

**«6D100200 – Ақпараттық қауіпсіздік жүйелері» мамандығы бойынша философия докторы (PhD) дәрежесіне іздену үшін ұсынылған
Болатбек Милана Асланбекқызының «Мәтіндегі экстремистік бағытты анықтау үшін веб-ресурстардағы семантикалық талдау
модельдерін әзірлеу және зерттеу» тақырыбындағы диссертациялық жұмысына ресми рецензенттің**

СЫН-ШІКІРІ

р/н №	Критерийлер	Критерийлер сәйкестігі	Ресми рецензенттің ұстанымы
1.	Диссертация тақырыбының (бекіту күніне) ғылымның даму бағыттарына және/немесе мемлекеттік бағдарламаларға сәйкес болуы	1.1 Ғылымның даму бағыттарына және/немесе мемлекеттік бағдарламаларға сәйкестігі: 1) <u>Диссертация мемлекет бюджетінен қаржыландырылатын жобаның немесе нысаналы бағдарламаның аясында орындалған (жобаның немесе бағдарламаның атауы мен нөмірі);</u> 2) Диссертация басқа мемлекеттік бағдарлама аясында орындалған (бағдарламаның атауы) 3) Диссертация Қазақстан Республикасының Үкіметі жанындағы Жоғары ғылыми-техникалық комиссия бекіткен ғылым дамуының басым бағытына сәйкес (бағытын көрсету)	Сәйкес келеді Берілген диссертациялық жұмыс ҚР цифрлық даму, инновациялар және аэроғарыш өнеркәсібі министрлігінің тапсырысы бойынша «Мәтіндегі экстремистік бағытты анықтау үшін веб-ресурстардағы семантикалық талдау модельдерін, алгоритмдерін құрастыру және кибер-криминалистика құрал-жабдықтарын әзірлеу», ЖТН АР06851248 жобасы аясында жазылды. Басым бағыт: Ұлттық қауіпсіздік және қорғаныс. Мамандандырылған бағыт: Ақпараттық қауіпсіздікті қамтамасыз ету.
2.	Ғылымға маңыздылығы	Жұмыс ғылымға елеулі үлесін <u>қосады/қоспайды</u> , ал оның маңыздылығы <u>ашылған/ашылмаған</u> .	Жұмыс ғылымға елеулі үлес қосады және оның маңыздылығы жұмыста жақсы ашылған. Диссертациялық зерттеу нәтижелерін пайдалану қазіргі ақпарат саны күн сайын артып келе жатқан уақытта веб-ресурстардағы экстремизмнің көріністерін анықтауға, оның алдын алуға және жолын кесуге мүмкіндік береді. Жоғарыда айтылғандар диссертациялық зерттеудің қолданбалы ғылымға енгізетін айтарлықтай үлесі бар екеніне қорытынды жасауға мүмкіндік береді.
3.	Өзі жазу принципі	Өзі жазу деңгейі: 1) жоғары; 2) орташа; 3) төмен; 4) өзі жазбаған	Диссертациялық жұмысты орындау барысында ізденуші жоғары дербестік дәрежесін көрсетті, зерттеу жұмысын жүргізу үшін қажетті теориялық және практикалық дереккөздерді зерттеді және талдады, ғылыми зерттеулерге тікелей қатысты, соның ішінде: жоспарлар құру, әдебиеттерді талдау, бағдарламалық жабдықтаманы құру. Диссертациялық жұмыс ізденушінің теориялық білімінің тереңділігі мен ғылыми эксперименттерге тікелей қатысуының, нәтижелерді ғылыми дәлелдеу мен оны ғылыми басылымдарда жариялаудың, сондай-ақ оны практикалық енгізудің нәтижесі болып табылады.

4.	Ішкі бірлік принципі	<p>4.1 Диссертация өзектілігінің негіздемесі:</p> <ol style="list-style-type: none"> 1) негізделген; 2) жартылай негізделген; 3) негізделмеген. 	<p>Диссертациялық жұмыстың өзектілігі кіріспе және бірінші бөлімде көрсетілген. Қазіргі ақпараттық технологиялар қарқынды даму үстінде. Жеке пайдаланушылардың ғана емес, сонымен қатар ақпараттық ұйымдардың да бүкіл әлемдік кеңістікке белсенді қатысуы ұлттық қауіпсіздікті қамтамасыз ету бойынша ақпараттық-коммуникациялық технологиялар дамуының қазіргі тенденцияларына сәйкес келетін іс-шараларды әзірлеу, атап айтқанда экстремизм мен терроризм идеяларының күшеюіне қарсы тұруға қатысты іс-шараларды ұйымдастыру қажеттілігін анықтайды. Ізденуші ұсынған семантикалық модельді қолдану осыған ұқсас класстар міндеттерін шешуде заманауи және өзекті тәсіл болып табылады.</p>
		<p>4.2 Диссертация мазмұны диссертация тақырыбын айқындайды</p> <ol style="list-style-type: none"> 1) айқындайды; 2) жартылай айқындайды; 3) айқындамайды 	<p>Диссертацияның мазмұны диссертация тақырыбы мен қорғалатын қағидаларды айқын көрсетеді. Диссертациялық жұмыс веб-ресурстардағы экстремистік бағыттағы мәтіндерді анықтауға арналған семантикалық талдау модельдерін әзірлеуге және ұсынылатын модельдің қазақ тіліндегі экстремистік бағыттағы мәтіндерді анықтау дәлдігін зерттеуге арналған. Аталған жұмыстардың барлығы диссертация тақырыбын толықтай айқындайды.</p>
		<p>4.3. Мақсаты мен міндеттері диссертация тақырыбына сәйкес келеді:</p> <ol style="list-style-type: none"> 1) сәйкес келеді; 2) жартылай сәйкес келеді; 3) сәйкес келмейді 	<p>Ғылыми жұмыста зерттеудің мақсаты мен міндеттері дұрыс көрсетілген және диссертация тақырыбына толық сәйкес келеді. Жұмыстың мақсаты – веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтау үшін семантикалық талдау моделін кешенді зерттеу және құру. Келесі зерттеу міндеттері қойылған:</p> <ul style="list-style-type: none"> – веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтау үшін машиналық және терең оқыту әдістерін оқытуға және тестілеуге арналған қазақ тіліндегі экстремистік мәтіндер корпусын құру; – веб-ресурстарда қазақ тіліндегі экстремистік мәтіндерді анықтау үшін семантикалық талдау моделін құру; – веб-ресурстарда қазақ тіліндегі экстремистік мәтіндерді анықтауға арналған әдістерді құру және зерттеу; – қазақ тіліндегі экстремистік түйінді сөздердің тізімін құру; – әзірленген модель мен әдістер негізінде веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтайтын бағдарламалық жабдықтаманы құру және тестілеу.
		<p>4.4. Диссертацияның барлық бөлімдері мен құрылысы логикалық байланысқан:</p> <ol style="list-style-type: none"> 1) толық байланысқан; 2) жартылай байланысқан; 3) байланыс жоқ 	<p>Диссертацияның барлық бөлімдері мен ережелері өзара байланысты, алынған нәтижелердің ішкі бірлігі диссертациялық жұмыстың мақсатына таңдалған әдістерге байланысты, бөлімдердің мазмұны дұрыс құрылымданған, логикалық байланысты және дәйекті баяндалады. Ұсынылған диссертациялық жұмыс диссертацияларға</p>

			қойылатын негізгі талаптарға сәйкес келеді, ондағы теориялық ережелер практикалық нәтижелермен байланысты.
		4.5 Автор ұсынған жаңа шешімдер (қағидаттар, әдістер) дәлелденіп, бұрыннан белгілі шешімдермен салыстырылып бағаланған: 1) сыни талдау бар; 2) талдау жартылай жүргізілген; 3) талдау өз пікірін емес, басқа авторлардың сілтемелеріне негізделген	Диссертацияда ұсынылған жаңа шешімдер дәлелденіп көрсетілген. Құрылған семантикалық талдау моделінің экстремистік мәтіндерді анықтау дәлдігінің жоғары екендігі басқа машиналық оқыту алгоритмдерімен салыстырыла отырып дәлелденіп көрсетілген. Экстремистік мәтіндерді анықтау үшін белгілер жиынтығын қалыптастырудың сөзді ендіру әдістері мен n-граммдарды комбинациялауға негізделетін әдіс құрастырылған және оның экстремистік мәтіндерді жіктеу сапасын арттыратындығы дәлелденіп көрсетілген.
5.	Ғылыми жаңашылдық принципі	5.1 Ғылыми нәтижелер мен қағидаттар жаңа болып табыла ма? 1) толығымен жаңа; 2) жартылай жаңа (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)	Жұмыстың ғылыми жаңалығы күмән тудырмайды және диссертацияда тұжырымдалған мақсаттарға қол жеткізу үшін бірегей тәсілдер мен шешімдерді әзірлеу болып табылады. Мұндай нәтижелерге, бірінші кезекте, алғаш рет қазақ тілінің ерекшеліктерін ескере отырып, LSTM желісінің сөзді ендіру қабатына алдын ала стемминг алгоритмі орындалған биграммдарға TF-IDF әдісін қолданумен ерекшеленетін және экстремистік мәтіндерді анықтау дәлдігін арттыратын семантикалық талдау моделін жатқызуға болады.
		5.2 Диссертацияның қорытындылары жаңа болып табыла ма? 1) толығымен жаңа; 2) жартылай жаңа (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)	Диссертациялық жұмыста теориялық және тәжірибелік зерттеулер жүргізу нәтижесінде веб-ресурстардағы қазақ тіліндегі экстремистік бағыттағы мәтіндерді анықтауға әдіс пен семантикалық талдау моделі құрастырылған. Алғаш рет қазақ тіліндегі экстремистік мәтіндерді анықтау үшін машиналық оқыту әдістерін оқытуға және тестілеуге арналған қазақ тіліндегі экстремистік мәтіндер корпусы құрылған. Әзірленген әдіс пен модельде қолданылған сөз ендіру әдістері сипатталып, әзірленген модель мен әдістер негізінде веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтайтын бағдарламалық жабдықтаманы құрылған және тестілеуден өткізілген.
		5.3 Техникалық, технологиялық, экономикалық немесе басқару шешімдері жаңа және негізделген бе? 1) толығымен жаңа; 2) жартылай жаңа (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)	Диссертациялық жұмыстың техникалық және технологиялық шешімдері жаңа және негізделген. Ақпараттық және коммуникациялық технологиялардың дамуымен қатар Интернет желісіндегі қауіпті ақпарат саны да артып келеді. Зерттеу жұмысында веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтауға арналған семантикалық талдау моделі құрылды.

6.	Негізгі қорытындылардың негізділігі	Барлық қорытындылар ғылыми тұрғыдан қарағанда ауқымды дәлелдемелерде <u>негізделген/негізделмеген</u> (qualitative research және өнертану және гуманитарлық бағыттары бойынша)	Диссертациялық жұмыстың қорытындылары жүргізілген ғылыми-зерттеу жұмыстарының нәтижелерімен негізделген. Алғаш рет қазақ тіліндегі экстремистік мәтіндерді анықтау үшін машиналық оқыту әдістерін оқытуға және тестілеуге арналған қазақ тіліндегі экстремистік мәтіндер корпусы құрылды. Алғаш рет қазақ тілінің ерекшеліктерін ескере отырып, LSTM желісінің сөзді ендіру қабатына алдын ала стемминг алгоритмі орындалған биграммдарға TF-IDF әдісін қолданумен ерекшеленетін және экстремистік мәтіндерді анықтау дәлдігін арттыратын семантикалық талдау моделі құрастырылды және оның қазақ тіліндегі экстремистік мәтіндерді анықтау дәлдігінің жоғары екендігі дәлелденіп көрсетілген. Экстремистік мәтінді анықтау дәлдігін тексеру кезінде машиналық оқытудағы келесі бағалау параметрлері қолданылған: Accuracy, Precision, Recall, F1-Score, AUC-ROC. Алынған нәтижелер барлық бағалау параметрлері бойынша жоғары көрсеткіш көрсеткен.
7.	Қорғауға шығарылған негізгі қағидаттар	<p>Әр қағидат бойынша келесі сұрақтарға жауап беру қажет:</p> <p>7.1 Қағидат дәлелденді ме?</p> <p>1) <u>дәлелденді</u>;</p> <p>2) шамамен дәлелденді;</p> <p>3) шамамен дәлелденбеді;</p> <p>4) дәлелденбеді</p> <p>7.2 Тривиалды ма?</p> <p>1) <u>ия</u>;</p> <p>2) <u>жоқ</u></p> <p>7.3 Жаңа ма?</p> <p>1) <u>ия</u>;</p> <p>2) <u>жоқ</u></p> <p>7.4 Қолдану деңгейі:</p> <p>1) <u>тар</u>;</p> <p>2) орташа;</p> <p>3) <u>кең</u></p> <p>7.5 Мақалада дәлелденген бе?</p> <p>1) <u>ия</u>;</p> <p>2) <u>жоқ</u></p>	<p>Диссертанттың жұмысы бойынша қорғауға шығарылатын негізгі келесі қағидаттарды атауға болады:</p> <p>– алғаш рет қазақ тіліндегі экстремистік мәтіндерді анықтау үшін машиналық оқыту әдістерін оқытуға және тестілеуге арналған қазақ тіліндегі экстремистік мәтіндер корпусы құрылды.</p> <p>Қағидат дәлелденген, тривиалды емес, жаңа, қолдану деңгейі кең және келесі мақалада дәлелденген:</p> <p>К. Шалабаев, К. Әліпбай, М. Болатбек, Ш. Мусиралиева, Вконтакте әлеуметтік желісіндегі экстремистік мәтіндерді анықтау және жіктеу, Вестник КазНИТУ №5 (135), 80-86 б., 2019</p> <p>– алғаш рет қазақ тілінің ерекшеліктерін ескере отырып, LSTM желісінің сөзді ендіру қабатына алдын ала стемминг алгоритмі орындалған биграммдарға TF-IDF әдісін қолданумен ерекшеленетін және экстремистік мәтіндерді анықтау дәлдігін арттыратын семантикалық талдау моделі құрастырылды.</p> <p>Қағидат дәлелденген, тривиалды емес, жаңа, қолдану деңгейі кең және келесі мақалада дәлелденген:</p> <p>Mussiraliyeva Sh., Omarov B., Yoo P., Bolatbek M. Applying Machine Learning Techniques for Religious Extremism Detection on Online User Contents // CMC – Computers, Materials & Continua. – 2021. – Vol. 70. No. 1. P. 915–934. ISSN:1546-2226 (SJR 2020 0.79, процентиль 72) https://doi:10.32604/cmc.2022.019189</p>

			<p>– белгілер жиынтығын калыптастырудың сөзді ендіру әдістері мен программдарды комбинациялауға негізделетін және экстремистік мәтіндерді жіктеудің сапасын арттыратын әдіс құрастырылды.</p> <p>Қағидат дәлелденген, тривиалды емес, жана, қолдану деңгейі кең және келесі мақалада дәлелденген:</p> <p>Mussiraliyeva Sh., Bolatbek M., Omarov B., Bagitova K., Alimzhanova Zh. Bigram based Deep Neural Network for Extremism Detection in Online User Generated Contents in the Kazakh Language // International Conference on Computational Collective Intelligence. – Greece, 2021. – P.559-570. https://doi.org/10.1007/978-3-030-88113-9_45</p> <p>– алғаш рет қазақ тілінде экстремистік түйінді сөздердің тізімі құрылды.</p> <p>Қағидат дәлелденген, тривиалды емес, жана, қолдану деңгейі кең және келесі мақалада дәлелденген:</p> <p>Bolatbek M.A., Mussiraliyeva Sh.Zh., Tukeyev U.A., Creating the dataset of keywords for detecting an extremist orientation in web-resources in the Kazakh language, AI-Farabi Kazakh National University, Journal of Mathematics, Mechanics and Computer Science, No 1 (97), pp.134-142, 2018.</p> <p>Диссертация тақырыбына байланысты 19 ғылыми баспа жұмыс жарияланды, оның ішінде ҚР БҒМ Білім және ғылым саласында бақылау комитеті ұсынған журналдарда 4 жұмыс, еліміздегі, алыс шетелдердегі халықаралық конференциялар баяндамаларында 14 жұмыс, Scopus деректер қорына кіретін баспада 1 жұмыс.</p>
8.	Дәйектілік принципі Дереккөздер мен ұсынылған ақпараттың дәйектілігі	<p>8.1 Әдістеменің таңдауы - негізделген немесе әдіснама нақты жазылған</p> <p>1) ия; 2) жоқ</p> <p>8.2 Диссертация жұмысының нәтижелері компьютерлік технологияларды қолдану арқылы ғылыми зерттеулердің қазіргі заманғы әдістері мен деректерді өңдеу және интерпретациялау әдістемелерін пайдалана отырып алынған:</p> <p>1) ия;</p>	<p>Диссертациялық жұмыста таңдалған әдістеме нақты жазылған және негізделген. Диссертациялық жұмыста мәліметтерді жіктеу мен ақпараттық, интеллектуалдық-программалық кешендерді құрудың заманауи және қолданылып жүрген тәсілдері, соның ішінде табиғи тілді өңдеу, нейрондық желілер, әлеуметтік желілерді талдау әдістері, статистикалық өңдеу әдістері, жүйелік талдау әдістері әдістері пайдаланылған.</p> <p>Диссертациялық жұмыстың негізгі бөлімдерінің бірі болып табылатын, құрылған семантикалық талдау моделінің қазақ тіліндегі экстремистік мәтіндерді анықтау дәлдігін зерттеу нәтижелері компьютерлік технологияларды қолдану арқылы орындалатын ғылыми зерттеулердің қазіргі заманғы әдістері мен деректерді өңдеу әдістерін пайдалана отырып алынған. Диссертациялық жұмыстың нәтижелері программалау және деректерді жіктеудің қазіргі заманғы компьютерлік технологияларын</p>

		2) жоқ	пайдалану нәтижесінде алынды. Зерттеу барысында жасанды зердені жобалауда кеңінен қолданылатын python тілі, NLTK (natural language toolkit) секілді табиғи тілді өңдеу пакеттері қолданылған.
		8.3 Теориялық қорытындылар, модельдер, анықталған өзара байланыстар және заңдылықтар эксперименттік зерттеулермен дәлелденген және расталған (педагогикалық ғылымдар бойынша даярлау бағыттары үшін нәтижелер педагогикалық эксперимент негізінде дәлелденеді): 1) ия; 2) жоқ	Әзірленген модельдің барабарлығы веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндердің үлесін анықтауға арналған бағдарламалық жабдықтаманы әзірлеу және оны өзекті ету эксперименттер жүргізу арқылы дәлелденді.
		8.4 Маңызды мәлімдемелер нақты және сенімді ғылыми әдебиеттерге сілтемелермен <u>расталған</u> / ішінара расталған / расталмаған	Диссертациялық жұмыстағы маңызды мәлімдемелер өзекті және сенімді, отандық және шетелдік нақты және сенімді ғылыми әдебиеттерге сілтеме келтірумен расталады. Пайдаланылған ғылыми әдебиеттер зерттеу саласына сәйкес.
		8.5 Пайдаланылған әдебиеттер тізімі әдеби шолуға <u>жеткілікті</u> /жеткіліксіз	Диссертациялық жұмыс бойынша ізденіс барысында пайдаланылған әдебиеттер тізімі әдеби шолуға жеткілікті. Олардың ішінде, Web of Science және SCOPUS халықаралық деректер қорына кіретін журналдар және машиналық оқыту саласы бойынша соңғы жылдары шыққан әдебиеттер бар.
9	Практикалық құндылық принципі	9.1 Диссертацияның теориялық маңызы бар: 1) ия; 2) жоқ	Диссертациялық жұмыстың теориялық маңыздылығы экстремистік іс-әрекеттер мен ұйымдарды анықтау әдістері мен алгоритмдері саласындағы білім жиынтығына негізделген. Алынған іргелі нәтижелерді әлемдік ғылыми қауымдастық пайдалана алады.
		9.2 Диссертацияның практикалық маңызы бар және алынған нәтижелерді практикада қолдану мүмкіндігі жоғары: 1) ия; 2) жоқ	Қолданбалы нәтижелерді ақпараттық қауіпсіздікті, сыни инфрақұрылымды қамтамасыз ету, интернет-экстремизммен күрес жөніндегі уәкілетті органдар пайдалануы мүмкін. Соңдықтан, құрылған семантикалық талдау моделін практикада қолдану мүмкіндігі жоғары.
		9.3 Практикалық ұсыныстар жаңа болып табылады? 1) толығымен жаңа; 2) жартылай жаңа (25-75% жаңа болып табылады):	Автор тұжырымдаған практикалық ұсыныстар жаңа болып табылады. Веб-ресурстардағы қазақ тіліндегі экстремистік бағыттағы мәтіндерді анықтауға арналған семантикалық талдау моделі мен әдісі құрастырылған және оны отандық ақпараттық қауіпсіздік құралдарын әзірлеуде кеңінен пайдалануға болады.

10.	Жазу және ресімдеу сапасы	<p>3) жаңа емес (25% кем жаңа болып табылады)</p> <p>Академиялық жазу сапасы:</p> <p>1) жоғары;</p> <p>2) орташа;</p> <p>3) орташадан төмен;</p> <p>4) төмен.</p>	<p>Ізденуші Болатбек Милана Асланбекқызының «Мәтіндегі экстремистік бағытты анықтау үшін веб-ресурстардағы семантикалық талдау модельдерін әзірлеу және зерттеу» тақырыбындағы диссертациялық жұмысы ғылыми тілде, сауатты жазылған, техникалық тұрғыдан дұрыс рәсімделген және өзектілігі, қойылған мақсаты мен міндеттері, ғылыми жаңалығы тәжірибелік маңыздылығы жағынан «6D100200 – Ақпараттық қауіпсіздік жүйелері» мамандығы бойынша философия докторы (PhD) дәрежесін алудың «Дәрежелер беру қағидалары» талаптарына толығымен сай келеді.</p> <p>Диссертациялық жұмыс бойынша келесі ескертулер мен ұсыныстарды атап өтуге болады:</p> <ul style="list-style-type: none"> - қазақша ғылыми терминдердің баламалары кейбір жерлерде дұрыс емес. Мысалы, классификация, шама және т.б. - жұмыста зерттеу саласы бойынша ғалымдардың әр түрлі тілдерге арналған еңбектері аталып өткен. Дегенмен де зерттеу жұмыстарында қарастырылатын тілдер мен қазақ тілі арасындағы айырмашылық талданып көрсетілсе, артық болмас еді. <p>Айтылған ескертулер жұмысты әрі қарай жетілдіруге нұсқаулық ретінде көрсетілген және олар диссертациялық жұмыстың құндылығын төмендетпейді.</p> <p>Болатбек Милана Асланбекқызы «6D100200 – Ақпараттық қауіпсіздік жүйелері» мамандығы бойынша философия докторы (PhD) дәрежесін алуға лайықты деп санаймын</p>
-----	---------------------------	---	--

Ресми рецензент:
 Л.Н.Гумилев атындағы Еуразия ұлттық университеті
 PhD, қауымдастырылған профессор

Т.К.Жукабаева

