекст, трансформер, деректер корпусы, локализация.

T. Ramazanov<sup>1\*</sup>, A.Zhanabekova<sup>1</sup>, A.Sadyk<sup>2</sup>

<sup>1</sup>A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan

<sup>2</sup>K. Sagadiyev University of International Business, Almaty, Kazakhstan

\*e-mail: tphd-92@mail.ru

### QazLLM vs. ChatGPT: Linguistic Similarities and Advantages in Conveying the Kazakh Ethnocultural Code

The rapid development of Large Language Models has expanded the range of Kazakh-language digital services; however, the heavy reliance of many systems on English-centric data often hampers the accurate rendering of nationally marked content, most notably the Kazakh ethnocultural code. This paper presents a comparative linguistic analysis of QazLLM and ChatGPT, examining their capacity to interpret Kazakh cultural-historical, legal, and idiomatic contexts.

The study is timely given the growing demand in public services, education, and media for Kazakh-language generation that is culturally appropriate and pragmatically robust. The empirical basis comprises 140,000 expert-annotated examples, and the results are additionally compared with OpenAI systems and solutions developed at Nazarbayev University (ISSAI).

Model outputs were evaluated against indicators including recognition of culture-specific realia, motivated explanations of phraseological units and fixed expressions, accurate delivery of evaluative meaning, and adherence to dialogic norms. A complementary error typology was constructed, covering cultural overgeneralization, concept substitution, and superficial contextual interpretation.

The analysis indicates that QazLLM more often aligns culturally specific notions, phraseology, and evaluative nuances with the surrounding context, whereas ChatGPT tends to be more consistent on gen-

ria for measuring cultural relevance and in substantiating the role of localized corpora in culture-sensitive AI. Practically, the findings can inform the improvement of Kazakh-language chatbots, instructional and methodological materials, translation assistants, and cultural reference services.

**Keywords:** large language model, artificial intelligence, Kazakh language, cultural context, transformer, data corpus, localization.

Т. Рамазанов<sup>1\*</sup>, А.Жанабекова<sup>1</sup>, А. Садық<sup>2</sup>

<sup>1</sup>Институт языкознания имени А. Байтурсынулы, Алматы, Казахстан

<sup>2</sup>Университет международного бизнеса имени К. Сагадиева, Алматы, Казахстан

\*e-mail: tphd-92@mail.ru

### QazLLM vs. ChatGPT: лингвистические сходства и преимущества в передаче казахского этнокультурного кода

Развитие больших языковых моделей расширило спектр цифровых сервисов на казахском языке, однако опора многих систем преимущественно на англоязычные данные создает трудности при точной передаче национально маркированного содержания, прежде всего казахского этнокультурного кода. В данной работе выполнен сравнительный лингвистический анализ ответов QazLLM и ChatGPT с точки зрения их способности интерпретировать культурно-исторический, правовой и идиоматический контекст казахского языка.

Актуальность исследования обусловлена растущим запросом в государственных сервисах, образовании и медиасреде на культурно корректную и прагматически устойчивую казахоязычную генерацию. Эмпирическую основу составили 140 000 экспертно размеченных примеров; результаты дополнительно сопоставлены с решениями OpenAI и разработками Назарбаев Университета (ISSAI).

В ходе исследования ответы моделей оценивались по показателям распознавания национально-культурных реалий, мотивированного объяснения фразеологизмов и устойчивых сочетаний, точной передачи оценочной семантики, а также соблюдения норм диалогового взаимодействия. Дополнительно была построена типология ошибок, включающая культурные обобщения, подмену понятий и поверхностную интерпретацию контекста.

Анализ показал, что QazLLM чаще согласует национально специфичные понятия, фразеологию и оценочные оттенки с контекстом, тогда как ChatGPT демонстрирует более стабильные результаты в универсальных информационных запросах. Научная значимость работы связана с уточнением лингвистических критериев измерения культурной релевантности и с обоснованием роли локализованных корпусов. Практическая ценность результатов заключается в возможности их применения при совершенствовании казахоязычных чатботов, учебно-методических материалов, переводческих ассистентов и культурно-справочных сервисов.

**Ключевые слова:** большая языковая модель, искусственный интеллект, казахский язык, культурный контекст, трансформер, корпус данных, локализация.

## Кіріспе

Соңғы жылдары үлкен тілдік моделдер (ҮТМ) жасанды зерденің тек «мәтін құрастыру» құралынан шығып, мағына өндірудің күрделі жүйесіне айналғанын көрсетті. Трансформерге негізделген үлгілердің қалыптасуы және контексті өзара «байланыстырып» есептейтін назар тетігінің (self-attention) енгізілуі мәтінді өңдеуді жаңа сапалық деңгейге көтерді. Нәтижесінде модель тек сөйлем ішіндегі байланыстарды емес, тұтас контекстің ішкі логикасын, стильдік реңкін және прагматикалық мүмкіндіктерін қатар ескеруге қабілетті болды. Бұл өзгеріс «модель сөзді жалғайды» деген түсініктің орнына «модель контексті есепке алып, мағыналық шешім қабылдайды» деген жаңа пайымға жол ашты. Алайда мұндай технологиялық ілгерілеу

ҮТМ-дің барлық тілдерде бірдей дәрежеде мәдени дәлдікпен жұмыс істейтінін білдірмейді. Көпшілік жаһандық үлгілердің оқыту деректері негізінен ағылшынтілді кеңістіктен жиналғандықтан, мәдени-тарихи ерекшелігі терең, идиомалық қоры бай, коммуникативтік этикасы нәзік тілдерде жауаптың «дұрыс сияқты көрінгенімен» ұлттық контексті толық игермей шығуы жиі байқалады.

Қазақ тілі осындай мәдени контекстке айрықша тәуелді тілдердің қатарына жатады. Мұнда ойды жеткізу тек синтаксистік құрылыммен шектелмейді: сөздің астарлы мағынасы, үлкен-кішіге қатынас нормасы, ишаралы тіркес, мақал-мәтелдің уәжі, тарихи жадқа байланған атаулар мен ұғымдар мәтіннің мазмұнын айқындайды. Сондықтан қазақша мәтіннің табиғи, жатық шығуы үшін моделдің сөздік мағынадан

бөлек, этномәдени кодты тануы қажет. Этномәдени код дегенде біз белгілі бір халықтың дүниені қабылдау тәсіліне, құндылық жүйесіне, салт-дәстүріне, әлеуметтік әдепке және тарихи тәжірибеге орныққан тілдік-мәдени белгілердің жиынтығын түсінеміз. Бұл код көбіне ашық айтылмай, контекст арқылы танылады. Бір сөздің мағынасы мен әсері тыңдаушының буынына, әлеуметтік ортасына, сөйлеу жағдаятына қарай құбылып отырады. Қалжың мен түспалдың межесі, келісім білдіру мен қарсылықты сыпайы түрде жеткізу үлгілері, сондай-ақ «жөн», «ұят», «киелі», «ырыс», «күт» тәрізді концептілерге тірелетін бағалау реңктері жауаптың табиғилығын айқындайды.

Осы тұрғыдан алғанда, ұлттық дерекке сүйеніп үйретілген қазақша моделдердің мәдени контексті дәл аңғару қабілеті мен мәнмәтінге үйлесуі ерекше мәнге ие. Мәселе тек «қазақша жауап беру» деңгейінде емес, жауаптың ішкі логикасында, дәлелдеу үлгісінде, сөйлеу әдебіне сай деңгейінде және тұрақты тіркестерді орнымен қолдануында жатыр. Кей жағдайда әмбебап модель грамматикалық тұрғыдан дұрыс, бірақ мәдени тұрғыдан жадағай сөйлемдер құрауы мүмкін; кейде ұлттық реалияны өзге мәдени кеңістікке тән ұғыммен алмастырып, мағыналық сәйкессіздік тудырады; енді бірде нақты тарихи-мәдени деталді үстірт жалпылап, жауаптың сенімділігін төмендетеді. Демек, YTM-ді бағалау тек дәлдік пен толықтық өлшемімен шектелмей, мәдени релеванттылық пен прагматикалық сәйкестік өлшемдерін де қажет етеді.

Осы мақала QazLLM vs ChatGPT: қазақ этномәдени кодын жеткізудегі лингвистикалық ұқсастықтар мен басымдықтар мәселесін салыстырмалы талдау арқылы қарастырады. Зерттеудің негізгі мақсаты – екі моделдің қазақ тіліндегі мәдени-тарихи, құқықтық және идиомалық контексті интерпретациялау қабілетін лингвистикалық көрсеткіштер арқылы сипаттау және этномәдени мазмұнды жеткізудегі артықшылықтарын айқындау. Талдау барысында ұлттық-мәдени реалияларды тану, фразеологизм мен тұрақты тіркестерді түсіндіру, бағалаушылық реңк пен жағдаятқа сай сөйлеу деңгейін сақтау, сондай-ақ диалогтік нормаларға сай жауап құрастыру секілді параметрлерге басымдық беріледі. Эмпирикалық негіз ретінде сарапшылар белгілеген көлемді дерек қолданылып, нәтижелердің тілдік және мәдени уәжі салыстырмалы түрде зерделенеді.

## Материалдар мен әдістер

Зерттеудің материалы ретінде қазақ тіліндегі тапсырмалар корпусына сүйенген деректер қолданылды. Корпус 13 негізгі тапсырма санаты бойынша құрастырылып, жалпы көлемі 140 000 бірегей мәтіндік базаны қамтиды. Бұл жинақта фактілік және ашық сұрақтар, гипотетикалық жағдаяттар, синоним мен парафраз жасау, мәдени контекст пен идиомаларды пайдалану, жауаптың тілдік дәлдігі, қысқа және кеңейтілген жауап түрлері, екіұшты сұраулар, күрделі жағдайлар мен қателерді өңдеу, сценарийлік диалог құру, диалект пен көнерген қолданыстар, семантикалық рөлдерді айқындау сияқты түрлер теңгерімді енгізілді. Аталған санаттар этномәдени кодтың тілдік көрінісін тек мазмұн деңгейінде емес, стильдік, прагматикалық және идиомалық қабаттарда да қамту үшін алынды.

Корпусты рәсімдеу мен құрылымдау бірізді стандартпен жүргізілді: әр үлгі «сұрау/нұсқау – күтілетін жауап – тақырыптық домен» қағидасымен белгіленіп, домендік және жанрлық әркелкілік сақталды.

### 1-кесте

*Сұрақ-жауап датасетінің үлгісі*

<SAMPLE>
<PROMPT> Осы мәтінге қысқаша шолу жасаңыз.
<RESPONSE> Мәтінде климаттың өзгеруінің ауыл шаруашылығына әсері және оны шешудің мүмкін жолдары туралы баяндалады...
<DOMAIN> Environmental Sciences and Ecology

*Ескерту: Кесте авторлар тарапынан құрастырылған*

Бұл тәсіл қазақ тілінің лингвомәдени ерекшеліктерін тек ресми не көркем стильмен шектемей, әртүрлі коммуникативтік жағдаяттарда қамтуға мүмкіндік берді. Сонымен қатар домендер мен жанрлар алдын ала белгіленуі кейінгі салыстырмалы талдауда этномәдени мазмұнның қай қабатта (тарихи, құқықтық, мәдени-идиомалық) көбірек бұрмаланатынын немесе дәл берілетінін нақтылауға жағдай жасады.

Әдістер кешені екі арнаны біріктірді: автоматтандырылған іріктеу және сарапшылық тексеріс. Алдымен Python негізіндегі сүзгілер мен статистикалық бақылау арқылы қате, орынсыз және мәдени тұрғыдан даулы құрылымдар та-

зартылды, кейін кәсіби тілші мамандар қолмен сараптап, әсіресе идиоматикаға, жергілікті реалияларға, диалектілік/көнерген қолданыстарға қатысты мысалдардың дәлдігі мен табиғилығын тексерді.

Салыстырмалы талдау үшін QazLLM және ChatGPT-ге бірдей типтегі күрделі қазақша сұраулар беріліп, нәтижелер мәдени-тарихи, құқықтық және идиомалық мазмұнды интерпретациялау қабілеті бойынша бағаланды. Бағалау лингвистикалық көрсеткіштерге сүйенді: ұлттық-мәдени реалияны тану және дәл атау, фразеологизмді уәжді түсіндіру, бағалауыштық реңк пен сөйлеу әдеп нормаларын, мәнмәтін тұтастығын сақтау. Нәтижелер қосымша түрде ChatGPT (OpenAI) жүйелері және Назарбаев Университеті жанындағы ISSAI әзірлемелерімен салыстырылып, этномәдени кодты жеткізудегі ұқсастықтар мен басымдықтар жүйелі түрде қорытындыланды.

### Әдебиеттерге шолу

Мемлекет басшысы Қ.Тоқаевтың жасанды интеллектіні елдің түбегейлі цифрлық жаңғыруының өзегіне айналдыру жөніндегі ұстанымы тіл технологияларын ұлттық деңгейде жүйелі дамытуға тікелей бағдар береді. Бұл бағыт «цифрлық мемлекет» идеясын іске асырумен бірге, қазақ тіліндегі дерек, корпус және модель сапасын көтеруді күн тәртібіне шығарады (Тоқаев, 2025, б. 4). 2026 жылғы 5 қаңтарда «Түркістан» газетіне берген сұхбатында да Президент «Қазақстан цифрлық державаға айналуы керек» деген тұжырымды нақты атап, жаңа технология дәуірінде бәсекеге қабілетті болудың шартын цифрлық экожүйе мен адами капиталдың сапасымен байланыстырды (Тоқаев, 2026а, б. 3).

Осы стратегиялық контекст QazLLM тәрізді ұлттық үлкен тілдік моделдерді тек инженерлік өнім ретінде ғана емес, «тіл-таным-мәдениет» бірлігін сақтайтын инфрақұрылым ретінде қарастыруды талап етеді. Қазақ этномәдени коды көбіне эксплицитті анықтамамен емес, сөйлеу жағдаяты, әдеп нормасы, бағалауыш реңк, ишаралы мағына арқылы танылады; демек, моделдің «дұрыс жауап» беруі ақпараттық дәлдікпен шектелмей, мәдени релеванттылықпен өлшенуі тиіс. Бұл ұстаным қазақ тіл біліміндегі антропоэзеттік және этнолингвистикалық бағыттардың теориялық негіздерімен сабақтас: тілдік бірліктің артында тұрған дүниетанымдық мазмұнды

ашу – ұлттық мазмұнды генерациялаудың ең күрделі бөлігі.

Қазақ этномәдени мазмұнын жүйелеуде Ә.Қайдардың этнолингвистикалық мектебі айрықша орын алады: этностың табиғатқа, кеңістікке, салтқа қатысты танымы тілдік дерек арқылы сипатталып, мәдени семантиканың «ішкі құрылымы» ашылады (Қайдар, 2013, б. 121). Этномәдени атаулардың танымдық негіздерін когнитивтік арнада талдаған ғалым Ж.Манкеева еңбегі де тілдік таңба мен мәдени білімнің байланысын нақты материалмен дәлелдейді (Манкеева, 2008, б. 91). Бұл зерттеулер QazLLM-нің тілтанымдық базасын құрастыруда «мәдени кодты» өлшем ретінде бекітудің ғылыми алғышартын қалыптастырады.

Отандық үлкен тілдік моделдің (QazLLM) тілтанымдық базасын қалыптастыруда Н.Уәлидің еңбектері айрықша ғылыми тірек бола алады. Ғалымның қамшыға қатысты зерттеу желісі қамшыны тек тұрмыстық зат ретінде емес, тілдік таңба мен мәдени таңбаның тоғысқан нүктесі ретінде қарастырып, оның тарихи қабаттағы әрі бүгінгі қолданыстағы семантикалық салмағын жүйелі ашады. Мұнда «қамшы» атауының артында тұрған әлеуметтік статус, билік рәмізі, тыйым мен жөн-жосық, әдеп пен қатынас мәдениеті сияқты көпқабатты мазмұн нақты дерек, мәтіндік айғақ, мысалдар арқылы негізделеді. Автор «алғаш қамшы кімге тиді?» деген мазмұнды сұрақты мифологиялық-танымдық арнаға шығарып, Адам ата мен Хауана туралы сюжет арқылы қамшының архетиптік мәнін талдайды. Бұл, бір жағынан, мәдени жадтағы бейнелердің тіл арқылы сақталу тетігін көрсетеді, екінші жағынан, этномәдени «жабық кодтың» қалай уәжделетінін айқындайды (Уәли, 2025, б. 148).

Н. Уәлидің еңбектеріндегі тағы бір маңызды арна – жұмбақ сандар (үш, төрт, алты, жеті, тоғыз, жүз), «жұмбақ жетілер» жүйесі және халықтың астрономиялық білімдерімен сабақтас мәдени семантиканы талдау. Ғалым бұл құбылыстарды жай нумеративтік бірлік деңгейінде қалдырмай, ұлттық мифологиялық танымдағы символика, дүниені жіктеу үлгілері, сакралды-әлеуметтік бағдарлармен байланыстыра отырып түсіндіреді. Осындай талдаулар этномәдени ақпараттың тілде қалай кодталатынын, қай жерде астарлап берілетінін және қандай контексте «ашылатын» тетіктерін нақтылауға мүмкіндік береді. Сондықтан Н. Уәли ұсынған тілтанымдық пайымдар мен деректік материалдар QazLLM-нің мәдени реңкті дұрыс ұғынуын күшейтетін, қазақ дүние-

танымына тән мағыналық қабаттарды дәл беретін білімдік негіз ретінде бағалануға тиіс. Бұл ерекшелік OpenAI жүйелеріндегі әмбебап үлгілерден ұлттық моделдің айырмасын айқындауға да қызмет етеді (Уәли, 2025а, б. 150).

Этномәдени кодты моделге жүйелі түрде енгізудің ең сенімді көзі – дәстүрлі ұғымдар мен категориялардың энциклопедиялық әрі сөздік сипаттамаларына сүйенген дереккөз. Мұндай ұжымдық еңбектерде бүкіл қазақ қауымына ортақ салт-дәстүр, наным-сенім, әдет-ғұрып, тұрмыс-тіршілікке қатысты атаулар бір ғана анықтамамен шектелмей, олардың астарындағы бүгінгі қоғамға бірден таныла бермейтін білім қабаты да қоса беріледі. Яғни «жабық код» нақты мысал, қолданым аясы, мәдени уәж арқылы ашылады (Әлімбаев, 2017, б. 196).

Осы сипаттағы энциклопедиялық материалдар отандық ChatGPT-дің тілтанымдық базасын толықтыруда айрықша қызмет атқарады. Өйткені модель үшін тек сөздің мағынасы емес, сол мағынаның қай жағдайда, қандай әлеуметтік ортада, қандай құндылық өлшемімен байланыса айтылатыны маңызды. Мұндай дерек мәдени контексті дәл ұғынуға, дәстүрлі категорияларды орынсыз жалпылаудан сақтауға және ұлттық мазмұнды орнықты түрде құрастыруға мүмкіндік береді.

Сонымен бірге өңірлік өмір салтынан жиналған, қазіргі қолданыста көмескіленіп кеткен этнографиялық мәліметтер де моделдің білім кеңістігін тереңдетеді: жергілікті реалиялар, рәсімдік қолданыстар, әдеп нормаларының нақтылы сипаттамасы мәтіннің прагматикалық дәлдігін күшейтеді (Қартаева, 2015, б. 135). Ал жылқыға қатысты мәдени білімді терең жүйелейтін зерттеулер қазақ дүниетанымындағы ұғымдық желілерді тілдік тұрғыдан дұрыс пайымдауға жағдай жасап, этномәдени мазмұнды генерациялаудағы қателікті азайтады (Тоқтабай, 2010, б. 103).

Халықаралық ғылыми кеңістікте ChatGPT тәрізді үлкен тілдік моделдердің әлеуеті ең алдымен зерттеу үдерісін жеделдету, мәтін өндірудің өнімділігін арттыру және ғылыми коммуникацияны оңтайландыру қырынан сипатталады. Мәселен, зерттеу мен жарияланым дайындауда моделді кезең-кезеңімен қолдану жолдары ұсынылып, сонымен бірге дереккөздің анықтығын тексеру, дәйексөзді дұрыс рәсімдеу, мәтіннің түпнұсқалылығын сақтау сияқты тәуекелдер нақты көрсетіледі (Biswas, 2023, р. 579). Осы бағыттағы еңбектер ChatGPT-ді «универсал көмекші» ретінде бағалай отырып, оның жауабы

ғылыми стандарттарға сай болуы үшін міндетті түрде адам тарапынан редакциялау мен верификация қажет екенін алға тартады.

Сонымен қатар, ғылыми жарияланым этикасы мен авторлық жауапкершілік мәселесінде редакциялық ұстанымдар күшейтілді. Генеративті моделді автор ретінде көрсетуге болмайтыны, мәтіннің мазмұнына жауапкершілік толықтай адамға тиесілі екені ашық айтылады (Thorp, 2023, р. 313). Журналдың саясаты деңгейінде де талаптар нақтыланып, қолжазба мен рецензия мәтіндерінің құпиялылығы, құралды қолдануды міндетті декларациялау, жауапты пайдалану қағидалары бекітіле бастады (Boruff, Kraft, 2023, р. 748). Бұл үрдіс генеративті жүйелердің ғылыми ортаға енуі «жылдамдық» пен «сапа – жауапкершілік» арасындағы теңгерімді қайта қарауды талап ететінін көрсетеді.

Ғылыми жарияланымның болашағы туралы талқылауларда генеративті моделдер мәтінді сауатты құрастыруға қабілетті болғанымен, дәлелдің тереңдігі, пәндік-әдіснамалық жауапкершілік және контекстік нюанстарды орнықты ұстау мәселелері шектеу ретінде аталады (Hill-Yardin, 2023, р. 152). Демек, LLM жауабын бағалау тек грамматикалық дұрыс құрылыммен шектелмеуі керек; оған қоса жауаптың контекстке сай келуі, келтірілген уәждің дерекке сүйеніп негізделуі және сөйлеу жағдаятына сәйкес берілуі сияқты өлшемдер ескерілуі тиіс.

Дәл осы жерде мәдени-тілдік релеванттылық мәселесі алдыңғы қатарға шығады: әмбебап моделдің жауабы грамматикалық жағынан мінсіз болғанымен, белгілі бір қоғамның құндылықтар жүйесіне, әдеп-нормаға, ишара мен астарға «үйлеспей» қалуы мүмкін. Кей зерттеулерде үлкен тілдік моделдердің жауаптары белгілі бір мәдени ортаға көбірек бейім болып шығатыны байқалады. Мысалы, кей жағдайда жауап мазмұны Батыс елдеріне тән құндылық өлшемдеріне жақындап, өзге мәдениеттердің ерекшелігін толық ескермей қалады. Сұрауды арнайы мәдени контекстпен нақтылап бергенде жауапты біршама түзетуге болады, бірақ бұл тәсіл мәдени алшақтықты түбегейлі жоймайды (Тао, 2024, р. 7). Ал LLM тәуекелдерін талдаған іргелі еңбектер моделді тек ауқымын ұлғайту арқылы жетілдіру жеткіліксіз екенін, мәдени және әлеуметтік факторлар ескерілмесе жауаптың мағыналық нақтылығы мен мәдени сәйкестігі төмендеуі мүмкін екенін ескертеді. (Bender, 2021, р. 615).

Сондықтан ChatGPT-дің жалпы ақпараттық міндеттердегі қуаты мойындалғанымен, қазақ

этномәдени кодының астарлы қабаттарын толық ашуда оның мүмкіндігі жиі шектеліп қалады. Мұндай қабаттарға дәстүрлі ұғымдар жүйесі, ишаралы мағына, ұлттық әдепке сай бағалау реңкі, идиомалар мен тарихи жадқа байланған түсініктер жатады. Осы себепті қазақ тілінің өз деректеріне сүйенген, ұлттық мазмұнды ескеріп құрылған тілдік база және соған бейімделген оқыту тәсілдері қажет екені халықаралық әдебиеттегі тұжырымдармен де үндеседі.

### Нәтижелер мен талқылау

Зерттеу барысында QazLLM, ChatGPT және ISSAI (HY) жүйелерінің қазақтың этномәдени кодын берудегі әлеуеті мен лингвистикалық ұқсастықтары талданады. Үш жүйенің жауаптарын салыстыру үшін этномәдени коды бар сауалдар берілді. Алғашқы сауал: *«күшік күйеу» тіркесінің мәдени мәнін ашыңыз.*

«Күшік күйеу» тіркесінің негізгі мағынасы қалай ашылады, оның шығу төркіні қалай түсіндіріледі, салт-дәстүрдегі орны мен бүгінгі қолданыстағы реңкі қаншалықты дәл беріледі. Осы сұрақтарға жауап беру барысында үш жүйенің мәтін құрау машығы ғана емес, этномәдени білімді тану және оны тілдік дәлдікпен жеткізу қабілеті айқын көрінді.

Алдымен негізгі мағына деңгейінде айырма анық байқалады. ChatGPT «күшік күйеу» дегенді қыздың төркіні жағында тұрып, сол үйдің шаруасына араласып, белгілі бір уақыт «қолда жүрген» күйеу ретінде түсіндіреді. QazLLM де осы мазмұнға жақын түсінікті береді, бірақ оны қалыңмал өтеу тәжірибесімен байланыстырып, нақты дерекке сүйеніп ашады. Ал ISSAI жауабы тіркесті «сүйікті, мықты күйеу» деп түсіндіріп, оны құрмет не еркелету реңкімен береді. Бұл түсіндіру халықтық қолданысқа да, этнографиялық дерекке де сәйкес келмейді. Осы жерде этномәдени коды бар тіркесті дұрыс тануда моделдердің мүмкіндігі әрқелкі екені көрінеді.

ChatGPT (OpenAI) жауабының күшті жағы – мазмұнды жүйелеп, тіркестің әлеуметтік рөлін және бүгінгі ауыспалы мағынасын қатар көрсетуі. Ол «күшік» сөзін кемсіту емес, тәуелділік, үлкен үйдің тәртібіне бағыну, үйреніс үстіндегі күй сияқты мағыналық ишара ретінде түсіндіреді. Сондай-ақ әзіл не сынап айту реңкі бар екенін атап өтеді. Бұған қоса мысал сөйлемдер ұсынып, тіркестің қолданысын көрсетеді. Дегенмен бұл жауапта бірқатар тұстар «жорамалға» сүйеніп берілген: «күшік» компонентін

дәл осылай уәждеу үшін нақты дерек, ескі мәтін немесе сөздік айғағы келтірілмейді; кей себептер (қыз жақта үл болмау т.б.) жалпы түсіндіру ретінде айтылғанымен, тіркестің шығу тегін нақты бекітпейді.

ISSAI жауабы этномәдени кодты ашуға келгенде ең әлсіз нәтиже көрсетті. «Күшік» сөзін «күш» түбірінен шыққан деп беруі және тіркесті «сүйікті, мықты күйеу» мағынасына бұруы тілдік дерекке үйлеспейді. Мұнда салт-дәстүр туралы түсінік жоқ, тіркестің қоғамдағы орны, әзіл-сықақ реңкі, тарихи қолданым қабаты айтылмайды. Ең маңыздысы – «күшік күйеу» тіркесінің мәдени мазмұны жоғалып, сөздің сыртқы тұлғасына сүйенген қате пайымға кетеді. Бұл нәтиже этномәдени мазмұны бар сұрақтарда дайын білім қоры мен дұрыс уәждеудің шешуші екенін көрсетеді.

QazLLM жауабының басымдығы – тіркесті жай сипаттап қана қоймай, оның шығу төркінін тарихи-этнографиялық дерекпен негіздеуі. Онда қалыңмалға шамасы келмеген жігіттің қыз әкесіне бірнеше жыл қызмет етіп, қалыңмалын «еңбекпен өтеуі» болғаны айтылады және осыған байланысты «күш күйеу» атауы қолданылғаны көрсетіледі. Бұған қоса қырғыз тіліндегі атаумен салыстырып, сөздіктегі деректі де келтіреді. «Күш күйеу» атауының кейін «күшік күйеу» түріне ауысуын дыбыстардың ілесе айтылуы және әзіл-оспаққа байланысты өзгеру мүмкіндігімен түсіндіреді. Мұндай баяндау этномәдени «жабық кодты» ашудың нақты тәсілін көрсетеді. Сөздің түбірі, тарихи жағдайы, түркілік салыстыруы және мәтіндік айғақ бір арнаға тоғысады.

Үш жауаптың лингвистикалық ұқсастығы да бар: OpenAI мен QazLLM тіркесті қоғамдағы «аралық мәртебені» білдіретін атау ретінде таниды және оның әзіл реңкі болуы мүмкін екенін жанама түрде мойындайды. Бірақ басымдық салыстырғанда айырма анық: OpenAI көбіне қазіргі түсіндіру жолын ұстанып, мағынаны сипаттаумен шектеледі; QazLLM мағынаны сипаттауды тарихи негіздеумен толықтырады. Ал ISSAI-де ортақ нүкте жоқ, өйткені тіркестің негізгі мағынасы басқа арнаға бұрылған.

Талқылау тұрғысынан ең маңызды мәселе – «күшік» компонентінің уәжі. OpenAI оны бейнелі теңеу ретінде түсіндіреді: «күшік» тәуелділік пен икемделгіштікті меңзейді дейді. QazLLM болса, бұл тіркес бастапқыда «күш күйеу» болғанын, кейін дыбыстық өзгеріс арқылы бүгінгі тұлғаға түскенін алға тартады. Этномәдени бірліктерде мұндай құбылыс жиі кездеседі:

сөздің сыртқы қалпы халықтық әзілмен, сөйлеу дағдысымен өзгеріп, бастапқы мән көмескіленеді. Осы тұрғыдан QazLLM ұсынған түсіндіру «неге дәл осылай аталған?» деген сұраққа дерекпен жақынырақ жауап береді және тіркестің «жабық» қабатын аша түседі.

Тағы бір өзекті түс – салттық тәжірибенің әлеуметтік себебі. OpenAI «қыз төркінінде тұру» жағдайын кең аяда сипаттайды: тұрмыстың дайын болмауы, белгілі әлеуметтік себептер сияқты жалпы факторларды атайды. QazLLM оны нақтырақ әлеуметтік-экономикалық себеппен байланыстырады: қалыңмалға шамасы жетпеу, қалыңмал жинау үшін жалдану, қыз әкесіне қызмет ету. Мұндай нақтылау этномәдени кодты дәл жеткізуге көмектеседі, өйткені тіркес сол тәжірибенің тілдегі атауы ретінде түсіндіріледі. Яғни сөздің мәні «тек қай жерде тұру» емес, «неге солай болған» деген білім жүйесімен ашылады.

Нәтижесінде этномәдени мазмұны бар тіркестерді түсіндіруде QazLLM-нің артықшылығы оның тілтанымдық базасының табиғатынан көрінеді. Ол тіркестің мағынасын тарихи дерекпен сабақтастырып, түркілік салыстыру арқылы бекітіп, өзгеру жолын түсіндіреді. ChatGPT болса мәтінді жинақы құрып, әлеуметтік мәні мен қазіргі қолданысын жақсы сипаттайды, бірақ уәждеу жағында деректік тірек аз. ISSAI үлгісі бұл мысалда қате түсіндіруге ұрынды. Демек, қазақ этномәдени кодын жеткізу міндетінде ұлттық дерекке сүйенген, қазақтың ішкі білім жүйесін қамтыған QazLLM бағыты басым екені осы салыстырудан анық көрінеді.

Сауал: *Екі кештің арасында» тұрақты тіркесінің мәдени мәнін ашыңыз. Неге екі кеш? Түсіндіріңіз.*

*OpenAI GPT-5 жауабы* тіркесті ымырт мезгілімен байланыстырып, «екіні» ақшам мен іңір тәрізді екі шекара арқылы түсіндіреді. Бұл жауаптың артықшылығы – түсіндіруі жинақы, оқырманға бірден ұғынықты және тіркестің уақыттық мағынасын нақтылап береді. Сонымен бірге ауыспалы қолданысты да (шапшаң, қапелімде) ескеріп, тіркестің дискурстағы қызметін көрсетеді. Алайда мұнда этномәдени кодтың терең қабаты толық ашылмайды: қазақтың дәстүрлі түн межелеуі «ақшам-іңір» ғана емес, ел тіршілігінің әлеуметтік ырғағымен өлшенетін «ел орынға отыру» және «ел жату» сияқты ұғымдарға да тіреледі. Яғни ChatGPT уақыттың сыртқы көрінісін дұрыс береді, бірақ «неге дәл екі?» деген сұрақтың мәдени негізін тұрмыс-әлеуметтік межемен байланыстырып бермейді.

*QazLLM жауабы: Қазақ халқының байырғы салт-дәстүрімен, таным-түсінігімен байланысты қолданылатын сан есімдердің бірі – екі. Екі кештің арасында деген фразеологизм түн ішінде беймезгіл жүрген адамға қарата айтылады. Екі кештің арасында ешкім жай жүрмейді. Тосын, күтпеген жердегі жүріс субъектінің анық-қанығына жеткенше, «не жағдай?», «әлдеқандай қиындық па?», «күтпеген қуанышты хабар ма?» деп әбігерленіп, әртүрлі имплицитті күйге түсіреді. Тосын жүрген субъект тосын жүрістің себебін айтқан соң ғана екі кештің арасындағы жүрістің мән-жайы экспликацияланады. Ал екі кеш деген сан есімді тіркес нені білдіреді? Бұндағы кеш қараңғы түскенге дейінгі уақыт емес, «түн», «түн баласы» деген мағынаны білдіреді. Туыстас түрік тілінде кеш сөзі «түн» мағынасын білдірген. Қазақы танымда түн немесе түн баласы ел орынға отырған шақ және ел жатқан шақ деп екі шаққа бөлінеді. Ел орынға отырған шақ ел-жұрттың мал-жананы жайғап болып, кешкі ас-суына отырған кезі, ал ел жатқан шақ жұрт аяғы басылып, түнгі ұйқыға жатқан кезі. Екі кеш арасындағы деген осы екі шақ. Ерекше жағдаймен емес, жай жүрген жолаушы ауыл-елге ел орынға отырғанша жетуге тырысады. Ондай жолаушыны «екіндіде елге сүйкімді қонақ келеді» деп мақалдайды.*

QazLLM жауабы этномәдени мазмұнды кеңірек ашады: ол тіркестің қолданысын «екі кештің арасында ешкім жай жүрмейді» деген прагматикалық пайымнан бастап, күтпеген жүріс-қимылдың ел ішінде әбігер, болжам тудыратын әлеуметтік әсерін көрсетеді. Бұл – тіркестің тек уақытты емес, «беймезгіл жүріс», «тосын әрекет» сияқты мәдени-бағалау реңкін де алып жүретінін білдіреді. Ең маңыздысы – «екі» санының уәжін қазақы түн межесімен байланыстырып береді: түн «ел орынға отырған шақ» және «ел жатқан шақ» болып екі кезеңге бөлінеді, ал «екі кештің арасында» сол екі меженің арасындағы уақытты меңзейді. Мұнда «кеш» сөзінің кей тілдерде «түн» мағынасымен қабаттасуы туралы ишара да беріледі, бұл тіркестің ішкі семантикасын кеңейтіп, мәдени кодтың тілдік ізін көрсетеді. QazLLM осылайша тіркесті тек «ымырт» ретінде тарылтпай, оны қазақтың әлеуметтік уақыт өлшемі арқылы түсіндіреді.

*ISSAI жауабы* бұл тіркесте этномәдени кодты жеткізу тұрғысынан әлсіз нәтиже көрсетті. Біріншіден, «екі кеш» дегенді «екі рет өткен кеш» немесе «екі күн бойы созылған кешкі уақыт» деп тура мағынада түсіндіруге ұмтылады. Бұл тір-

кестің тұрақты сөз орамы ретіндегі табиғатын көмескілейді және қазақы уақыт межелеудің мәдени логикасын ашпайды. Екіншіден, жауапта нақты дәстүрлі ұғымдар жоқ: «ел орынға отыру», «ел жату», беймезгіл жүрудің әлеуметтік бағалануы сияқты мәдени тіректерсіз түсіндірме жалпы сипатта қалып қояды. Үшіншіден, «екі көш» туралы қосымша жорамал енгізуі жауаптың фокусын шашыратып, ғылыми талдауға тән дәлдікті әлсіретеді.

Лингвистикалық ұқсастық тұрғысынан ChatGPT мен QazLLM-де ортақ белгі бар: екеуі де тіркестің негізгі өзегі уақыт межесіне қатысты екенін мойындайды және «екі» санының уәжін белгілі бір екі межемен байланыстырады. Бірақ басымдық QazLLM жағында айқынырақ көрінеді, өйткені ол уақыт межесін табиғи құбылыспен ғана емес, қазақ қоғамының түнгі тіршілік ырғағымен, әлеуметтік тәртіппен ұштастырады. Бұл – этномәдени кодтың өзегі: уақытты «сағатпен» емес, қауымның әдеттік режимімен өлшеу. ChatGPT мұны жалпы түсіндіруге жақындатқанымен, дәл осы әлеуметтік межені толық енгізбейді. Сондықтан оның жауабы әмбебап әрі ықшам, бірақ мәдени қабатты ашуы жеткіліксіз екені көрінеді.

Тіркестің прагматикалық қыры да салыстыруда маңызды. QazLLM «екі кештің арасында жүру» әрекетін тосындықпен, елдің аландауымен байланыстырып, тіркестің бағалаушытық реңкін (жай жүріс емес, себебі бар жүріс) нақты көрсетеді. ChatGPT бұған жақын мағынаны тек ауыспалы қолданыс арқылы (қапелімде) шолып өтеді, ал ISSAI мүлде жеткізбейді. Этномәдени бірліктің өміршеңдігі көбіне осындай прагматикалық реңкпен өлшенетінін ескерсек, бұл жерде QazLLM-нің мәдени мазмұнды нақты беруі оның басымдығын күшейтеді.

Осы салыстырудан байқалатыны: этномәдени мазмұны бар тұрақты тіркестерде дұрыс жауап беру үшін сөздің сөздіктегі мағынасын ғана емес, уақыт межелеудің дәстүрлі үлгісін, қоғамдағы әдептік норманы және сөз орамының қалыптасқан бағалау реңкін қатар ұстау қажет. QazLLM осы үш қабатты бір арнаға тоғыстырып береді; ChatGPT уақыттық өзекті дәл жеткізгенімен, әлеуметтік-мәдени уәжді толық тереңдетпейді; ISSAI тұрақты орамды тура мағынаға ығыстырып, мәдени логиканы аша алмай қалады.

Сауал: «*Жүрек жұтқан*» тұрақты тіркесінің мәдени мәнін ашыңыз.

Бұл сұрақта үш жүйенің жауабы бір тіркестің мағынасын түсіндірудің үш түрлі деңгейін

көрсетеді: бірі – қате тану, екіншісі – жалпы тілдік түсіндіру, үшіншісі – этномәдени білімнің ішкі қабатын ашу. «Жүрек жұтқан» тіркесі қазақ тілінде батырлық пен қайсарлыққа қатысты қалыптасқан бағалаушы мағына береді, ал оның мәдени түбірі кей жағдайда халықтық танымдағы арнайы түсініктермен байланысып жатады. Сондықтан бұл жерде салыстырудың өзегі – тіркестің тұрақты мағынасын дәл табу және соны ұлттық дүниетанымға сай уәждеу.

*ISSAI жауабы* ең алдымен тіркесті дұрыс танымай, оны «өте таңғалу, шошу» мағынасына бұрып жібереді. Қазақ тілінде қатты шошу, жүрек дүрсілдеу сияқты күйді білдіретін басқа тұрақты орамдар бар, ал «жүрек жұтқан» дәл осы мағынада орнықпаған. Мұнда жүйе сөздің сыртқы тұлғасына сүйеніп, сөйлем үлгісін қолдан жасап тұр: «жүректі жұтты» деген тіркес қазақ тілінің қалыпты фразеологиялық қоры ретінде дәлелденбейді. Соның салдарынан этномәдени мазмұн ашылмақ түгілі, негізгі мағына да бұзылады.

*ChatGPT жауабы* тіркестің тілдік қолданыстағы негізгі мағынасын дұрыс береді: «өте батыл, өжет, тәуекелі күшті адам» деген бағалау. Бұдан бөлек, «жүрек» сөзінің қазақ танымында ерік-жігер мен батыллықтың белгісі екендігін, «жүту» етістігі арқылы әсірелеу жасалатынын түсіндіреді. Бұл – ғылыми тұрғыдан ықшам, түсінікті, кең аудиторияға ыңғайлы талдау. Яғни модель тіркестің қазіргі семантикасын дұрыс беріп, оны жалпы метафоралық тәсілмен түсіндіреді.

*QazLLM жауабы* осы дұрыс мағынаны сақтай отырып, оны этномәдени контекстпен тереңдетеді. Мұнда «жүрек жұтқан» тіркесінің астарында әйелдің жерік болуына байланысты халықтық түсінік жатқанын көрсетеді: «жүрекке жерік болу», «жерікті басу үшін жүрек әкелу», содан туған баланың қайсар, батыл болып өсуін сол әрекетпен байланыстыру. Бұл түсіндіру тіркесті жай ғана «әсірелеу» деп шектемей, мәдени жадтағы түсінік жүйесіне кіріктіреді. Нәтижесінде тіркестің «жабық» қабаты ашылып, сөздің бағалаушы мағынасы халықтық уәжбен бекітіледі.

Екі дұрыс жауаптың (ChatGPT және QazLLM) арасындағы басты айырма – уәждеу деңгейі. ChatGPT тіркесті тілдік бейнелеу тәсілі ретінде сипаттайды: «жүрек» – батыллықтың белгісі, «жүту» – сол қасиетті бойға сіңіруді әсірелеу. QazLLM осы бейнелеудің қайдан шыққанын нақтырақ мәдени сценарий арқылы көр-

сетеді: «жерік болу – жоралғы атқару – қасиет қону» тәрізді дәстүрлі ойлау тізбегі беріледі. Мұндай тәсіл этномәдени кодты жеткізуде маңызды, өйткені қазақтағы көптеген тұрақты тіркес белгілі бір салттық тәжірибемен, наныммен, фольклорлық сюжетпен астасып жатады.

QazLLM мәтінінде фольклордан мысал келтіруі де осы артықшылықты күшейтеді: жыр-ертегілерде «жерік» мотивінің кездесуі арқылы «жүрек» ұғымының күш-қуат, ерлік қасиеттерімен байланысы бекітіледі. Бұл жерде модель тек анықтама бермей, мәдени білімнің қалай таралатынын – әңгіме, жыр, ертегі арқылы ұрпақ жадында сақталатынын аңғартады. Ал ISSAI мұндай қабатты мүлде байқамайды, өйткені ол тіркесті өзге эмоциялық күйге теліп, қазақы дүниетанымдағы уәждерді айналып өтеді.

Лингвистикалық ұқсастық ретінде екі жүйе де «жүрек» соматизмі арқылы бағалау жасалатынын ортақ: жүрек – адамның ішкі қуатының, қайратының таңбасы. Бірақ этномәдени кодты дәл беру тұрғысынан QazLLM басым түседі. Ол тіркестің тек қазіргі мағынасына тоқталмай, оның мәдени түбірін, салттық түсініктермен байланысын түсіндіреді. OpenAI-дың жауабы тілдік тұрғыдан дұрыс әрі жинақы болғанымен, мәдени негізді «жалпы метафора» деңгейінде қалдырады.

ChatGPT жалпы тілдік түсіндіруде тұрақты әрі ұғынықты, ал QazLLM этномәдени мазмұнды тереңірек, ұлттық танымға жақын етіп береді. ISSAI-дың бұл кейсте жаңылысуы мәдени кодты тануда нақты ұлттық дерек пен ішкі мәдени сценарийлердің маңызын көрсетеді. Яғни «жүрек жұтқан» сияқты бір ғана тіркестің өзі моделдің этномәдени білім қорын, тұрақты тіркестерді дұрыс тану қабілетін және түсіндірудің қандай деңгейде жүргізілетінін айқын ажыратады.

Сауал: *«Өлі күйеу» тіркесінің мағынасын ашыңыз.*

ISSAI жауабы бұл тіркесті нақты этномәдени бірлік ретінде тани алмай, жалпы жорамалға сүйенеді. Бірде «өте шаршаған» деген мағына ұсынып, бірде «рухани қайта оралу» сияқты қисынсыз бағытқа кетеді. Мұндай түсіндірулерде қазақтың неке, құда түсу, қалыңмалға қатысты дәстүрлі ұғымдар мүлде жоқ, сондықтан тіркестің өмірлік контексі ашылмай қалады. Ең басты қате – «өлі» сөзін кез келген мәдениетке ортақ метафора ретінде алып, қазақ тіліндегі нақты дәстүрлік мәнін айналып өту. Нәтижесінде жауап тұтасып тұрмайды, тіркестің мағынасын дәл бермей, тек ықтимал нұсқаларды тізбектеп қояды.

ChatGPT жауабы тіркестің әлеуметтік мағынасын әлдеқайда нақты береді: «аты бар, заты жоқ», формалды күйеулік, әйел / қыз тағдыры байлаулы, бірақ күйеуден тірек жоқ деген ахуалды сипаттайды. Мұнда «өлі» компоненті «әрекетсіз, іске аспай тұрған» деген ауыспалы реңкте тұрғанын дұрыс түсіндіреді. Бұл – қазіргі түсіндірме үшін өте ұтымды, өйткені оқырманға бірден түсінікті және тіркестің дискурстық қызметін көрсетеді. Сөз орамы адамның әлеуметтік жағдайын, оның әділетсіздігін немесе қиындық сипатын ықшам түрде білдіреді. Бірақ жауаптың әлсіз тұсы – дәстүрлі құқықтық-ритуалдық негізді жалпылап береді: «келісім бар, міндет орындалмаған» деңгейінде қалады да, «өлі» аталуына себеп болған нақты қазақы рәсім атауларын, ішкі ережені ашпайды.

QazLLM жауабы: *Қазақ тілінде «өлі күйеу» деген тұрақты тіркес қолданылады. Оның мәні – дәстүрлі ортада құдалық бекіп, қалыңмалдың қомақты бөлігі берілгенімен, «өлтірісі» тапсырылмаған кезеңде қыз ауылына барған күйеу жігіттің жағдайын атау. Бұл атаудың астарында мәдени ұғым жатыр: қалыңдық ауылына келген күйеудің еркіндігі шектеулі болған, ол қайын жұрттағы үлкендердің көзіне түспеуге тырысқан әрі жеңгелердің жөн-жосығына бағынып, айтқанын орындаған. Өлтірісі берілгеннен кейін ғана дұға оқылып, неке қиылып, күйеу жігіттің мәртебесі өзгерген, осы сәттен бастап «өлі күйеу тірілгендей» саналған. Халық арасында «Өлтірісін берген соң, өлі күйеу жатар ма» деген сөздің айтылуы да осы түсінікті бекітеді. Сондай-ақ өлтіріге сойылған малдың етінен күйеу мен қыздың ауыз тимеуі – рәсімге байланысты сақталатын шарттардың бірі (Хабиева, 2025, б. 317).*

QazLLM жауабы тіркестің этномәдени кодын нақты дәстүрмен байланыстырады. Мұнда «өлі күйеу» деп баталасып, қалыңмалдың бір бөлігі берілгенімен, «өлтірісі» берілмеген кезеңдегі күйеу жігіт айтылып отыр. Яғни тіркес біреуді жамандау үшін ғана емес, дәстүрдегі мәртебелік жағдайды атау үшін де қолданылғаны көрсетіледі. «Өлтірісін берген соң өлі күйеу тіріледі» деген мәтел арқылы «өлі / тірі» оппозициясы тек ауыспалы бейне емес, рәсімге байланған түсінік екені ашылады. QazLLM-нің мұндағы басымдығы – терминнің мәнін салттық-құқықтық сценарий арқылы түсіндіруі және тіркестің пайда болу логикасын қазақы дүниетанымға сай бекітуі.

Үш жауапты салыстырғанда, ChatGPT мен QazLLM-нің ортақ тұсы – «өлі» компонен-

тін тура мағынада емес, мәртебе мен әрекеттің жоқтығын білдіретін таңба ретінде түсіндіруі. Алайда бұл ұқсастық әртүрлі деңгейде көрінеді: ChatGPT «өлі» сөзін жалпылама «іске аспай тұрған» мағынасында алады, QazLLM болса «іске аспау» неден екенін нақты рәсіммен дәлелдейді. Этномәдени кодты жеткізудегі басымдық дәл осындай жерде білінеді – ұлттық модель тіркестің ішкі мәдени қисынын көрсете алады.

Тіркестің мәнін жеткізуде тағы бір маңызды қыр бар: «өлі күйеу» бір ғана «күйеуі жоқ» деген сөз емес, салттық-әлеуметтік қатынастағы өтпелі мәртебені білдіреді. QazLLM мұны күйеу жігіттің құқығы шектеулі болғанын, үлкендердің көзіне түспеуге тырысатынын, белгілі тәртіпке бағынатынын айту арқылы мәдени кеңістікке орналастырады. ChatGPT да әйел / қыздың «байлаулы» күйін көрсетеді, бірақ күйеудің дәстүрлі мәртебесін, оның «неге толық күйеу саналмай тұрғанын» мәдени тұрғыдан тарқатпайды. ISSAI-де бұл қабаттың бәрі көмескі.

Жүргізілген салыстырмалы талдау этномәдени мазмұны бар тұрақты тіркестерді түсіндіруде жүйелердің айырмасы ең алдымен «мағынаны дәл тану» және «мәдени уәжді дұрыс ашу» деңгейінен көрінетінін көрсетті. ChatGPT көп жағдайда тіркестің қазіргі қолданыстағы негізгі мағынасын дұрыс беріп, оны жалпы тілдік заңдылықтармен ұғынықты түрде түсіндіруге бейім екені анықталды. Алайда этнографиялық деректер, салттық мәртебе, әлеуметтік уақыт межесі сияқты қазақ дүниетанымына тән ішкі білім жүйесі қажет болатын тұстарда жауап көбіне жалпылама сипатта қалып, тіркестің шығу төркінін дәлелдейтін ұлттық дерекқор толық көрінбей қалды. ISSAI үлгісі кейбір мысалдарда тұрақты тіркестің мағынасын қате бағытқа бұрып, сөздің сыртқы тұлғасына сүйенген жорамал ұсынды. Бұл этномәдени кодты тануда нақты тілдік база мен дәстүрлі білім қабатының маңызы жоғары екенін айқындады.

QazLLM жауаптары талданған үш кейсте де этномәдени кодтың «жабық қабатын» ашуға жақынырақ екені байқалды: «екі кештің арасында» тіркесін қазақы түн межелеу үлгісімен байланыстыруы, «жүрек жұтқан» орамын халықтық дүниетанымдағы жерік туралы түсінікпен уәждеуі, «өлі күйеу» тіркесін салттағы мәртебелік кезең мен рәсімдік атаулар арқылы негіздеуі осыны дәлелдейді. Осы нәтижелер қазақ тіліндегі этномәдени бірліктерді сапалы генерациялау үшін моделдің тек лингвистикалық дұрыстығы емес, ұлттық дерекке сүйенген білім қоры,

дәстүрлі ұғымдар жүйесі және мәдени сценарийлермен жұмыс істей алу қабілеті шешуші екенін көрсетті. Демек, QazLLM-ді дамытуда фольклор, этнография, этнолингвистика, тарихи мәтіндер және түсіндірме-сөздік деректерін жүйелі кіріктіру ұлттық мазмұнды дәл әрі табиғи жеткізудің негізгі шарты ретінде ғылыми тұрғыдан негізделді.

### Қорытынды

Осы мақалада QazLLM, ChatGPT және ISSAI жауаптары қазақ этномәдени кодын жеткізу тұрғысынан салыстырмалы лингвистикалық талдаудан өтті. Нәтижелер ұлттық мазмұны бар тұрақты тіркестерді түсіндіруде айырма, ең алдымен, мағынаны дәл тану және оны мәдени уәжбен негіздеу деңгейінен байқалатынын көрсетті. ChatGPT көп жағдайда тіркестің қазіргі қолданыстағы жалпы мағынасын түсінікті әрі жүйелі жеткізіп, мәтінді ықшам құру, ақпаратты реттеу және әмбебап сұрақтарда тұрақты жауап беру қабілетімен ерекшеленеді. Алайда «күшік күйеу», «екі кештің арасында», «жүрек жұтқан», «өлі күйеу» сияқты этномәдени мазмұны қалың бірліктерде негізгі мағынаны дұрыс берудің өзі жеткіліксіз болып, мәдени семантиканы, дәстүрлі уақыт межесін, астарлы бағалау реңкін дәл ашу шешуші мәнге ие екені айқындалды.

Талдаулар QazLLM-нің басымдығы дәл осы «жабық кодты» ашуда көрінетінін негіздеді. Ол тіркесті тек қазіргі мағынамен шектемей, тарихи-этнографиялық деректерді, салт-дәстүрге тән түсініктерді және фольклорлық танымды тірек етіп, мағынаның мәдени уәжін нақтырақ көрсетеді. Осы тұрғыдан алғанда, QazLLM үшін арнайы жасалған тілтанымдық база этномәдени ақпаратты «жай сипаттап беру» деңгейінен «дұрыс түсіндіріп, орнықты уәждеу» деңгейіне көтеретін тірек ретінде көрінді.

Мақаланың тұжырымы Мемлекет басшысы Қасым-Жомарт Тоқаевтың цифрлық технологияларды, креативті индустрияны, жасанды интеллектіні дамыту және QazLLM әзірлеу жөніндегі стратегиялық бағытымен тікелей сабақтас. Президенттің «Ана тілі» газетіне берген сұхбатындағы «Цифрлық технологияларды қолдануға баса мән берген жөн, өйткені қазақ тілінің алдағы тағдыры, еліміздің жаһандық бәсекеге қабілеті көбіне осыған байланысты болады» (Тоқаев, 2025b, б. 1) деген ұстанымы тілдік технологияны ұлттық деңгейде дамытуға негіз болатын бағдар ретінде осы зерттеудің өзектілігін күшейтеді.

А. Байтұрсынұлы атындағы Тіл білімі институты 30-дан астам AI-өзірлеме, зерттеу мен цифрлық ресурс ұсынып, қазіргі кезде әл-Фараби атындағы ҚазҰУ, Тіл-Қазына орталығы және Ақпараттық есептеуіш технологиялар институтымен бірлескен консорциум аясында QazLLM өзірлеу жұмысына қатысып отыр. Институт ғалымдары осы жобаның маңызды құрамдасы ретінде QazLLM-нің тілтанымдық базасын өзірледі, ал ұлттық колоритті ақпарат бойынша моделді оқытуға арналған промттар жүйесін жасау міндеті де Институт мамандарына жүктелді. Бір жыл ішінде 140 мыңнан аса ұлттық үлгідегі сұрақ-жауап датасетінің дайындалуы осы бағыттағы нақты нәтижені көрсетеді.

QazLLM қарапайым тілмен айтқанда, қазақша ойлап, қазақша сөйлейтін жасанды интеллект қазақ «чат джипитиінің» «миы». Оның қазақша нейрондарын қазақ тіліндегі идеотіл – фразеологизмдер, бейнелі қолданыстар, мақал-мәтелдер, этномаркерлі лексика құрайды. Бұл база қазақтың этномәдени ерекшеліктерін дәл беруге ғана емес, оны ЖИ жүйелеріне ғылыми негізде кіріктіруге, мәдени тұрғыдан орнықты әрі табиғи қазақша генерацияға қол жеткізуге мүмкіндік береді. Демек, QazLLM-нің артықшылығы әмбебап ақпаратты қайталауда емес, қазақтың мәдени жадында орныққан ұғымдар жүйесін, астарлы мағына мен бағалау реңкін контекспен бірлікте түсіндіруінде. Ал бұл бағыт ұлттық цифрлық егемендікті нығайтып, мемлекеттік сервистер, білім беру, медиакөністік және мәдени-анықтамалық сервистер үшін сапалы қазақша тілдік технологиялардың жаңа деңгейін қалыптастыруға негіз болады.

## Қаржыландыру

Бұл мақала Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің ЖТН BR24993001 «Қазақ тілін қолдау және технологиялық прогресті дамыту мақсатында үлкен тілдік модельді (LLM) құру» гранттық немесе бағдарламалық-нысаналы қаржыландыру жобасы аясында дайындалды.

## Мүдделер қақтығысы

Авторлар осы зерттеуге қатысты қандай да бір кәсіби, қаржылық немесе жеке мүдделер қақтығысының жоқ екенін мәлімдейді.

## Мақала авторларының үлесі

Т. Рамазанов мақаланың негізгі ғылыми тұжырымдамасын өзірледі, зерттеу мақсаты мен міндеттерін айқындады, QazLLM, ChatGPT және ISSAI жүйелерінің жауаптарын салыстырмалы талдау өлшемдерін белгіледі, негізгі мәтінді дайындады. А. Жаңабекова қазақ этномәдени коды, фразеологизмдер, тұрақты тіркестер, ұлттық дүниетаным және мәдени семантикаға қатысты теориялық-әдіснамалық негізді жүйелеуге, талдау нәтижелерін лингвистикалық тұрғыдан интерпретациялауға үлес қосты. А. Садық жасанды интеллект, үлкен тілдік модельдер, цифрлық трансформация және деректер корпусына қатысты ғылыми әдебиеттерді саралауға, эмпирикалық материалдарды құрылымдауға және мақала мәтінін ғылыми-редакциялық тұрғыдан жетілдіруге қатысты.

## Әдебиеттер

Тоқаев, Қ. (2025, 9 қыркүйек). Жасанды интеллект дәуіріндегі Қазақстан: өзекті мәселелер және оны түбегейлі цифрлық өзгерістер арқылы шешу. *Мемлекет басшысы Қасым-Жомарт Тоқаевтың Қазақстан халқына Жолдауы. Қазақстан Республикасы Президентінің ресми сайты. Ақорда*. <https://www.akorda.kz/kz/memleket-basshysy-kasym-zhomart-tokaevty-n-kazakstan-halkyna-zholdauy-zhasandy-intellekt-daurinderi-kazakstan-ozekti-maseleler-zhane-ony-tubegeyli-cifrluk-ozgerister-arkyly-sheshu-881957>

Тоқаев, Қ. (2026, 5 қаңтар). Қазақстан жаңғырудың жаңа кезеңіне қадам басты. *Мемлекет басшысы Қасым-Жомарт Тоқаевтың сұхбаты. «Turkistan» газеті*. <https://www.akorda.kz/kz/kazakstan-respublikasynyn-prezidenti-kasym-zhomart-tokaev-kazakstan-zhangyru-dyn-zhana-kezenine-kadam-basty-401240>

Қайдар, Ә. (2013). *Қазақтар ана тілі әлемінде. 3-т. Табиғат*. Алматы: Сардар.

Манкеева, Ж. (2008). *Қазақ тіліндегі этномәдени атаулардың танымдық негіздері*. Алматы: Жібек жолы.

Уәли, Н. (2025). *Тілтанымдық зерттеулер. Бірінші кітап*. Алматы: Асыл ЖК.

Уәли, Н. (2025). *Сөз мәдениетінің теориялық негіздері (2-басылым)*. Алматы: Асыл.

Әлімбаев, Н. (2017). *Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. Энциклопедия. 1-т.* Алматы: Алем Даму Интеграция.

Қартаева, Т. (2015). *Сыр өңірінің дәстүрлі этнографиясы*. Алматы: Арыс.

Токтабай, А. (2010). *Қазақ жылқысының тарихы*. Алматы: Алматыкітап.

- Biswas, S. (2023). ChatGPT for research and publication: A step-by-step guide. *Journal of Pediatric Pharmacology and Therapeutics*, 28(6), 576–584. <https://doi.org/10.5863/1551-6776-28.6.576>
- Thorp, H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
- Boruff, T., & Kraft, M. (2023). Introducing the Journal of the Medical Library Association's policy on the use of generative artificial intelligence in submissions. *Journal of the Medical Library Association*, 111(4), 747–749. <https://doi.org/10.5195/jmla.2023.1826>
- Hill-Yardin, E., & Hutchinson, M. (2023). A Chat(GPT) about the future of scientific publishing. *Brain, Behavior, and Immunity*, 110, 152–154. <https://doi.org/10.1016/j.bbi.2023.02.022>
- Tao, Y. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), 1–9. <https://doi.org/10.1093/pnasnexus/pgae346>
- Bender, E., & Gebru, T. (2021). On the dangers of stochastic parrots: Can language models be too big? *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Хабиева, А., & Уәли, Н. (2025). Қазақ тұрақты сөз тіркестерінің мәдени-танымдық мазмұны. *Абылай хан атындағы ҚазХҚжӘТУ Хабаршысы. Филология сериясы*, 1(76), 311–324. <https://doi.org/10.48371/PHILS.2025.1.76.021>
- Тоқаев, Қ. (2025, 3 қаңтар). Мақсатым – экономиканы және егемендікті нығайту. *Мемлекет басшысы Қасым-Жомарт Тоқаевтың сұхбаты. «Ана тілі» газеті*. <https://anatili.kazgazeta.kz/news/63674>

## References

- Tokayev, Q. (2025, September 9). *Zhasandy intellekt dauirindegi Qazaqstan: Ozekti maseleler zhane ony tubegeili tsifirlyk ozgerister arqyly sheshu* [Kazakhstan in the era of artificial intelligence: Current issues and their solution through radical digital transformation]. *Address of the Head of State Kassym-Jomart Tokayev to the people of Kazakhstan. Official website of the President of the Republic of Kazakhstan. Akorda*. <https://www.akorda.kz/kz/memleket-bashshysy-kasym-zhomart-tokayevty-n-kazakstan-halkyna-zholdauly-zhasandy-intellekt-dauirindegi-kazakstan-ozekti-maseleler-zhane-ony-tubegeyli-cifirlyk-ozgerister-arkyly-sheshu-881957> (In Kazakh)
- Tokayev, Q. (2026, January 5). *Qazaqstan zhangyrudyn zhana kezenine qadam basty* [Kazakhstan has entered a new stage of modernization]. *Interview of the Head of State Kassym-Jomart Tokayev. Turkistan newspaper*. <https://www.akorda.kz/kz/kazakhstan-respublikasynyn-prezidenti-kasym-zhomart-tokayev-zhangyrudyn-zhana-kezenine-kadam-basty-401240> (In Kazakh)
- Qaidar, A. (2013). *Qazaqtar ana tili aleminde. 3-t. Tabigat* [Kazakhs in the world of their native language. Vol. 3. Nature]. Sardar. (In Kazakh)
- Mankeeva, Zh. (2008). *Qazaq tilindegi etnomadeni ataulardyn tanymdyq negizderi* [Cognitive foundations of ethnocultural names in the Kazakh language]. Zhibek Zholy. (In Kazakh)
- Uali, N. (2025). *Tiltanymdyq zertteuler. Birinshi kitap* [Linguistic studies. Book one]. Asyl ZhK. (In Kazakh)
- Uali, N. (2025). *Soz madenietinin teoriialyq negizderi* [Theoretical foundations of speech culture] (2nd ed.). Asyl. (In Kazakh)
- Alimbai, N. (2017). *Qazaqtyn etnografiialyq kategoriialar, ugymdar men ataularynyn dasturli zhuyesi. Entsiklopediia. 1-t.* [The traditional system of Kazakh ethnographic categories, concepts, and names. Encyclopedia. Vol. 1]. Alem Damu Integratsiia. (In Kazakh)
- Qartaeva, T. (2015). *Syr onirinin dasturli etnografiiasy* [Traditional ethnography of the Syr region]. Arys. (In Kazakh)
- Toqtabai, A. (2010). *Qazaq zhylqysynyn tarikhy* [The history of the Kazakh horse]. Almatykitap. (In Kazakh)
- Biswas, S. (2023). ChatGPT for research and publication: A step-by-step guide. *Journal of Pediatric Pharmacology and Therapeutics*, 28(6), 576–584. <https://doi.org/10.5863/1551-6776-28.6.576>
- Thorp, H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
- Boruff, T., & Kraft, M. (2023). Introducing the Journal of the Medical Library Association's policy on the use of generative artificial intelligence in submissions. *Journal of the Medical Library Association*, 111(4), 747–749. <https://doi.org/10.5195/jmla.2023.1826>
- Hill-Yardin, E., & Hutchinson, M. (2023). A Chat(GPT) about the future of scientific publishing. *Brain, Behavior, and Immunity*, 110, 152–154. <https://doi.org/10.1016/j.bbi.2023.02.022>
- Tao, Y. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), 1–9. <https://doi.org/10.1093/pnasnexus/pgae346>
- Bender, E., & Gebru, T. (2021). On the dangers of stochastic parrots: Can language models be too big? *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Khabieva, A., & Uali, N. (2025). Qazaq turatqy soz tirkesterinin madeni-tanymdyq mazmuny [Cultural and cognitive content of Kazakh phraseological units]. *Abylai Khan KazUIR&WL Bulletin. Philology Series*, 1(76), 311–324. <https://doi.org/10.48371/PHILS.2025.1.76.021> (In Kazakh)
- Tokayev, Q. (2025, January 3). *Maqsatym – ekonomikany zhane egemendikti nygaitu* [My goal is to strengthen the economy and sovereignty]. *Interview of the Head of State Kassym-Jomart Tokayev. Ana tili newspaper*. <https://anatili.kazgazeta.kz/news/63674> (In Kazakh)

**Авторлар туралы мәлімет:**

Рамазанов Талгат Бекболатұлы – PhD, А. Байтұрсынұлы атындағы Тіл білімі институты (Алматы, Қазақстан, e-mail: tphd-92@mail.ru).

Жаңабекова Айман Әбділдәқызы – филология ғылымдарының докторы, қауымдастырылған профессор, А. Байтұрсынұлы атындағы Тіл білімі институты (Алматы, Қазақстан, e-mail: aiman\_miras@mail.ru).

Садық Айнагүл Мақсатқызы – магистр, К. Сағадиев атындағы Халықаралық бизнес университеті (Алматы, Қазақстан, e-mail: s.ainagul\_94@mail.ru).

**Information about the authors:**

Ramazanov Talgat – PhD, A. Baitursynuly Institute of Linguistics (Almaty, Kazakhstan, e-mail: tphd-92@mail.ru).

Zhanabekova Aiman – Doctor of Philological Sciences, Associate Professor, Akhmet Baitursynuly Institute of Linguistics (Almaty, Kazakhstan, e-mail: aiman\_miras@mail.ru).

Sadyk Ainagul – Master, K. Sagadiyev University of International Business (Almaty, Kazakhstan, e-mail: s.ainagul\_94@mail.ru).

**Сведения об авторах:**

Рамазанов Талгат Бекболатович – PhD, Институт языкознания имени А. Байтұрсынұлы (Алматы, Қазақстан, e-mail: tphd-92@mail.ru).

Жанабекова Айман Абдильдаевна – доктор филологических наук, ассоциированный профессор, Институт языкознания имени Ахмета Байтұрсынұлы (Алматы, Қазақстан, e-mail: aiman\_miras@mail.ru).

Садық Айнагуль Максатовна – магистр, Университет международного бизнеса имени К. Сағадиева (Алматы, Қазақстан, e-mail: s.ainagul\_94@mail.ru).

Келіп түсті: 13 қаңтар 2026 жыл

Қабылданды: 4 маусым 2026 жыл