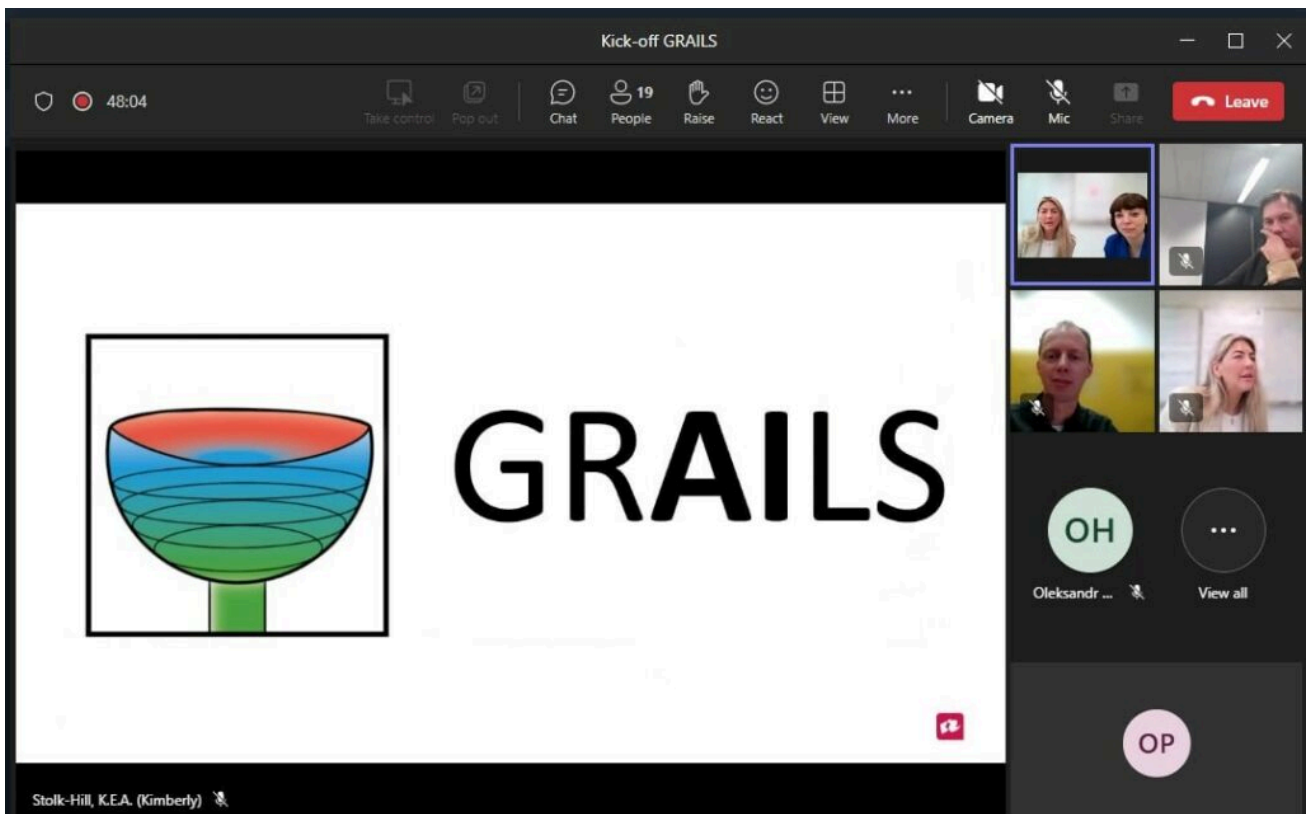




## Перемога студентів Хмельницького національного університету в міжнародному конкурсі Grails

19.06.2024

З гордістю повідомляємо про блискучий успіх студентів Хмельницького національного університету на престижному міжнародному конкурсі Grails – Global Responsible AI League for Students. Цей конкурс, організований Роттердамським технічним університетом у Нідерландах, став майданчиком для презентації відповідальних IT-рішень із застосуванням штучного інтелекту. Хмельницький національний університет був представлений двома командами, які зайняли призові місця, продемонструвавши високий рівень професіоналізму та інноваційності.



Конкурс Grails орієнтований на розвиток відповідальних та етичних IT-рішень із використанням штучного інтелекту. Участь у конкурсі складалася з декількох етапів, зокрема:

- Kick-off зустрічі з засновником конкурсу Пітером ван Богеманом, де учасники були ознайомлені з правилами та етичними принципами застосування ШІ.
- Лекції щодо ролі ШІ у суспільстві.
- Підготовки проєктів та запису демонстраційних відеороликів.
- Захисту проєктів перед членами журі.

Такий комплексний підхід дозволив учасникам глибоко зануритись у тему відповідального використання штучного інтелекту та розробити унікальні проєкти

Перше місце здобула команда кафедри комп'ютерних наук «Vector-KhNU» у складі:

Мазурець Олександр (к.т.н., доцент кафедри комп'ютерних наук – голова команди, координатор від ЗВО)

Молчанова Марина (студентка КНас-21-2, викладач кафедри комп'ютерних наук)

Собко Олена (студентка КНас-21-2, викладач кафедри комп'ютерних наук)

Залуцька Ольга (студентка КНМ-22-1, магістр)

Гардиш Дарина (студентка КН-21-1)

Шебетко Ольга (студентка КН-21-1)

Ільяна Марія (студентка КН-22-3)

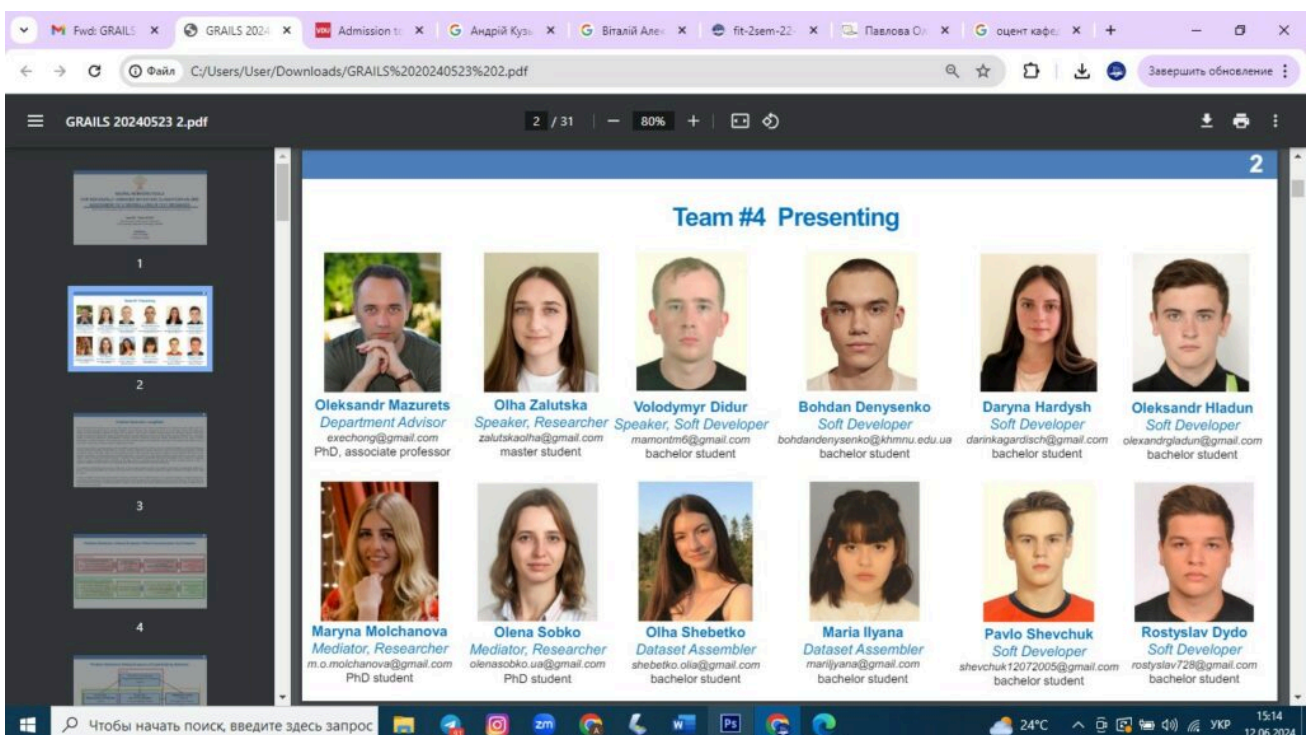
Шевчук Павло (студент КН-22-2)

Денисенко Богдан (студент КН-22-2)

Дідур Володимир (студент КН-22-1)

Гладун Олександр (студент КН-22-1)

Дидо Ростислав (студент КН-22-1)



Команда представила проект «Нейромережеві інструменти для індивідуально неупередженого виявлення, класифікації та оцінки кібербулінгу в текстових повідомленнях». Метою цього проекту є створення безпечного та толерантного комунікативного середовища, спрямованого на запобігання агресивним проявам та конфліктним ситуаціям.

**Problem Statement: LongRead**

High intelligence technologies pose unique challenges and risks that require careful consideration to protect the rights, safety, and well-being of people affected by their activities. Recent neural networks have become a source of unprovoked content that affects the mental and communication abilities of people who perceive it. As a result, further communication and activities, people involuntarily experience the personal and ethical semantic content and thus add an additional negative content to their communication. Such content generated by artificial systems. The consequences of the cognitive aspect of the abstract content of the report are well known and will continue to be in the context of use of a digital aggressive content.

The origin of content in communication and information exchange is usually hybrid. It can be content generated by artificial intelligence, content artificial generated and adjusted by a person, content created by a person taking into account the remembered semantic results of artificial origin. In the latter case, the basic artificial content of the message may contain aspects of human elements and have its own semantic orientation. It demonstrates how artificial intelligence with its growing capabilities significantly affects various aspects of human society, culture, thinking and communication. Therefore, the problem of automated checking of messages for the presence of negative content is urgent. Considering the complexity of the task, when it is necessary to track both the author's intention and the content of semantic models of artificial origin. It is a complex task that can be solved only with artificial intelligence.

It is possible to distinguish the sources of influence of the socially negative content. The negative component is deliberately added by a person, involuntarily added by a person as a result of the reproduction of artificial semantic models, artificially generated content used by a person.

In order to identify and evaluate the content in the textual content of an artificial content, it is relevant to develop innovative methods or approaches for identifying, checking and evaluating the signs associated with manifestations of bullying and abuse. This will allow not only to target the influence of artificial semantic models on the mental and communication abilities of people, but also to determine the subjects of such influence.

**Problem Statement: Ethical AI aspect of Non-Discrimination Text Datasets**

**Problem Statement: Ethical AI aspect of Cyberbullying Detection**

**Literature Review**

Author	Title	Year	Summary
Smith, J. & Doe, A.	AI in Cyberbullying Detection	2023	Exploring the use of neural networks for identifying harmful content in digital communication.
Chen, L. & Kim, S.	Ethical AI and Bias in Text Classification	2022	Discussing the challenges of ensuring fairness and transparency in machine learning models.
Patel, R. & Garcia, M.	Deep Learning for Sentiment Analysis	2021	Reviewing the effectiveness of deep neural networks in understanding user emotions.
Johnson, K. & Lee, H.	Generative AI and Content Moderation	2024	Investigating the risks and benefits of AI-generated content in online environments.
Williams, T. & Brown, D.	AI Ethics and Data Privacy	2020	Examining the intersection of artificial intelligence and user data protection.

**The Team Project Aim**

Our project aim is to promote creation of safe and tolerant communication environments that targeted the potential risks of aggressive manifestations and conflict situations, which is a useful goal for social and professional needs of human communication.

**Evaluation of the representativeness of test datasets for training neural networks according to the FATE principle of fairness - non-discrimination on various grounds (like religion, age, gender etc.)**

- Evaluation and correction of test datasets for ethically sensitive grounds: Age, gender, race, religion, etc.

**Detection and classification of cyberbullying in text content for automatic detection of ethically incorrect influences, including human and artificial, intentional and unconscious.**

- Ethically incorrect content for detection: Human generated, Human generated using AI, Human generated under AI influence.

**Project Overview: Completed Research Tasks**

1. Analysis of methods, tests and technologies for automated detection of cyberbullying in text content.
2. Development of method for evaluating the representativeness of test datasets according to the FATE principle of justice - ethical non-discrimination based on age, race, religious etc.
3. Creation of informative system for validation of method for evaluating the representativeness of test datasets.
4. Research of the effectiveness of the developed method for evaluating the representativeness of test datasets.
4. Applied study of Machine Learning classifier models for cyberbullying detection, to select the most effective models for detecting various types of cyberbullying.
5. Development of the method for detecting and identifying cyberbullying in text content.
6. Creation of informative system for validation of method for detecting and identifying cyberbullying in text content.
7. Research of the effectiveness of the developed method for detecting and identifying cyberbullying in text content.

**Method of Evaluating the Text Datasets Representativeness**

**Method of Evaluating the Text Datasets Representativeness**

Dataset Name	Used Datasets	Base language	Total count of sample	Class
Openaer	Twitter Files for Gender Gleaning	English	34 148	Women, Men
Religion	Cyberbullying Classification and Cyberbully Detection Dataset	English	8000	Anti-Male, Christian, Muslim, Hindu
Age	Multi-Dataset Distribution	Indian	15 000	18-24, 25-34, 35-44, 45-54, 55-64

**Method of Evaluating the Text Datasets Representativeness**

**Applied Study of Machine Learning Classifier Models for Classification Types:**

Machine Learning Model	Dataset Type	Results of Machine Learning	Assessment
Feedforward Neurons Machine Learning Algorithm	Gender	Accuracy 0.88, Precision 0.92, F1 Score: 0.90	Good
SVM	Religion	Accuracy 0.85, Precision 0.88, F1 Score: 0.86	Good
Feedforward Neurons Machine Learning Algorithm	Age	Accuracy 0.82, Precision 0.85, F1 Score: 0.83	Good
Feedforward Neurons Machine Learning Algorithm	Religion	Accuracy 0.81, Precision 0.84, F1 Score: 0.82	Good

**Method of Evaluating the Text Datasets Representativeness**

**Example: classification by Openaer**

**Method of Evaluating the Text Datasets Representativeness**

**Example: classification by Age**

**Method of Evaluating the Text Datasets Representativeness**

**Example: classification by Religion**

**Method of Evaluating the Text Datasets Representativeness**

**Soft for View Dataset Labels**

**Method of Evaluating the Text Datasets Representativeness**

**Balance of distribution by feature: Cyberbullying type**

Судді високо оцінили інноваційний підхід та ретельне опрацювання технічних та етичних аспектів.

“Команда-переможець зосередилася на виявленні та оцінці кібербулінгу, що є неабияким викликом. Команді вдалося розробити новий підхід, який демонструє вражаючі інноваційні та творчі здібності студентів” – підкреслив засновник конкурсу Пітер ван Богеман.

Команда кафедри комп'ютерної інженерії ХНУ «EcoMindInnovators» посіла почесне друге місце. До складу команди входили:



– Віталій Алексейко (студент KI2M-22-1, магістр)

– Андрій Кузьмін (студент KI2-19-1)

– Валерія Швайко (студентка KI2-20-1)

– Юлія Ільчишина (студентка KI2-20-1)

Науковими керівниками команди були доктор філософії, доцент кафедри комп'ютерної інженерії Ольга Павлова та старший викладач Єгор Іштван.






Команда представила проєкт «EarthSaverAI» – веб-орієнтований застосунок на базі ШІ для прогнозування змін температурного режиму у різних кліматичних зонах планети.



Dataset	
Main Climate Zone	Number of countries
A	94
B	45
C	64
D	26
E	6

Important aspects:

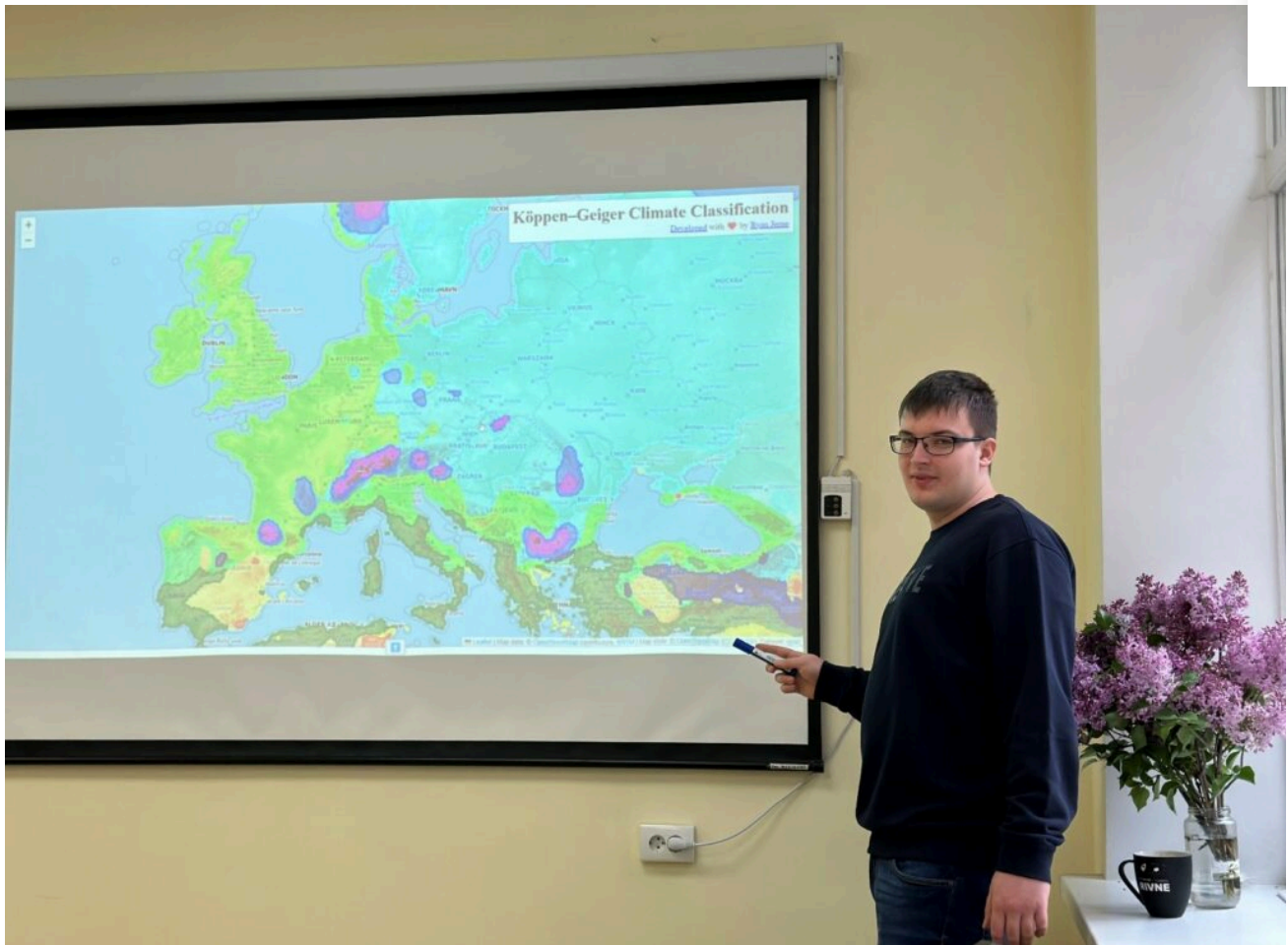
- Area of Countries;
- Geopolitical factors;
- Selecting Data Sources.






Проект вражаюче інтегрує сучасні технології та етичні аспекти, відповідаючи на актуальні глобальні виклики, що підкреслено і в коментарях керівника GRAILS Пітера ван Богемана:

*“Судді були дуже задоволені досягненнями команд з Хмельницького національного університету. Команда Earthsave, яка посіла друге місце, очевидно, доклала багато зусиль для вивчення етичних аспектів проекту та неймовірно ефективно перетворила суспільну потребу у функціональний додаток.”*



Загальні питання: [centr@khnmu.edu.ua](mailto:centr@khnmu.edu.ua)

Подача новин та анонсів: [press@khnmu.edu.ua](mailto:press@khnmu.edu.ua)

Центр кар'єри

Скринька довіри

Цивільний захист

Пожежна безпека

Охорона праці



Хмельницький національний університет, 2024