Brief information about the project

Name of the project	IRN AP19576868 Development of models and methods to identify youth extremism and ensure the safety of youth in the modern information space
Relevance	Extremist speech on the Web can serve as a tool for planning and carrying out extremist actions, including terrorist acts. The definition of such speech allows you to identify potential security threats and prevent their implementation. The Internet is often used to spread ideas of extremism and radicalization. The definition of extremist speech allows early detection of signs of radicalization and warns against possible negative consequences. The definition of extremist speech is necessary to balance freedom of speech and the protection of society from potential danger. This makes it possible to distinguish between legitimate statements and those that may pose a threat to public safety. Many countries have laws regulating extremist activity and speech. The definition of extremist speech helps to comply with laws and prevent illegal actions. Defining extremist speech helps create a safe online space for users, especially young people, who may be more vulnerable to the impact of extremist ideas. Extremist speech can provoke social tension, strife and conflict. Defining such speech helps to prevent the spread of hatred and contributes to the creation of a more coherent society. Identifying extremist speech on the Internet requires cooperation between countries and organizations. This helps to effectively control and counter global threats. All these aspects emphasize the importance of defining extremist speech on the Internet to ensure the safety of society, prevent radicalization, and maintain a balance between freedom of speech and the duty to ensure public safety.
Purpose	The aim of the project is to research and develop models and methods of semantic analysis to identify and counter the spread of violent, national extremism, racism, bullying among young people, methods of monitoring and analyzing traffic on the network to counter the spread of illegal ideology among young people, create a list of potentially dangerous web resources for young people, adapt methods of psycho-emotional analysis for the Kazakh language.
Objectives	**1.** Development of new models and methods for defining texts of national, violent extremism, bullying and racism aimed at young people 1.1 Analysis of available texts in the chosen field and identification of the main sources of information 1.2 Development of a parser for collecting data from web resources 1.3 Building a corpus of texts of national, violent extremism, bullying and racism aimed at young people 1.4 Data preprocessing in the case 1.5 Defining a set of features to improve the task of detecting national, violent extremism, bullying and racism on web resources

	1.5 Development of new models and methods of semantic analysis to identify texts of national, violent extremism, bullying and racism aimed at young people in the Kazakh language 1.6 Adaptation of methods of analysis of psychoemotional analysis of texts for the Kazakh language 2. Development of new methods for analyzing and monitoring network traffic 2.1 Development of a network data collection module. 2.2 Development of a module for analyzing processed traffic logs. 2.3 Development of a method for analyzing and monitoring network traffic based on machine learning. Creating a list of potentially dangerous websites for young people 3. Development of software to identify and counter the spread of violent, violent extremism, racism and bullying among young people 3.1 Architecture Design 3.2 Implementation of the server and front end 3.3 Software product testing
Expected and achieved results	Achieved results: New models and methods have been developed to identify texts of national, violent extremism, bullying and racism aimed at young people. A review of new literature in domestic and foreign publications on the identification of extremist texts on Web resources has been conducted. New models and methods have been developed to identify national, violent extremist, bullying and racist texts aimed at young people. One article was published on the use of machine learning methods such as support vector machines, naive Bayesian classifiers, random tree methods, decision tree, k nearest neighbor algorithm, logistic regression, gradient boosting, to detect extremist texts. An extensive review was conducted of existing methods of classifying texts related to national, violent extremism, bullying and racism aimed at young people on the Internet. The review includes recent publications published in highly regarded scientific journals such as Springer, Elsevier and others included in the Web of Science and Scopus databases. The analysis of the literature helped to determine the current state of research in this area and identify the current directions of our project. This analysis and review of the current state of methods for detecting extremist texts will be useful for the community of researchers and engineers working in this field. This allows them to apply these methods more effectively in their work and contribute to the development of this important area. Traditional machine learning methods, methods and models based on transformers have been created to identify texts related to national, violent extremism, bullying and racism on the Internet. With the help of search engines, studies of publicly available texts on various web resources (social networks Vkontakte, Twitter, YouTube, Telegram, blogs,

forums, news articles) were identified and conducted. As a result of this research, key phrases and primary sources of texts related to national, violent extremism, bullying and racism were identified.

A parser has been developed to collect texts from the web resources of the social networks Vkontakte, Twitter, YouTube, Telegram using the identified keywords. The domain name of the source, the storage location and the review period are given for the input of the parser. As a result, the text content of the specified web resources is loaded. API technologies were used.

As a result of the compiled parser, a text corpus was created, compiled from the content of groups and channels on social networks Vkontakte, Twitter, YouTube, Telegram. The corpus includes 5 categories: national extremism, violent extremism, racism, bullying and texts of neutral categories, each category is assigned appropriate designations (from 0 to 4). The linguistic and statistical analysis of the corpus texts was carried out. The total volume of the corpus is about 10,000 texts.

Preprocessing algorithms were performed for the texts of the collected text corpus: including tokenization, morphological analysis of texts, stemming, removal of punctuation marks, removal of numeric values and hyperlinks in the text, removal of stop words.

The signs that increase the accuracy of the definition of texts of national, violent extremism, bullying and racism aimed at young people such as tf-idf, tf-idf-bigram, bag-of-words are identified. These features are used in the compilation of models and methods related to the identification of destructive content in the text.

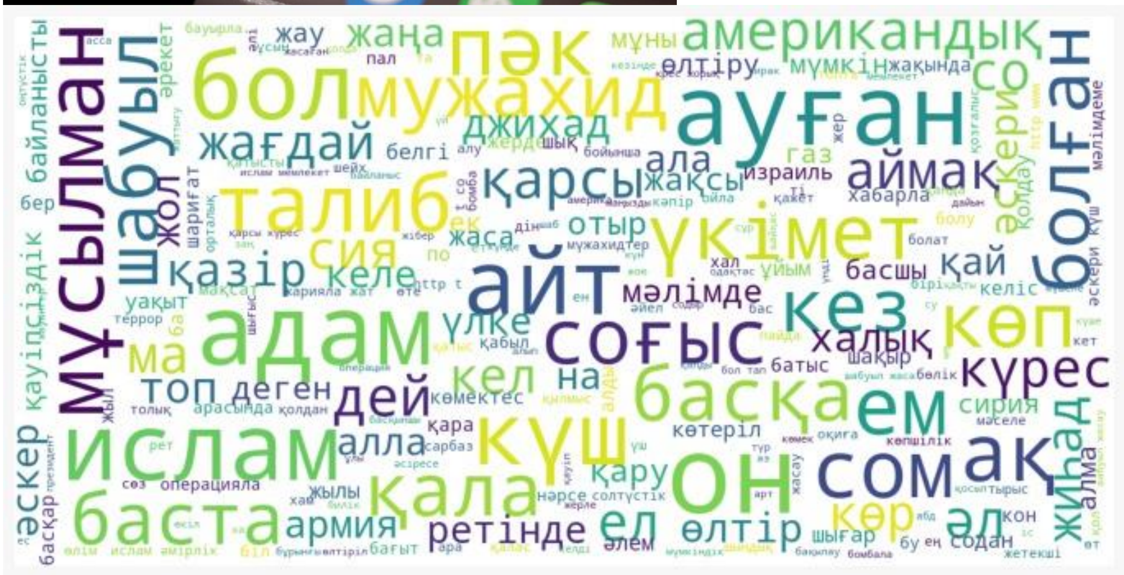
Based on the machine learning methods of Decision Trees, Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine, LSTM, BiLSTM, work was carried out to develop new methods and models of semantic data analysis to identify national, violent extremism, bullying and racism aimed at young people. A model based on Stemming+TF-idf+BERT has been created. In addition, a model of psychoemotional analysis has been built to improve the accuracy of determining texts in this category. A model was created based on transformers such as DistilBert and Roberta. RoBERTa improves BERT by carefully and intelligently optimizing reading hyperparameters. The RoBERTa model was developed in Pytorch. Model hyperparameters: Input = 128 words or tokens, 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 The Distilbert-based semantic analysis model for identifying texts of national, violent extremism, bullying and racism in the Kazakh language is aimed at optimizing learning by reducing size and increasing speed, all this was done in order to preserve productivity. Hyperparameters of the model: Input = 128 words or tokens, DistilBERT = 768 vectors, 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. An MLM model (masked language model -"masked language modeling") was also created to identify national, violent extremism, bullying and racism aimed at young people. Model hyperparameters: Input = 128 words or tokens, 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 Well-known methods of psychoemotional analysis of texts have been adapted for the Kazakh language, an analyzer of psychoemotional lexemes in texts of national, violent extremism, bullying and racism aimed at young people has been developed. During the research work, the extremist linguistic corpus was analyzed using the word counting strategy and the liwc closed dictionary method. The task of the proposed method is to search and count words belonging to psychological categories in a set of text data. In total, more than 80 categories were identified. The result of processing a text file in the program are the following output variables: the number of words, cumulative language variables (analytical thinking, influence, text specificity and emotional tone) and the percentage of words in the text, which represents the percentage of words in the text (for example, pronouns, articles, auxiliary verbs, etc.), categories, influencing psychological structures (for example, affect, cognition, biological processes, impulses), the category of personal interests (for example, work, home, rest), informal language markers (for example, curses). The results indicate the benefits of general theoretical knowledge about the expression of levels of personality development in the ways of using words. Using the developed method, an LSTM-based model was created for classification by national extremism, violent extremism, racism and Bullying texts. The author's certificate for the compiled module has been received. Expected results: New methods will be developed to analyze and monitor network traffic in order to identify web resources that are dangerous for young people. Software will be developed to identify web resources with content of national, violent extremism, bullying and racism aimed at young people.
Research team members with their identifiers (Scopus Author ID, Researcher ID, ORCID, if available) and links to relevant profiles	1. Bolatbek Milana, ORCID: https://orcid.org/0000-0002-2153-180X , Scopus profile: https://www.scopus.com/authid/detail.uri?authorId=57202834055 , Web of Science профайл сілтемесі: https://www.webofscience.com/wos/author/record/GZL-7318-2022 2. Baisylbayeva Kymbat, ORCID: https://orcid.org/0000-0001-9753-0398, Web of Science profile: https://www.webofscience.com/wos/author/record/N-9664-2017 3. Sagynay Moldir, ORCID: https://orcid.org/0009-0004-1377-5742 4. Yeltay Zhastay, 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. Akhmed Gulmaral, ORCID: https://orcid.org/0000-0002-4464-9544 6. Meirbekova Bibinur, ORCID: https://orcid.org/0000-0001-9215-9382 , Scopus profile: https://www.scopus.com/authid/detail.uri?authorId=57212476113 ,

	Web of Science profile: https://www.webofscience.com/wos/author/record/ABD-4499-2021 7. Shayzat Medet, ORCID: https://orcid.org/0000-0002-1651-8205 , Scopus profile: https://www.scopus.com/authid/detail.uri?authorId=57216968174 8. Raiymkulova Alima
List of publications with links to them	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
Patents	-

facebook
YouTube
Microsoft

Social
Clubhouse
FaceTime
Instagram
Messag
Messenger
Phone
Signal
Skype
15


американдық
пәк
мұжаһид
бол
ауған
болған
мұсылман
шабуыл
жағдай
джихад
қарсы
аймақ
талиб
үкімет
әскери
қазір
келе
айт
кез
адам
соғыс
көп
халық
күрес
баста
ислам
күш
басқа
ем
сомақ
он
қала
армия
ретінде
өлтір
ел
көр
әл
жиһад
сирия
дей
топ
деген
жаңа
жау
өлтіру
мүмкін
жақында
израиль
мәлімде
үлке
кел
на
алла
қару
әскер
қауіпсіздік
байланысты
жол
шариғат
сия
ма
уақыт
террор
отыр
жақсы
ала
белгі
хабарла
газ
қай
басшы
келіс
батыс
шақыр
көмектес
көтеріл
жылы
операцияла
арасында
қолдан
ұйым
мұжахидтер
жаса
сарбаз
қара
соңдан
шығар
бу
ең
кон