

ЭОЖ 81.116.6

ГТАХР 16.21.35

<https://doi.org/10.48371/PHILS.2026.2.81.032>

## ТІЛДІ МОДЕЛДЕУ: ЖАСАНДЫ ИНТЕЛЛЕКТ ПЕН ТАБИҒИ ТІЛДІҢ СЕМАНТИКАЛЫҚ ЖӘНЕ МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫНДАҒЫ ЕРЕКШЕЛІКТЕР

\*Шенгелбай М.Қ.<sup>1</sup>, Алкебаева Д.А.<sup>2</sup>

\*<sup>1,2</sup>әл-Фараби атындағы Қазақ ұлттық университеті  
Алматы, Қазақстан

**Аңдатпа.** Цифрлану дәуірі әлемге үлкен өзгерістер әкелді. Бұл ғылым саласының дамуында елеулі екені белгілі. Лингвистика саласында да цифрландыру үдерісінің әкелген жаңалығы мол және ықпалы зор. Цифрландырудың жемісі саналатын жасанды интеллекті (ЖИ) әлемнің, ғылымның үлкен жаңалығы ретінде танылып, бұл технологиялар жиынтығын қолдану адам өмірінің тиімділігін арттыруға бағытталды. Жасанды интеллект ұсынған мәтіндердің лингвистикалық өзгешеліктерінің түрлі тетіктерін ашу, адам интеллектісі мен жасанды интеллектінің жазу мен сөйлеудегі ерекшеліктерін жүйелеу тіл ғылымы үшін өзекті. Жасанды интеллект дайындаған мәтіндерге ғылыми филологиялық зерттеулер мен сараптама жасау ұлттық тіл үшін өте маңызды. Адам тілі мен жасанды интеллект тілінің құрылымында едәуір айырмашылықтар бар. Жасанды интеллектінің күнделікті өмірге, коммуникацияға тез ену үдерісі тілге де елеулі өзгерістер әкеліп жатыр. Зерттеудің мақсаты: Тілдің моделдену сипаты мен құрылымдық айырмашылықтарды тану, семантикалық мағынаның өзгерістерін ашу.

Бұл мақалада нейрожелілердің және тілдің даму кезеңдері, өзгеруі зерделенеді. ChatGPT дайындаған мәтіндер мен табиғи тілде жасалған мәтіндердің айырмашылықтары, ерекшеліктері сараланады. Жасанды интеллектінің семантикалық мәнді түсіну қабілеті мен морфологиялық құрылымдарды тани алу мүмкіндігі айқындалады.

Зерттеу барысында тіл эволюциясына қатысты, машиналық оқытуға байланысты теориялық тұжырымдар негізге алынды. Мақаланың әдіснамалық бөлімінде ғылыми пайымдау, жинақтау, салыстыру, моделдеу, талдау әдістері қолданылды. Зерттеу нәтижесінде семантикалық мағынаны мысалдар арқылы талдап, тілді моделдеудің шектеулері мен қиындықтары анықталып, нейрожелілерді моделдеудегі этикалық қарама-қайшылықтардың себептері нақтыланды.

Зерттеудің құндылығы – жасанды интеллект тілін талдау арқылы цифрландыру үдерісінде ұлттық тілдің бояуын, ерекшелігін сақтаудың маңызын көрсету. Жұмыстың практикалық маңыздылығы мақаланы

жоғары оқу орындарында оқытылатын «Қазіргі қазақ тілі лексикологиясы», «Қазіргі қазақ тілі морфологиясы» пәндеріне материал ретінде қолдануға болады.

**Тірек сөздер:** жасанды интеллект тілі, тіл эволюциясы, табиғи тіл, семантика, фразеологизмдер, мақал-мәтелдер, этнографиялық атаулар, тілдік модель, мәдени код, морфологиялық құрылым

### **Кіріспе**

Жаһандану дәуірі тіл эволюциясына зор әсерін тигізді. ХХІ ғасырдың ширегінде жасанды интеллекті (ЖИ) әлемдегі озық технология жаңалығы деп қарастырылды. Бұл ғылымның әртүрлі салаларында атап айтқанда, медицина, құқық, экономика, инженерия, қаржы салаларында танымал. Тіл ғылымы саласында да жасанды интеллекті өзіндік орнын белгіледі.

Говард Гарднер интеллект түрлерін бірнеше салаға бөледі. Г. Гарднер адам интеллектісін сипаттай отырып, физикалық, эмоционалдық, шығармашылық, кинестетикалық, лингвистикалық интеллект, логика-математикалық сынды түрлерін бөліп көрсетеді [1, 12 б.]. Осы интеллект шеңберінде ғалымдар робототехника, жасанды интеллект, нейрожелі секілді құрылғыларды жасап шығарды. Адам алдымен табиғатты тану арқылы ұшақ, көлік түрлерін жасады. Оған әсер еткен жан-жануарлар, жәндіктер, құстар болды. Адам өзін тани түсу арқылы адамға ұқсас құралдарды ойлап тапты. Адам жұмысын автоматтандыратын жүйе жасанды интеллектіні шығарды.

Интеллект – ақпаратты қабылдау, өңдеу, талдау және шешім қабылдау қабілеті. Адамның интеллектісі әртүрлі қабілет пен ойлау жүйесіне қатысты, ал ЖИ – адамның ойлау үдерістерін имитациялайтын технологиялық жүйе. Жасанды интеллектінің қазіргі лингвистикалық деңгейіне жетуі біршама уақыттың, дамытудың нәтижесінде болды.

Эволюция – бұл өзгеру, даму, өсудің кезеңдері. Ноам Хомскийдің жасаған тілдің әмбебап грамматикасы (генеративті лингвистикасы) ХХ ғасырдың орта шенінде әлемдегі озық бір жаңалық ретінде танылды. Америка ғалымы Н.Хомский революциясы тіл ғылымын «басынан аяғына дейін» таныған теория болды. Бұл теория бойынша мәтін пайда болу процесінде фонологиядан бастап синтаксиске қарай жүретінін негіздейді [2, 518 б.]. Ал бұған дейін синтаксистен фонологияға қарай қарастырылатын. Дәл осы теория машиналық аудармаға жол ашты.

Н. Хомский теориясында: Баланың сөйлеу кезеңінде бала өзіне қонымды сөйлеу конструкцияларын орта арқылы жасайды. Ғалымның теориясы бойынша «сөз сыйы» яғни балада универсалды тілдік құрылым болады деп тұжырымдайды. Хомскийдің зерттеулерінде грамматика адамның санасында туа біткен жүйе деп қарастырылады. Ал тілдік модельдерде ондай жүйе болмайды. Ол қондырғы ретінде танылады [3, 27 б.].

## Тілді моделдеу: жасанды интеллект пен табиғи тілдің семантикалық және..

Хомскийдің генеративті лингвистикасы негізінде компьютерлік лингвистика мен табиғи тілді өңдеу саласы дамыды. Бұл теориялық ілім бойынша сөйлем грамматикалық дұрыс құрылуы мүмкін, бірақ сөйлем мағынасыз бола алады [2, 504 б.]. Жасанды интеллект тілін талдау барысында осы теорияның негізі басшылыққа алынды.

Д. Журафский мен Ж. Мартин өздерінің зерттеуінде жасанды интеллектінің қалай сөйлем құрайтынын көрсетеді [4]. Осы зерттеуде жасанды интеллект тіліне тән терминдерге анықтама береді. Мына терминдерді қарастырдық:

Токенизация (Tokenization – сөйлемді сөздерге бөліп тастау). Сөйлемді немесе мәтінді жеке сөздерге, белгілерге, кейде морфемаларға бөлу. Мысалы: Мен / кітап/ оқыдым. Сөйлем сөздерге бөлінген.

Лемметизация (Lemmatization) Мәтіндегі сөздерді негізгі түбір қалпына келтіру. Мақсаты – бір мағынадағы әртүрлі сөз формаларын (етістік шақтары, көпше түр) бір леммаға келтіру. Мысалы: «оқыдым», «оқиды», «оқып жатыр» → «оқу». Негізгі түбір «оқу» сөзінен бірнеше сөз формалары шығады.

LLM – тілдерді бірнеше деректер арқылы оқыту. Бұл модельдер миллиардтаған сөздер мен сөйлемдерден үйреніп, тілді қолданудың үлгілерін меңгереді. Олар контексті сақтап, күрделі сұрақтарға жауап береді [4].

Тіл эволюциясы мен жасанды интеллект эволюциясын қарастырған уақытта, тіл қолданушылардың үлкен тілдік модельдерге жүгінуі бағдарға алынды. Жасанды интеллект тілін моделдеу токендер арқылы іске асады. Сөз формасы лемматизация арқылы айқындалады. Семантикалық ерекшелігі жайлы айтсақ, табиғи тіл яғни адам тілінде контекст, интенция, прагматика бар, ал жасанды интеллект тілі контекстуалды ықтималдық арқылы мәтін жасайды.

Үлкен тілдік модельдерде ең жиі қолданылатын декодтау әдісі «іріктеу» саналады. Іріктеу – модель ықтималдықтары бойынша кездейсоқ сөз таңдау. Әрбір қадамда сөздерді модель контекстке қатысты берген ықтималдыққа сәйкес таңдайды. Модель қайсыбір сөзге жоғары ықтималдық көрсетсе, сол сөздің таңдалуы жиі болады, керісінше ықтималдығы төмен сөздер сирек таңдалады [4, 276-277 бб.]. Біз осы зерттеу арқылы жасанды интеллектінің тілді қалай өңдейтінін аңғардық, бұл біздің зерттеу мақаламызда негізгі өзек болады. Салыстыру жұмысында осы жасанды интеллект жасаған жауаптардың қалай жасалауына қатысты мән беріледі.

ЖИ қысқа уақыт ішінде компьютерлік көру және табиғи тілді өңдеудің жетілдірілген әдістері сияқты көптеген технологиялық жетістіктерге жетті. Қазіргі таңда ол көп жағдайларда адам интеллектісінің баламасы ретінде қарастырылады. Бүгінгі даму қарқынына қарай, ЖИ-дің адам интеллектісінен асып түсу болжамы бар. Сондықтан ЖИ-дің түрлі

салалардағы қолданылуын қарастыру өте өзекті мәселе болып отыр [5]. Әсіресе, лингвистика саласындағы ерекшеліктерін, ақпаратты беру әдісін, кемшіліктері мен жетістіктерін таныту басты назарға алыну қажет.

### **Материалдар мен әдістердің сипаттамасы**

Зерттеу мақаламызда жасанды интеллект негізінде жасалған үлкен тілдік модельдердің қазақ тіліндегі семантикалық және морфологиялық құрылымды игеру деңгейі талданды. Зерттеу барысында қазақ тіліндегі 50 тұрақты тіркес, 50 мақал-мәтел, 30 этнографиялық атаулар және морфологиялық құрылымды зерделеуге байланысты 30 сөйлем материал ретінде алынды. Осы тілдік бірліктерге қатысты жасанды интеллектіге әртүрлі тапсырмалар берілді:

- Фразеологизмнің мағынасын түсіндіру;
- Мақалдың екінші бөлігін аяқтау;
- Этнографиялық атаудың түсіндірмесін беру;
- Әртүрлі тақырыпқа сөйлемдер құрау.

Зерттеу материалы негізінде ғылыми пайымдаулар жасалып, жинақтау, салыстыру, талдау, моделдеу әдістері қолданылды.

### **Нәтижелер мен талқылау**

Табиғи тіл мен ЖИ тілі берген ақпаратты салыстыру барысында ұқсастықтар мен айырмашылықтар анықталды. Салыстыру жұмысы семантикалық және морфологиялық құрылымдық ерекшеліктерді тануға бағытталды. ЖИ тілінің сұрауға жауап беруі тез, табиғи тілден керек ақпаратты іздеу үшін біршама уақыт кетеді. Жасанды интеллектінің мәтінді интерпретациялауы, өз нұсқасын жасай алуы ақпараттың нақты, шындық екенін тексеруді талап етеді.

#### **Мақал-мәтелді салыстыру.**

*Ел жасымен көрікті.*

*Тау тасымен көрікті,*

*Аяқ асымен мықты [6, 6 б.].*

ЖИ тілі:

*Ел жасымен мықты.*

*Жер шөбімен мықты.*

Берілген мақал-мәтелдің бірінші бөлігі айтылып, жалғастырылуы сұралды. Қазақ мақал-мәтел жинағындағы нұсқамен сәйкес келмейді. Қазақ мақал-мәтелі дыбыстық үйлесімділікке құрылғанын байқаймыз, «жас», «тас», «ас» сөздерінің өзара дыбыстық байланысы, алдыңғы сөздің мәнімен үйлесуі танылады. Ал ЖИ тілінде ықтимал үлгі ретінде екінші жалғасы берілген. «Жанбыр жауса жер ырысы, батыр туса ел ырысы» деген мақал-мәтелдің ықтималдығы негізінде «Ел жасымен мықты, жер шөбімен мықты» деген нұсқаны ұсынған. Бұл ел, жер сөздерінің қатар қолданысына қатысты түзілу орын алған.

*Елдің көркі – сақал*

*Сөздің көркі – мақал [6, 108 б.].*

ЖИ жауабы:

*Ердің көркі – сақал,*

*Елдің көркі – ақыл.*

Тағы бір нұсқа (ЖИ):

*Ердің көркі – сақал,*

*Етектің көркі – шұғала.*

Жасанды интеллект жаңа мақал жасауға шебер екені көрінеді. ЖИ ықтималдық негізінде мақалға арналған сөз таңдап, мақалға жаңа өңдеу жасайды. Бұл тіптен мақал-мәтелдің мәнін, өңін құбылтып, өзгертіп жібереді. Біз қазақ тілінің ұлттық мәдени мазмұнды моделдейтін үлкен тілдік модельдерді жасауға жұмыс жасалуды құп көреміз. Қазақ халқының ерекшелігі – айтыс өнері. Әр сөзді ұйқасқа құруы, әрбір мақалы белгілі бір ұйқастың, дыбыстардың үндесуі ғана емес, сөздің, мағынаның үйлесімін дәл береді. «Сақал, мақал» сөздері де дыбыстық үйлесімділікке сай құрылған. Жасанды интеллект берген нұсқада сақал мен ақыл сөздері дыбыстық үйлесімге келмейді, ал екінші нұсқасындағы «Етектің көркі – шұғала» деген жалғасының беретін мәні мүлдем анық емес.

Келесі бір мақалды талдап көрейік:

*Қанағат қарын тойғызар,*

*Қанағатсыз жарлының жалғыз атын сойғызар [7, 33 б.].*

ЖИ жауабы:

*Қанағат – қарын тойғызар,*

*Құныққан – өлім тойғызар.*

Қазақ мақал-мәтелдеріндегі ерекшелік – екі құбылысты салыстыру, салғастыру. Бұл тұста да қанағат пен қанағатсыздықтың арасындағы құбылыстар салыстырылады. Ал ЖИ берген жауап қанағатсыз сөзін «құныққан» сөзімен алмастырған, «өлім тойғызар» тіркесі грамматикалық жағынан қате құрылған тіркес екені танылады, мағыналық жақтан да берілген мақалдың мәнін аша алмайды. «Өлім тойғызар» деген тіркес қазақ тілінде мүлдем кездеспейді. Қазақ мақал-мәтелі – қаншама жылдардан бері бекіген, халық аузынан шыққан сөздің халық танымында орныққан асыл сөз екені айқын. Демек, жасанды интеллект жасайтын нұсқалардың ешқайсысы мақал-мәтелдің өз жалғастығының орнын өзгерте алмайды.

**Фразеологизмдерді салыстыру.** І. Кеңесбаев редакторлығымен шыққан фразеологиялық сөздіктегі 50 фразеологизмнің мағынасы жасанды интеллектіден сұралды. Жауаптардың жартысынан көбі семантикалық мағынаны қате етіп моделдеді.

Фразеологиялық сөздік ұсынған мәлімет:

**Апталығы шықты.** *Әбден қалжырады, шаршады, діңкесі құрыды.* Апталығым шығып, эзер бітірдім [8, 43 б.].

ЖИ жауабы: *Апталығы шықты. Фразеологизм қазақ тілінде адамның немесе жануардың қатты арықтап, жүдеп кеткенін білдіреді. Бұл тіркес адамның немесе жануардың денесіндегі қабырғаларының айқын көрініп тұруын сипаттайды, яғни олардың арық екенін көрсетеді. Мысал: Ауыр жұмыс пен тамақтанбаудың салдарынан ол қатты жүдеп, апталығы шықты.*

Салыстыра қарасақ, ЖИ тіліндегі фразеологизмнің мәні мен табиғи тілдегі мән сәйкес емес. Екеуі екі түрлі мағынаны көрсеткен. Қазіргі сандық технология ақпаратты жылдам іздеуге бағытталған. Адам бұл фразеологизмді іздеуде ЖИ ұсынатын мәліметтерге бірден сүйенеді. Кейбір ақпараттың дәл берілмеуі қарама-қайшылық туғызып қана қоймай, нақты ақпаратты білмеуге әкеледі.

Табиғи тілде *«Араны жүрді»* тіркесі *«айтқаны болды»* деген мағынаны береді [8, 45 б.].

ЖИ жауабы: *«Араны жүрді»* фразеологизмі қазақ тілінде адамның немесе жануардың ашқарақтық, тойымсыздық қасиетін білдіреді. Бұл тіркес біреудің ашкөздікпен, қомағайлықпен әрекет етіп жатқанын сипаттау үшін қолданылады.

Қазақ тілінде *«араны ашылды»* деген фразеологизм бар. Оның мағынасы *«ашкөздікті»* білдіреді. Фразеологизм құрамындағы бірліктер, етістіктер бірігіп бір мағынаны береді, ал ЖИ белгілі бір тіркесті ғана айқындай алады, сұрауға жауап беру үшін соған ұқсас фразеологизмнің беретін мағынасын ұсынған. *«Араны ашылды»* фразеологизмінің беретін мағынасын *«араны жүрді»* деген фразеологизмі орнына қолданған.

Табиғи тіл: *Азу шалды. Ұлғайды, егде тартты, қартайды* [8, 15 б.].

ЖИ тілі: *«Азу шалды»* фразеологизмі адамның егде тартқанын, жасының ұлғайғанын білдіреді. *Dositel.prib*

Бұл сұрауда фразеологизмнің мағынасын ашу үшін белгілі бір дереккөз қолданылған, сілтемесі көрсетілген. Соған сай, фразеологизмнің мағынасы да дұрыс берілген.

Табиғи тіл: *Арда бала. Әкесінің үйінен екіқабат болып келген келіншектен туған бала немесе неке қиылғаннан кейін аз уақыттың ішінде туған бала* [8, 46 б.].

ЖИ тілі: *«Арда бала»* – қазақ тіліндегі тұрақты тіркес, бірақ ол классикалық фразеологизм ретінде кең таралмаған. Мағынасы (фразеологиялық тұрғыдан): *Құрметке лайық, асыл, үлгілі бала, ел үшін бағалы, қадірлі жас ұрпақ.*

Табиғи тіл: *Арда емді. 1. Тай шыққанша, есейгенше емді. Ақ тай енесінің асаулығынан былтыр байланған. Биыл арда емді. 2. Оңаша, еркін иемденді. Магниткадан көтеріліп ұшқан бозғылт түтін аспанның төсін арда емген қаз бауыр бұлт мәнзелдес* [8, 46 б.].

ЖИ тілі: *«Арда емді»* – қазақ тіліндегі тұрақты тіркес. Мағынасы: *өте қадірлі, қымбат бағалы, ерекше құрметке лайық. Толығырақ түсінік:*

*Арда – құрметті, қымбат мағынада.*

*Емді – белгілі бір нәрсеге лайықты, сай, тең.*

ЖИ тілінде «арда бала» және «арда емді» фразеологизмдерінің мағынасы қате көрсетілді. Бұл ықтимал үлгіге салып, әрі фразеологизмнің әр компонентін бөлшектеп көрсетуден де туды. «Арда» сөзін «ардақты» сөзіне үйлестіру негізінде фразеологизмнің мән-мағынасын басқа арнаға бұрып жіберді.

Табиғи тіл: *Арқадан бүрді. Суыққысты, тоңдырды. – Мынау аяз қойды қинап тұрған жоқ. Өстіп, міне, өзімізді де арқадан бүріп, бүршеңдетіп тұр* [8, 47 б.].

ЖИ тілі: *«Арқадан бүрді» – қазақ тіліндегі фразеологизм, яғни тұрақты тіркес. Оның мағынасы: Күтпеген жерден бірдеңе болу немесе орын алу.*

*Қысқаша:*

*«Арқадан» – кенеттен, күтпеген жерден*

*«Бүрді» – шықты, пайда болды.*

*Мысалы: Ол жұмысқа кешігіп, бастығы «арқадан бүрді» деді.* Берілген мысал да фразеологизмнің мәнін аша алмайды. Мысал тұрғысынан сөйлемнің семантикалық мәнін түсініксіз етіп берген.

Берілген салыстыру жұмысынан шығатын қорытынды:

1. Фразеологизмдер құрамындағы сөздер бірігіп, тұтас бір мағынаны береді. ЖИ тілінде фразеологизмдік құрамды бөлшектеп, әр сөздің мағынасын беру арқылы мәнді ашқан. Бұл тіпті берілген фразеологизмнің беретін мағынасына мүлдем сай келмейді.

2. Берілген мысалды да ЖИ өзі құрастырады. Бұл мысалдардан байқайтынымыз, сөйлемдегі құрылымның дұрыс еместігін, мағынасы үйлесімсіз құрылғанын байқауға болады.

Фразеологизмдер ұлттық мәдениетті және дүниетанымды бейнелейтін ұлттық кодтың бөлігі ретінде қызмет етеді. Олар ұлттың мінез-құлқын, құндылықтарын және дүниетанымын көрсетеді, сондықтан фразеологизмдердің маңыздылығы олардың ұлтты танытатын, мәдениетті жалғастыратын қызметінде жатыр [9, 145 б.]. Жасанды интеллекті ұлттық кодты тануы мүмкін емес, ұлттық код ол ұлттың санасында орныққан мәдениетін бейнелейді және таңбалайды.

Жасанды интеллект тілі табиғи тілдің кейбір құрылымдық және семантикалық ерекшелігін қайталауға тырысса да, деректер жиынтығына тәуелді екені айқындалады. Р. Бинс өз мақаласында жасанды интеллектіге қатысты философиялық тұрғыда сараптама жасайды. Бинстың негіздеуінше, техникалық шешімдерде яғни ЖИ жауабында этикалық бейтараптық жоқ. Біз де бұл пікірді басшылыққа ала отырып, ЖИ тілінің құрылымы бейтарап емес екенін атап өтеміз және жасанды интеллект мәтіні деректерді таңдау мен өңдеу нәтижесінде қалыптасатынын негіздейміз. Тұжырымдай келе,

табиғи тіл эволюциялық, мәдени, прагматикалық тұрғыда қалыптасса, ал ЖИ мақсатты түрде моделденетін, бұл моделдену алгоритмдік логикаға сүйенетіні танылды. Берілген мысалдардан ЖИ тілінің тек деректерді таңдап, өңдеп беретіні айқындалды, бұл этикалық мәселелердің туындауына әкеледі. Яғни қате ақпараттың берілуі ЖИ тілінде өте жиі кездеседі. Бұл қазақ тіліндегі деректердің ғаламтор желісінде аз болуына да қатысты.

ЖИ-дің сұрауға жауап беруін талдайық. ЖИ-ға берілген промт мынадай:

«Аманай мен Заманай» шығармасының авторы кім?

ЖИ жауабы: *«Аманай мен Заманай» — бұл шығарма қазақтың белгілі жазушысы Дулат Исабековтің туындысы.*

*Негізгі кейіпкерлер:*

• *Аманай — адал, шынайы сезім иесі, әділеттілікті іздейтін жас жгііт*

• *Заманай — атастыру арқылы басқа адамға тұрмысқа шығуға мәжбүр болған, бірақ Аманайды сүйетін қыз*

Тілдік модельдер шындыққа жанаспайтын, жалған ақпарат айтуы мүмкін. Бұл құбылыс «галлюцинация» деп аталады. Бұндай модельдер мәтіннің болжамға негізделуіне үйретіледі, бірақ оқыту алгоритмдері мәтіннің дұрыс немесе шынайы болуын қамтамасыз ете алмайды. Бұл әсіресе фактіге негізделген ақпараттық қорларға кері әсер етпек [10, 280 б.]. Осы тұста шығатын қорытынды: Мәтіннің болжамға негізделуі шынайы ақпаратты тануға кедергі келтіреді. Chatgpt – бірнеше функция атқара алатын, көбіне мәтіндік деректерді жеткізетін, ақпараттық көмекші рөлді атқаратын құрал. Бірақ тіл қолданудағы семантикалық дәлдіктің жоқтығы, ақпаратты қатесіз беруге қатысты өзіндік шектеулері бар. ЖИ моделі «Аманай мен Заманай» туындысын белгілі жазушы Дулат Исабековке теліп, Сәкен Жүнісов авторлығын ескермей жауап беруі жасанды интеллектінің тілдік білім қоры тек статистикалық ықтималдықтарға негізделетінін анықтайды. Жасанды интеллект өз жауабында қазақ әдеби шығармаларындағы ықтимал нұсқаны берген. Бұл жерде авторды ғана қате айтқан жоқ, шығармадағы кейіпкерлердің образдары, олардың кім екендігінен бейхабар, шынайы ақпаратты бере алмайды.

Ж. Манкееваның «Қазақ тіліндегі этномәдени атаулардың танымдық негіздері» еңбегінен қазақ тіліндегі этнолексикалық ұғымдардың беретін мағынасын жасанды интеллектіге сұрау жіберу арқылы іздестірдік. ЖИ жауабында *көнек, тегене, шығыр, қауға, шалғы, желінқап, кебеже* сынды біршама этнографиялық атаулардың мән-мағынасы дәл берілді. Ал мына мысалдарда жасанды интеллекті ұғымдардың беретін мағынасын мүлдем басқа бағытқа бұрмалап жіберген.

*Сыпыра – қамыр илейтін жұқа жарғақ тері* [11, 104 б.].

ЖИ жауабы: *Зат есім тұрғысынан «сыпыра» сөзі – көбінесе көне немесе әдеби тілде қолданылатын сөз және ол мынадай мағына береді:*

«Сыпыра» – зат есім. Мағынасы: түгел, жаппай, барлығы яғни, жалпылықты білдіретін сөз. Мысал: «Сыпыра қазақ» бүкіл қазақ халқы деген мағынада.

**Түймеш** – туралып, жанышылып, кішірек етіп доғалақталынып майға қуырылған ет тағам [11, 116 б.].

ЖИ жауабы: Тағам атауы ретінде «түймеш» сөзі кейбір өңірлік нұсқаларда кездеседі. Бұл жағдайда түймеш – ұннан немесе еттен жасалатын ұсақ түйірлі тағам түрін білдіруі мүмкін. Енді ықтимал мағынасы:

Түймеш – ұннан жасалатын ұсақ түйірлі тағам (ботқа түрі) Ұн, май, су немесе сүт қосылып, қуырылады немесе қайнатылады. Етті тағамдар қатарына емес, көбіне қарапайым, тоқ тағам ретінде дайындалған.

**Сарбөртпе** – екі-үш рет суға шайып, біршама қан-сөлінен арылтқан қыран құсқа беретін ас [11, 116 б.].

ЖИ жауабы: Сарбөртпе – бұл сөз қазақтың көне тұрмысында сирек қолданылатын, мағынасы ұмыт бола бастаған этнографиялық атау. Оның нақты мағынасы өңірлік диалект пен контекстке байланысты әртүрлі болуы мүмкін, бірақ ең ықтимал түсіндірмелер мыналар: Киіз үй жабдықтарына қатысты атау (болжамды)

Кейбір деректерде «сарбөртпе» сөзі киіз үйдің ішкі сәндік жабдықтарының бірі, яғни өрнектелген киіз немесе алаша түскиіз іспеттес бұйым ретінде айтылған.

ChatGPT-дің ең үлкен ерекшелігі – ол әрдайым үйренуші. Ол өзі жасаған қатені түзетуге ұмтылады, мәтін стилін есте сақтай алады, мәтін аудиториясын таңдап, соған арнап мәтін құра алады. Жасанды интеллектіге тек дұрыс бағдар жасап отыру маңызды, қате ақпаратты бірден дұрыстауды да сұрау қажет. Жалған ақпаратты берген кезде ықтимал түрде немесе ескі, көнеден келе жатқан, диалекті деген ақпараттарды қосымша айту ЖИ тілінде кездеседі, бұл берілген ақпараттың шынайы емес екеніне дәлел бола алады, мәтінді тексеруді ұсынады. «Сыпыра», «түймеш», «сарбөртпе» сынды этнолексиканы түсіндіруді сұрағанда болжамды жауап екенін көрсеткен, бұл ақпаратты тексеруге бағыттайды.

LLM модельдері жіктеу тапсырмаларында жақсы нәтиже көрсетіп, генеративті тапсырмаларда қиындықтарға тап болды. Кейбір тапсырмалар мен модельдер үшін ағылшын тіліндегі нұсқаулар нәтиженің сапасын жақсартады [12]. Классификация – бір жауапты таңдау, ал генеративті тапсырма – жаңа мәтін жасау, сұраққа жауап бере отырып, мәтінді ары қарай жалғастыру, жалғастыруды сұрау. Генеративті тапсырмаларда ЖИ модельдері қазақ тілінің грамматикасы, морфологиясы, стильдік ерекшеліктерін түсінбей жиі қателеседі. Сондықтан генерация нәтижесі бұрмалануы мүмкін. Ал жіктеу тапсырмаларында жауаптар шектеулі болғандықтан, модельге жеңіл болады.

**Морфологиялық құрылымды жасанды интеллекті қалай танитынын, қандай шектеулері бар екенін қарастырайық:**

ЖИ тіліне берілген сұрау: Аяқ бау ма, аяқбау ма?

*Дұрыс жазылуы: аяқбау(біріктіріліп жазылады)*

*Түсіндірме:*

**«Аяқбау» сөзі екі түбірден құралған: «аяқ» және «бау»**

*Бұл сөз мағыналық тұтастықты білдіреді(аяқты байлау немесе бекіту үшін қолданылатын бау) сондықтан бірге жазылады.*

*Ереже: Қазақ тілінің орфографиялық нормаларына сәйкес, мағыналық және қызметтік бірлікті білдіретін сөздер біріктіріліп жазылады.*

**Р.Сыздық өз еңбегінде:**

*Ереже: «Екі (кейде одан да көп) сөз қатар айтылып, бір ғана ұғымды білдіретін, яғни бір ғана затты, құбылысты, қимылды, түр-түсті, санды атайтын күрделі сөздердің әрбір компоненті бөлек жазылады.»*

Ғалым өз еңбегінде екінші компоненті бау, жіп, сап, киім, қалта, қағаз, ағаш сияқты зат есімдерден тұратын және олардың түрлік ерекшелігін айқындайтын сапалық не қатыстық мәндегі есім сөздерді қарастырады. *Мысалы, бау сөзімен келетін: аяқ бау, балақ бау, жел бау, қарғы бау, тамақ бау деген күрделі сөздердің барлығында да алдыңғы компонент баудың түрлерін атайды [13, 42 б.].* Дұрысында, аяқбау болып жазылу қажет. Ал жасанды интеллекті морфологиялық ерекшелікті ескермегендіктен, бірге жазылады деген нұсқаны ұсынған.

ЖИ-ға берілген сұрау: Жолсерік пе әлде жолсерік пе?

ЖИ жауабы: *Дұрыс нұсқасы – жолсерік.(біріктіріліп жазылады)* Бұл сөз – біріккен сөз, яғни екі сөздің бірігіп, жаңа мағына беріп тұрған түрі. «Жол» - бағыт, сапар. «Серік» - бірге жүретін адам, жолдас. *Жолсерік – дұрыс, орфографиялық сөздіктерде де осылай жазылады.*

Р.Сыздық еңбегінде: *Әр алуан мамандық, кәсіп иелерінің аттары және белгілі бір іске (әрекетке, жұмысқа т.б.) бейім адамдардың атаулары да екі сөз арқылы білдірілсе, біріккен сөз болады: алыпсатар, атқамінер, теміржолшы, жолсерік, өнертапқыш [13, 47 б.].*

Бұл сұрауға ЖИ дұрыс жауап берді, орфографиялық сөздікті негізге алғаны байқалды. ЖИ жауабында үнемі екі сөздің мағынасын бөлек алып қарастыру жиі ұшырайды. «Жол» және «серік» сөздерін бөліп қарауды негізге алған, ал қазақ тілінде екі сөздің бірігіп, бір мағынаны беретінін көрсетеді.

ЖИ дайындаған сөйлемдерге талдау:

*1. Қорқыныш сезімі кейде адамның дұрыс шешім қабылдауына кедергі келтіреді, бірақ оны басқаруға үйренуге болады.* «Басқаруды үйренуге болады» деген тіркеспен сөйлем аяқталуы қажет. ЖИ септік жалғауларының ықтималдыққа салғанда тұйық етістік формалы сөздерге қаз-қатар жалғаған. Бұл тұста бірінші тұйық етістік формасына табыс септік формасын жалғау мағыналық, грамматикалық жағынан дұрыс келеді.

2. *Егер таң ерте тұрсаң, күндізгі жұмысқа көп үлгеруге болады.* Бұл сөйлемде де септік жалғауы қалыс қалған. «Таңда ерте тұрсаң» деген грамматикалық нұсқа дұрыс болады. «Күндізгі жұмысқа көп үлгеруге болады» деген сөйлем бөлігінде де грамматикалық қателер көрінеді. «Күндіз көп жұмыс істеп үлгеруге болады» деген нұсқаны ұсынамыз. «Көп үлгеруге болады» тіркесінде баяндауыштың анықтауышы болмайтыны ескеріледі. Демек «көп» сөзі «үлгеруге болады» тіркесінің алдында тұра алмайды.

3. *Егер біз бір-біріміздің көзқарасын тыңдай білсек, өзара түсіністік артады.* «Көзқарасымызды» деген тәуелдік формасының көпше формасы қолданылуы қажет. Өйткені, көзқарасын сөзінің алдында *бір-біріміздің* деген сөз тәуелдік септігінің көпше формасында айтылған.

4. *Армандарсыз өмір – түссіз және мәнсіз сияқты көрінеді.* «Арман, үміт, қуаныш, мұң» сынды дерексіз зат есімдерге көптік жалғауы жалғанбайды. Ал бұл сөйлемде «армандарсыз» деп беру қате, «Армансыз адам қанатсыз құспен тең» деген мақалда көптік форма жалғанбаған, осы сынды нұсқалар ЖИ тіліне ықтимал формаларды ұсыну тұрғысынан көмектесуі қажет.

Жасанды интеллект негізіндегі тіл модельдері табиғи тілді өңдеуде біршама жетістіктерге жеткенімен, күрделі мағыналық құрылымдар мен фразеологизмдерді түсіну барысында шектеулері анықталды.

Біріншіден, фразеологизмдердің, этнографиялық атаулардың астарлы, мәдени мазмұнға негізделген мағынасын дәл жеткізе алмады. Бұл тілдің семантикалық тереңдігі мен когнитивті құрылымдарының толық моделденбегенін көрсетеді. Тұрақты тіркестің беретін мағынасын дұрыс интерпретация жасай алмады, ЖИ фразеологиялық бірліктерді тұтас мағыналық бөлік ретінде емес, жеке-жеке сөздерін бөліп қарастыратынын көрсетті. Тілдік модель мақал-мәтелдің құрылымдық тұтастығын айқындай алмады. ЖИ жауабында мақалдың логикалық жалғасы, екінші бөлігі дұрыс танылмады.

Екіншіден, тіл модельдері морфологиялық құрылымдарды беру кезінде белгілі бір деңгейде қателер жіберді. Көптеген мысалдарда көптік жалғаулары, септік формаларын, етістік шақтарын дұрыс қолданбауы қазақ тілінің табиғатын толық игермегенін аңғартады. Бұл әсіресе қазақ тілі секілді жалғамалы тілдер үшін өзекті мәселе болмақ. Зерттеу нәтижелері тілді моделдеу саласында семантикалық және морфологиялық деңгейдегі жетілдірудің қажеттілігін көрсетеді. Әсіресе фразеологизмдер мен ұлттық мәдени мазмұндағы сөз қолданыстарын моделдеу бағытында терең лингвистикалық әдіс пен қосымша дереккөз енгізу маңызды болып отыр.

Маңызды мәселе – LLM-модельдердің қандай тілдерді өңдей алатынын қарастыру. Себебі бұл модельдер тілдік гегемонияны күшейтіп, лингвистикалық әрі мәдени әртүрлілікке кері әсер етуі мүмкін. Зерттеу LLM-модельдердің ағылшын және тағы бірнеше тілде өте жақсы жұмыс істейтінін көрсеткенімен, жаһанның оңтүстігіндегі елдердің тілдеріне келгенде бұл көрсеткіш айтарлықтай төмен екенін дәлелдейді. Ал аз ұлттар тілі, байырғы

және жойылып бара жатқан тілдер туралы айтудың өзі қиын [14]. Тілдік гегемония – белгілі бір тілдің басқатілден үстемтұруы, ғылымда, технологияда, білім беру мен халықаралық қатынаста басымдыққа ие болуы. Қазақ тілі – жалғамалы тіл. Ағылшын тілі – флективті тіл. Екі тілдің құрылымы бір-біріне сәйкес емес. Ағылшын тілінің үстемдігі үлкен тілдік модельдерде айқын көрінеді. Үлкен тілдік модельдердегі айырмашылықтар лингвистикалық бөлінуге әкелуі мүмкін. Бұл технологиялық мәселе ғана емес, мәдени, әлеуметтік, идеологиялық мәселе болатыны айқын. Жаһандану кезеңінде технологиялардың тоғысуы, ұлт мәдениеті мен ұлттық сананы, ұлттық тілді өзгертуге күш салып жатыр. Бұл тіпті ана тілінің басты мәселелеріне айналып отыр. Тіл тек құрылымын ғана сақтап қоймайды, ұлттық мазмұн мен мәдени кодтарды сақтау назарға алынады. Агглютинативті тілдерде морфемалар түбірге жалғанып, семантикалық қана емес, синтаксистік қатынастарды да білдіреді. Сондықтан мұндай тілдерде морфологиялық көпмағыналылық әртүрлі формада көрінеді және кей жағдайларда оны шешу үшін әрі синтаксистік, әрі семантикалық талдау қажет.

Табиғи тілде түбір мен қосымшаға ажырайды, сөз формалары шақ, жақ, рай арқылы, септік, тәуелдік сынды жалғаулар арқылы көрінеді, қазақ тілінің семантикалық деңгейі мен морфологиялық құрылымындағы қателер, этикалық қарама-қайшылықтар лингвистикалық түсіндіру мен дереккөздің көп енгізілуіне де қатысты мәселе болмақ. Жасанды интеллект тілі түбір мен қосымшаны ажыратпастан, деректер қорына сүйенеді. Қазақ тілін сақтап қалу әрі әлемдік деңгейде дамуы үшін цифрландыру үрдісін сапалы әрі жылдам жүргізу талап етіледі. Тілдің үстемдікке ие болуы мақсатында тілдің ұлттық бояуы жоғалмайтындай дереккөздерді көбейту, тілдің грамматикалық ережелерге бағынатындай дәрежесін жасанды интеллектіге енгізуді қолға алу шарт. Дархан Ахмет-Зәки бастаған отандық ғалымдардың мақаласында қазақ тілінің медиа-корпусы үшін мәтінді автоматты алдын ала өңдеудің үш құралы ұсынылған: Сөз формаларын генерациялау құралы, морфологиялық талдағыш, морфологиялық көпмағыналылықты ажырату құралы. Морфологиялық көпмағыналылықты ажырату әзірге эксперименттік сипатта екені көрсетіледі. Оған себеп, белгіленген корпус болмағандықтан, сөйлем құрылымының үлгілік моделіне сүйене отырып жасалған. Зерттеу барысында контекстік және статистикалық-ықтималдық тәсілдерді біріктіруге талпыныстар жасалған [15]. Бұл зерттеу қазақ тілінің корпусық белгіленімі үшін өзекті жұмыс екені танылады. Аталған тәсілдер сөз формаларының түрлі нұсқаларын ұсына алады. ЖИ тек қана статистикалық-ықтималдық тәсілді пайдаланса, қазақ тілінің кейбір заңдылықтарын ескерместен жауап ұсынуы мүмкін. Сол себепті контекстік тәсілді біріктіру оңтайлы ұсыныстардың бірі деп танылмақ.

### **Қорытынды**

Жасанды нейрондық желілер бүгінгі таңда табиғи тілді өңдеу бойынша жоғары нәтижелерге қол жеткізді, біршама шектеулері де

## Тілді моделдеу: жасанды интеллект пен табиғи тілдің семантикалық және..

танылды. Зерттеу нәтижелері бойынша, ЖИ модельдері адам тілінің күрделі мағыналық, грамматикалық ерекшеліктерін игеруде әлі де мүмкіндіктерінің жетіспеушілігі байқалады. ЖИ тілінің ерекшелігін саралай отырып, мынадай нәтижелерді атап көрсетуге болады:

– ЖИ берген жауаптар фразеологиялық сөздік, мақал-мәтелдер жинағы, ғалымдардың еңбегі негізінде салыстырылып, жауаптың дәлдігі мен сәйкестігі анықталды. Салыстыру нәтижесінде тілді моделдеудің сипаты табиғи тіл мен ЖИ тілінде танылып, лексикалық, морфологиялық айырмашылықтары айқындалды.

– Машиналық оқыту мен адам тілі мағына және форма жағынан ерекшеленетіні негізделді.

– Үлкен тілдік модель дайындаған жауаптардағы мәдени мазмұнға сай келмейтін тұстар, морфологиялық қателер жеке-жеке белгіленді.

– Зерттеу негізінде тілдік модельдің қазақ тіліне тән мағыналық құрылымдарды, тұрақты тіркес және грамматикалық ерекшеліктерді тану деңгейі жүйелі түрде бағаланды.

ЖИ тілі – мәтіндік дереккөздегі жиілік пен үлгіні қолданады, сол себепті қате жауап жиі кездеседі. Семантикалық деңгейде тіл модельдерінің тұрақты тіркестер мен этнографиялық атаулардың мағынасын дұрыс өңдеу жасай алмағаны мысалдардан көрінді. Ал кейбір лексикология саласының элементтерін дұрыс тани алады, бұл дереккөздерді айқын пайдалана алудың нәтижесін көрсетеді. ЖИ тілі лексикалық үйлесімділік, жинақталған мағынаны болжау сияқты қабілеттерге ие болғанымен, адам тіліне тән семантикалық мазмұнды, мағыналық түрленуді, тілдік вариативтілік және морфологиялық күрделілік сияқты терең құрылымдарды толық игере алмайды. Тілдік модельдер үлкен мәтіндік корпус негізінде статистикалық заңдылықтарды меңгергенімен, оларда мән мен мағынаға бағытталған когнитивтік қабілет жоқ. Дегенмен жасанды тілдік модельдер моделдеу құралы ретінде лингвистикалық зерттеулерге тың серпін беріп отыр. Тіл мен интеллектінің түйісу нүктесінде жатқан бұл мәселе қазіргі пәнаралық зерттеулер үшін өте өзекті, әрі лингвистикалық теория мен технологияны байланыстыратын көпір қызметін атқармақ.

## ӘДЕБИЕТ

[1] Гарднер Г. Структура разума: теория множественного интеллекта. – Москва: И.Д.Вильямс, 2007. – 512 с.

[2] Хомский Н. Синтаксические структуры. – М., 1962. – 527 с.

[3] Millière R. Language Models as Models of Language. – Oxford University Press, 2024. – 55 p.

[4] Jurafsky D., Martin J. H. Speech and language processing. – 2024. <https://web.stanford.edu/~jurafsky/slp3/>. 12.01.2025

[5] Rahman A., Raj A., Tomy P. et al. A comprehensive bibliometric and content analysis of artificial intelligence in language learning: tracing between

the years 2017 and 2023. *Artif Intell Review*. – 2024. – Vol. 57. – 107. DOI: <https://doi.org/10.1007/s10462-023-10643-9>

[6] Қазақтың мақал-мәтелдері / Құрастырған: Смайлова А.Т. – Алматы: Көшпенділер, 2003. – 312 б.

[7] Қазақ халқының мақалдары мен мәтелдері / Құрастырған: С.Садырбайұлы. – Алматы: Қазақ университеті, 2012. – 230 б.

[8] Қазақ тілінің фразеологиялық сөздігі / Ред: Кеңесбаев І.К. – Алматы: Ғылым, 1977. – 712 б.

[9] Қоянбекова С.Б., Сатинбекова М.М., Кубдашева К.Б., Нұрмұхаметова Қ.Т. Экспрессивтік стилистика: қазақ тіліндегі фразеологизмдердің ұлттық кодтағы лингвомәдени парадигмалар // Абылай хан атындағы ҚазХҚ және ӘТУ Хабаршысы, Филология ғылымдары сериясы. – 2024. – №3 (72). – Б. 498-512. DOI: <https://doi.org/10.48371/PHILS.2024.3.74.008>

[10] Binns R. Fairness in Machine Learning: Lessons from Political Philosophy // *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* – 2020. – № 81. – Pp. 149-159.

[11] Манкеева Ж.А. Қазақ тіліндегі этномәдени атаулардың танымдық негіздері. – Алматы: Жібек жолы, 2008. – 356 б.

[12] Maxutov A., Myrzakhmet A., Braslavski P. Do LLMs Speak Kazakh? A Pilot Evaluation of Seven Models // *Proceedings of the First Workshop on Natural Language Processing for Turkic Languages*. – 2024. – Pp.81-91.

[13] Сыздық Р. Емле және тыныс белгілері. – Алматы: Рауан, 1996. – 285 б.

[14] Choudhury M. Generative AI has a language problem // *Nature Human Behavior*. – 2023. – Vol. 7. – Issue 11. – Pp. 1802-1803. DOI: <https://doi.org/10.1038/s41562-023-01716-4>

[15] Darhan A., Mansurova M., Madiyeva G., Kadyrbek N., Kyrgyzbayeva M. Development of the Information System for the Kazakh Language Preprocessing. *Cogent Engineering*. – 2021. – Vol. 8 (1). DOI: <https://doi.org/10.1080/23311916.2021.1896418>

## REFERENCES

[1] Gardner G. *Struktura razuma: teorija mnozhestvennogo intellekta* [The structure of the mind: the theory of multiple intelligences]. – Moskva: I.D.Vil'jams, 2007. – 512 s.[in Rus.]

[2] Homskij N. *Sintaksicheskie struktury* [ Syntactic structures]. – М., 1962. – 527 s. [in Rus.]

[3] Millière R. *Language Models as Models of Language*. – Oxford University Press, 2024. – 55 p.

[4] Jurafsky D., Martin J. H. *Speech and language processing*. – 2024. <https://web.stanford.edu/~jurafsky/slp3/>. 12.01.2025

[5] Rahman A., Raj A., Tomy P. et al. A comprehensive bibliometric and content analysis of artificial intelligence in language learning: tracing between

the years 2017 and 2023. Artif Intell Review. – 2024. – Vol. 57. – Article 107.

DOI: <https://doi.org/10.1007/s10462-023-10643-9>

[6] Qazaқтың maqal-mätelderi [Kazakh proverbs and sayings]. Qūrastyrған: Smailova A.T. – Almaty: Köşpendiler, 2003. – 312 b.[in Kaz.]

[7] Qazaq halqynyñ maqaldary men mätelderi [Proverbs and sayings of the Kazakh people]. Qūrastyrған: S.Sadyrbaiüly. – Almaty: Qazaq universiteti, 2012. – 230 b.[in Kaz.]

[8] Qazaq tılınñ frazeologialyq sözdıǵı [Phraseological dictionary of the Kazakh language]. Red: Keñesbaev I.K. – Almaty: Ğylym, 1977. – 712 b. [in Kaz.]

[9] Qoianbekova S.B., Satinbekova M.M., Kubdaşeva K.B., Nürmühametova Q.T. Expressive Stylistics: linguocultural paradigms in the National Code of phraseology in the Kazakh language // Abylai han atyndaǵy QazHQ jáne ÄTU Habarşysy, Filologia ğylymdary seriesy. – 2024. – №3 (74). – B.498-512. – DOI: <https://doi.org/10.48371/PHILS.2024.3.74.008> [in Kaz.]

[10] Binns R. Fairness in Machine Learning: Lessons from Political Philosophy // Proceedings of the 1st Conference on Fairness, Accountability and Transparency. – 2020. – Vol 81, – Pp. 149-159.

[11] Mankeeva J.A. Qazaq tılındegı etnomädeni ataulardyñ tanymdyq negızderi [Cognitive foundations of ethno-cultural terms in the Kazakh language]. – Almaty: Jibek joly, 2008. – 356 b. [in Kaz.]

[12] Maxutov A., Myrzakhmet A., Braslavski P. Do LLMs Speak Kazakh? A Pilot Evaluation of Seven Models // Proceedings of the First Workshop on Natural Language Processing for Turkic Languages, 2024. – Pp. 81-91.

[13] Syzdyq R. Emle jáne tynys belgileri [Spelling and punctuation marks]. – Almaty: Rauan, 1996. – 285 b. [in Kaz.]

[14] Choudhury M. Generative AI has a language problem // Nature Human Behavior. – 2023. – Vol. 7. – Issue 11. – Pp. 1802-1803. DOI: <https://doi.org/10.1038/s41562-023-01716-4>

[15] Darhan A., Mansurova M., Madiyeva G., Kadyrbek N., Kyrgyzbayeva M. Development of the Information System for the Kazakh Language Preprocessing. Cogent Engineering. – 2021. – Vol. 8 (1). DOI: <https://doi.org/10.1080/23311916.2021.1896418>

## **МОДЕЛИРОВАНИЕ ЯЗЫКА: ОСОБЕННОСТИ СЕМАНТИЧЕСКОЙ И МОРФОЛОГИЧЕСКОЙ СТРУКТУРЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ЕСТЕСТВЕННОГО ЯЗЫКА**

**\*Шенгелбай М.К.<sup>1</sup>, Алкебаева Д.А.<sup>2</sup>**

**\*<sup>1,2</sup>Казахский национальный университет имени аль-Фараби  
Алматы, Казахстан**

**Аннотация.** Эпоха цифровизации принесла миру значительные изменения. Известно, что это значимо в развитии отрасли науки. В

области лингвистики процесс цифровизации также характеризуется новыми научными направлениями и значительным влиянием на развитие исследований. Искусственный интеллект (ИИ), являющийся результатом цифровизации, признан одним из значимых достижений современной науки и технологий, а применение этих технологий направлено на повышение эффективности жизни человека. Выявление различных механизмов лингвистических особенностей текстов, создаваемых искусственным интеллектом, а также систематизация различий между человеческим интеллектом и искусственным интеллектом в письменной и устной речи являются актуальными задачами языкознания. Научно-филологическое исследование и экспертиза текстов, подготовленных искусственным интеллектом, имеют большое значение для национального языка. Между структурой человеческого языка и языка искусственного интеллекта существуют существенные различия. Быстрое внедрение искусственного интеллекта в повседневную жизнь и коммуникацию также оказывает заметное влияние на язык. Цель исследования: распознать моделирующий характер языка и структурные различия, выявить изменения смыслового значения.

В статье рассматриваются этапы развития и изменения нейронных сетей и языка. Проводится анализ различий и особенностей текстов, созданных ChatGPT, и текстов, написанных на естественном языке. Определяется способность искусственного интеллекта понимать семантическое значение и распознавать морфологические структуры.

В ходе исследования были использованы теоретические положения, связанные с эволюцией языка и машинным обучением. В методологической части статьи применяются методы научного рассуждения, обобщения, сравнения, моделирования и анализа. В результате исследования семантические значения анализируются на основе примеров, выявляются ограничения и трудности языкового моделирования, а также уточняются причины этических противоречий в моделировании нейронных сетей.

Научная ценность исследования заключается в анализе языка искусственного интеллекта и обосновании важности сохранения национальной языковой специфики и культурной окраски языка в условиях цифровизации. Практическая значимость исследования заключается в том, что его результаты используются в качестве учебного материала для дисциплин «Лексикология современного казахского языка», «Морфология современного казахского языка» изучаемых в вузах.

**Ключевые слова:** язык искусственного интеллекта, эволюция языка, естественный язык, семантика, фразеологизмы, пословицы, этнографические названия, языковая модель, культурный код, морфологическая структура

## LANGUAGE MODELING: FEATURES OF THE SEMANTIC AND MORPHOLOGICAL STRUCTURE OF ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE

\*Shengelbay M.K.<sup>1</sup>, Alkebaeva D.A.<sup>2</sup>

\*<sup>1,2</sup>al-Farabi Kazakh National University, Almaty, Kazakhstan

**Abstract.** The era of digitalization has brought significant changes to the world. It is known that this is significant in the development of the branch of science. In linguistics, the process of digitalization is also characterized by the emergence of new research directions and a significant impact on scientific studies. Artificial intelligence (AI), which is the result of digitalization, is recognized as one of the important achievements of modern science and technology, and the use of these technologies is aimed at improving the efficiency of human life. Identifying the mechanisms of linguistic features of texts generated by artificial intelligence, as well as systematizing the differences between human intelligence and artificial intelligence in written and spoken communication, are relevant tasks for linguistic research. Scientific and philological analysis and examination of texts produced by artificial intelligence are of great importance for the national language. There are significant differences between the structure of human language and the language generated by artificial intelligence. The rapid integration of artificial intelligence into everyday life and communication also has a noticeable impact on language. The aim of the study is to identify the characteristics of language modeling and structural differences, as well as to reveal changes in semantic meaning.

The article examines the stages of development and transformation of neural networks and language. It analyzes the differences and specific features of texts generated by ChatGPT and texts written in natural language. The ability of artificial intelligence to understand semantic meaning and recognize morphological structures is also determined.

The study is based on theoretical concepts related to language evolution and machine learning. In the methodological section of the article, the methods of scientific reasoning, generalization, comparison, modeling, and analysis are used. As a result of the study, semantic meanings are analyzed using examples, the limitations and challenges of language modeling are identified, and the causes of ethical contradictions in neural network modeling are clarified.

The scientific value of the study lies in analyzing the language of artificial intelligence and substantiating the importance of preserving the national linguistic specificity and cultural coloring of language in the context of digitalization. The practical significance of the research lies in the fact that its results can be used as teaching material for the disciplines “Lexicology of the Modern Kazakh Language” and “Morphology of the Modern Kazakh Language”, studied in universities.

**Keywords:** artificial intelligence language, language evolution, natural language, semantics, phraseological units, proverbs, ethnographic names, language model, cultural code, morphological structure

*Мақала түсті / Статья поступила / Received: 05.09.2025.  
Жариялауға қабылданды / Принята к публикации / Accepted: 25.06.2026.*

***Авторлар туралы мәлімет:***

Шенгелбай Мәдина Қайратбекқызы – докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, e-mail: madinashengelbay@mail.ru, <https://orcid.org/0000-0001-7289-5948>

Алкебаева Дина Акбергенқызы – филология ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, e-mail: alkebaeva@mail.ru, <https://orcid.org/0000-0003-2348-4047>

***Информация об авторах:***

Шенгелбай Мадина Кайратбековна – докторант Казахского национального университета имени аль-Фараби, Алматы, Казахстан, e-mail: madinashengelbay@mail.ru, <https://orcid.org/0000-0001-7289-5948>

Алкебаева Дина Акбергеновна – доктор филологических наук, профессор, Казахский Национальный университет имени аль-Фараби, Алматы, Казахстан, e-mail: alkebaeva@mail.ru, <https://orcid.org/0000-0003-2348-4047>

***Information about the authors:***

Shengelbay Madina Kairatbekkyzy – doctoral student, al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: madinashengelbay@mail.ru, <https://orcid.org/0000-0001-7289-5948>

Alkebaeva Dina Akbergenkyzy – Doctor of Philological Sciences, Professor, al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: alkebaeva@mail.ru, <https://orcid.org/0000-0003-2348-4047>