

М.Қ. Шенгелбай¹, Д.А. Алкебаева¹, Ж.С.Таласпаева²

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, 050000, Қазақстан

²М. Қозыбаев атындағы Солтүстік Қазақстан университеті, Петропавл, 150000, Қазақстан

ЖАСАНДЫ ИНТЕЛЛЕКТ ТІЛІ: ЛИНГВИСТИКАНЫҢ ЖАҢА КӨКЖИЕКТЕРІ

Бұл мақалада жасанды интеллектің тілдік мүмкіндіктері мен табиғи тілдің мәдени және когнитивтік ерекшеліктері арасындағы айырмашылықтар жан-жақты талқыланып, қазіргі заманғы ЖИ жүйелерінің адам тілін имитациялаудағы жетістіктері мен шектеулері зерттелді. Мақаланың өзектілігі қазіргі ақпараттық дәуірде ЖИ технологияларының кеңінен қолданылуы, аударма мен мәтін генерациясы саласындағы алға жылжуымен тығыз байланысты, өйткені бұл жүйелердің өнімдері тілдің сыртқы формасын дәйекті түрде жеткізсе де, ұлттық мәдениет пен эмоционалдық мазмұнды жеткізуде әлсіздік танытады. Зерттеу мақсаты – жасанды интеллект генерациялаған мәтіндер мен адам тіліне тән табиғи шығармаларды салыстыра отырып, лексикалық, семантикалық, синтаксистік және стилистикалық деңгейдегі ерекшеліктерді айқындау және ЖИ жүйелерінің нақты тілдік тәжірибеге қаншалықты жақын келетінін бағалау. Зерттеу әдістері ретінде корпуслық лингвистика, лексикалық және синтаксистік талдау, сондай-ақ стилистикалық талдау қолданылды; GPT және басқа үлгілерден генерацияланған мәтіндер мен ғылыми мақалалар, әдеби шығармалардан құралған корпус арқылы AntConc, Sketch Engine секілді құралдар пайдаланылды. Алынған нәтижелер көрсеткендей, ЖИ мәтіндері нақты, логикалық және формальды құрылымға ие болғанымен, оларда көркемдік, идиомалық және ұлттық дәстүрлерге тән экспрессивті элементтер жеткіліксіз екені байқалды, ал табиғи тілде жазылған мәтіндер бай лексикалық әртүрлілік, күрделі синтаксистік құрылым және эмоционалдық-идеологиялық контекстке ие. Осы тұрғыдан, мақала ЖИ жүйелерінің тілдік құбылыстарды тек формальды деңгейде ғана емес, терең мағыналық және мәдени контексті ескере отырып дамыту қажеттігін, сондай-ақ философиялық және когнитивтік аспектілерді интеграциялаудың маңыздылығын айқындап, болашақта ЖИ тіл модельдерін жетілдіруде көпсалалы тәсілдерді қолданудың тиімділігін дәлелдейді.

Кілт сөздер: жасанды интеллект, тіл моделі, прагматика, табиғи тіл, мәтін.

НЕГІЗГІ ЕРЕЖЕЛЕР

Жасанды интеллект (ЖИ) – цифрлық технологияның басты бір бөлігі. Бұл жаһандану дәуірінде адам өміріне түрлі өзгерістер әкеліп, жаңа бір деңгейге шығаратын қуатты көз болмақ. Адам қолынан жасалған ешбір технология қауіп-қатер әкелмейді. Тек қана оны қолдану барысында қауіп келуі әбден мүмкін. Бұл мақалада жасанды интеллект тілінің қолданудағы ерекшеліктер қарастырылады, табиғи тілден өзгешелігі айқындалады, табиғи тілге жақындатудың мүмкіндіктері анықталады. Жасанды интеллектің тілін зерттеу өзектілігі мынадай: ЖИ жүйелері адам ұсынған нұсқауларды орындау арқылы жүзеге асады. Демек, барлық нұсқауды беруші адамға ЖИ тілін түсіну маңызды, оның барлығы сөз арқылы орындалады. Олай болса, жасанды интеллектіге өз тілінде нұсқаулықтарды орындатқызып сөз айту маңызды болады, сол алгоритмді машина түсінетіндей ЖИ жүйесін қамтамасыз ету талап етіледі. Жасанды интеллект тілін зерттеу тұрғысынан қазақ тілінде ғылыми зерттеу еңбегі толықтай жазыла қойған жоқ. Тек жасанды интеллектің мүмкіндіктері, сұраулар (промттар) жіберу жайына қатысты, ЖИ-ді қолданудың түрлі әдіс-тәсілдері тұрғысынан зерттеліп жатқан еңбектерді ғана атап өтуге болады. Алан Тьюринг, Гаспарян, Рудер сынды шетелдік ғалымдардың еңбектерінен жасанды интеллектің тілін зерттеуге қатысты тұжырымдары мақалада назарға алынды. Табиғи тілдің ерекшеліктері, функционалдық стильдерге байланысты өзгешеліктер жайлы М.Серғалиев, Р.Сыздықова

еңбектері басшылыққа алынды. Бұл еңбектер ЖИ тілінің стильдік ерекшелігін зерттеу аспектісінде табиғи тілге жақындатуға қатысты негіз болды.

КІРІСПЕ

Жасанды интеллект (ЖИ) өміріміздің бір бөлшегіне айналып келеді. ЖИ мүмкіндіктері күн сайын жаңартылып, адамзатты өзіне тартуда. Заманның бет-бейнесі тез ауысып жатқан уақытта жасанды интеллектіні тез шешім қабылдау үшін, тез тапсырма орындау үшін тұтыну белең алды. Алайда кез келген жаңалықтың біздің санамызға, тілімізге, әдеттерімізге тигізер әсері бары заңдылық. Осы орайда жасанды интеллект біз тұтынып жүрген тілге қаншалықты әсерін тигізеді, оның тілі қандай деген сауал туындайды. Ең алдымен, жасанды интеллект тілі аударма саласында қарқынды түрде қолданылатынын айта кетейік. Машиналық аударма бұрынғымен салыстырғанда әлдеқайда жүйелі, синтаксистік құрылымы жинақы мәтіндер беретін болған. Бұның бірнеше себебі бар: қазақ тілді материал интернет жүйесінде кең көлемде таралып жүр және оны ЖИ жақсы оқиды. Екіншіден, аудармашылар қазақ тілді аударманың дұрыс болуы үшін дәл қазіргі уақытта ЖИ-дің берген мәтінін ала салмай, тіліміздің табиғатына жақын аудармаға төсеу ісімен айналысып жүр. Жасанды интеллект біз сұраған ақпаратты өңдеуді, талдауды, жүйелеп бір шаблонға салуды жақсы атқарады. Алайда бұл тілдің сыртқы формасына қатысты тиянақтылық деп түсінеміз. Ал, тіл – өн бойында бір ұлттың тарихын, ділін, дінін, мәдениетін, дүниетанымын сақтайтын мәдени мұра болғандықтан, ЖИ бұл ерекшелікті жеткілікті деңгейде сезіне алмайды немесе оның маңыздылығын түсінбейді. Осы тұрғыдан алғанда, қазақ тілінде мәтін жазып беретін жасанды интеллект тілінің табиғи тілге жақындығы мен айырмашылығын анықтау өзектілігі туындайды. Зерттеудің мақсаты ЖИ тілі мен табиғи тілді салыстыра отырып, ерекшеліктерін айқындау. Зерттеудің міндеттері:

– Жасанды интеллект генерациялаған мәтін мен табиғи тілде жазылған мәтінді жинақтау, салыстыру;

– Лексикалық, семантикалық, синтаксистік, стилистикалық деңгейлерде ЖИ тілі мен табиғи тілді салыстырмалы талдау;

– ЖИ мәтіндерінің логикалық құрылымы мен тілдік ерекшеліктерді айқындау.

Жасанды интеллект тіліне қатысты алғашқы зерттеулер 1950 жылдан бастау алған. Авторы – атақты ғалым, заманының данышпаны саналған Алан Тьюринг. «Computing Machinery and Intelligence» мақаласында машина интеллектісін талқылап, тілді өңдеу мен адам ойының логикасын имитациялау мәселелерін көтереді [1, 437]. Аланның концепциясы дәл қазіргі уақытта жеңіске жетті. Мәселен, П.Н. Барышников мақаласында сананы сипаттауда қолданылған «компьютерлік метафораның» дамуын қарастырады. Өте ертедегі механикалық ұқсастықтардан бастап XX ғасырдың ортасындағы «ақыл-ой –компьютер» концепциясына дейінгі даму жолын тандайды. Ол когнитивтік ғылымдағы ақыл-ойды машина не бағдарламалық код түрінде қарастыруға қатысты бірқатар пайдалы модельдер ұсынғанымен, психикалық күйлердің семантикалық қырын түсіндіре алмайтынын көрсетті. Функционалдық есептеу тәсілінде модельдер сананы тек бағдарламалық үрдіс ретінде сипаттап, субъективтік мағынаны есепке алмайтын шектеулерге ие екені мәлім болды. Сонымен қатар, автор өз зерттеуінде «нейрондық желі» метафораларын талдайды. Нейроғылым деректеріне сүйеніп, байланысушылық тәсіл когнитивтік функцияларды модельдеуге тырысқанымен классикалық есептеу теориясының әлсіз тұсы бетпе-бет келіп жүр. Мысалы, нейрондық желілердің салмақ коэффициенттерін реттеу арқылы «үйренуі» адам білімінің символдық әрі мағыналық сипатта екенін толық қайталай алмайды деп көрсетеді [2, 29]. Сондықтан П.Н. Барышников сананы тек алгоритмдер жиынтығы не статистикалық үлгілер деңгейінде теңеу оның мәнін тым қарапайымдандырып жіберетінін ескертеді, өйткені мұндай тәсіл саналы тәжірибе мен мағынаны түсіндіре алмайды.

Д.Э. Гаспарян өз еңбегінде тіл семантикасын модельдеуде жасанды интеллект әдістерін құрылымдық және постқұрылымдық лингвистиканың призмасы арқылы талдайды. Автор

философиядағы сана мәселелерінде көбіне елеусіз қалған құрылымдық семиотика тұжырымдарын негізге алып, машинаға мағына мен сезімді еліктету мәселелерін қарастырады. Құрылымдық лингвистиканың «белгі» және «мағына» ұғымдарын қайта зерделей отырып, адам миы мен машинаның таңбаны өңдеу мүмкіндіктері түбегейлі айырмашылыққа ие екендігін көрсетеді [3, 720]. Зерттеуде заманауи ЖИ жүйелері синтаксистік деңгейде күрделі есептерді шеше алғанымен, семантикалық «түсіну» тұрғысынан шектеулі екені атап өтіледі. Компьютер таңбалардың формальды сәйкестіктерін үйренгенімен, олардың мәнін адамша түсінбейді. Сол себепті Д.Э. Гаспарян тілдік мағынаның контекстік өзгергіштігін және мәдени-танымдық қабаттарын ескеру қажеттігін баса айтады. Ол «мағынаны формализациялау» мәселесін талдай отырып, мағынаны толық ресми алгоритмдерге сыйғызу мүмкін еместігін, оның трансценденттік сипаттары барын алға тартады. Тіл семантикасын модельдеу барысында формалды талдауды адамзаттың когнитивтік тәжірибесі және мәдени контекстімен біріктіру қажет деген қорытынды жасайды, яғни тілдің мәнін түсіну үшін тек статистикалық әдістер емес, көпқырлы тәсіл керек.

У. Крофт пен Э. Дж. Вуд тілдік мағынаны түсіндірудегі когнитивтік категориялардың (метафора, фрейм, концепт ұғымдары) рөлін және олардың жасанды интеллектідегі көрінісін зерттейді. Олар тіл мағынасы адам санасындағы концептуалдық құрылымдарға негізделетінін, мысалы, сөзді не сөйлемді түсіну үшін оның тиісті frame-үлгісі белгілі бір ситуациялық білім құрылымы ескерілуі керегін айтады. Бұл тұрғыда, тек үлкен деректердегі жиілікке сүйенген статистикалық модельдер мұндай ұғымдық байланыстарды толық игере алмайтыны көрсетіледі. Авторлардың пайымдауынша, қазіргі таза деректік әдістермен жұмыс істейтін ЖИ жүйелері адамның ойлауындағы категориялар мен тұжырымдамаларды өз жұмыс үлгілеріне толық енгізе алмайды, сондықтан тілдің астарындағы терең мағыналық құрылымды бере алмайды. Яғни метафоралық мағыналарды, фреймдік білімді және энциклопедиялық ақпаратты ескермеген жағдайда, үлкен көлемдегі мәліметтерге негізделген статистикалық тәсілдер тілдік коммуникацияның күрделі қырларын түсіндіруге жеткіліксіз болып қала бермек. Нәтижесінде, У. Крофт пен Э.Дж. Вудтың жұмысы тіл мен ойлаудың концептуалды негіздерін ЖИ-де моделдеу қажеттігін және деректерге сүйенген тәсілдің шектеулі екенін көрсетеді [4, 60].

Г.А. Кемпен редакторлығымен жарық көрген еңбектер жинағында табиғи тіл генерациясы (яғни адамның сөйлеуін модельдеу) саласындағы жаңа ғылыми үрдістер ЖИ, психология және лингвистика тұрғысынан қарастырылады [5, 100]. Бұл еңбектерде жасанды жүйелердің адамша мәтін немесе сөйлем шығаруы ерекше қиын мәселе екендігі атап көрсетіледі. Нақтырақ айтқанда, компьютерлік модель бір мезгілде тілдің синтаксистік және морфологиялық дұрыстығын сақтап қана қоймай, сонымен бірге семантикалық тұтастық пен прагматикалық сәйкестікті қамтамасыз етуі керек. Адам сөйлеуінде бұлардың бәрі табиғи түрде үйлеседі. Адамның сөз шығару процесі лексикалық таңдау, грамматикалық құрылым түзу, сөз формаларын түрлендіру және контексті ескере отырып мәнді хабарды жеткізуден тұратын күрделі жүйе екені белгілі; мұны толықтай имитациялау қазіргі ЖИ үшін оңай емес. Дегенмен соңғы нейрондық тіл модельдерінің пайда болуы бұл бағытта елеулі жетістіктер әкелді: олар үлкен мәтін корпусынан үйрену арқылы грамматикалық заңдылықтарды және белгілі бір семантикалық байланыстарды өз бойына сіңіре алды. Мұндай модельдер генерациялау барысында бұрынғы ережеге негізделген жүйелерге қарағанда түрлі деңгейлерді (морфология, синтаксис, семантика) біріктіріп, адам тіліне ұқсас мәтіндерді құрастырып жүр. Алайда, бұған қарамастан, нейрожелілік модельдердің шектеулері де бар: оларда тілдік мазмұнды «түсіну» қабілеті болмағандықтан, кейде қисынсыз немесе контекске сай емес сөйлемдер пайда болуы мүмкін. Г.А. Кемпен және өзге зерттеушілер табиғи тіл тудырудың толыққанды моделін жасау үшін осы аталған барлық деңгейлерді (синтаксис, семантика, прагматика) бір жүйеге интеграциялау қажеттігін және бұл мақсатта адам ойын, дүниетанымын модельдеумен ұштастырылған көпсалалы тәсіл қажет екенін алға тартады.

В.А. Ладовтың «Проблема зомби в контексте философии ИИ» еңбегінде сананы түсіну тұрғысынан қызықты ой-тұжырым – «философиялық зомби» концепциясы – жасанды

интеллект философиясы контекстінде талданады. Философиялық зомби дегеніміз – сырттай қарағанда саналы адам сияқты барлық әрекеттерді орындайтын, бірақ ішкі саналы тәжірибесі жоқ гипотетикалық тіршілік иесі. В.А. Ладов осы ұғымды қолдана отырып, қазіргі жасанды интеллект жүйелері де белгілі бір дәрежеде сондай «зомби» тәрізді болуы мүмкін бе деген сұрақ қояды. Яғни компьютерлік алгоритмдер адам тілінде жауап беріп, саналы істеп жатқандай әсер қалдырғанымен, шынында олар ақпаратты механикалық өңдеп, ешнәрсені сезінбейді деуге болады. Автор саналы түсіну мен алгоритмдік орындалудың арасындағы айырмашылықты айқындап, ЖИ жүйелері мағынаны шынымен ұғына ала ма деген мәселеге назар аударады. Егер жасанды агент сыртқы ортаға адамша әрекет еткенімен, оның «ішкі» санасы болмаса, оны саналы деуге бола ма? В.А. Ладов бұл сұрақ төңірегіндегі пікірталастарды талдап, қазіргі ЖИ көбіне функционалдық деңгейде ғана іске асатынын, ал субъективті тұрғыда «түсіну» қабілеті күмәнді екенін көрсетеді. Демек, жасанды интеллект нақты мағынада саналы түсінікке ие болмай, тек философиялық зомби тәрізді, яғни мінез-құлық бар да, бірақ саналы түйсік жоқ жүйе болып қалу қаупі бар деп түйінделеді [6, 44].

Рудер және әріптестері өз зерттеулерінде табиғи тілді өңдеудегі трансферлік оқытудың мүмкіндіктерін жан-жақты қарастырады. Үлкен көлемдегі мәтіндер корпусында алдын ала дайындықтан өткен модельдердің айрықша артықшылықтары бары көрсетілген. Атап айтқанда, миллиардтаған сөздерден тілдік заңдылықтарды үйренген ірі нейрондық тіл модельдері (мысалы, BERT, ELMo сияқты) кейін оларды шағын мәліметтер жиынтығына бейімдеу арқылы жоғары нәтижелер көрсетеді. Мұндай алдын ала оқытылған жүйелер жалпы тілдік құрылымдар мен семантикалық байланыстарды өздерінің ішкі көрінісінде сақтайды. Зерттеулерге сәйкес, тілдік модель ретінде алдын ала жаттықтыру нейрондық желіге тілдің синтаксистік ережелері мен құрылымдық заңдылықтарын терең меңгеруге мүмкіндік береді. Тіпті «тіл моделі ретінде алдын ала оқыту аудармамен оқытуға қарағанда көбірек синтаксистік білім береді» деген нәтижелер де бар [7, 15]. Осындай универсал білімнің арқасында модельдер жаңа тапсырмаларды аз ғана қолтаңба (fine-tuning) арқылы үйренгенде деректер тапшылығына қарамастан жоғары өнімділікке қол жеткізеді. Яғни, трансферлік оқыту тәсілі модельдің бұрын көрмеген міндеттерді де игеруіне жол ашып, лингвистикалық заңдылықтарды ұтымды пайдалануға мүмкіндік береді. Нәтижесінде, алдын ала үлкен деректер (үйретілген модельдер) аз мәліметпен жұмыс істегеннің өзінде анағұрлым дәл және тиімді жұмыс істей алады. Бұл әдіс деректер көлемі шектеулі тілдік міндеттерде де үздік нәтижелерге жетудің жолын көрсетті.

К. Арнет, Б. Бергэн көптілді тілдік талдауды (парсингті) ағылшын тілі аясынан кеңейту мәселелерін зерттеп, түрлі тілдердің синтаксистік ерекшеліктерін ескермейінше, әмбебап ЖИ модельдерінің шектеулі болатынын алға тартады. Әсіресе морфологиялық тұрғыдан күрделі тілдер үлкен көлемді деректерде үйретілген қазіргі модельдер үшін қиындық туғызады. Мысалы, агглютинативті тілдерде бір сөз көптеген морфемалардың тіркесуінен құралып, өте көп дербес сөз формалары пайда болады. Мұның нәтижесінде сирек кездесетін үлгілер саны артып, статистикалық модельдердің өнімділігі төмендейді. Зерттеулер 22 тілге жүргізілген салыстырмалы талдау негізінде морфологиялық бай тілдерде тіл модельдерінің көрсеткіші морфологиясы жеңілдеу тілдерге қарағанда нашарлау болатынын дәлелдеген [8, 16]. Мұның себебі – күрделі сөз түрленімдері деректерде шашыраңқылық тудырып, үлгінің ықтималдықпен үйренуіне кедергі келтіруі мүмкін. Сондықтан авторлар көптілді ортада жоғары нәтижеге жету үшін тілдердің морфологиялық және синтаксистік ерекшеліктерін ескеру керек дейді. Ағылшын тілі негізінде дайындалған модельдерді тікелей басқа тілдерге қолданғанда, сол тілдің құрылымдық айырмашылықтары назардан тыс қалса, қателіктер көбейетіні анықталды. Бұл мәселені шешу үшін әр тілге бейімделген тәсілдер мен үлкен трансформер модельдерін үйлестіру қажет. Лингвистикалық факторларды есепке ала отырып, көптілді синтаксистік өңдеуді жетілдіру – тілдік әртүрлілікті қамтитын әдістер мен типологиялық білімдерді ЖИ модельдеріне енгізуді талап етеді.

Жоғарыда қарастырылған барлық зерттеулерде ортақ пікір – адам санасы мен тілінің бай күрделілігін модельдеу үшін тек формалды немесе статистикалық тәсілдердің мүмкіндіктері

шектеулі екендігі. Яғни дәстүрлі алгоритмдік не деректік үлгілер тілдік құбылыстың белгілі бір жақтарын сәтті имитациялағанымен, «түсіну» ұғымын толық көлемде іске асыра алмайды. Жасанды интеллект қазіргі деңгейінде тілдің беткі құрылымын меңгергенімен, оның терең мағыналық және саналы қырларын дөп басуы екіталай. Сондықтан тіл туралы білімді, сананың философиялық қырларын және когнитивтік психологияны біріктіретін көпсалалы тәсілдер ғана жасанды интеллектінің адам секілді түсіну қабілетіне жақындауын қамтамасыз ете алады деп тұжырымдалады. Бұл әдеби шолу тіл теориялары мен ЖИ технологияларын интеграциялаудың маңыздылығын айқындайды.

МАТЕРИАЛДАР МЕН ТӘСІЛДЕР

Зерттеуімізде ЖИ тіліне қатысты мақалалар мен еңбектерге шолу жасалды. Жасанды интеллект тілін зерттеу әлемдік тенденция болғандықтан алыс-жақын шетелдегі авторлардың еңбектері қолданылды. Сонымен қатар, ЖИ тілі мен табиғи тілді салыстыру үшін корпустық лингвистика әдісі, лексикалық және синтаксистік талдау, стилистикалық талдау қолданылды. Корпустық лингвистика әдісінде ChatGPT генерациялаған мәтіндерден корпус жасалды. Адам жазған мәтіндерден (мақалалар, әдеби шығармалар) корпус құрылды. AntConc және Sketch Engine құралдары арқылы мәтіндер салыстырылды. Лексикалық және синтаксистік талдауда ЖИ-генерацияланған және табиғи тілдегі мәтіндердің сөз жиілігі, синтаксистік құрылымы және сөз тіркестерінің қолдану жиілігі зерттелді. Лексикалық әртүрлілік (Type-Token Ratio) өлшенді. Стилистикалық талдау кезінде ЖИ және табиғи мәтіндердің логикалық байланысы, прагматикалық ерекшеліктері қарастырылды. Бағыныңқылы құрылымдардың қолданылуы зерттелді. Зерттеу нәтижелері бойынша ЖИ мәтіндері формалды тұрғыда логикалық құрылымға ие, дегенмен, лексикалық тұрғыда шектеулі ғана сөздерді пайдаланады. Себебі, мәдени контексті түсіну қиындық туғызады.

НӘТИЖЕЛЕР

Жасанды интеллект тілін корпустық лингвистика әдісімен зерттеудің мақсаты – жасанды интеллект жүйелерінің өндіретін және өңдейтін тілдік мәліметтердің құрылымдық, семантикалық және прагматикалық ерекшеліктерін анықтап, оларды табиғи тілдің заңдылықтарымен салыстыру арқылы жүйелердің шынайы тілдік тәжірибеге қаншалықты жақын келетінін бағалау болып табылады. Бұл зерттеу әдісі арқылы жасанды интеллект моделдерінің синтаксистік, морфологиялық және стилистикалық үлгілерін талдап, олардың қалыптасқан мәтіндеріндегі ықтимал қателер, тұрақсыздықтар мен ықшамдылық мәселелерін анықтауға мүмкіндік туады. Сонымен қатар, корпустық лингвистика әдісі жасанды интеллект технологияларын жетілдіру мақсатында, модельдерге енгізілетін деректердің сапасын арттыруға және қолданылатын алгоритмдердің табиғи тілдік құбылыстарды толық әрі дәл бейнелеу деңгейін жоғарылатуға үлес қосады. Осылайша, зерттеу тек технологиялық жетістіктерді бағалаумен шектелмей, жасанды интеллект жүйелерінің адам тілін түсінуі мен имитациялауының терең философиялық және лингвистикалық аспектілерін де айқындауға мүмкіндік береді. Зерттеу барысында табиғи тілде жазылған эссе мен жасанды интеллект жазып берген эссені салыстырдық. Тақырыбы «Балаларды оқытудағы ата-ананың рөлі», көлемі – 200 сөз. Жасанды интеллект жазып берген эссе:

Балаларды табысты оқытуда ата-аналардың рөлі ерекше маңызға ие. Ата-аналар – баланың алғашқы ұстаздары, олар үй жағдайында тәрбиелік орта қалыптастырып, білімге деген қызығушылығын арттырады. Баланың алғашқы қадамдары, сөздері, әдет-ғұрыптары ата-аналар арқылы қалыптасады, сондықтан да ата-аналар білім беру процесіне белсенді қатысуы қажет.

Ата-аналардың баланы қолдау мен ынталандыруы оқу үлгерімін арттыруда маңызды фактор болып табылады. Олар баланың жетістіктерін бағалап, табысты болуына сенімділік оятады, сонымен бірге қиындықтар туындаған жағдайда психологиялық көмек

көрсетеді. Ата-аналардың баламен тілдесуі, қызығушылықтарын тыңдауы, оның ойын құрметтеуі баланың өзін-өзі бағалауын арттырады және білім алуға деген ынтасын күшейтеді.

Сонымен қатар, ата-аналар білім беру мекемелерімен тығыз байланыс орнатуы керек. Мұғалімдермен кездесу, баланың сабақтары туралы ақпарат алу, оқу үрдісіне қатысты ұсыныстар беру арқылы ата-аналар білім сапасын жақсартуға үлес қосады. Осының нәтижесінде, бала өз мүмкіндіктерін толық іске асырып, табысты тұлға болып қалыптасады.

Қорыта келгенде, баланы табысты оқытуда ата-аналардың қолдауы мен белсенділігі үлкен рөл атқарады. Олар баланың білімге деген қызығушылығын арттырып, оның болашақ табысына негіз қалаушы ең басты фактор болып табылады.

Табиғи тілде жазылған эссе:

Мектеп қабырғасы – баланың білім, білік дағдыларын қалыптастырып, оқудың негізгі нәрін бойына дарытатын орда. Ал, ұстаз – оқушының сапалы білім алуы үшін барын салып, педагогикалық мүмкіндіктерін толықтай пайдалануға күш салатын тұлға. Дегенмен, баланың үлгерімін тек мектеп пен мұғалімнің ғана жауапкершілігіне артып қоюға болмайды. Әрбір ата-ана да бала оқыту процесіне қатысып, оның жетілуіне барынша жәрдем беріп отырғаны абзал. Ата-ана өз перзентінің зерек, оқу үлгерімі өте жақсы болғанын қалайды. Сол үшін мектептен тыс уақытта балаға қолдау білдіріп, үйде алған білімін пысықтауға жан-жақты жағдай жасауы керек. Мұғаліммен де тұрақты байланыс орнатып, оқушының сабаққа деген қызығушылығы мен ынтасын әркез оятып отыру маңызды. «Балаға күштеп білім беруден гөрі, баланың білімге деген құштарлығын ояту ең маңызды мақсат» деген К.Ушинский. Тапсырманы өз бетінше орындаудан қиналған балаға міндетті түрде жәрдем беріп, жол көрсетіп жіберген жөн. Себебі, көпшілік ішінде, яғни сыныптастарының алдында тақырыпты түсінбегенін мойындау, мұғалімнен қайталап сұрау кейбір оқушылар үшін қиынға соғуы мүмкін. Сондай сәттерде ол өзінің қамқоршысы ата-анасынан еш қысылмай көмек сұрай алуы қажет. Әлбетте, ата-ана да бұл жағдайда балаға ұрсып, зекімей, алдымен бұл қиындыққа қалай тап болғанын сұрап, іш тартып сөйлескені құба құп. Баланың көңіл күйі жайланғанда ғана тапсырманы бірге орындауға кірісіп, оның әр жетістігін мақтап-мадақтау арқылы ынтасын арттыра аламыз. Білім ордасы мен ата-ана арасындағы алтын көпірді орнатып, ынтымақтастықта жұмыс атқаруды ұйымдастыра алатын бірден бір тұлға – сынып жетекшісі. Әр ата-ананың баламен қарым-қатынасы қандай деңгейде екенін анықтап, оны жетілдіре түсуге жәрдем бере алады. «Тәрбие – тал бесіктен» деп дәріптеген атам қазақ баласының дамуы мен мықты азамат болуындағы жауапкершілікті отбасындағы құндылықтарға тіке байланыстырған. Содан кейінгі жауапкершілікті көрген ортасы, алдынан өткен ұстазына қалдырған. Қорыта айтқанда, мектепті ешқашан қоғамнан бөліп қарауға, баланың білімді әрі жарқын болашағына қатысты міндеттемелерді тек білім беру мекемесіне артып қоюға болмайды. Балам қатарынан озық, ұлтының қалаулысы болсын десеңіз ата-ана да оған 50% өз үлесін қосуы керек деп санаймын. «Тәрбиесіз берілген білім адамзаттың қас жауы» деп Әл-Фараби айтқандай, білім мен жан тәрбиесі ұштасқанда ғана біз озық ойлы ұрпақ өсіре аламыз.

1-кесте –Лексикалық және семантикалық ерекшеліктер бойынша салыстыру

Көрсеткіш	Жасанды интеллектпен жазылған эссе	Табиғи тілдегі мәтін
Лексикалық қор	Нақты, жинақы және академиялық мәнмәтінге бағытталған сөздер мен тіркестер (мысалы, «ата-аналардың рөлі», «қолдау», «білім беру процесі»). Бұл сөздер ғылыми және жүйелі ақпарат ұсынуға арналған.	Әдеби-мәдени және идиомалық сөз тіркестері кеңінен қолданылған («мектеп қабырғасы», «алтын көпір», «тәрбие – тал бесіктен»). Бұл тіркестер мәдени және тарихи контексті терең ашады, оқырманның эмоциясын қозғайды және ұлттық құндылықтарды бейнелейді.

Көрсеткіш	Жасанды интеллектпен жазылған эссе	Табиғи тілдегі мәтін
Семантикалық бағыт	Негізгі идея – ата-аналардың баланың білім алу процесіндегі қолдауының маңыздылығын көрсету. Мәтін формальды стильде жазылып, логикалық және нақты деректерге негізделген. Семантикалық өрісі тар, тек басты ұғымдарға (қолдау, білім, ата-ана рөлі) шоғырланған.	Ата-аналардың, ұстаздардың және жалпы білім беру процесінің әлеуметтік, эмоционалдық және мәдени-идеологиялық аспектілері жан-жақты қарастырылады. Мәтін оқырманға оқиға, жеке тәжірибе мен ұлттық дәстүрлерге негізделген терең ой салуға мүмкіндік береді.
Лексикалық әртүрлілік	Қарапайым және түсінікті сөздер жиі қолданылған, күрделі әдеби немесе идиомалық конструкциялар сирек кездеседі. Сөздер мен тіркестердің әртүрлілігі салыстырмалы түрде шектеулі, академиялық стильге сәйкес келеді.	Күрделі құрылымдар, метафоралар, нақыл сөздер және мақал-мәтелдер көптеп кездеседі. Бұл лексикалық әртүрлілік мәтіннің байлығын арттырып, көркемдік және эмоционалдық әсерін күшейтеді, оқырманның ұлттық сана-сезіміне және мәдени мұраға тікелей әсер етеді.

ЖИ-дың тілі негізінен нақты, бейтарап, логикалық құрылымға негізделген, ал адам тілі эмоционалды және экспрессивті элементтерді жиі қолданады.

2-кесте –Синтаксистік құрылым мен стилистикалық өзгешелікті айқындау

Көрсеткіш	Жасанды интеллект мәтіні	Табиғи тілдегі мәтін
Сөйлем құрылымы	Негізінен қарапайым, қысқа сөйлемдерден тұрады; Сөйлемдер логикалық байланыспен, дәлелді және түсінікті құрылымда ұсынылған; Әр сөйлем өз алдына айқын мағына береді.	Көпқабатты, күрделі синтаксистік құрылымдарға ие; Жанама сөйлемдер және қосымша ақпарат беретін сөйлемдер жиі кездеседі; Толыққанды мәтін терең ойлануға әсер етеді, мәтін күрделі синтаксистік тұтастыққа ие.
Стиль	Ғылыми стильде жазылған; Дәлелді аргументацияға ие, объективті ақпарат беру басым. Эмоционалдық реңктер мен көркемдік фигуралар сирек кездеседі.	Публицистикалық стильде жазылған; Эмоционалдық және экспрессивті элементтерге ие; Риторикалық сұрақтар, нақыл сөздер, мақал-мәтелдер арқылы байытылған. Оқырманның жеке тәжірибесіне және ұлттық дәстүрлерге қатысты ақпарат беріп, эмоционалдық байланыс орнатады.

ЖИ сөйлемдері көбіне қарапайым, бірақ ұзын болуы мүмкін. Табиғи тілде күрделі құрылымдар мен бағыныңқылы сөйлемдер жиі кездеседі. Жасанды интеллектпен генерацияланған мәтіндердегі синтаксистік құрылым бірқалыпты болып көрініп, күрелілік деңгейі табиғи мәтіндерге қарағанда төмен болады. Бұл айырмашылықтар ЖИ архитектурасының локалдық байланыстарды басым ұстайтындығымен және оқыту деректерінің ерекшеліктерімен түсіндіріледі. ЖИ генерациялаған мәтіндерде логикалық байланыстар әлсіздеу болып шықты. Табиғи мәтіндерде себеп-салдарлық байланыстар айқынырақ көрінеді.

М.Серғалиев өзінің зерттеуінде «Стильдік белгі – белгілі бір функциональдық стильдің даралығы, басқа стильден өзгешелігін танытатын қасиеті» деп айтады.[9,39] Демек, әр стильді

тану үшін өзіндік стильдік белгі қажет. Жасанды интеллект мәтіні өзіндік бір стильдік белгілерге ие және ол белгілер ғылыми стильдің көрсеткіштерін танытып отырады.

Р. Сыздықова өз еңбегінде: «Етістіктің ауыспалы шақ деп аталатын а+ды қосымшалы тұлғасы мезгілге бейтарап, яғни әрдайым бола беретін іс-әрекетті, қимыл қозғалысты білдіретіндіктен, бұл да ғылыми стильге тән болып келеді» деп тұжырымдайды [10,204] ЖИ ұсынған мәтіндердің деншілігі ғылыми стильді ұстанады. Басқа функционалдық стильдердің формасы кездеспейді. Табиғи тіл мәтінінде а+ды формасынан бөлек, -ған, жіктік тұлғалары (жекеше және көпше түрі), модаль сөздер, есім сөздер аралас түрде қолданылған.

Жасанды интеллект тілі табиғи тілден өзгешілігі анықталды, генерациялау кезінде тілдің «түсіну», «сезіну» сынды мүмкіндіктері жасанды интеллектіде көрініс таппайды. Ол өзінің алгоритмдік жүйесі негізінде жұмыс жасайды. Адамның сөзді шығару процесі күрделі сипатқа ие, табиғи тілдің логикалық мазмұны мен ЖИ тілінің логикалық құрылымында сыртқы формасы жағынан ұқсастық болғанымен, ішкі мазмұнды беру тұрғысынан ерекшеліктер айқындалады. Зерттеу объектісі ретінде жасанды интеллектіні алып, семантикалық, прагматикалық тұрғыда мәтіндерді салыстыру, лингвистикалық талдау жұмыстары ұсынылады. ЖИ ұсынған мәтіндерге лингвистикалық болжау жасау маңызды, өйткені алдағы уақытта ЖИ-дің жетілген жұмыстары мен бастапқы жұмыстарын салыстыруға болады. Мақаланың практикалық маңыздылығы табиғи тіл мен жасанды интеллект тілін ажыратуға көмектеседі, тілдік ерекшеліктерді талдау негізінде айырмашылықтардың танылуына мүмкіндік береді.

ТАЛҚЫЛАУ

Жасанды интеллект (ЖИ) генерациялаған мәтіндердің лексикалық ерекшеліктері бойынша зерттеулер көрсеткендей, ЖИ тілі негізінен ақпараттық және формальды сипатқа ие, ал табиғи мәтіндерде көркемдік құралдар, эмоциялық және экспрессивті элементтер жиі кездеседі. Мысалы, Холцман және әріптестерінің зерттеулерінде нейрондық тіл үлгілері жиі қауіпсіз, стандартты сөздерді қолданатыны анықталды, себебі олар ықтималдық үлгілеріне негізделген [11, 275]. Бұл модельдер оқыту кезінде ең жиі кездесетін, жалпылама және бейтарап сөз тіркестерін үйренеді, осылайша эмоционалды сөздердің қолданылуы айтарлықтай төмен болады. Эмоционалды сөздер мен метафораларды қолдану ЖИ үшін қиын, себебі олар субъективті тәжірибеге, әлеуметтік-мәдени контекстке және когнитивтік бейнелеуге байланысты. Лакофф пен Джонсонның теориясы бойынша, метафоралар біздің әлемді түсінуіміздің терең когнитивтік құрылымын бейнелейді, [12, 55] ал ЖИ жүйелері тек статистикалық үлгілерді үйренеді. Сондықтан олар табиғи тілде кездесетін шығармашылық, контекстуалдық және эмоционалды мән-мағынаға ие метафораларды қалыптастыруда шектеулерге тап болады. Сонымен қатар, Чартэрис Блэк еңбегінде нейрондық модельдердің контекстуалдық нақтылықты және бай лексикалық вариативтілікті жеткізу қабілеті шектеулі екені атап көрсетіледі. [13, 250] Осыған байланысты, ЖИ мәтіндері техникалық терминдер мен нақты, логикалық құрылымға негізделсе, табиғи мәтіндер эмоционалды, экспрессивті және мәдени-идеологиялық ерекшеліктерімен ерекшеленеді.

Жасанды интеллект (ЖИ) генерациялаған мәтіндердің синтаксистік құрылымын талдаған зерттеулер көрсеткендей, мұндай мәтіндерде қысқа, қарапайым сөйлемдер басым, ал күрделі және бағыныңқы сөйлемдер салыстырмалы түрде аз кездеседі. Мысалы, «Neural machine translation» зерттеу еңбегінде нейрондық тіл үлгілерінің негізі локалдық контекстке сүйенетіні анықталған. [14, 281] Бұл үлгілер келесі сөзді болжау алгоритміне негізделгендіктен, олар ұзақ және көпқабатты құрылымдарды ұстап тұруда қиындыққа тап болады. Сонымен қатар, Сатскевер [15, 3105] зерттеуінде көрсетілгендей, реттік модельдер (sequence-to-sequence models) қысқа және қарапайым сөйлемдерді қалыптастыруда тиімді болса да, синтаксистік құрылымның әртүрлілігін жеткілікті дәрежеде қамтамасыз ете алмайды. ЖИ жүйелері оқыту деректерінен статистикалық үлгілерді үйренгендіктен, олар

белгілі бір құрылымдарды қайталамақ, ал табиғи тілде көрініс табатын күрделі синтаксистік конструкциялар мен бағыныңқы сөйлемдер, контекстуалдық нақтылықты қамтамасыз етуде әлдеқайда икемді болады.

ҚОРЫТЫНДЫ

Қорыта келе, жасанды интеллект жүйелері табиғи тілдің синтаксистік құрылымдарын дәл имитациялай алатын болса да, олардың лексикалық әртүрлілігі, семантикалық тереңдігі және дискурстық байланыстары адам тіліне тән экспрессивтілік пен мәдени байлықты жеткізуде шектеулі болып қала береді. Осыған байланысты, болашақта ЖИ технологияларын жетілдіру барысында тек статистикалық және деректік үлгілерге сүйенуден гөрі, философиялық, когнитивтік және лингвистикалық аспектілерді де интеграциялау қажет екендігі анықталды. Бұл көпсалалы тәсіл жасанды интеллектінің табиғи тілге жақындатылуына, сонымен қатар эмоционалдық және мәдени контекстерді толық әрі дәл көрсетуіне мүмкіндік туғызады, осылайша адам мен машина арасындағы тілдік қарым-қатынастың сапасын арттыруға септігін тигізбек.

ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Turing A. Computing Machinery and Intelligence.// Mind, New series. – 1950. – Vol 59 №236, pp. 433-460.
- 2 Барышников П. Н. Компьютерные метафоры в теориях сознания // Философские проблемы информационных технологий и киберпространства. [Электрон. ресурс]. – 2011. – URL:<https://cyberleninka.ru/article/n/kompyuternye-metafor-y-v-teoriyah-soznaniya/viewer> (қол жеткізу күні: 01.04.2025)
- 3 Gasparyan D. Artificial Intelligence and Semantics through the Prism of Structural, Post-Structural and Transcendental Approaches. //Integrative Psychological & Behavior Science. – 2016. – Vol 50.4. pp. 704-743.
- 4 Albertazzi L. Meaning and cognition. – Amsterdam: «Benjamins», 2000. – 269 p.
- 5 Kempen G.. Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics. – Boston, 1987. – 466 p.
- 6 Ладов В. А. Проблема зомби в контексте философии искусственного интеллекта // Гуманитарная информатика.[Электрон. ресурс]. – 2018. –URL: <https://cyberleninka.ru/article/n/problema-zombi-v-kontekste-filosofii-iskusstvenno-go-intellekta/viewer> (қол жеткізу күні: 01.04.2025).
- 7 Ruder S. et al. Transfer learning in natural language processing //Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials, Minneapolis, Minnesota, 2019. – Minneapolis, 2019. pp. 15-18.
- 8 Arnett C., Bergen B. K. Why do language models perform worse for morphologically complex languages? [Электрон.ресурс]. – 2024.–URL:<https://arxiv.org/abs/2411.14198> (дата обращения:02.04.2025)
- 9 Серғалиев Р. Стилистика негіздері. – Астана, 2006. – 273 б.
- 10 Сыздықова Р. Сөз құдіреті. – Алматы : «Атамұра», 2005. – 272 б.
- 11 Holtzman A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2019). The Curious Case of Neural Text Degeneration.[Электрон.ресурс]. – 2020. –URL:<https://arxiv.org/abs/1904.09751> (қол жеткізу күні:02.04.2025)
- 12 Lakoff G. , Johnson, M. Metaphors We Live By.London: «University of Chicago Press», 2003. – 386 pp.
- 13 Charteris-Black J. Corpus approaches to critical metaphor analysis. – London : «Palgrave Macmillan UK», 2004. – 243 pp.
- 14 Koehn P. Neural Machine Translation. –Cambridge University Press, 2020. – 394 pp.

15 Sutskever I., Vinyals O., and Le Q. V. . Sequence to Sequence Learning with Neural Networks. // Proceedings of the 27th International Conference on Neural Information Processing Systems, Massachusetts, MIT Press, 2014. pp 3104-3112.

Материал 02.06.2025 баспаға түсті

Язык искусственного интеллекта: новые горизонты лингвистики

М.К. Шенгелбай¹, Д.А. Алкебаева¹, Ж.С.Таласпаева²

¹Казахский национальный университет имени Аль-Фараби, Алматы, 050000, Казахстан

²Северо-Казахстанский университет имени М.Козыбаева, Петропавл, 150000, Казахстан

В данной статье всесторонне обсуждаются различия между языковыми возможностями искусственного интеллекта и культурными и когнитивными особенностями естественного языка, а также изучаются достижения и ограничения современных систем искусственного интеллекта в моделировании человеческого языка. Актуальность статьи неразрывно связана с широким применением технологий ИИ в современную информационную эпоху, прогрессом в области перевода и генерации текста, поскольку продукты этих систем, хотя и последовательно передают внешнюю форму языка, проявляют слабость в передаче национальной культуры и эмоционального содержания. Цель исследования-выявить особенности на лексическом, семантическом, синтаксическом и дискурсивном уровнях и оценить, насколько близко системы ИИ подходят к реальному языковому опыту, сравнивая тексты, созданные искусственным интеллектом, и природные произведения, характерные для человеческого языка. В качестве методов исследования использовались корпусная лингвистика, лексический и синтаксический анализ, а также дискурсивный анализ; использовались такие инструменты, как AntConc, Sketch Engine, через корпус, состоящий из текстов и научных статей, созданных из GPT и других образцов, литературных произведений. Полученные результаты показали, что, хотя тексты ИИ имеют четкую, логическую и формальную структуру, было замечено, что в них недостаточно выразительных элементов, характерных для художественных, идиоматических и национальных традиций, а тексты, написанные на естественном языке, имеют богатое лексическое разнообразие, сложную синтаксическую структуру и эмоционально-идеологический контекст. В этом смысле статья подчеркивает необходимость развития языковых явлений систем ИИ не только на формальном уровне, но и с учетом глубокого смыслового и культурного контекста, а также важность интеграции философских и когнитивных аспектов, доказывая эффективность использования многопрофильных подходов в совершенствовании моделей ИИ языка в будущем.

Ключевые слова: искусственный интеллект, модель языка, прагматика, естественный язык, текст

СПИСОК ЛИТЕРАТУРЫ

- 1 Turing A. Computing Machinery and Intelligence.// Mind, New series. – 1950. – Vol 59 №236, pp. 433-460.
- 2 Барышников П. Н. Компьютерные метафоры в теориях сознания // Философские проблемы информационных технологий и киберпространства. [Электрон. ресурс]. – 2011. – URL:<https://cyberleninka.ru/article/n/kompyuternye-metafor-y-v-teoriyah-soznaniya/viewer> (дата обращения: 01.04.2025)
- 3 Gasparyan D. Artificial Intelligence and Semantics through the Prism of Structural, Post-Structural and Transcendental Approaches. //Integrative Psychological & Behavior Science. – 2016. – Vol 50.4. pp. – 704-743.
- 4 Albertazzi L. Meaning and cognition. – Amsterdam: «Benjamins», 2000. – 269 p.

- 5 Kempen G.. Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics. – Boston, 1987. – 466 p.
- 6 Ладов В. А. Проблема зомби в контексте философии искусственного интеллекта // Гуманитарная информатика.[Электрон.ресурс]. – 2018.
–URL: <https://cyberleninka.ru/article/n/problema-zombi-v-kontekste-filosofii-iskusstvenno-go-intellekta/viewer> (дата обращения: 01.04.2025).
- 7 Ruder S. et al. Transfer learning in natural language processing //Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials, Minneapolis, Minnesota,2019. – Minneapolis, 2019. – pp. – 15-18.
- 8 Arnett C., Bergen B. K. Why do language models perform worse for morphologically complex languages? [Электрон.ресурс]. – 2024. –URL:<https://arxiv.org/abs/2411.14198>(дата обращения:02.04.2025)
- 9 Сергалиев Р. Основы стилистики. – Астана, 2006. – 273 с.
- 10 Сыздыкова Р. Сила речи. – Алматы : «Атамұра», 2005. – 272с.
- 11 Holtzman A., Buys J., Du L., Forbes M., & Choi Y. (2019). The Curious Case of Neural Text Degeneration.[Электрон.ресурс]. –2020. – URL:<https://arxiv.org/abs/1904.09751> (дата обращения: 02.04.2025)
- 12 Lakoff G., Johnson M. *Metaphors We Live By*. London: «University of Chicago Press», 2003. – 386 pp.
- 13 Charteris-Black J. Corpus approaches to critical metaphor analysis. – London: «Palgrave Macmillan UK», 2004. – 243 pp.
- 14 Koehn P. Neural Machine Translation. – Cambridge University Press, 2020. – 394 pp.
- 15 Sutskever I., Vinyals O., and Le Q. V. . *Sequence to Sequence Learning with Neural Networks*. // Proceedings of the 27th International Conference on Neural Information Processing Systems, Massachusetts, MIT Press, 2014. pp 3104–3112

Материал поступил в редакцию журнала 02.06.2025

Artificial Intelligence Language: New horizons of linguistics

M. Shengelbay¹, D. Alkebaeva¹, Zh. Talaspaeva²

¹Al-Farabi Kazakh National University, Almaty, 050000, Kazakhstan

²M.Kozybaev North Kazakhstan University, Petropavl,150000, Kazakhstan

This article comprehensively discusses the differences between the linguistic capabilities of artificial intelligence and the cultural and cognitive features of natural language, as well as examines the achievements and limitations of modern artificial intelligence systems in modeling human language. The relevance of the article is inextricably linked to the widespread use of AI technologies in the modern information age, progress in translation and text generation, since the products of these systems, although consistently conveying the external form of language, show weakness in conveying national culture and emotional content. The aim of the study is to identify features at the lexical, semantic, syntactic, and discursive levels and assess how closely AI systems approach real language experience by comparing texts created by artificial intelligence and natural works characteristic of human language. The research methods used were corpus linguistics, lexical and syntactic analysis, as well as discursive analysis.; Tools such as AntConc and the Sketch Engine were used through a corpus consisting of texts and scientific articles created from GPT and other samples of literary works. The results showed that although AI texts have a clear, logical and formal structure, it was noticed that they lack expressive elements characteristic of artistic, idiomatic and national traditions, and texts written in natural language have a rich lexical diversity, complex syntactic structure and emotional-ideological context. In this sense, the article emphasizes the need to develop linguistic phenomena of AI systems not only at the formal level, but also taking into account the deep semantic and cultural context, as well as the importance of integrating philosophical and cognitive

aspects, proving the effectiveness of using multidisciplinary approaches in improving AI language models in the future.

Key words: artificial intelligence, language model, pragmatics, natural language, text

REFERENCES

- 1 Albertazzi L. (ed) (2000), *Meaning and cognition*. Amsterdam.
- 2 Arnett C., Bergen B. K (2024). "Why do language models perform worse for morphologically complex languages?" available at: <https://arxiv.org/abs/2411.14198>
- 3 Baryshnikov P. N.(2011) "*Komp'yuternye metafory v teoriyah soznaniya*", Filosofskie problemy informacionnyh tekhnologij i kiberprostranstva [Computer metaphors in theories of consciousness, Philosophical problems of information technology and cyberspace], available at: <https://cyberleninka.ru/article/n/kompyuternye-metafory-v-teoriyah-soznaniya/viewer>
- 4 Charteris-Black J.(2004) *Corpus approaches to critical metaphor analysis*. London : «Palgrave Macmillan UK».
- 5 Gasparyan D. (2016) "*Artificial Intelligence and Semantics through the Prism of Structural, Post-Structural and Transcendental Approaches*", Integrative Psychological & Behavior Science, vol.50 no. 4. pp.704-743.
- 6 Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). "*The Curious Case of Neural Text Degeneration*", available at: <https://arxiv.org/abs/1904.09751>
- 7 Kempen G.. (1987) *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics*. Boston.
- 8 Koehn, P. (2020) *Neural Machine Translation*. Cambridge University Press.
- 9 Ladov V. A.(2018) "*Problema zombi v kontekste filosofii iskusstvennogo intellekta*", Gumanitarnaya informatika.[The zombie problem in the context of the philosophy of artificial intelligence, Humanitarian Informatics], available at: <https://cyberleninka.ru/article/n/problema-zombi-v-kontekste-filosofii-iskusstvennogo-intellekta/viewer>
- 10 Lakoff, G. and Johnson, M. (2003) *Metaphors We Live By*. London
- 11 Ruder S. (ed) (2019), "*Transfer learning in natural language processing*", Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials, Minnesota,. Minneapolis, 2019 pp.–15– 18.
- 12 Serġaliev R. (2006) *Stilistika negızderi*. [Basics of stylistics]. Astana.
- 13 Syzdyqova R. (2005) *Söz qüdireti* [Power of speech]. Almaty.
- 14 Sutskever, I., Vinyals, O., and Le, Q. V.(2014) . "*Sequence to Sequence Learning with Neural Networks*", Proceedings of the 27th International Conference on Neural Information Processing Systems, Massachusetts, 2014 pp. 3104–3112
- 15 Turing, A.(1950) *Computing Machinery and Intelligence*. Mind, New series, vol. 59 no. 236, pp 433-460.

Received: 02.06.2025