

УДК 004.8

ДОСЛІДЖЕННЯ СПОСОБІВ ВИЗНАЧЕННЯ ТЕКСТУ ЗГЕНЕРОВАНОГО ШТУЧНИМ ІНТЕЛЕКТОМ

Олександр Сорокопуд

Національний авіаційний університет, Київ

Науковий керівник – Владислав Сорокопуд, PhD, ст. викладач

Ключові слова: штучний інтелект, нейронна мережа, NLP, Велика мовна модель

Вступ

Штучний інтелект (ШІ) - це галузь інформатики, яка займається розробкою інтелектуальних машин, здатних виконувати завдання, які зазвичай потребують людського інтелекту [1]. В повсякденному житті люди все частіше звертаються до штучного інтелекту з різними запитами. У зв'язку зі зростанням кількості інформації, згенерованої ШІ та використання мовних моделей, таких як ChatGPT, постає питання щодо достовірності отриманої інформації. Не виключенням є і використання ШІ студентами для написання статей, тез доповідей, курсових робіт, тощо. Метою роботи є дослідження методів та ресурсів, реалізованих на їх основі для визначення текстів, згенерованих штучним інтелектом.

Методи виявлення тексту згенерованого ШІ та результати дослідження

Беручи до уваги те, що детальний аналіз методів штучного інтелекту ускладнюється необхідністю їх реалізації, було вирішено проаналізувати якість роботи даних методів з використанням онлайн ресурсів, які ці методи реалізують.

Спершу було визначено методи виявлення тексту згенерованого ШІ:

1. Велика мовна модель або велика модель мови (ВММ або LLM від англ. Large language model) [2] – алгоритм, який знає, наскільки ймовірно послідовність слів є дійсним реченням
2. NLP – моделювання людської мови на основі правил, зі статистичними моделями, моделями машинного навчання та глибокого навчання. Разом ці технології дозволяють комп'ютерам обробляти текст і «розуміти» його повне значення, враховуючи наміри та настрої мовця чи письменника.
3. Генеративні змагальні мережі (GAN) – це клас алгоритмів ШІ, що використовуються в навчанні без вчителя та реалізовані системою двох штучних нейронних мереж, які змагаються одна з одною в рамках гри з нульовою сумою [3].

Отже, для дослідження було обрано онлайн ресурси:

- Smodin [4] – сервіс, який використовує NLP;
- Scribbr [5] – сервіс, який використовує ВММ;

- UNDETECTABLE AI [6] – сервіс, який використовує GAN

Та тексти, згенеровані ШІ «CHAT-GPT 3.5» та «Gemini» та людиною, на спільну тему «Python – мова програмування».

Результати дослідження подано в таблиці (Табл. 1) у вигляді відсотків – середня точність, з якою тексти класифіковані як «згенеровані штучним інтелектом».

Таблиця 1

Результати дослідження

Метод генерування тексту Сервіс, який використовує методи ШІ	CHAT-GPT 3.5	Gemini	Людина
Smodin	100%	54%	56,5%
Scribbr	87%	100%	56,5%
UNDETECTABLE AI	100%	100%	90%

З таблиці 1 видно, що достовірним сервісом для класифікації контенту, створеного штучним інтелектом є «Scribbr», він використовує метод ВММ. Сервіс «UNDETECTABLE AI», який використовує метод GAN, з низькою точністю класифікує тексти створені людиною, а «Smodin», який використовує NLP, з низькою точністю класифікував тексти написані за допомогою штучного інтелекту «Gemini».

Висновок

Було проаналізовано методи виявлення тексту згенерованого штучним інтелектом та виявлено найкращий метод, а саме метод ВММ для класифікації оригінальності тексту, створеного людиною. Дана інформація допоможе максимально точно визначати оригінальність робіт студентів та збільшити відсоток академічної доброчесності у ВНЗ.

Список використаних джерел:

1. Вікіпедія. Штучний інтелект [Електронний ресурс]. URL: https://uk.wikipedia.org/wiki/Штучний_інтелект
2. Заболоцький, М. О. Магістерська дисертація на тему «Створення навчальних даних для виявлення згенерованих новин» [Електронний ресурс]. URL: <https://ela.kpi.ua/server/api/core/bitstreams/b256b401-f099-4ce5-ae6-47ae001c663a/content>
3. Філіпенко, Л.В. Академічна доброчесність в науковому та освітньому середовищі закладів освіти України: погляд крізь призму наявності штучного інтелекту [Електронний ресурс]. URL: <https://academy-vision.org/index.php/av/article/view/380/336>
4. Scribbr [Електронний ресурс]. URL: <https://www.scribbr.com/>
5. Smodin [Електронний ресурс]. URL: <https://app.smodin.io/>
6. UNDETECTABLE AI [Електронний ресурс]. URL: <https://undetactable.ai/>