

ТЕЛЕГРАМ-БОТ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ “ЦИФРОВОЙ ПОМОЩНИК БРГТУ”

Палто Е. С.

Кафедра интеллектуальных информационных технологий,
Брестский государственный технический университет
Брест, Республика Беларусь
E-mail: alena_sokolovskau@mail.ru

Телеграм-боты являются неотъемлемой частью современных цифровых сервисов, и их роль в упрощении общения, автоматизации задач и предоставлении ответов в реальном времени постоянно растет. Использование искусственного интеллекта в таких ботах обеспечивает их способность справляться с более сложными задачами и предоставлять персонализированные ответы. В данной статье рассматривается разработка телеграм-бота с возможностями искусств для Брестского государственного технического университета, названного "Цифровой помощник БРГТУ". Бот основан на системе генерации с поддержкой поиска - Retrieval-Augmented Generation, которая позволяет ему эффективно искать и извлекать информацию из различных источников, а также генерировать релевантные ответы.

ВВЕДЕНИЕ

Искусственный интеллект [1] сегодня является важнейшей составляющей цифровых технологий, активно применяемых в различных сферах, включая образование, бизнес и повседневную жизнь. Разработка телеграм-ботов с ИИ позволяет автоматизировать процессы взаимодействия с пользователями, обеспечивать оперативную поддержку и предоставление информации в реальном времени. Телеграм-боты могут не только отвечать на типовые запросы, но и адаптироваться под нужды пользователей, анализируя запросы и предоставляя уникальные ответы. Цифровой помощник БРГТУ – это специализированный телеграм-бот, разработанный для предоставления информации и поддержки студентам, преподавателям и сотрудникам университета. Основная цель проекта – автоматизация ответов на вопросы, предоставление актуальной информации и облегчение взаимодействия между пользователями и учебным заведением.

I. ПОСТАНОВКА ПРОБЛЕМЫ

Цифровой помощник БРГТУ был создан для решения ряда проблем, связанных с коммуникацией и доступом к информации в университете. Одной из основных трудностей, с которой сталкиваются пользователи, является затрудненный доступ к актуальной информации. Студенты, абитуриенты и их родители часто испытывают сложности при поиске данных, связанных с учебным процессом, расписанием, мероприятиями и административными вопросами. Традиционные способы получения информации через веб-сайты или обращения к сотрудникам иногда оказываются неудобными и требуют временных затрат.

Кроме того, университетская среда характеризуется большим количеством однотипных запросов от студентов и сотрудников, что создает дополнительную нагрузку на персонал, особенно в периоды высокой активности, такие как сес-

сии или приемная кампания. Также существует потребность в обеспечении круглосуточной поддержки, которая могла бы оперативно реагировать на запросы пользователей, предоставляя точные и своевременные ответы. В этой связи автоматизированное решение становится необходимым для оптимизации взаимодействия и повышения эффективности коммуникации.

II. МЕТОДОЛОГИЯ РАЗРАБОТКИ И ПРИНЦИП РАБОТЫ СИСТЕМЫ

Цифровой помощник БРГТУ представляет собой телеграм-бот с искусственным интеллектом, предназначенный для помощи студентам, преподавателям и сотрудникам Брестского государственного технического университета. Основная цель его создания – автоматизация взаимодействия между пользователями и университетом, предоставление актуальной информации, а также упрощение доступа к необходимым данным. Технология Retrieval-Augmented Generation (RAG) [2] была выбрана для разработки цифрового помощника БРГТУ благодаря её уникальной комбинации методов поиска, извлечения информации и генерации ответов с использованием нейронных сетей. Основным элементом работы бота является способность быстро находить релевантные данные в базе, которая содержит сведения об университете, учебном процессе, кафедрах, мероприятиях и других аспектах жизни БРГТУ. Это обеспечивает минимальное время отклика и точные ответы на запросы пользователей, такие как расписание занятий, информация о структуре университета или последние новости.

В тех случаях, когда необходимая информация отсутствует в базе данных или не полностью релевантна, бот использует нейронные сети и языковые модели для генерации ответов на основе уже имеющихся данных. Такой подход позволяет боту адаптироваться к нестандартным запросам и предоставлять уникальные и контекстуальные ответы.

Основное преимущество RAG заключается в сочетании высокой скорости поиска с гибкостью в адаптации ответов, что делает эту технологию особенно эффективной в образовательной среде, где информация может часто обновляться. Телеграм-бот способен обслуживать большое количество пользователей одновременно, эффективно справляясь с запросами и большими объёмами данных.

III. ОБОСНОВАНИЕ ВЫБОРА ТЕХНОЛОГИИ RETRIEVAL-AUGMENTED GENERATION

Технология Retrieval-Augmented Generation (RAG) [2] представляет собой современный гибридный метод, сочетающий механизмы информационного поиска и генерации текста на основе нейронных сетей, что делает её особенно полезной в задачах, требующих точного извлечения информации и контекстуально релевантных ответов. Основная идея RAG [2] заключается в интеграции двух ключевых компонентов: системы поиска и генеративной модели, что позволяет не только находить существующие данные, но и генерировать новые ответы на основе найденной информации.

Первый этап работы RAG [2] включает в себя поиск релевантных данных. Для этого используется информационно-поисковая система, которая извлекает информацию из заранее подготовленных и структурированных источников, таких как базы данных, документы и другие текстовые корпуса. Этот процесс осуществляется с помощью методов ранжирования и поиска по ключевым словам или другим запросам, что обеспечивает высокую скорость и точность получения релевантных данных. В контексте использования RAG [2] для создания цифрового помощника БрГТУ, это позволяет эффективно находить информацию об учебных расписаниях, структуре университета, кафедрах, мероприятиях и других аспектах университетской жизни.

После того как релевантные данные найдены, происходит второй этап – генерация ответа на основе извлечённой информации. Здесь используются нейронные языковые модели, обученные на больших корпусах текстов, что позволяет им генерировать осмысленные и контекстуально корректные ответы, даже если извлечённая информация является неполной или не совсем релевантной. Генеративная модель может комбинировать извлечённые данные и, используя свои внутренние языковые представления, формировать ответы, которые соответствуют запросу пользователя. Это особенно полезно в тех случаях, когда конкретного ответа нет в базе данных, но он может быть сгенерирован на основе похожих данных или контекста запроса.

Особое внимание при разработке было уделено снижению вероятности появления так называемых "галлюцинаций" – ситуации, когда языковая

модель генерирует неправдоподобные или ошибочные ответы на запросы пользователей. Для этого в систему были внедрены специальные механизмы контроля качества генерации, которые обеспечивают большую точность и релевантность предоставляемых данных. Модель проходит проверку на наличие информации в базе данных и стремится предоставлять ответы только на основе проверенных и доступных данных.

Кроме того, в процессе разработки бота были специально настроены элементы вежливого общения. Эта функция позволяет создавать комфортное общение, где пользователи, независимо от ситуации, получают вежливые и точные ответы на свои запросы. Данный аспект важен для обеспечения высокого уровня доверия и удобства использования системы, особенно в образовательной среде.

Основным преимуществом RAG [2] является её способность сочетать точность поиска с гибкостью генерации. В отличие от традиционных систем поиска, которые просто возвращают информацию без какой-либо интерпретации, RAG [2] может адаптировать ответы, что делает её подходящей для сложных задач, где требуется не только быстро найти информацию, но и предоставить её в структурированной и понятной форме. Эта гибридная модель также способна обновляться в режиме реального времени, поддерживая актуальность предоставляемых данных.

ЗАКЛЮЧЕНИЕ

Таким образом, технология RAG [2] является эффективным решением для разработки интеллектуальных систем, таких как цифровой помощник БрГТУ. Комбинируя высокоскоростной поиск информации с гибкостью генерации контекстуальных ответов, она позволяет не только оперативно предоставлять точные данные, но и адаптироваться под нестандартные запросы. Особое внимание уделено снижению вероятности генерации ошибочных ответов и поддержанию вежливого общения с пользователями, что делает бота удобным и надежным инструментом в образовательной среде, способным обрабатывать большие объёмы запросов. Автоматизация взаимодействия с пользователем позволяет экономить время пользователя и повысить эффективность коммуникации.

IV. СПИСОК ЛИТЕРАТУРЫ

1. Kaplan, A. Artificial Intelligence, Business and Civilization: Our Fate Made by Machines / A. Kaplan // New York: Routledge. – 2022. – 2nd ed. – P. 456-470.
2. Lewis, P., D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks / P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – Curran Associates, Inc.: P. 9459–9474..