Краткая информация о проекте

Наименование	AP19576868 Разработка моделей и методов для выявления молодежного экстремизма и обеспечения безопасности молодежи в современном информационном пространстве
Актуальность	Экстремистская речь в сети может служить инструментом для планирования и проведения экстремистских действий, в том числе террористических актов. Определение такой речи позволяет выявлять потенциальные угрозы безопасности и предотвращать их реализацию. Интернет часто используется для распространения идей экстремизма и радикализации. Определение экстремистской речи позволяет рано выявлять признаки радикализации и предостерегать от возможных негативных последствий. Определение экстремистской речи необходимо для баланса между свободой слова и защитой общества от потенциальной опасности. Это позволяет различать между законными высказываниями и теми, которые могут представлять угрозу общественной безопасности. Многие страны имеют законы, регулирующие экстремистскую деятельность и речь. Определение экстремистской речи помогает соблюдать законы и предотвращать незаконные действия. Определение экстремистской речи способствует созданию безопасного онлайн-пространства для пользователей, особенно молодежи, которая может быть более уязвимой перед воздействием экстремистских идей. Экстремистская речь может провоцировать социальную напряженность, раздоры и конфликты. Определение такой речи помогает предотвращать распространение ненависти и способствует созданию более согласованного общества. Идентификация экстремистской речи в интернете требует сотрудничества между странами и организациями. Это помогает эффективному контролю и противодействию глобальным угрозам. Все эти аспекты подчеркивают важность определения экстремистской речи в интернете для обеспечения безопасности общества, предотвращения радикализации и поддержания баланса между свободой слова и обязанностью обеспечения общественной безопасности.
Цель	Целью проекта является исследование и разработка моделей и методов семантического анализа для определения и противодействия распространения насильственного, национального экстремизма, расизма, буллинга среди молодежи, методов мониторинга и анализа трафика в сети для противодействия распространения среди молодежи идеологии противоправного характера, создание списка потенциально опасных веб-ресурсов для молодежи, адаптация методов психоэмоционального анализа для казахского языка.
Задачи	1. Разработка новых моделей и методов для определения текстов национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь 1.1 Анализ доступных текстов по выбранному направлению и определение основных источников информации

	1.2 Разработка парсера для сбора данных из веб-ресурсов 1.3 Построение корпуса текстов национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь 1.4 Предобработка данных в корпусе 1.5 Определение набора признаков для улучшения задачи обнаружения национального, насильственного экстремизма, буллинга и расизма на веб-ресурсах 1.5 Разработка новых моделей и методов семантического анализа для определения текстов национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь на казахском языке 1.6 Адаптация методов анализа психоэмоционального анализа текстов для казахского языка 2. Разработка новых методов для анализа и мониторинга сетевого трафика 2.1 Разработка модуля сбора сетевых данных. 2.2 Разработка модуля анализа обработанных журналов трафика. 2.3 Разработка метода анализа и мониторинга сетевого трафик на основе машинного обучения. Создание списка потенциально опасных для молодежи веб-сайтов 3. Разработка программного обеспечения определения и противодействия распространения насильственного, насильственного экстремизма, расизма и буллинга среди молодежи 3.1 Дизайн архитектуры 3.2 Реализация серверной и лицевой части 3.3 Тестирование программного продукта
Ожидаемые и достигнутые результаты	Достигнутые результаты: Были разработаны новые модели и методы для определения текстов национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь. Проведен обзор новой литературы в отечественных и зарубежных изданиях по выявлению экстремистских текстов на Веб-ресурсах. Разработаны новые модели и методы выявления национальных, насильственных экстремистских, буллинговых и расистских текстов, ориентированных на молодежь. Была опубликована одна статья, посвященная использованию методов машинного обучения, таких как опорные векторные машины, наивные байесовские классификаторы, методы случайного дерева, дерево решений, алгоритм k ближайших соседей, логистическая регрессия, градиентный бустинг, для обнаружения экстремистских текстов. Был проведен обширный обзор существующих методов классификации текстов, связанных с национальным, насильственным экстремизмом, буллингом и расизмом, ориентированных на молодежь в интернете. Обзор включает последние публикации, опубликованные в высоко оцененных научных журналах, таких как Springer, Elsevier и других, включенных в базы данных Web of Science и Scopus. Анализ

	литературы помог определить текущее состояние исследований в этой области и определить актуальные направления нашего проекта. Этот анализ и обзор современного состояния методов выявления экстремистских текстов будут полезны для сообщества исследователей и инженеров, работающих в этой области. Это позволяет им более эффективно применять эти методы в своей работе и вносить свой вклад в развитие этой важной области. Традиционные методы машинного обучения, методы и модели на основе трансформеров были созданы для выявления текстов, связанных с национальным, насильственным экстремизмом, буллингом и расизмом в интернете. С помощью поисковых машин были выявлены и проведены исследования общедоступных текстов на различных веб-ресурсах (социальные сети Вконтакте, Twitter, YouTube, Telegram, блоги, форумы, новостные статьи). В результате данного исследования были выявлены ключевые фразы и первоисточники текстов, связанных с национальным, насильственным экстремизмом, буллингом и расизмом. Разработан парсер для сбора текстов с веб-ресурсов социальных сетей Вконтакте, Twitter, YouTube, Telegram с использованием выявленных ключевых слов. На вход парсера даются доменное имя источника, место хранения и срок рассмотрения. В результате загружается текстовый контент указанных веб-ресурсов. Использовались технологии API. В результате скомпилированного парсера был создан текстовый корпус, собранный из контента групп и каналов в социальных сетях Вконтакте, Twitter, YouTube, Telegram. В корпус включены 5 категорий: национальный экстремизм, насильственный экстремизм, расизм, буллинг и тексты нейтральных категорий, каждой категории присвоены соответствующие обозначения (от 0 до 4). Проведен лингвистический и статистический анализ текстов корпуса. Общий объем корпуса составляет около 10 000 текстов. Для текстов собранного текстового корпуса были выполнены алгоритмы препроцессинга: включая токенизацию, морфологический анализ текстов, стемминг, удаление знаков препинания, удаление числовых значений и гиперссылок в тексте, удаление стоп-слов. Определены признаки, повышающие точность определения текстов национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь такие как, как tf-idf, tf-idf-bigram, bag-of-words. Данные признаки используются при составлении моделей и методов, касающихся выявления деструктивного содержания в тексте. На основе методов машинного обучения Decision Trees, Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine, LSTM, BiLSTM была проведена работа по разработке новых методов и моделей семантического анализа данных для выявления национального, насильственного экстремизма, буллинга и расизма, ориентированного на молодежь. Создана модель на основе Stemming+TF-idf+BERT. Кроме того, построена модель проведения

	психоэмоционального анализа для повышения точности определения текстов данной категории. Была создана модель на основе таких трансформеров, как DistilBert и Roberta. RoBERTa улучшает BERT за счет тщательной и разумной оптимизации гиперпараметров чтения. Модель RoBERTa была разработана в Pytorch. Гиперпараметры модели: Input = 128 слов или токенов, RoBERTa = 1280 vector, Linear = 768, DropOut = 0.1, linear Classification = 5. model_name = "xlm-roberta-base", num_classes = 5, max_length = 128, batch_size = 64, num_epochs = 20, learning_rate = 2e-5, val_size=0.2, test_size=0.2 Модель семантического анализа для выявления текстов национального, насильственного экстремизма, буллинга и расизма на казахском языке на основе Distilbert направлена на оптимизацию обучения за счет уменьшения размера и увеличения скорости, все это было сделано с целью сохранения производительности. Гиперпараметры модели: Input = 128 слов или токенов, DistilBERT = 768 векторов, Linear = 768, DropOut = 0.1, linear Classification = 5. model_name = "distilbert-base-uncased", num_classes = 5, max_length = 128, batch_size = 64, num_epochs = 20, learning_rate = 2e-5, val_size=0.2, test_size=0.2. Также была создана модель MLM (модель маскированного языка-"моделирование маскированного языка") для выявления национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь. Гиперпараметры модели: Input = 128 слов или токенов, MLM = 1280 vector, Linear = 768, DropOut = 0.1, linear Classification = 5. model_name = "xlm-mlm-100-1280", num_classes = 5, max_length = 128, batch_size = 64, num_epochs = 20, learning_rate = 2E-5, val_size=0.2, test_size=0.2 Известные методы психоэмоционального анализа текстов были адаптированы для казахского языка, был разработан анализатор психоэмоциональных лексем в текстах национального, насильственного экстремизма, буллинга и расизма, ориентированных на молодежь. В ходе исследовательской работы был проанализирован экстремистский лингвистический корпус с использованием стратегии подсчета слов и метода закрытого словаря liwc. Задача предлагаемого метода-поиск и подсчет слов, относящихся к психологическим категориям, в наборе текстовых данных. Всего было выделено более 80 категорий. Результатом обработки текстового файла в программе являются следующие выходные переменные: количество слов, совокупные языковые переменные (аналитическое мышление, влияние, специфичность текста и эмоциональный тон) и процент слов в тексте, который представляет процент слов в тексте (например, местоимения, статьи, вспомогательные глаголы и т. д.). д.), категории, влияющие на психологические структуры (например, аффект, познание, биологические процессы, импульсы), категория личных интересов (например, работа, дом, отдых), неформальные языковые маркеры (например, проклятия). Результаты свидетельствуют о пользе общетеоретических знаний о выражении уровней развития личности в способах употребления слов. С помощью разработанного метода была создана модель на основе LSTM для классификации по национальному экстремизму, насильственному экстремизму, расизму и буллинским текстам. Получено авторское свидетельство на составленный модуль.

	Ожидаемые результаты: Будут разработаны новые методы для анализа и мониторинга сетевого трафика с целью определения опасных для молодежи веб-ресурсов. Будет разработано программное обеспечение, позволяющее определить веб-ресурсы с контентом национального, насильственного экстремизма, буллинга и расизма, направленных на молодежь.
Имена и фамилии членов исследовательской группы с их идентификаторами (Scopus Author ID, Researcher ID, ORCID, при наличии) и ссылками на соответствующие профили	1. Болатбек Милана Асланбекқызы, ORCID: https://orcid.org/0000-0002-2153-180X , Scopus профайл: https://www.scopus.com/authid/detail.uri?authorId=57202834055 , Web of Science профайл: https://www.webofscience.com/wos/author/record/GZL-7318-2022 2. Байсылбаева Кымбат Данияровна, ORCID: https://orcid.org/0000-0001-9753-0398, Web of Science профайл: https://www.webofscience.com/wos/author/record/N-9664-2017 3. Сағынай Мөлдір, ORCID: https://orcid.org/0009-0004-1377-5742 4. Елтай Жастай Ыбрайұлы, Researcher ID: https://www.webofscience.com/wos/author/record/JNR-6763-2023 , ORCID: https://orcid.org/my-orcid?orcid=0000-0002-9275-7582 Scopus author ID: https://www.scopus.com/authid/detail.uri?authorId=57237959800 5. Ахмед Гүлмарал Жалғасбекқызы, ORCID: https://orcid.org/0000-0002-4464-9544 6. Мейрбекова Бибинур Калдыбаевна, ORCID: https://orcid.org/0000-0001-9215-9382 , Scopus профайл: https://www.scopus.com/authid/detail.uri?authorId=57212476113 , Web of Science профайл: https://www.webofscience.com/wos/author/record/ABD-4499-2021 7. Шайзат Медет Жанболатұлы, ORCID: https://orcid.org/0000-0002-1651-8205 , Scopus профайл: https://www.scopus.com/authid/detail.uri?authorId=57216968174 8. Райымкулова Алима Мухамбеткалиевна
Список публикаций со ссылками на них	1. Scientific Journal of Astana IT University ISSN (P): 2707-9031 ISSN (E): 2707-904X VolUmE 14, JUNE 2023, COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORTMS TO IDENTIFY EXTREMIST TEXTS IN THE KAZAKH LANGUAGE,DOI: 10.37943/14DKRN4681, Shynar Mussiraliyeva , Milana Bolatbek ,Aigerim Zhumakhanova ,Zhanar Medetbek , Moldir Sagynay https://journal.astanait.edu.kz/index.php/ojs/article/view/344 2. Болатбек М.А., Сағынай М., Мусиралиева Ш.Ж., Байсылбаева К.Д., Шайзат М.Ж. Қазақ тіліндегі мәтінге психо-эмоционалдық талдау жүргізуге арналған әдісті құру және зерттеу, VIII — Международная научно-практическая конференция «Информатика и прикладная математика» https://conf.iict.kz/wp-content/uploads/2023/11/collection_CSAM_VIII_2023_2.pdf 3. Shynar Mussiraliyeva, Milana Bolatbek, Aygerim Zhumakhanova, Moldir Sagynay, Development of a software module for collecting and analyzing web content to determine extremist direction in the text принята к публикации в 17th International Conference on Information Technology and Applications (ICITA2023)

	https://link.springer.com/book/9789819983230 4. М.А.Болатбек, К.Д.Байсылбаева, М.Сағынай, Ш.Ж. Мусиралиева, А.Н.Жумаханова, Интернет кеңістігіндегі жастарға бағытталған деструктивті мәтіндерді жинақтауға қажетті парсер бағдарламасын әзірлеу, Известия НАН РК. Серия физико-математическая, №4, 2023 г. https://journals.nauka-nanrk.kz/physics-mathematics/article/view/5925 5. Bolatbek, Milana, and Shynar Mussiraliyeva. “Detection of Extremist Messages in Web Resources in the Kazakh Language.” Lodz Papers in Pragmatics, vol. 19, no. 2, Dec. 2023, pp. 415–425, doi:10.1515/lpp-2023-0020. https://journals.scholarsportal.info/details/18956106/v19i0002/415_doemiwritkl.xml
Информация о патентах	-

мұсылман
шабуыл
бол
мужахид
пәк
жаңа
жау
американдық
ауған
болған
талиб
үкімет
айт
адам
соғыс
кез
көп
ислам
күш
баста
қала
он
сом
ақ
басқа
ем
күрес
жиһад
джихад
жағдай
қарсы
аймақ
сия
қазір
келе
ма
топ
деген
дей
кел
на
халық
алла
армия
ретінде
ел
өлтір
әл
көр
содан
әскер
қайындық
байланысты
әрекет
жол
өлтіру
мүмкін
жақында
ала
жақсы
газ
израиль
хабарла
әскери
қолдау
басшы
қай
мәлімде
үлке
отыр
жаса
белгі
шариғат
уақыт
террор
арасында
қолдан
толық
операцияла
жылы
кара
сарбаз
қару
көтеріл
сирия
мәселе
көмектес
батыс
шақыр
бөлік
келіс
бер
басқар
сөз
армия
нәрсе
солтүстік
әлем
бағыт
ара
шығар
бу
ең
кон
тырыс
жетекші
көпшілік
оқиға
алма
мүны
болат
болу
хал
ұйым