«8D06301 – Ақпараттық қауіпсіздік жүйелері» білім беру бағдарламасы бойынша (PhD) философия дәрежесін алу үшін дайындалған Сағынай Мөлдірдің «Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған тереңдетіп оқыту үлгілерін жасау» тақырыбындағы диссертациялық жұмысына

АҢДАТПА

Зерттеудің өзектілігі. Қазіргі цифрлық ортада деструктивті веб-контентті автоматты түрде анықтау ақпараттық қауіпсіздік пен қоғамдық тұрақтылық үшін қажет ғылыми және практикалық бағытқа айналды. Интернетте таралатын кибербуллинг, өшпенділікке құрылған пікірлер, агрессиялық үндеулер және құқыққа қайшы контент тек жеке адамға зиян келтіріп қоймайды, қоғамдағы әлеуметтік қатынастарға да әсер етеді. Сол үшін мұндай контентті уақытында байқап, ажырата алатын интеллектуалды жүйелерді жасау бүгін өте қажет міндеттердің бірі. Зерттеудің өзектілігі заңнамалық талап, технологиялық әлсіз тұстар және лингвистикалық қиындықтар сияқты үш бағыттан көрінеді.

Заңнамалық қырына келсек, Қазақстан Республикасының Қылмыстық кодексінің 174-бабында әлеуметтік, ұлттық, этностық, нәсілдік немесе діни алауыздықты қоздыруға байланысты әрекеттердің құқықтық қауіптілігі нақты көрсетілген. Бұдан бөлек, «Байланыс туралы» Қазақстан Республикасының заңында байланыс және интернет кеңістігінде құқыққа қайшы ақпараттың таралуын реттейтін нормалар бар. Демек, деструктивті веб-контентті дер кезінде анықтау тек техникалық мәселе емес, ұлттық ақпараттық қауіпсіздік пен құқықтық тәртіп үшін де қажет шарттардың бірі.

Технологиялық тұрғыдан алғанда көп зерттеу жұмыстары ағылшын, орыс және т. б. ресурсттық тілдерге қатысты жүргізіледі. Қазақ тіліндегі деструктивті веб-контентке қатысты дайын классификаторлар, жеткілікті аннотацияланған корпус немесе белгілі бір міндеттерге бағытталған жүйелер жоқ. Мәтіндерді автоматтандырылған түрде талдау үшін жаңа әдістер мен модельдер құру қажет екенін білдіреді.

Лингвистикалық жақтан мәселе одан да күрделірек. Қазақ тілі агглютинативті, морфологиясы да ауыр тілдердің бірі. Сөздің мағынасы түбірге жалғанған қосымшалар арқылы өзгеріп отырады, ал сөйлемнің реңкі көбіне контекстке, семантикалық байланысқа тәуелді. Оның үстіне деструктивті веб-контент көп жағдайда тура айтылмайды: астармен беріледі, эмоциялық реңкке жасырынады, кейде жанама тілдік құрылымдармен жеткізіледі. Сондықтан қазақ тіліндегі веб-контентті талдайтын модель тек контексті емес, морфологиялық және семантикалық ерекшеліктерді де қатар ескеруі керек.

Сол себепті берілген диссертациялық жұмыс тақырыбының өзектілігі деструктивті интернет-контенттің қоғамға қауіп төндіретіндігімен, оны құқықтық тұрғыдан дер кезінде анықтау қажеттілігімен, қазақ тіліне арналған технологиялық шешімдер жеткіліксіздігімен түсіндіріледі. Аталған жағдайлар қазақ тіліндегі деструктивті веб-контентті анықтауға арналған жетілдірілген модель жасаудың ғылыми да, практикалық та мәні бар екенін көрсетеді.

Зерттеу мақсаты – қазақ тіліндегі деструктивті веб-контентті классификациялау үшін тереңдетіп оқыту модельдерін кешенді зерттеу және құру.

Зерттеудің міндеттері:

1. Әлеуметтік желі платформаларынан қазақ тіліндегі деструктивті мәтіндердің деректер жинағын құру;
2. Қазақ тіліндегі деструктивті веб-контентті классификациялау үшін мәтінді ендіру әдістері мен классикалық машиналық оқыту және терең оқыту әдістерін зерттеу;
3. Қазақ тілі үшін морфологиялық белгілер модулін әзірлеу;
4. Қазақ тіліндегі деструктивті веб-контентті классификациялауда онтологияға негізделген семантикалық белгілерді әзірлеу;
5. Қазақ тілінің ерекшеліктерін ескеретін деструктивті веб-контенттің түрлері мен белгілерін классификациялайтын гибридті архитектураны құру;
6. Әзірленген модельді эксперименталды ортада сынау.

Зерттеу объектісі: Әлеуметтік желілер мен веб-ресурстардағы қазақ тіліндегі деструктивті мәтіндік контент.

Зерттеу пәні: Қазақ тіліндегі деструктивті веб-контентті көпкласты классификациялауға арналған контекстік, морфологиялық және онтологияға негізделген семантикалық белгілерді біріктіру әдістері мен модельдері.

Зерттеу әдістері: Машиналық және терең оқыту әдістері, мәтінді жіктеу әдістері, табиғи тілді өңдеу әдістері, әлеуметтік желілерді талдау әдістері, онтологиялық жіктеу әдістері.

Диссертациялық жұмыстың ғылыми жаңалығы:

1. Қазақ тіліндегі деструктивті веб-контентті классификациялау үшін деструктивті веб-контенттің көп классты аннотацияланған корпус құрылды.

2. Қазақ тіліндегі деструктивті веб-контентті классификациялау үшін қазақ тіліндегі деструктивті веб-контенттің түрін, ашықтығын және нысанасын анықтайтын онтологияға негізделген семантикалық белгілер әзірленді.

3. Қазақ тіліндегі деструктивті веб-контентті классификациялау үшін контексттік эмбеддингтерді, қазақ тілінің морфологиялық белгілерін және онтологияға негізделген семантикалық белгілерді біріктіретін гибридті классификациялық архитектура әзірленді.

Қорғауға шығарылған негізгі тұжырымдама:

1. Қазақ тіліндегі деструктивті веб-контентті көпкласты классификациялау міндетін шешу үшін деструктивті веб-контент корпусы әзірленіп, контенттің түрін, ашықтығын және нысанасын сипаттайтын онтологияға негізделген семантикалық белгілерді KazBERT эмбеддингтерімен және морфологиялық белгілермен біріктіру әдісі ұсынылды, ұсынылған әдісті қолдану базалық KazBERT-пен салыстырғанда Accuracy көрсеткішін 88,015%-дан 90,492%-ға, ал Macro F1-score көрсеткішін 88,086%-дан 90,567%-ға дейін арттыруға мүмкіндік берді.

2. Қазақ тіліндегі деструктивті веб-контентті көпкластық классификациялауға арналған KazBERT эмбеддингтері, морфологиялық белгілер, онтологияға негізделген семантикалық белгілер және BiLSTM негізіндегі тізбектік талдау қабаты біріктірілген гибридті архитектура әзірленіп, базалық KazBERT-пен салыстырғанда Accuracy көрсеткіші 88,015%-дан 91,045%-ға, ал Macro F1-score көрсеткіші 88,086%-дан 91,103%-ға дейін арттырылып, Accuracy бойынша салыстырмалы қате классификациялау үлесі 25,28%-ға төмендетілді.

Зерттеудің теориялық маңыздылығы: қазақ тіліндегі деструктивті веб-контентті классификациялау саласының теориялық және әдіснамалық негіздерін дамытуымен сипатталады. Зерттеу барысында деструктивті веб-контенттің түрін, ашықтығын және нысанасын сипаттайтын онтологияға негізделген семантикалық белгілер жүйесі ұсынылып, олардың мәтінді жіктеудегі рөлі теориялық тұрғыдан негізделді. Сонымен қатар қазақ тілінің морфологиялық ерекшеліктері мен контекстік семантикасын біріктіретін гибридті тәсілдің ғылыми негіздері қалыптастырылды. Ұсынылған тәсіл қазақ тіліндегі табиғи тілдерді өңдеу (NLP), деструктивті контентті автоматты анықтау және семантикалық мәтіндік талдау бағыттарындағы ғылыми зерттеулерді дамытуға теориялық үлес қосады. Зерттеу нәтижелері қазақ тіліндегі контентті талдау модельдерін жетілдіруге, сондай-ақ басқа түркі тілдеріне бейімделетін интеллектуалды мәтіндік өңдеу әдістерін әзірлеуге теориялық негіз бола алады.

Зерттеудің практикалық құндылығы: модельдің қазақ тіліндегі деструктивті веб-контентті автоматты түрде анықтау және жіктеу жүйелерінде қолданылу мүмкіндігімен айқындалады. Зерттеу нәтижелері әлеуметтік желілерді, веб-платформаларды және ақпараттық жүйелерді мониторингтеу, заңсыз немесе деструктивті мазмұнды ерте анықтау, сондай-ақ контентті модерациялау үдерістерінің тиімділігін арттыру үшін пайдаланылуы мүмкін. Ұсынылған әдістер мен әзірленген деректер жиынтығы қазақ тіліндегі табиғи тілдерді өңдеу, ақпараттық қауіпсіздік және жасанды интеллект салаларындағы қолданбалы зерттеулерде әрі қарай дамытуға негіз болады. Зерттеу нәтижелері бойынша «Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған BERT негізіндегі терең оқыту

модельдерінің тиімділігін Supervised Contrastive Learning (SupCon) әдісі арқылы арттыру» тақырыбындағы компьютерлік бағдарламаға №69171 мемлекеттік тіркеу куәлігі 2026 жылғы 27 наурызда алынды.

Ізденушінің жеке үлесі. Қорғауға алынған диссертациялық жұмыстың барлық нәтижелерін докторант өзі алды. Негізгі нәтижелер қатарына: қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған деструктивті веб-контенттің түрін, ашықтығын және нысанасын сипаттайтын онтологияға негізделген семантикалық белгілер жүйесі; қазақ тілінің морфологиялық белгілерін анықтау және оларды контекстік KazBERT эмбеддингтерімен біріктіретін гибридті классификациялық архитектура; KazBERT эмбеддингтері, морфологиялық және онтологиялық семантикалық белгілерді біріктіру арқылы деструктивті веб-контентті жіктеудің жетілдірілген терең оқыту моделі; ұсынылған модельдің тиімділігін бағалауға арналған бағдарламалық қамтамасыз ету, эксперименттік зерттеулердің нәтижелері және оларды статистикалық өңдеу жатады.

Диссертацияда келтірілген алгоритмдерді әзірлеу, бағдарламалық іске асыру, модельдерді оқыту және тестілеу, гиперпараметрлерді оңтайландыру, эксперименттік зерттеулерді жүргізу, алынған нәтижелерді салыстырмалы талдау, ғылыми қорытындыларды қалыптастыру және диссертация мәтінін дайындау жұмыстарын докторант өзі орындады. Ғылыми жарияланымдар мен компьютерлік бағдарламаға мемлекеттік тіркеу куәлігін алу жұмыстары ғылыми кеңесшілермен бірлесіп жүзеге асырылды.

Диссертация тақырыбының ғылыми-зерттеу жұмыстарының жоспарларымен байланысы. Берілген зерттеу Қазақстан Республикасы Ғылым және жоғары білімминистрлігінің Ғылым комитеті қаржыландыратын “Жастар экстремизмін анықтау және заманауи ақпараттық кеңістікте жастардың қауіпсіздігін қамтамасыз етуге арналған модельдер мен әдістерді әзірлеу” жобасы аясында орындалды (грант AP19576868).

Диссертация нәтижелерін апробациялау. Ұсынылған диссертациялық жұмыстың зерттеу нәтижелері әл-Фараби атындағы Қазақ Ұлттық университеті Ақпараттық технологиялар және жасанды интеллект факультетінің, Киберқауіпсіздік және криптология кафедрасынның ғылыми семинарларында бірнеше рет тыңдалып, талқыланды. Сонымен қатар, халықаралық және отандық ғылыми конференцияларда баяндама жасалып, ғылыми қауымдастық тұрғысынан оң бағаланды.

Әрбір басылымды дайындауға докоранттың қосқан үлесі. Диссертациялық жұмыстың нәтижелері 10 ғылыми басылымда жарияланды. 1 мақала Scopus базасында индекстелетін басылымда, 3 мақала Scopus базасында индекстелетін конференцияларда, 2 мақала ҚР БҒМ Білім және ғылым саласындағы бақылау комитеті ұсынған журналдарында жарық көрді. 4 мақала халықаралық конференциялар материалдарында жарияланып, ҚР авторлық құқықпен қорғалатын объектілерге құқықтарды мемлекеттік тізілімге мәліметтерді енгізу туралы 1 куәлік алынды.

Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым және жоғары білім саласындағы сапаны қамтамасыз ету комитеті ұсынатын журналдардағы мақалалар:

1. Болатбек М.А., Сағынай М., Мусиралиева Ш.Ж. Қазақ тіліндегі деструктивті веб-контентті анықтау үшін машиналық оқыту әдістерін қолдану //Известия НАН РК. Серия физико-математическая. – 2024. – №. 4. – С. 99–111.

2. Сағынай М., Болатбек М. Қазақ тіліндегі деструктивті веб-мазмұнды жіктеудің морфологиялық және семантикалық тұрғыдан жетілдірілген моделі //Вестник КазАТК. – 2026. – Т. 144. – №. 3. – С. 151-166.

Скопуста индекстелетін журналдардағы ғылыми мақалалар:

1. Bolatbek M, Sagynay M, Mussiraliyeva S, Yeltay Z. 2025. Detection of offensive content in the Kazakh language using machine learning and deep learning approaches. PeerJ Computer Science 11:e3027 https://doi.org/10.7717/peerj-cs.3027 (Indexed in Web of Science: Q2, Indexed in SCOPUS, SJR 2024 равен 0.719, процентиль 80)

Скопуста индекстелетін конференциялардағы ғылыми мақалалар:

1. Milana Bolatbek, Moldir Sagynay, Shynar Mussiraliyeva, Yeltay Zhastay, Machine learning methods for identifying destructive texts in the Kazakh language 2024 3rd International Conference on Algorithms, Data Mining, and Information Technology (ADMIT 2024) https://doi.org/10.1145/3701100.3701121.

2. Milana Bolatbek, Moldir Sagynay, Shynar Mussiraliyeva, Kymbat Baisylbayeva, Zhastay Yeltay. Kazakh language dataset for hate speech detection on social media text, 2024 9th International Conference on Computational Intelligence and Applications (ICCIA 2024) 10.1109/ICCIA62557.2024.10719327.

3. Moldir S., BolatbekM., Mussiraliyeva S., & Baisylbayeva, K. (2025, March). Natural Language Processing and Machine Learning for Detecting Harmful Speech in Kazakh Social Media. In 2025 11th International Conference on Computing and Artificial Intelligence (ICCAI) (pp. 347-351). IEEE 10.1109/ICCAI66501.2025. 00060

Халықаралық конференция материалындағы мақалалар:

1. Сағынай М., Болатбек М. Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған терең оқыту үлгілерін жасау, Материалы международной конференции студентов и молодых учёных «Фараби әлемі»

2. Болатбек, М. Сағынай М., Ш. Мусиралиева Веб ресурстағы зиянды веб контентті анықтау мәселесі және оның өзектілігі Материалы IX Международной научно-практической конференции «Информатика и прикладная математика», посвященной 90-летию Казахского национального университета имени аль-Фараби

3. Сағынай М., Болатбек М., Қазақ тіліндегі қауіпті веб-контентті анықтау және жіктеу үшін машиналық оқыту модельдерін құру Фараби әлемі 2025, 265 бет.

4. Bolatbek M., Mussiraliyeva S., Sagynay M. Machine Learning-Based Hate Speech Detection in the Kazakh Language // Research in Language. – 2025. – Vol. 23. – P. 311–319.

Авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы куәліктер:

1. Болатбек М.А., Сағынай М. «Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған BERT негізіндегі терең оқыту модельдерінің тиімділігін Supervised Contrastive Learning (SupCon) әдісі арқылы арттыру», №69171, күні: 27.03.2026 ж.

Жұмыстың көлемі мен құрылымы. Диссертациялық жұмыс кіріспеден, 4 бөлімнен, қорытынды, әдебиеттер тізімі және қосымшалардан тұрады. Диссертацияның толық көлемі: 98 бет машиналық мәтіні, оның ішіне 26 сурет, 21 кесте, 107 пайдаланылған әдебиеттер тізімі және 1 қосымша кіреді.

Кіріспеде диссертациялық зерттеу тақырыбының өзектілігі негізделіп, жұмыстың мақсаты мен міндеттері, зерттеу объектісі мен пәні айқындалды. Зерттеу әдістері, жұмыстың ғылыми жаңалығы және нәтижелерінің практикалық маңызы сипатталды. Қорғауға ұсынылатын негізгі тұжырымдар келтіріліп, зерттеу нәтижелерінің апробациясы мен жарияланымдар туралы мәліметтер берілді. Сонымен қатар диссертацияның көлемі мен құрылымы көрсетілді.

Диссертацияның бірінші бөлімінде қазақ тіліндегі деструктивті веб-контентті анықтаудың теориялық және әдістемелік негіздері қарастырылды. Деструктивті веб-контент ұғымы, оның негізгі түрлері мен ақпараттық қауіпсіздікке ықпалы сипатталып, қазақ тіліндегі веб-мәтіндердің тілдік ерекшеліктері талданды. Деструктивті мәтіндерді автоматты түрде анықтауда қолданылатын табиғи тілді өңдеу тәсілдеріне, дәстүрлі машиналық оқыту және терең оқыту әдістеріне салыстырмалы шолу жасалды. Қазақ тілінің агглютинативті құрылымы, деструктивті мағынаның контекске тәуелділігі және аннотацияланған деректердің жеткіліксіздігіне байланысты ғылыми мәселелер айқындалды. Осы талдау негізінде зерттеу есебі қойылып, контекстік, морфологиялық және онтологияға негізделген семантикалық белгілерді біріктіру қажеттілігі негізделді.

Диссертацияның екінші бөлімінде қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған әдістің негіздері әзірленді. Telegram, YouTube және VKontakte платформаларынан мәтіндерді жинау, сүзу, алдын ала өңдеу және сараптамалық аннотациялау

кезеңдері сипатталды. Нәтижесінде бейтарап мәтіндер, буллинг, нацизм және зорлық-зомбылық экстремизмі кластарына бөлінген 12 059 мәтіннен тұратын корпус қалыптастырылып, оның статистикалық және тілдік ерекшеліктері талданды. Сөз табы, шақ, жақ және рай категорияларын қамтитын морфологиялық белгілерді қалыптастыру тәсілі ұсынылды. Деструктивті веб-контентті түрі, ашықтығы және нысанасы бойынша сипаттайтын онтологиялық білім моделі әзірленіп, лексикон мен граф құрылымы негізінде семантикалық белгілерді алу тәртібі анықталды. Аталған белгілерді KazBERT қалыптастыратын контекстік көріністермен біріктірудің әдістемелік негізі сипатталды.

Диссертацияның үшінші бөлімінде қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған терең оқыту модельдері мен гибридті архитектура әзірленді. KazBERT базалық контекстік модель ретінде алынып, оны морфологиялық белгілер модулімен және онтологиялық-графтық модульмен толықтыру тәсілдері сипатталды. Контекстік, морфологиялық және онтологияға негізделген семантикалық көріністерді біріктірудің математикалық моделі ұсынылды. Мәтінді классификациялауда [CLS] көрінісін қолданатын нұсқа және токендер тізбегін LSTM, BiGRU немесе BiLSTM арқылы өңдейтін баламалы конфигурациялар қарастырылды. KazBERT, MFA, онтологиялық семантикалық белгілер және BiLSTM компоненттерін қамтитын гибридті архитектураның құрылымы, белгілерді өңдеу тәртібі және класс ықтималдықтарын есептеу жолы сипатталды.

Диссертацияның төртінші бөлімінде ұсынылған модельдердің тиімділігі эксперименттік зерттеулер арқылы бағаланды. Деректерді оқыту, валидация және тест жиындарына бөлу тәртібі, эксперименттік орта және бағалау метрикалары сипатталды. Модельдер бірыңғай бағалау хаттамасы бойынша бес түрлі кездейсоқ бастапқы мәнмен іске қосылып, нәтижелердің орташа мәндері мен стандарттық ауытқулары есептелді. Морфологиялық, онтологиялық және тізбектік өңдеу компоненттерінің классификациялау сапасына қосқан үлесі анықталып, қателесу матрицалары негізінде қате болжамдар талданды. Ұсынылған модель мен базалық KazBERT нәтижелерінің айырмашылығы McNemar тесті арқылы бағаланды. Қарастырылған модельдер арасында BiLSTM негізіндегі гибридті архитектура ең жоғары нәтижеге қол жеткізіп, Accuracy көрсеткіші 91,045%, ал Macro-F1 көрсеткіші 91,103% болды. Модельдің практикалық қолданылу мүмкіндігі веб-интерфейс прототипі арқылы көрсетілді.

Қорытындыда диссертацияда алынған негізгі нәтижелер көрсетіледі.