

УДК 004.89:001.89

## **ПІДХІД ДО СТВОРЕННЯ ІНТЕЛЕКТУАЛЬНОГО ПОМІЧНИКА ДЛЯ НАУКОВИХ КОНФЕРЕНЦІЙ НА ОСНОВІ LLM ТА PFE**

Теплюк М.В., Гриньова О.Є.

e-mail: maksym.tepliuk@nure.ua, olena.hrynova@nure.ua

Харківський національний університет радіоелектроніки, каф. ШІ  
м. Харків, Україна

The article describes the development of an intelligent assistant for the website of the International Forum of Young Scientists «Radio Electronics and Youth in the XXI Century» to automate the search for similar scientific papers and answer frequently asked questions. A combination of NLU and PFE is used. The system is based on the SBERT model for fast semantic text comparison, Ukr-RoBERTa-base for Ukrainian language processing, and ChatGPT for checking the structure of papers. This solution will enhance the forum's efficiency and the quality of research.

Міжнародний молодіжний форум (ММФ) виступає важливою платформою для обміну ідеями, презентації наукових досліджень та налагодження співпраці між молодими дослідниками з різних країн, де організатор – Харківський національний університет радіоелектроніки. Щороку захід збирає велику кількість учасників, які подають свої наукові роботи для обговорення та рецензування. Однак зі зростанням масштабів форуму виникають нові виклики, пов'язані з ефективною організацією процесу перевірки та оцінювання представлених матеріалів. У зв'язку з цим актуальним стає розробка інтелектуального асистента на сайті форуму. Такий помічник зможе автоматизувати пошук схожих робіт, надавати учасникам відповіді на типові запитання, а також допомагати організаторам у швидкому та об'єктивному відборі досліджень.

Ручна перевірка та порівняння великого обсягу текстів є трудомістким і часозатратним процесом, який до того ж може супроводжуватися людським фактором та суб'єктивністю оцінки. Це знижує ефективність проведення конференції та може негативно позначитися на її науковій цінності [1].

Для вирішення проблеми виявлення схожих наукових робіт пропонується впровадження сучасних технологій штучного інтелекту, зокрема комбінації Natural Language Understanding (NLU) та Prompt Flow Engineering (PFE). NLU забезпечує розуміння змісту наукових текстів, а PFE дозволяє ефективно керувати великими мовними моделями (LLM) шляхом створення послідовностей промптів для виявлення плагіату та відповідей на типові запитання.

PFE є інноваційним підходом до розробки чат-ботів та асистентів, який базується на створенні послідовностей промптів для керування поведінкою мовних моделей. Це дозволяє розробникам створювати більш

складні та контекстно-залежні діалогові системи. NLU є ключовою технологією для створення ефективних чат-ботів, що здатні розуміти та інтерпретувати людську мову.

Поєднання NLU та PFE дозволяє створювати більш інтелектуальні та адаптивні чат-боти, здатні ефективно взаємодіяти з користувачами та вирішувати складні завдання в різних галузях, від обслуговування клієнтів до освіти.

У PFE використовуються такі технології машинного навчання: великі мовні моделі (LLM) – застосовуються для генерації тексту та ведення діалогу; нейронні мережі (зокрема, трансформери) – для обробки природної мови та розуміння контексту; мультимодальне навчання – обробляти дані різних типів (текст, зображення, аудіо). Методи навчання: Few-shot learning – для налаштування моделі на конкретні завдання; Chain-of-thought – для покрокового виведення відповідей. PFE поєднує ці технології для створення ефективних послідовностей промптів та керування складними діалоговими системами на основі LLM.

Замість того, щоб відправляти одну велику і складний промпт моделі, завдання розбивається на кілька менших частин. Кожна частина є окремим етапом у процесі роботи з моделлю. Залежно від типу користувача, можна виділити наступні типи запитів: пошук, перевірка на плагіат та AI Content Detector, FAQ (рисунок 1).

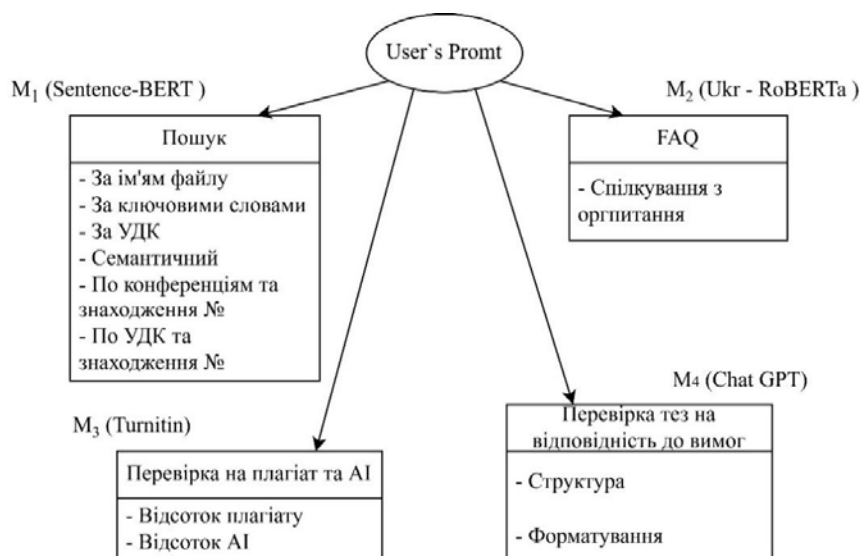


Рисунок 1 – Prompt Flow Engineering

Пошук та перевірка на плагіат. Для автоматичного виявлення схожих наукових робіт на вебсайті ММФ обрано модель M<sub>1</sub> – (SBERT). Основною причиною вибору саме цієї моделі є її здатність ефективно виконувати семантичне порівняння текстів. Стандартна модель BERT демонструє високу якість при розв'язанні різних завдань обробки природної мови, однак вона не оптимізована для порівняння великих обсягів текстових даних. SBERT долає це обмеження, використовуючи модифікацію

архітектури BERT, що дозволяє отримувати компактні та інформативні векторні представлення (ембедінги) речень і текстів. Завдяки цьому обчислення схожості між текстами за допомогою SBERT відбувається у сотні разів швидше порівняно з оригінальним BERT.

Крім високої продуктивності, SBERT забезпечує точність у виявленні як очевидних, так і семантично схожих робіт, що не обов'язково містять однакові слова, але мають подібний зміст. Це особливо важливо для наукових конференцій, де однакові дослідницькі ідеї можуть бути описані різними термінами. На відміну від інших моделей, таких як Universal Sentence Encoder (USE) або GPT-4, SBERT забезпечує баланс між точністю та продуктивністю, не вимагаючи значних обчислювальних ресурсів. Наприклад, GPT-4 хоч і демонструє виняткову точність у розумінні контексту, проте її використання може бути надмірним для завдання пошуку схожості через високу вартість обчислень та складність інтеграції. Universal Sentence Encoder, у свою чергу, поступається SBERT у точності при складніших семантичних завданнях. Turnitin може бути використаний для семантичного аналізу в моделі M<sub>3</sub>. Для комплексної перевірки робіт на відповідність вимогам буде інтегрована модель M<sub>4</sub>, а саме ChatGPT, яка аналізуватиме тексти на наявність структурних елементів, форматування та стилістичних норм. Використання моделі M<sub>4</sub> забезпечить швидкий та об'єктивний відбір матеріалів для ММФ.

Для вирішення організаційних питань (FAQ) у проведенні ММФ запропонована модель Ukr-RoBERTa-base (модель M<sub>2</sub>), це модель обробки природної мови, спеціально навчена для української мови [2]. Базується на архітектурі RoBERTa. Має 12 шарів, 768 прихованих нейронів, 12 головок уваги. Загальна кількість параметрів – 125 мільйонів.

Впровадження інтелектуального асистента на вебсайт ММФ дозволить значно оптимізувати роботу організаторів та рецензентів, а також надаватиме допомогу потенційним учасникам. Завдяки аналізу семантики тексту та виявленню прихованих зв'язків система швидко визначатиме ступінь схожості матеріалів, допомагаючи уникнути дублювання тематики та підвищуючи унікальність представлених досліджень. Такий підхід зменшить час рецензування, дозволяючи рецензентам зосередитися на оцінюванні справді релевантних робіт.

Список використаних джерел:

1. В.І. Жеребкін, С.Г. Удовенко, Л.Е. Чала, О.Є. Гриньова. Аналіз складності та побудова концептуальних графів масових відкритих онлайн-курсів// Біоніка інтелекту. – 2023. – Вип. 1 (99). С.26-37. [https://doi.org/10.30837/bi.2023.1\(99\).04](https://doi.org/10.30837/bi.2023.1(99).04) (дата звернення: 23.02.2025).

2. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084. URL: <https://aclanthology.org/D19-1410.pdf> (дата звернення: 23.02.2025).