Краткая информация о проекте

Наименование	AP19676342, “Мультиклассификация идеологических направлений кибер экстремизма на казахском языке методами искусственного интеллекта”
Актуальность	Экстремисты могут использовать интернет для распространения своих идей, набора новых сторонников, координации действий и даже для планирования и совершения преступлений. Интернет предоставляет им широкие возможности для коммуникации и пропаганды, что делает важной работу по контролю за контентом и противодействию экстремистским группировкам в онлайн среде. Определение идеологии экстремистской речи имеет большое значение в современном мире, особенно в контексте онлайн-коммуникаций. Актуальность этого определения обусловлена рядом факторов. Правительства и общественные организации стремятся противодействовать экстремизму и терроризму, и для этого важно четко определять идеологии, лежащие в их основе. В интернете распространение экстремистских идей может привести к радикализации и негативным последствиям. Определение и борьба с такой речью помогают создать безопасное онлайн-пространство. Сотрудничество между странами и международными организациями в борьбе с экстремизмом требует ясного определения идеологий, чтобы обеспечить единые стандарты и подходы. Таким образом, актуальность определения идеологий экстремистской речи неоспорима в современном информационном обществе.
Цель	Цель проекта-изучение распространения киберпропаганды деструктивного характера с помощью методов искусственного интеллекта в социальных сетях и мессенджерах, создание моделей мультиклассификации в идеологических направлениях экстремистского содержания в текстовых, аудио-и видео публикациях и создание наиболее активной киберпропаганды религиозного контента, моделей и методов выявления деструктивных сообществ.
Задачи	1. Разработка модуля сбора данных. 2. Создание корпуса для мультиклассификации идеологического направления экстремистского содержания. 3. Создание модели мультиклассификации, определяющей идеологическую направленность экстремистского содержания в текстовых публикациях социальных сетей и мессенджеров (пропаганда деструктивных религиозных течений, радикализация и вовлечение в экстремистские и террористические организации). 4. Создание моделей мультиклассификации для определения идеологической направленности экстремистского содержания в аудио-и видеопубликах социальных сетей и мессенджеров 5.Создание модели мультиклассификации социальных сетей и мессенджеров, определяющей экстремистское (религиозный

	экстремизм, политический экстремизм, ксенофобия) содержание в текстовых, аудио, видео публикациях. 6. Создание гибридной модели выявления наиболее активных киберпропаганд в социальных сетях и мессенджерах 7. Разработка модели, алгоритмов и методов выявления сообществ в социальных сетях на основе заданного набора параметров. 8. Создание чат-бота с диалогом на казахском языке для консультирования по вопросам религии. 9. Разработка программного обеспечения, реализующего разработанные методы и модели.
Ожидаемые и достигнутые результаты	Достигнутые результаты: Разработан модуль сбора данных. Модуль сбора данных разработан с использованием технологий API для поиска данных в социальных сетях Телеграм, ВКонтакте, Твиттер, Youtube. В выбранных социальных сетях проанализировано более 400 групп с признаками деструктивных убеждений. В модуле сбора данных реализован функционал пополнения базы данных в соответствии со списком ключевых слов и на основе выбранного временного интервала, запроса идентификатора группы. Создан корпус мультиклассификации экстремистского содержания и идеологической направленности. Для определения идеологической экстремистской направленности, собирающей текстовые данные из социальных сетей и новостных сайтов, были определены правила и категории экстремистских и нейтральных текстов. Составлены списки ключевых слов для каждого класса, правила, списки групп для сбора данных. Корпус отсортирован вручную по правилам. В результате аннотации классам были присвоены следующие обозначения: Propaganda (0), Radicalization (1), Recruitment (2), Neutral (3). Собранный корпус был разделен на наборы для обучения и тестирования в соотношении 80% и 20% с целью применения методов машинного обучения. В каждом классе собрано более 2000 текстов. К данным корпуса применялись такие алгоритмы препроцессинга, как токенизация, очистка от пунктуации, очистка от наиболее распространенных слов, очистка от стоп-слов. Составлена статистика и визуализация корпуса. Использовались алгоритмы и модели Word2vec, bag of words и n-gram, и корпус был подготовлен к машинному обучению. Создана модель мультиклассификации идеологического направления экстремистского содержания в текстовых публикациях социальных сетей и мессенджеров (пропаганда деструктивных религиозных течений, радикализация и вовлечение в экстремистские и террористические организации). К данным корпуса применялись алгоритмы word2vec, tf-idf. Поскольку тексты в веб-ресурсах в основном находятся в неструктурированном состоянии и заполняются различными пользователями, существует

	множество орфографических ошибок. Поэтому в первую очередь был предложен метод на основе Spell Checker для исправления орфографических ошибок в казахском языке. Упомянутый метод является очень полезной функцией любой поисковой системы. Самое простое решение – отсортировать все позиции словаря по мере увеличения редакционного расстояния и отображать только первые несколько позиций. В качестве редакционного расстояния можно взять расстояние Левенштейна. Расстояние Левенштейна-показывает минимальное количество попыток ввода/удаления/изменения символов, необходимых для преобразования исходной строки в целевую строку. Для определения идеологического направления в тексте был проведен сравнительный анализ классических машинных алгоритмов, таких как Logistic Regression, KNN, SVM, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting. На базе Spellcheker+Stemming+TF-idf+LSTM+BERT была создана новая модель. Гиперпараметры модели Берта: вход = 128 слов или токенов, линейная классификация = 4, bert_model_name = 'Bert-base-multilingual-uncased', num_classes = 4, max_length = 128, batch_size = 64, num_epochs = 10, learning_rate = 2e-5. Кроме того, для решения задачи мультиклассификации идеологических текстов были объединены 2 алгоритма глубокого обучения (BERT и LSTM, Bert+linear). Для LSTM модель принимает последовательность текста в качестве входных данных вместе с соответствующими длинами каждой последовательности. Он встраивает текст (вложенный размер = 20), обрабатывает его через слой LSTM (размер = 64), передает последнее скрытое состояние через полностью добавленные слои с активациями ReLU и, наконец, использует сигмовидную активацию для получения единственного выходного значения. Гиперпараметры объединенной модели: вход = 128 слов или токенов, BERT = 768, LSTM = 256, DropOut = 0.2, linear Classification = 4, bert_model_name = 'Bert-base-multilingual-uncased', num_classes = 4, max_length = 128, batch_size = 64, num_epochs = 20, learning_rate = 2E-5. Ожидаемые результаты: Модели машинного и глубокого обучения для мультиклассификации экстремистской направленности текстовых, аудио-и видеоматериалов в социальных сетях и мессенджерах (религиозный экстремизм, политический экстремизм, ксенофобия и др.), гибридная модель, которая представляет собой комбинацию различных глубоких нейронных сетей, выбираемых в зависимости от типа меток, предназначенную для выявления наиболее активной киберпропаганды в социальных сетях и мессенджерах.
Имена и фамилии членов исследовательской группы с их идентификаторами (Scopus Author ID, Researcher ID,	1. Мусиралиева Шынар Женисбековна, ORCID: https://orcid.org/0000-0001-5794-3649 , Scopus профайл: https://www.scopus.com/authid/detail.uri?authorId=57202216979 , Web of Science профайл: https://www.webofscience.com/wos/author/record/2394890 2. Болатбек Милана Асланбеккызы, ORCID: https://orcid.org/0000-0002-2153-180X , Scopus профайл

ORCID, при наличии) и ссылками на соответствующие профили	сілтемесі: https://www.scopus.com/authid/detail.uri?authorId=57202834055 , Web of Science: https://www.webofscience.com/wos/author/record/GZL-7318-2022 3. 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 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. Нарбаева Салтанат Муратбековна, ORCID: https://orcid.org/0000-0001-5230-3781 , Scopus: https://www.scopus.com/authid/detail.uri?authorId=57216484412 , Web of Science профайл: https://www.webofscience.com/wos/author/record/KBA-1599-2024 6. Медетбек Жанар ORCID: http://orcid.org/0000-0001-7536-5889 7. Беккожаева А. 8. Аетова Багдат
Список публикаций со ссылками на них	1. Ш.Ж. Мусиралиева, Р.Қ. Оспанов, М.А. Болатбек, Ж.Б. Медетбек Профилирование пользователей социальных сетей на основе демографических данных. (КОКСОН) Журнал Вестник АУЭС, Том 3 № 62(2023), стр. 133-144. https://vestnik.aues.kz/index.php/none/issue/view/95/116 2. Мусиралиева Ш.Ж., Байспай Г.Б., Болатбек М.А., Сағынай М., Терейковский И.А. Веб-ресурстардағы экстремисттік мәліметтерді анықтауға арналған машиналық әдістерді оқыту және сынау үшін қазақ тіліндегі мәтіндер корпусын құру. Журнал Труды университета (КарГТУ), №3, 2023, стр 453-458. Рекомендовано КОКСОН, tu.kstu.kz/archive 3. Ш.Ж. Мусиралиева, М.А. Болатбек, М. Сағынай, Ж.Ы. Елтай, К.Б. Багитова. Экстремистік мәліметтер түсінігі және экстремизмге қарсы күрес жобаларына жүйелік шолу. (КОКСОН) . Журнал NEWS OF THE NATIONAL ACADEMY OF SCIENCESOF THE REPUBLIC OF KAZAKHSTAN PHYSICO-MATHEMATICAL SERIES, ISSN 1991-346X Volume 3. № 347 (2023). 112–130. https://journals.nauka-nanrk.kz/physics-mathematics/article/view/5792 4. Ш.Ж.Мусиралиева, Ж.Б.Медетбек , М.А.Болатбек , Жумаханова А.Н., Ж.Ы.Елтай. ӘЛЕУМЕТТІК МЕДИАДАҒЫ ЭКСТРЕМИСТІК МАЗМҰНДЫ АНЫҚТАУ ЖӘНЕ ЗЕРТТЕУ. В материалах конференции. VIII Международная научно-практическая конференция «Информатика и прикладная математика» посвященной памяти 85 д.т.н, профессора Бияшева Р.Г. , 26 – 27 октября 2023 г., 265-272стр., Алматы, Казахстан

	5. Shynar Mussiraliyeva, Milana Bolatbek, Moldir Sagynay, Aygerim Zhumakhanova, Zhastay Yeltay, and Zhanar Medetbek. 2024. Identifying Cyber-Threatening Texts in the Kazakh Segment of Web Resources. In Proceedings of the 2023 7th International Conference on Advances in Artificial Intelligence (ICAAI '23). Association for Computing Machinery, New York, NY, USA, 68–72. https://doi.org/10.1145/3633598.3633610
Информация о патентах	-