

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ ТА RAG ДЛЯ ПІДВИЩЕННЯ ПЕРСОНАЛІЗАЦІЇ СИСТЕМ НАВЧАННЯ

Полуектова Н.Р.

Запорізький інститут економіки та інформаційних технологій,
Україна

E-mail: n.poluektova@econom.zp.ua

Abstract

This paper explores the use of artificial intelligence (AI) to enhance personalized learning through Retrieval-Augmented Generation (RAG). RAG combines the internal knowledge of large language models (LLM) with specialized external information, increasing accuracy in generating personalized questions and explanations. The proposed system uses Python, LangChain, and the ChatGPT API to create a learning assistant that generates random questions and explains incorrect answers. Key improvements involve optimizing question templates, adjusting chunk sizes for relevant text, and enhancing data quality. Metrics such as Groundedness and Answer Correctness are employed using libraries like Trulens and RAG-X.

Сучасний ландшафт освіти дедалі більше переплітається з новітніми технологічними інноваціями. Серед них штучний інтелект (ШІ) став рушійною силою для персоналізації навчального досвіду, що призводить до зміни парадигми у традиційних методах навчання. У своїй суті персоналізоване навчання означає адаптацію освітніх досвідів для врахування унікальних потреб, стилів навчання та темпів кожного учня [1]. Зі зростанням кількості масових онлайн-курсів, цифрових класів та платформ електронного навчання, освітяни та технологи усвідомили труднощі підходу "один розмір для всіх", що призвело до дослідження потенціалу ШІ у покращенні персоналізації.

Природна здатність систем ШІ аналізувати великі набори даних і генерувати тексти пропонує безпрецедентну можливість зрозуміти учнів на детальному рівні. Однак існує проблема поєднання розподілених в мережі знань та унікальних навчальних програм та курсів, розроблених у навчальному закладі, коли увага студента повинна зосереджуватись саме на матеріалі, викладеному в курсі.

Для створення таких адаптивних персоналізованих технологій навчання можуть використовуватись методи та засоби RAG (Retrieval-Augmented Generation) або «доповненої генерації пошуку». Ця концепція, в спрощеному вигляді, полягає в тому, щоб для генерації відповіді на спеціалізований запит користувача, доповнювати «базу знань» моделі додатковою спеціалізованою інформацією [2]. Це посилює точність і достовірність моделей, особливо для наукомістких завдань, і дозволяє постійно оновлювати знання та інтегрувати актуальну особисту, корпоративну або дослідницьку інформацію з загальнодоступною інформацією LLM. RAG синергетично поєднує внутрішні знання LLM з динамічними сховищами зовнішніх баз даних.

В контексті проблеми підвищення якості персоналізованого навчання технологія RAG може використовуватись для розв'язування наступних задач:

1. автоматична генерація персоналізованих навчальних матеріалів для кожного студента, враховуючи його рівень знань та інтереси;
2. надання студентам персоналізованих рекомендацій щодо додаткових матеріалів або тем, на яких слід зосередитися для кращого розуміння.

3. генерація персоналізованих тестів та завдань, які адаптовані до індивідуального рівня знань і прогалин у розумінні.

4. генерація миттєвих відповідей на запитання студентів тощо.

Технологія RAG може бути інтегрована у віртуальних помічників або чат-ботів для надання допомоги студентам у режимі реального часу.

В даній роботі пропонується реалізація RAG помічника з вивчення матеріалів навчального курсу, до якого генеруються випадкові питання. Якщо студент відповідає невірно, система генерує пояснення.

Архітектура додатку представлена на рис. 1. Вона включає наступні компоненти.

- Інтерактивний інтерфейс для студента (консольний або веб-додаток).
- Python-програма, що керує основною логікою та використовує:
 - LangChain - фреймворк для розробки додатків із застосуванням великих мовних моделей (LLM), таких як GPT. Основна ідея LangChain полягає в інтеграції LLM з різними джерелами даних, наприклад локальними документами або базами даних створюючи ланцюжки обробки (chains) які можуть включати генерацію тексту, пошук інформації, аналіз тощо.
 - API до ChatGPT.
 - ChatGPT, який використовується саме для генерації питання, перевірки відповіді студента та формування пояснень
- Зовнішні системи:
 - База даних або JSON-файли для зберігання інформації про лекції, запитання, відповіді та пояснення.
 - Можливо API або файлова система для завантаження додаткових навчальних матеріалів.

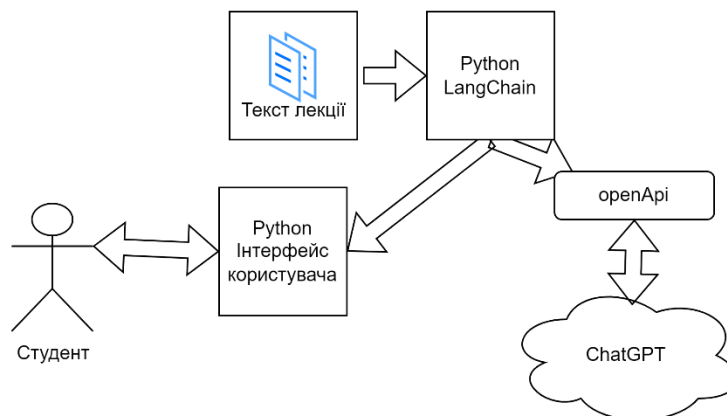


Рис. 1. Загальна архітектура пропонованого додатку

Приклад виводу додатку який тестувався на одній з лекцій курсу «Програмування на мові Python»

```

Чи можна видалити елемент зі списку за його значенням? (так/ні): ні
Невірно.
Правильна відповідь: так, можна видалити елемент зі списку за його значенням за допомогою методу remove(). тому правильна відповідь на запитання є 'так'.
Пояснення:
Відповідь студента є неправильною, оскільки вона не відповідає тексту з лекції. У лекції зазначено, що елемент можна видалити і за індексом, і за значенням.
  
```

Рис. 2. Приклад генерації запитання, перевірки та пояснення

Покращення якості генерації відповідей в подібних системах відбувається за рахунок впровадження наступних методів:

1. Вдосконалення шаблонів запитів. Наприклад, застосування шаблону запиту "Студент дав неправильну відповідь на запитання: {question}. Відповідь студента: {user_answer}. На основі тексту з лекції: {text}, поясніть, чому відповідь студента є неправильною." привело до генерації більш релевантних пояснень.

2. Підбір чанків. Під чанком в RAG-моделях мається на увазі блок тексту (символів або слів) які можуть використовуватись для генерації релевантних відповідей. Кращий результат може бути досягнутий через зміну розмірів, або накладання чанків один на інший.

3. Підвищення якості даних. Використання більш структурованих матеріалів лекцій дозволить моделі краще «розуміти» контекст та генерувати більш корисні відповіді.

Шляхи вдосконалення подібних систем також пролягають через попереднє налаштування моделі через навчання з вчителем. Для цього системі пропонуються правильні відповіді та оцінюється якість видачі завдяки спеціальним метрикам RAG: Groundness - обґрунтованість, відповідність відповіді контексту, Answer Correctness – відповідність відповідей та правильних відповідей з контрольного прикладу, Answer Relevance – релевантність відповідей та Context Relevance – релевантність контексту. Для розрахування подібних показників використовуються спеціальні Python-бібліотеки, такі як Trulens або RAG-X.

Література

1. Pane, J. F. et al. Informing Progress: Insights on Personalized Learning Implementation and Effects. / J. F. Pane, , E. D. Steiner,. M. D. Baird, L. S. Hamilton, [Електронний ресурс]. Режим доступу: https://www.researchgate.net/publication/318340596_Informing_Progress_Insights_on_Personalized_Learning_Implementation_and_Effects
2. Y Gao et al. Retrieval-augmented generation for large language models: A survey/Y Gao, Y Xiong, X Gao, K Jia, J Pan, Y Bi, Y Dai, J Sun, H Wang//arXiv preprint arXiv:2312.10997, 2023
3. The RAG Triad. [Електронний ресурс]. Режим доступу: https://www.trulens.org/trulens_eval/core_concepts_rag_triad/