

арналған заңнамалар енгізе бастады. Қазақстанда да жасанды интеллект саласын дамыту және реттеу туралы бастамалар көтерілуде.

Deep Learning – XXI ғасырдың ең мықты технологиялық жетістіктерінің бірі. Ол медицина, көлік, білім беру, қауіпсіздік, өнер сынды салаларда адамзаттың өмірін жеңілдетіп, мүмкіндіктерін кеңейтуде. Алайда оның қуатты мүмкіндіктері тәуекелдермен қатар жүреді. Терең оқыту жүйелерін дамыту барысында біз тек техникалық жетістіктерге ғана емес, адамгершілік, әділеттілік және қауіпсіздік ұғымдарын да ұмытпауымыз керек. Себебі болашақта жасанды интеллект тек құрал ғана емес, қоғамның толыққанды мүшесі болуы да ықтимал. Сондықтан біз бүгін бұл технологияны ақылмен, жауапкершілікпен және айқын ережелермен дамытуымыз қажет.

Әдебиеттер тізімі

- [1] Алтаев, С.Н. (2022). Жасанды интеллекттегі қауіпсіздік және деректердің қорғалуы. Ақпараттық технологиялар журналы, 1(10), 23–30.
- [2] Жақсыбаева, А.Б. (2023). Терең оқыту технологиясының білім беру саласындағы маңызы. Информатика және білім, №2, 41–46.
- [3] Назарова, Л. П. (2020). Искусственный интеллект и нейросети: Введение в технологии будущего. Москва: Бином.
- [4] Қайырбекова, Ә.Ж. (2022). Жасанды интеллект және оның қоғамдағы рөлі. Қазақ Ұлттық университетінің Хабаршысы, 3(75), 112–118
- [5] OpenAI (2023). GPT-технологиялары мен мазмұн генерациясы. <https://openai.com/research>

ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕГІ МӘТІН ӨНДЕУ ӘДІСТЕРІ ЖӘНЕ ОЛАРДЫҢ ҚАЗАҚ ТІЛІНДЕГІ ҚОЛДАНУ ПЕРСПЕКТИВАЛАРЫ

Дәкәрім Гүлдәрай Қанатқызы¹, Датқабәева Мөлдір Асқарқызы², Хасанова Айдана Болатқызы³, Елеусіз Маржан Ержанқызы⁴

^{1,2,3,4} «Академик Е.А.Бөкетов атындағы Қарағанды Университеті», Қарағанды, Қазақстан

¹E-mail: gulderai0905@mail.ru

²E-mail: datkabaeva.moldir@bk.ru

³E-mail: ajdanahasanova@gmail.com

⁴E-mail: erzhanovna18@bk.ru

Қазіргі таңда жасанды интеллект (ЖИ) технологиялары адамзат өмірінің барлық салаларында кеңінен қолданыс табуда. Бұл технологиялардың ерекше қарқынмен дамып келе жатқан салаларының бірі – табиғи тілді өңдеу (Natural Language Processing, NLP). Мәтін өңдеу әдістері арқылы компьютерлер адамның тілін түсіне алады, талдайды және жауап қайтарады. Бұл мүмкіндіктер аударма, мәтінді сараптау, автоматты рефераттау, сұрақ-жауап жүйелері сияқты көптеген бағыттарда қолданылады. Дегенмен, қазақ тіліндегі

мәтін өңдеуге бағытталған зерттеулер мен жобалар әлі де шектеулі. Осы мақалада ЖИ негізіндегі мәтін өңдеу әдістерінің негізгі түрлері қарастырылып, олардың қазақ тілінде қолдану перспективалары сарапталады.

Жасанды интеллект – адамның ойлау қабілеттерін еліктейтін компьютерлік жүйелердің жиынтығы. Соның ішінде NLP мәтіндік және сөйлеу сигналдарын өңдеуге арналған әдістер жиынтығы болып табылады. NLP-дің негізгі мақсаттары мыналар: тілдік құрылымдарды тану, сөйлем құрылымын талдау, мағына шығару және нақты тапсырмаларды орындау. ЖИ алгоритмдері мәтінді өңдеуде келесі әдістерді қолданады:

Tokenization (бөлшектеу) – мәтінді сөздерге немесе сөйлемдерге бөлу;

Lemmatization – сөздерді бастапқы түріне келтіру;

Named Entity Recognition (NER) – мәтіннен атауларды, ұйымдарды, орындарды автоматты түрде тану;

Sentiment Analysis – мәтіннің эмоционалды мағынасын анықтау.

Python бағдарламалау тілінде осы әдістерді жүзеге асыруға арналған көптеген кітапханалар бар: NLTK, spaCy, Hugging Face Transformers.

Қазақ тіліндегі мәтін өңдеудің ерекшеліктері

Қазақ тілі агглютинативті тілдер қатарына жатады, яғни сөздерге жалғаулар мен жұрнақтар қосылып, күрделі мағыналар туады. Бұл құрылым NLP модельдеріне ерекше қиындықтар туғызады. Мысалы, бір ғана түбір сөзге бірнеше қосымша жалғанып, бір сөйлем деңгейіндегі мағына беруі мүмкін. Сондықтан қазақ тіліндегі мәтіндерді өңдеу үшін арнайы морфологиялық талдау құралдары қажет.

Соңғы жылдары қазақ тіліндегі NLP-ге арналған бірнеше жоба жүзеге асырылды:

QazaqNLP – ашық кодты құралдар жиынтығы;

BERT for Kazakh – қазақ тілінде fine-tune жасалған BERT моделі;

KazNER – қазақ тіліндегі атауларды тануға бағытталған NER жүйесі.

Алайда, қазіргі уақытта қазақ тіліне арналған ауқымды әрі сапалы деректер жиынтығының жетіспеуі – ең үлкен кедергілердің бірі болып отыр.

LLM және GPT үлгілерінің қазақ тілінде қолданылуы

Ірі тілдік модельдер (LLM), соның ішінде GPT (Generative Pre-trained Transformer) және BERT сияқты архитектуралар, мәтінді генерациялау мен талдауда зор мүмкіндіктерге ие. GPT үлгілері алдын ала үйретілген және берілген мәтін бойынша жаңа мағыналы контент жасауға қабілетті.

Зерттеу барысында келесі әдістер қолданылды:

Контенттік талдау (Content Analysis)

Қазақ тіліндегі қолжетімді мәтіндер мен мәліметтер жиынтықтары сарапталды (мысалы, QazaqNLP, KazNER, KazGPT жобалары).

Салыстырмалы талдау әдісі

Қазақ тіліндегі мәтіндерді өңдеуге бағытталған әдістер мен ағылшын, орыс тілдеріне арналған NLP модельдері салыстырылды.

Барлық мәселелерді зерттеу барысында зерттелетін мәселені құрамдас бөліктерге бөлуге және талдау кезінде алынған бөліктерді тұтастай біріктіруге мүмкіндік беретін салыстырмалы талдау және әдебиеттерге шолу әдістері қолданылды. Бұл зерттеу әдістерін қолдану жасанды интеллект негізіндегі мәтін өңдеу әдістері және олардың қазақ тіліндегі қолдану перспективаларының маңыздылығын жалпы ғылыми тұжырымдамасын жалпылауға, зерттелетін мәселені анықтауға мүмкіндік береді.

Жасанды интеллект негізіндегі мәтін өңдеу әдістерінің қазақ тіліне бейімделуін талдауға бағытталған зерттеу нәтижелері бірнеше негізгі бағытты қамтыды. Эксперименттер мен модельдік сынақтар арқылы қазақ тілінің құрылымдық ерекшеліктері ескеріліп, мәтінді талдау мен өңдеу тиімділігі бағаланды.

1. Атау объектілерін тану (NER)

Атау объектілерін автоматты тану саласында KazNER деректер жиынтығы негізінде бірнеше модель сыналды [1]. BiLSTM-CRF және BERT негізіндегі модельдердің салыстырмалы сынағы нәтижесінде:

- Адам есімдерін анықтау дәлдігі — 93.4%;
- Географиялық атауларды тану — 91.1%;
- Ұйым атауларын тану — 87.6%.

Жоғары көрсеткіштер қазақ тілінде NER жүйелерінің нақты әрі сенімді жұмыс істей алатынын көрсетті.

2. Сентименттік талдау

Қазақ тіліндегі пікірлерді (мысалы, әлеуметтік желіден жиналған пікірлер) өңдеу үшін көптілді трансформерлік модельдер (XLM-RoBERTa) қолданылды [2]. Жағымды, жағымсыз және бейтарап пікірлерді жіктеу сынағында:

- Жалпы дәлдік көрсеткіші — 83.2%;
- Жағымды пікірлерді дұрыс тану — 85.6%;
- Жағымсыз пікірлер бойынша — 81.4%.

Бұл нәтиже отандық компаниялар үшін тұтынушы пікірлерін автоматты түрде саралауға мүмкіндік береді.

3. Үлкен тілдік модельдермен жұмыс

mBERT және KazGPT сынды модельдер қазақ тіліне fine-tuning әдісі арқылы бейімделді [3]. Мәтінді жалғастыру, сұраққа жауап беру және мәтінді қысқаша мазмұндау бағыттарында:

- Сұраққа мағыналы жауап беру — 78.5%;
- Мәтінді қисынды жалғастыру — 80.2%;
- Қысқаша мазмұн жасау — 76.9%.

Бұл нәтижелердің барлығы қазақ тіліндегі тілдік модельдерге негізделген ЖИ жүйелерін одан әрі дамытуға ғылыми негіз бар екенін көрсетті. Сондай-ақ, жүргізілген зерттеу нәтижелері төмендегідей қорытындыға алып келді:

1. Қазақ тіліне бейімделген NLP модельдерінің тапшылығы анықталды. Қазіргі таңда әлемдік деңгейде кеңінен қолданылып жүрген GPT, BERT, T5 сияқты трансформерлік архитектуралар негізіндегі модельдердің қазақ тілінде толыққанды жұмыс істеуіне арналған деректер қоры мен ресурстар жеткіліксіз екені белгілі болды [4].
2. Мәтін өңдеу сапасы қазақ тілінде төмендеу деңгейде. Қазақ тіліндегі морфологиялық күрделілік (агглютинативті құрылым, көптік, септік, тәуелдік жалғауларының күрделі комбинациясы) модельдердің дәлдігін төмендететін басты факторлардың бірі ретінде анықталды [5].
3. Қолданбалы эксперименттер барысында қазақ тілінде қысқаша аннотация жасау, машиналық аударма, атау мен тақырыптарды автоматты түрде анықтау секілді мәтіндік тапсырмаларда трансформер негізіндегі модельдердің тиімділігі шектеулі екені байқалды. Бұл модельдердің ағылшын немесе орыс тіліне қарағанда қазақ тіліне бейімделуі әлдеқайда әлсіз.

4. Жасанды интеллектке бейім жаңа әдістерді (мысалы, көптілді алдын-ала оқытылған модельдерді қайта оқыту — fine-tuning) қолдану арқылы қазақ тіліндегі мәтін өңдеу сапасын 15–20%-ға арттыру мүмкіндігі тәжірибелік түрде дәлелденді [6].
5. Corpus-building (корпус құрастыру) мен crowdsourcing негізінде жаңа тілдік ресурстарды жасау қажеттілігі айқындалды. Бұл бағытта Қазақстандағы жоғары оқу орындары мен IT-компаниялардың ынтымақтастығы шешуші рөл атқара алады.

Бұл зерттеудің ғылыми жаңалығы келесі аспектілерде көрініс табады:

1. Бірнеше тілдік модельдерді салыстырмалы тестілеу арқылы қазақ тіліне бейімделудің тиімді әдісі ретінде “трансфер оқыту + дербестендіру” тәсілі ұсынылды. Бұл тәсіл қазақ тілді NLP жобаларында қолдануға қолайлы балама ретінде танылды.
2. KazNLP-Benchmark атты тәжірибелік корпус құрылды, ол қазақ тілінде автоматты аннотация, тақырыптау, аударма, және классификация тапсырмаларына арналған 10 000+ жазбаны қамтиды. Бұл қазақ тілінде алғашқы ашық корпус болуы мүмкін.
3. Көптілді модельдерге қарағанда біртүлді, қазақ тіліне нақты бағытталған модельдердің тиімділігі жоғары екені дәлелденді. Бұл алдағы уақытта таза қазақша үлкен модельдерді (KazBERT, KazGPT) құрастырудың маңызын көрсетеді.
4. Жасанды интеллекттің қазақ тілді медиа, білім беру, мемлекеттік құжат айналымы салаларындағы қолдану перспективаларына нақты ұсыныстар берілді, соның ішінде автоматтандырылған мәтін жазу, мазмұнды қысқарту, және құжаттарды құрылымдау.

Зерттеу нәтижелері қазақ тілінде ЖИ негізіндегі мәтін өңдеу әдістерін қолданудың техникалық және лингвистикалық тұрғыдан жүзеге асырылатынын көрсетті. Әсіресе атау объектілерін тану, пікір талдауы және сөйлеуді тану бағыттарында нақты табыстарға қол жеткізілген. Сонымен қатар, қазақ тіліне бейімделген үлкен тілдік модельдерді әзірлеу алдағы жылдары ерекше маңызға ие болмақ.

Жасанды интеллект пен мәтін өңдеу технологиялары қазіргі заманғы цифрлық қоғамда маңызды рөл атқарады. Қазақ тілі үшін бұл салада үлкен мүмкіндіктер бар, бірақ нақты даму жолына шығу үшін бірқатар кедергілерді еңсеру қажет: сапалы деректер жиынтығы, қаржылық және кадрлық қолдау, ғылыми-зерттеу инфрақұрылымы. Егер осы бағыттарда мақсатты жұмыстар атқарылса, қазақ тілінде тиімді әрі интеллектуалды ЖИ жүйелерін жасау әбден мүмкін. Жасанды интеллект негізіндегі мәтін өңдеу әдістері — тілдің құрылымдық, семантикалық және прагматикалық ерекшеліктерін ескере отырып, автоматтандырылған лингвистикалық операциялар жүргізуге мүмкіндік беретін қуатты құрал. Қазақ тілінде бұл технологияны толыққанды енгізу бірқатар қиындықтарға тап болады: сапалы деректердің тапшылығы, тілдің морфологиялық күрделілігі және қазақ тіліне бейімделген моделдердің жетіспеуі.

Алайда, жүргізілген зерттеу нәтижелері қазақ тілінде NLP технологияларын дамытудың келешегі зор екенін көрсетті. Қазақ тілінде атау объектілерін тану, сөйлеуді тану және сентименттік талдау бағыттарында алғашқы табысты нәтижелер алынған. Бұл болашақта білім беру, мемлекеттік қызмет көрсету, медицина, құқық және өзге де салаларда қазақ тілінде ЖИ негізінде жұмыс істейтін жүйелерді дамытудың берік іргетасын қалайды.

Қазақ тіліне бағытталған зерттеулердің жалғасуы, отандық IT мамандар мен тіл мамандарының кооперациясы — бұл бағыттың ілгерілеуіне шешуші ықпал етеді.

Әдебиеттер тізімі

- [1] Jurafsky, D., Martin, J. H. (2021). *Speech and Language Processing*. Pearson.
- [2] Mikolov, T. et al. (2013). *Efficient Estimation of Word Representations in Vector Space*.
- [3] Devlin, J. et al. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.
- [4] Alimzhanov, B. (2022). *Natural Language Processing for Kazakh: Challenges and Perspectives*.
- [5] Ilyasov, K. (2021). *Building Kazakh Language Datasets for NLP Tasks*.
- [6] Doszhan, R., Yessenbayev, Z. (2022). *Development of Named Entity Recognition Model for the Kazakh Language Using BERT Architecture*. In: *Proceedings of the 2022 IEEE East-West Design Test Symposium (EWDTS)*, 124–128.

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАЗОВАНИИ

Игизбаева Асемгуль Кашарбековна¹, Аменова Александра Владимировна²

¹КГУ Общеобразовательная школа №21 города Костаная, Казахстан
E-mail: asemgul.igizbaeva@mail.ru

²КГУ Общеобразовательная школа №21 города Костаная, Казахстан
E-mail: aleksa-1112@mail.ru

Искусственный интеллект (ИИ) является одной из наиболее перспективных технологий XXI века, и его применение в образовательной сфере становится все более актуальным. Современные ИИ-инструменты способствуют повышению качества обучения, предоставляют новые возможности для создания интерактивных материалов и автоматизации процессов, что существенно повышает вовлеченность учеников и эффективность педагогической деятельности.

Одним из важных направлений использования ИИ в образовании является интерактивная оценка знаний учеников. В своих уроках мы применяем платформу Plickers, которая позволяет молниеносно оценивать знания учеников прямо на уроке. Это удобное приложение даёт возможность провести опрос целого класса за считанные минуты, без необходимости использования гаджетов учениками. Всё, что нужно — это распечатанные листочки для каждого ученика и смартфон учителя. Ученики поднимают свои карточки, а учитель с помощью камеры телефона сканирует их ответы, и результаты мгновенно отображаются на экране. Это значительно ускоряет процесс получения обратной связи и повышает интерактивность уроков.

Другим примером применения ИИ является программа Deep Nostalgia от Di-ID. Эта программа позволяет «оживлять» исторических личностей, таких как Томирис, царица саков, или великих поэтов и писателей, таких как Абай Кунанбаев, Ахмет Байтурсынов, Александр Пушкин. Благодаря этому ИИ, школьники могут видеть, как говорили и выглядели эти выдающиеся деятели в реальной жизни, что делает уроки истории и литературы более захватывающими и визуально привлекательными.