

УДК 004.4

Медведчук В.Ю., Багрій Р.О., Скрипник Т.К.

Хмельницький національний університет

МЕТОД ГЕНЕРАЦІЇ ВІДПОВІДЕЙ ДЛЯ ДОПОМІЖНОЇ КОМУНІКАЦІЇ

Проблема комунікації є надзвичайно важливою в сучасному світі, де велика кількість людей стикається з труднощами в спілкуванні через вади слуху чи мовлення. Виявлення потреб таких людей та розробка рішень для полегшення комунікації має вирішальне значення для їхнього соціального включення та доступу до інформації. Зроблено огляд існуючих методів генерації відповідей для допоміжної комунікації та запропоновано метод для створення ефективних комунікаційних інструментів, здатних адаптуватися до потреб користувачів.

The problem of communication is extremely important in today's world, where a large number of people face difficulties in communication due to hearing or speech impairments. Identifying the needs of such people and developing solutions to facilitate communication is crucial for their social inclusion and access to information. An overview of existing response generation methods for assistive communication is made and a method for creating effective communication tools capable of adapting to user needs is proposed.

Комунікація є однією з найважливіших потреб людини, але для багатьох людей з вадами слуху чи мовлення це стає великою проблемою. За даними ВООЗ, близько 430 мільйонів людей у світі мають значні порушення слуху чи мовлення, що ускладнює їхню соціальну інтеграцію та повсякденне життя [1]. Допомога людям з вадами слуху та мовлення є важливим кроком до забезпечення їхнього повноцінного доступу до спілкування та інформації, сприяючи їхній соціальній і професійній інтеграції. Штучний інтелект відкриває нові можливості, дозволяючи створювати додатки для підтримки комунікації людей з такими вадами. Перспективним підходом є поєднання генеративних моделей з інформаційним пошуком, що дозволяє швидко знаходити необхідну інформацію та адаптувати відповіді до потреб користувачів [2].

Допоміжна комунікація — це підхід, що використовує різноманітні засоби та методи для полегшення спілкування осіб, які мають труднощі в усному чи письмовому спілкуванні. Допоміжна комунікація використовує різні засоби, зокрема жестову мову, комунікаційні пристрої, картки із зображеннями та текстові додатки, для полегшення спілкування осіб з порушеннями слуху, мовлення або когнітивними обмеженнями. Сучасні мобільні додатки та спеціалізовані пристрої

забезпечують інтерактивне й зручне спілкування, враховуючи індивідуальні потреби користувачів [3].

У системах автоматизованої допоміжної комунікації використовують різні методи генерації відповідей, кожен з яких має свої переваги та обмеження.

Методи на основі правил базуються на заздалегідь визначених умовах, що дозволяє отримати точні й передбачувані відповіді. Цей підхід особливо підходить для задач, де важлива стабільність і контрольованість результатів, наприклад, в інформаційних системах з обмеженими варіантами відповідей. Однак, такі методи мають низьку гнучкість, не підходять для адаптації до нових запитів і важко масштабуються для складніших сценаріїв [4].

Методи на основі шаблонів використовують готові шаблони, які дають можливість швидко формувати осмислені відповіді. Це корисно для чат-ботів, які обробляють рутинні запити, оскільки їхня реалізація проста та економить ресурси. Проте ці методи мають обмежену адаптивність, не здатні створювати оригінальні відповіді та з часом втрачають ефективність при розширенні кількості запитів [5].

Методи пошуку відповіді зосереджені на виборі найбільш підходящих відповідей із бази даних або текстового корпусу. Вони добре працюють у випадках, коли є велика кількість даних, оскільки дозволяють швидко знайти відповіді на повторювані питання, але сильно залежать від якості й обсягу бази знань. Такий підхід обмежує здатність системи адаптуватися до нових запитів, оскільки вона не може створювати нові відповіді [6].

Методи на основі великих мовних моделей (LLM), таких як GPT чи BERT, використовують нейронні мережі для створення унікальних відповідей залежно від контексту запиту. Ці моделі забезпечують високу гнучкість і можуть адаптуватися до різноманітних запитів, оскільки здатні "навчатися" на великих масивах даних, а також легко масштабуються. Незважаючи на це, LLM є ресурсоемними, потребують значних обчислювальних ресурсів і чутливі до якості даних для тренування, що може призвести до появи непередбачуваних або неточних відповідей при недостатньому навчанні [7].

Метою роботи є розробка методу генерації відповідей для допоміжної комунікації. Робота запропонованого методу полягає в генерації відповідей для допоміжної комунікації шляхом інтеграції великих мовних моделей (LLM) з механізмами інформаційного пошуку для адаптації відповідей до контексту і потреб користувача, забезпечуючи їхню релевантність, точність та інформаційну насиченість. На рисунку 1 представлено схему роботи методу.

Етап 1 – Отримання даних. На цьому початковому етапі відбувається отримання даних з різних джерел. Це можуть бути документи, статті, чи інші форми вмісту, які слугуватимуть основою для подальшої роботи. Дана інформація збирається і зберігається для подальшого оброблення.

Етап 2 – Індексация. Отримані дані індексуються, що дозволяє швидко їх знаходити під час запитів. Індексация передбачає структурування даних у такий спосіб, щоб забезпечити ефективний доступ до них. Це може включати створення словників, баз даних або інших структур даних.

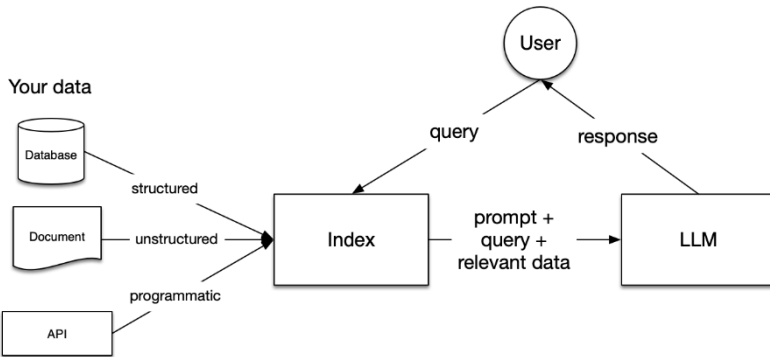


Рисунок 1 – Схема роботи методу генерації відповідей

Етап 3 – Запит. Коли користувач формулює запит, система аналізує його, щоб знайти відповідні дані з індексованих джерел. Це може включати обробку природної мови для кращого розуміння намірів користувача і вибору найбільш релевантних результатів.

Етап 4 – Генерація зразка. На цьому етапі система формує зразок відповіді, використовуючи інформацію, отриману з індексованих даних. Це може включати генерацію тексту, який надає користувачу корисну інформацію, базуючись на знайдених даних.

Етап 5 – Генерація відповіді. Фінальний етап передбачає формування остаточної відповіді на запит користувача. Це може включати об'єднання згенерованих зразків і даних, що були отримані на попередніх етапах, для створення змістовної і зрозумілої відповіді.

Етап 6 – Висновок. Після генерації відповіді, система може надавати рекомендації або додаткові ресурси, що допомагають користувачу краще зрозуміти тему або задовольнити його потреби. Це завершує процес, але система може також зберігати отриману інформацію для покращення майбутніх запитів.

Отже, запропонований метод генерації відповідей із доповнюючим інформаційним пошуком є ефективним інструментом для покращення комунікації людей з порушеннями мовлення або слуху. Він дозволяє автоматично генерувати відповіді з використанням додатковою інформації, що значно підвищує точність і

релевантність відповідей. Інтеграція великих мовних моделей із системами пошуку дозволяє створити гнучку систему, здатну адаптуватися до нових запитів та оновлювати базу знань для підтримки актуальної інформації. Це сприятиме покращенню комунікації та соціальної взаємодії людей з особливими потребами.

Перелік посилань

1. Вади слуху URL: https://uk.wikipedia.org/wiki/Вади_слуху
2. Shamane Siriwardhana, Rivindu Weerasekera, Elliott Wen, Tharindu Kaluarachchi, Rajib Rana, Suranga Nanayakkara; Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering. Transactions of the Association for Computational Linguistics 2023
3. "Everyone Deserves AAC": Preliminary Study of the Experiences of Speaking Autistic Adults Who Use Augmentative and Alternative Communication Amy L. Donaldson , endeavor* corbin and Jamie McCoy
4. Eftimov T, Koroušić Seljak B, Korošec P (2017) A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations. PLoS ONE
5. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean; Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems, 2013.
6. Shamane Siriwardhana, Rivindu Weerasekera, Elliott Wen, Tharindu Kaluarachchi, Rajib Rana, Suranga Nanayakkara; Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering. Transactions of the Association for Computational Linguistics 2023
7. Zhicheng Guo, Sijie Cheng, Yile Wang, Peng Li, Yang Liu; Retrieval-Augmented Generation for Large Language Models: A Survey.