COLLEGE OF SIGNALS

(A Campus of NUST)

REVIEW OF THE PhD DISSERTATION

Applicant: Sagynay Moldir

Title: Development of Deep Learning Models for the Classification of Destructive Web Content in the Kazakh Language

Educational Program: 8D06301 – Information Security Systems

The dissertation entitled "Development of Deep Learning Models for the Classification of Destructive Web Content in the Kazakh Language" addresses a highly relevant scientific and practical problem in the fields of information security, artificial intelligence, and natural language processing. The rapid growth of harmful online content, including bullying, Nazism, and violent extremism, has created a significant need for automated content monitoring systems. While substantial progress has been made for high-resource languages, research on destructive content detection in the Kazakh language remains limited due to the lack of annotated datasets and language-specific intelligent models.

The main objective of the dissertation is to develop effective deep learning models for the classification of destructive web content in the Kazakh language. To achieve this goal, the author constructed a large annotated dataset, investigated machine learning and deep learning approaches, developed a morphological feature extraction module, introduced ontology-based semantic features, and proposed a hybrid classification architecture that incorporates the linguistic characteristics of the Kazakh language.

The scientific novelty of the dissertation is evident in several important contributions. First, the author developed a multi-class annotated corpus of destructive Kazakh-language texts covering bullying, Nazism, violent extremism, and neutral content. Second, ontology-based semantic features describing the "Type", "Severity", and "Target" of destructive content were designed and integrated into the classification process. Third, a novel hybrid architecture combining contextual KazBERT embeddings, morphological features, and ontology-based semantic representations was proposed and experimentally validated.

The research methodology is comprehensive and scientifically sound. The author conducted a detailed review of existing approaches, developed a corpus collection and annotation framework, and performed extensive experiments using classical machine learning algorithms and modern deep learning architectures. The proposed model was evaluated using widely accepted metrics, including Accuracy, Precision, Recall, Macro F1-score, ROC-AUC, PR-AUC, and MCC, ensuring the reliability of the obtained results.

The experimental findings demonstrate the effectiveness of the proposed approach. Compared with the baseline KazBERT model, the integration of morphological and ontology-based semantic features significantly improved classification performance, increasing the Accuracy from approximately 88% to over 91% and the Macro F1-score from 88% to 91%. The conducted ablation study further confirms the positive contribution of each component of the proposed architecture.

Department of Information Security Military College of Signals, NUST, Humayun Road, Rawalpindi, Pakistan

The practical significance of the dissertation is substantial. The developed models can be applied in automated systems for monitoring social networks, detecting harmful online content, supporting content moderation processes, and enhancing information security measures. The created annotated dataset also represents a valuable resource for future research in Kazakh-language natural language processing and other Turkic languages.

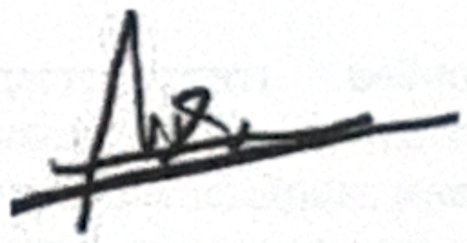
The results of the dissertation have been sufficiently validated through scientific publications, including articles in Scopus-indexed journals and international conference proceedings. The author has also obtained a copyright certificate for software developed within the scope of the research, demonstrating the practical implementation of the proposed methods.

Among the strengths of the dissertation are the relevance of the topic, the originality of the proposed hybrid approach, the creation of a valuable linguistic resource, comprehensive experimental validation, and the practical applicability of the obtained results. Minor improvements could include a more detailed discussion of annotation quality assessment, model explainability, and ethical considerations related to automated content moderation. However, these comments do not diminish the overall scientific value of the research.

The dissertation *"Development of Deep Learning Models for the Classification of Destructive Web Content in the Kazakh Language"* is a completed scientific work that contains original research results of theoretical and practical significance. The author has demonstrated a high level of scientific competence and successfully solved an important research problem in the field of information security and artificial intelligence.

The dissertation meets the requirements for doctoral research, and its author, Sagynay Moldir, deserves to be awarded the degree of Doctor of Philosophy (PhD) in the educational program 8D06301 – Information Security Systems*.

Sincerely,



Assoc Prof Mian Muhammad Waseem Iqbal
Dept of Information Security
MCS – NUST
+92-333-9113007
waseem.iqbal@mcs.edu.pk

(логосы: ҰЛТТЫҚ ҒЫЛЫМ ЖӘНЕ ТЕХНОЛОГИЯ УНИВЕРСИТЕТІ (NUST), ПӘКІСТАН)

БАЙЛАНЫС КОЛЛЕДЖІ (CS)

(NUST кампусы)

ДОКТОРЛЫҚ ДИССЕРТАЦИЯҒА ШОЛУ

Өтініш беруші: Сағынай Мөлдір

Атауы: Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған тереңдетіп оқыту үлгілерін жасау

Білім беру бағдарламасы: 8D06301 – Ақпараттық қауіпсіздік жүйелері

"Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған тереңдетіп оқыту үлгілерін жасау" атты диссертация ақпараттық қауіпсіздік, жасанды интеллект және табиғи тілді өңдеу саласындағы аса өзекті ғылыми және практикалық проблемаға арналған. Құдалау, нацизм және зорлық-зомбылық экстремизмін қоса алғанда, зиянды интернет-контент көлемінің қарқынды өсуі контентті бақылаудың автоматтандырылған жүйелерін құрудың айтарлықтай қажеттілігін тудырдыа. Ресурстары мол тілдер үшін қол жеткізілген елеулі прогреске қарамастан, қазақ тіліндегі деструктивті контентті анықтау жөніндегі зерттеулер белгіленген деректер жиынтығының және осы тіл үшін мамандандырылған зияткерлік модельдердің болмауына байланысты шектеулі болып қалады.

Диссертацияның негізгі мақсаты қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған тереңдетіп оқыту үлгілерін жасау болып табылады. Осы мақсатқа жету үшін автор ірі белгіленген деректер жиынтығын қалыптастырды, машиналық оқыту және терең оқыту тәсілдерін зерттеді, морфологиялық белгілерді алу модулін әзірледі, онтологиялық негізделген семантикалық белгілерді енгізді және қазақ тілінің лингвистикалық ерекшеліктерін ескеретін жіктеудің гибридті архитектурасын ұсынды.

Диссертацияның ғылыми жаңалығы бірнеше маңызды ғылыми нәтижелерде көрінеді. Біріншіден, автор құдалауды, нацизмді, зорлық-зомбылық экстремизмін және бейтарап контентті қамтитын деструктивті контентті қамтитын қазақ тіліндегі мәтіндердің көп деңгейлі белгіленген корпусын әзірледі. Екіншіден, деструктивті контенттің "Түрін", "Ауырлығын" және "Мақсатын" сипаттайтын онтологиялық негізделген семантикалық белгілер әзірленді және жіктеу процесіне біріктірілді. Үшіншіден, KazBERT контекстік эмбеддингтерін, морфологиялық белгілерді және онтологиялық негізделген семантикалық түсініктерді біріктіретін жаңа гибридті архитектура ұсынылды және эксперименталды түрде расталды.

Зерттеу әдістемесі жан-жақты және ғылыми негізделген. Автор қолданыстағы тәсілдерге егжей-тегжейлі шолу жасады, корпусты жинау және аннотациялау жүйесін жасады, сонымен қатар классикалық машиналық оқыту алгоритмдері мен заманауи терең оқыту архитектураларын қолдана отырып кең тәжірибелер жасады. Ұсынылған модель Accuracy (жіктеу дәлдігі), Precision (дәлдік), Recall (толықтық), Macro F1-score, ROC-AUC, PR-AUC және MCC сияқты жалпы қабылданған көрсеткіштерді қолдана отырып бағаланды, бұл нәтижелердің сенімділігін қамтамасыз етеді.

Эксперименттік нәтижелер ұсынылған тәсілдің тиімділігін көрсетеді. KazBERT базалық моделімен салыстырғанда морфологиялық және онтологиялық негізделген семантикалық белгілердің интеграциясы жіктеу сапасын едәуір арттырып, Accuracy көрсеткішін шамамен 88%-дан 91% — ға дейін, ал Macro F1-score көрсеткішін 88% - дан 91% - ға дейін арттырды. Абляцияны зерттеу ұсынылған архитектураның әрбір компонентінің оң үлесін одан әрі растайды.

Ақпараттық қауіпсіздік кафедрасы, Әскери байланыс колледжі, Ұлттық ғылым және технологиялар университеті (NUST), Хумаюн жолы, Равалпинди, Пәкістан

Диссертацияның практикалық маңыздылығы елеулі. Әзірленген модельдер әлеуметтік желілерді бақылаудың, зиянды интернет-контентті анықтаудың, контентті модерациялау процестерін қолдаудың және ақпараттық қауіпсіздік деңгейін арттырудың автоматтандырылған жүйелерінде қолданыла алады. Құрылған белгіленген деректер жиынтығы қазақ тілінде және басқа да түркі тілдерінде табиғи тілді өңдеу саласындағы болашақ зерттеулер үшін құнды ресурс болып табылады.

Диссертацияның нәтижелері Scopus-та индекстелген журналдардағы мақалалар мен халықаралық конференциялардың материалдарын қоса алғанда, ғылыми жарияланымдар арқылы жеткілікті түрде расталды. Сондай-ақ, автор осы зерттеу аясында жасалған бағдарламалық жасақтамаға авторлық құқық туралы куәлік алды, бұл ұсынылған әдістердің практикалық іске асырылуын көрсетеді.

Диссертацияның күшті жақтарына тақырыптың өзектілігі, ұсынылған гибридтік тәсілдің өзіндік ерекшелігі, құнды лингвистикалық ресурсты құру, жан-жақты эксперименттік валидация және алынған нәтижелердің практикалық қолданылуы жатады. Жетілдірудің мүмкін бағыттары аннотация сапасын бағалауды, модельдердің интерпретациясын және контентті автоматтандырылған модерациялауға байланысты этикалық аспектілерді егжей-тегжейлі қарастыруды қамтиды. Алайда бұл ескертулер жүргізілген зерттеудің жалпы ғылыми құндылығын төмендетпейді.

"Қазақ тіліндегі деструктивті веб-контентті классификациялауға арналған тереңдетіп оқыту үлгілерін жасау" диссертациясы теориялық және практикалық маңызы бар зерттеулердің өзіндік нәтижелерін қамтитын аяқталған ғылыми жұмыс болып табылады. Автор ғылыми құзыреттіліктің жоғары деңгейін көрсетті және ақпараттық қауіпсіздік пен жасанды интеллект саласындағы маңызды зерттеу міндетін сәтті шешті.

Диссертация докторлық зерттеулерге қойылатын талаптарға сәйкес келеді, ал оның авторы Сағынай Мөлдір 8D06301 - "Ақпараттық қауіпсіздік жүйелері" білім беру бағдарламасы бойынша философия докторы (PhD) дәрежесін беруге лайық.

Құрметпен,

/қол қойылған/
Доцент Миан Мухаммад Васим Икбал
Ақпараттық қауіпсіздік кафедрасы
Әскери байланыс колледжі (MCS)-Ұлттық ғылым және технологиялар университеті (NUST)
+92-333-9113007
waseem.iqbal@mcs.edu.pk