

УДК 004.8:004.934

ДОСЛІДЖЕННЯ ВИКОРИСТАННЯ GPT-МОДЕЛЕЙ OPENAI ДЛЯ СТВОРЕННЯ ВІДПОВІДЕЙ НА ОСНОВІ ВЛАСНОГО КОНТЕНТУ

Широкорад К.А.

Науковий керівник - доц. Яковлева О.В.

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

e-mail: kseniia.shyrokorad@nure.ua

This research explores the application of OpenAI's GPT models in education, focusing on retraining them to provide customized responses aligned with instructional material. By leveraging GPT models, the study aims to enhance teaching quality and student engagement by addressing queries and clarifying educational topics. It examines two retraining approaches and suggests the use of the LangChain framework for optimization. Overall, the research demonstrates the practicality of utilizing GPT models for tailored educational support, offering potential advancements in web development and educational tools.

Прогрес в галузі Computer Vision та розвиток Language Model (LLM) відкривають безліч можливостей для впровадження їх у різноманітні сфери [1]. Зокрема, використання GPT-моделей OpenAI в освіті виявляється дуже актуальним напрямком інновацій. Перш за все, великий обсяг контенту, що генерується такими моделями, може служити цінним інструментом для вивчення та навчання. Однак основним перевагам є можливість вивільнення часу викладача, особливо у ситуаціях з великими потоками студентів. GPT-моделі можуть ефективно взаємодіяти зі студентами, надаючи відповіді на їх запитання та допомагаючи розкрити теми, що залишилися непоясненими після лекцій. Такий підхід сприяє покращенню якості освіти та сприяє розвитку інновацій у сфері освіти.

Робота присвячена дослідженню моделей GPT від OpenAI для створення відповідей на основі власного контенту з метою використання даних моделей у навчальному процесі. Тобто метою є «донавчити» модель GPT відповідати згідно навчально-методичного матеріалу, що супроводжує дисципліну під час її викладання. Це дозволить студентам швидко орієнтуватися як в безпосередньо в навчальному матеріалі, так і мати відповіді щодо організаційної інформації. Наприклад, студент може отримати відповіді на такі організаційні питання:

- що має містити в собі звіт з лабораторної роботи;
- на які питання потрібно вміти відповідати, щоб захистити лабораторну роботу за конкретною темою;
- скільки сторінок повинна мати записка до курсової роботи;
- яке програмне забезпечення буде необхідне на лабораторних роботах.

Звісно, відповіді на такі питання можна знайти витративши час та передивившись багато матеріалу, але можна донавчити GPT модель і отримати відповідь миттєво.

Існує два варіанти донавчання моделей GPT [2]:

- класичне донавчання моделі GPT із зміною вагів нейронної мережі;
- «донавчання» на основі передачі власного контенту (базі знань) під час звертання до моделі GPT.

Обидва підходи мають переваги та недоліки. Часто обирається підхід «донавчання» на основі власного контенту за такими причинами: зберігається конфіденційність власних даних; не потрібні витрати гроші та час на створення великого датасету для навчання; не потрібно платити за донавчання на платформі вартість OpenAI; можливість використання в режимі реального часу найактуальнішої інформації для генерації відповідей.

В роботі був використано другий підхід, суть якого полягає у звертанні до GPT моделі через API та передачі моделі заздалегідь підготовленого контенту, на основі якого вона сформує відповідь [3].

Для роботи з моделями GPT через API існують наступні ролі: system (промпт, який пояснює моделі, що їй потрібно робити); user (запит користувача); assistant (відповідь моделі). Для передачі власного контенту моделі можна використати роль system, та попросити модель для відповіді використовувати тільки наданий документ. Але моделі GPT мають обмеження на об'єм загальної інформації, що може міститися у ролях system, user, assistant. Наприклад, для моделі GPT-3.5 Turbo-1106 цей об'єм складає 8,192 токенів, для GPT-4 Turbo – 128 токенів, але використання моделі GPT-4 Turbo приблизно у 10 разів дорожче. Для скорочення об'єму в роботі запропоновано використовувати фреймворк LangChain – фреймворк з відкритим вихідним кодом, який реалізує велику множину задач для спрощення розроблення застосунків із використанням моделей LLM, таких як GPT, BERT, T5 та інші.

Таким чином, алгоритм «донавчання» за власним контентом буде складатися з таких кроків:

- зчитується запит користувача та записується до «user»;
- за допомогою LangChain знаходяться фрагменти з власного контенту, які якнайкраще відповідають запиту;
- налаштовуються параметри для формування відповіді (temperature, verbose); формується промпт для GPT;
- додаються знайдені фрагменти та промпт до «system»;
- відправляється до ChatGPT «system» та «user»;
- отримується відповідь «assistant» від моделі GPT;
- відповідь надається користувачу.

В роботі можливість використання моделі GPT генерувати відповідь протестована для дисципліни «Бази даних та інформаційні системи», яка викладається у Харківського Університету Радіоелектроніки. Було спробу-

вано декілька варіантів промптів, один з яких показаний на рис. 1. На рис. 2 наведено приклад відповіді моделі GPT з використанням промпту з рис. 1.

Ти пютер (викладач) для студентів, які вивчають дисципліну "Бази даних та інформаційні системи" на кафедрі Інформатики Харківського Університету Радіоелектроніки. Дисципліна складається з лекцій, практичних занять, лабораторних робіт та курсової роботи. У тебе є всі матеріали: лекційний матеріал, методичні вказівки з лабораторних робіт, курсового проекту, в яких міститься безпосереднього сам матеріал за змістом дисципліни, а також організаційна інформація. Тобі студент ставить запитання. Дай йому відповідь, спираючись на документ. Відповідай максимально точно за документом, не випадай нічого від себе. Не згадувай документ під час відповіді, студент нічого не повинен знати про документ, за яким ти відповідаєш. Не відповідай "За даними з документу". Відповідай згідно такого контенту:

Рисунок 1 – Приклад промпту

```
temperature=1
verbose=0
# питання студента
topic='скільки сторінок повинна мати записка до курсової роботи та з яких розділів складатися?'
ans=answer_user_questionNew(system_doc_url, knowledge_base_url, topic)
ans
```

```
' Записка до курсової роботи повинна мати від 40 до 55 сторінок. Вона повинна складатися з таких розділів: титульний аркуш, завдання на дипломний проект, реферат, зміст, \n перелік умовних по значень, вступ, змістовну частину (суть пояснювальної записки), висновки, перелік посилань і до датки.'
```

Рисунок 2 – Приклад відповіді GPT моделі

Таким чином, результати досліджень показали доцільність використання моделей GPT від OpenAI для створення відповідей на основі власного контенту. Таке рішення може бути використано у веб-застосунках або telegram ботах з метою супроводження навчального процесу.

Список використаних джерел:

1. Yakovleva, O., Kovač, M., Ardasov, V. & Yeremenko, I. (2023). Study on adding functionality to the Zoom online conference system for monitoring the participant activities. *Public Administration and Regional Development*, 19(1), pp. 158–184.
2. Yakovleva, O., Nebeský, L., Kirichenko, A. (2023) Using the GPT models for responses based on custom content to develop neural consultant for university applicants. *Abstracts of V International Scientific and Practical Conference*. Madrid, Spain. pp. 172–178.
3. OpenAI platform. (2023), from <https://platform.openai.com/docs/quickstart>