

## **ВИКОРИСТАННЯ ТЕХНОЛОГІЇ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ЗАДЛЯ ВИЗНАЧЕННЯ ДОСТОВІРНОСТІ ПУБЛІЧНОЇ ІНФОРМАЦІЇ В МЕДІА**

**Редозуб І.А., Хайрова Н.Ф.**

*Національний технічний університет  
«Харківський політехнічний інститут», м. Харків*

Розвиток штучного інтелекту, особливо генеративного штучного інтелекту, а також глобалізація світу призводять до безпрецедентного збільшення публічної інформації в інтернеті. З кожним днем ми все частіше зустрічаємось з великою кількістю новин. Людина без спеціальних знань або освіти не може критично обробити такий великий об'єм інформації. У нашому дослідженні ми пропонуємо використати сучасні великі мовні моделі задля виявлення недостовірної маніпулятивної інформації.

В нашій роботі ми ставимо перед собою такі задачі як:

1. Створення датасету європейських, українських та російських воєнних новин об'єднаних за подією.
2. Створення таксономії типізації маніпулятивної інформації у медіа.
3. Створення розміченого синтетичного датасету прикладів
4. Проведення аналізу джерел публічної інформації на наявність маніпулятивної інформації.

На першому етапі нашої роботи для досягнення мети дослідження було розроблено датасет RUWA(Russian-Ukrainian War)[1], який налічує більше 16500 статей з відомих медіаресурсів. Датасет розподілений за 13 подіями 2022 року російсько-української війни. Крім того даний датасет включає в себе такі аспекти як: ключові слова, джерело, опис, заголовки, дата та субподії.

Після цього за допомогою великих мовних моделей було розроблено ієрархічну таксономію типізації маніпуляцій у медіа. Ця таксономія включає в себе 5 основних видів маніпулятивної інформації, а саме: емоційні маніпуляції, логічні маніпуляції, лінгвістичні та стилістичні маніпуляції, структурні та форматні маніпуляції, інституційні маніпуляції. В свою чергу кожен вид залучає декілька типів маніпуляцій. Використовуючи створену таксономію ми розробили розмічений синтетичний датасет прикладів маніпуляцій. Загалом датасет налічує 330 прикладів маніпуляцій(по 10 на кожен тип) з визначенням виду і типу кожної.

У подальшій роботі ми плануємо натренувати мовну модель на нашому датасеті прикладів та провести аналіз використання маніпуляцій.

### **Література:**

1. Khairova, N., Galassi, A., Scudo, F. L., Ivasiuk, B., & Redozub, I. (2024). Unsupervised approach for misinformation detection in Russia-Ukraine war news. CLW-CoLInS 2024, Computational Linguistics Workshop at Colins 2024: Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume IV: Computational Linguistics Workshop, Lviv, Ukraine, April 12-13, 2024, IV, 21–36. URL: <https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-227968>