

УДК 004.032.26

В. В. Орлов

г. Гомель, ГГУ имени Ф. Скорины

КУРСОВОЕ И ДИПЛОМНОЕ ПРОЕКТИРОВАНИЕ: РАЗРАБОТКА ЦИФРОВОГО АССИСТЕНТА НА ОСНОВЕ GPT ТЕХНОЛОГИИ

В современном мире информация является ценнейшим ресурсом и ее объемы растут с невероятной скоростью. Использование систем на основе машинного обучения способно кардинально упростить процесс анализа данных, если сделать его более эффективным и доступным для широкого круга специалистов. Такие системы позволят автоматизировать сбор, обработку и интерпретацию данных, выявлять закономерности и предоставлять обоснованные рекомендации, что является несомненным прорывом в области аналитики и может найти применение в различных секторах экономики, науки и образования.

Студенты факультета математики и технологий программирования IT-специальностей на протяжении всего срока обучения осваивают дисциплины учебного плана, ориентированные на их будущую профессиональную деятельность. Важную роль в процессе такой подготовки играет курсовое и дипломное проектирование. Студентам предлагаются актуальные темы проектов, связанные с разработкой компьютерных приложений для решения прикладных задач в различных сферах жизнедеятельности человека.

Одним из актуальных направлений научно-исследовательской работы студентов в рамках курсового и дипломного проектирования может быть разработка приложений на основе искусственных нейронных сетей, в том числе генеративных предобученных трансформеров (GPT – Generative Pre-Trained Transformer). Generative означает, что модель генерирует новый текст; pre-trained – что модель была обучена на больших объемах данных.

Трансформер – это особый вид нейронной сети, модель машинного обучения (Machine Learning Model). Существует широкое разнообразие моделей, которые можно построить с использованием трансформеров: преобразование голоса в текст (voice-to-text), текста в голос (text-to-voice), текста в изображение (text-to-image), машинный перевод и другое. Конкретный вариант, который лежит в основе таких инструментов, как ChatGPT, – это модель, обученная принимать фрагмент текста, возможно, даже с сопутствующими изображениями или звуком, а затем выдавать предсказание того, что будет дальше, в форме распределения вероятностей по всем фрагментам текста, которые могут следовать.

Технология GPT от OpenAI представляет собой мощную модель для генерации текста, основанную на архитектуре трансформеров. Модели GPT обучаются на больших объемах текстовых данных и могут выполнять различные задачи обработки текста на естественном языке, включая перевод, ответы на вопросы и создание текста. Эти модели способны генерировать осмысленный и связный текст, что делает их полезными для множества приложений.

В условиях стремительного развития технологий и возрастания объемов данных способность компаний адаптироваться и принимать обоснованные решения является ключевым фактором успеха. Аналитики играют в этом процессе центральную роль, предоставляя критически важные рекомендации, которые помогают организациям ориентироваться в сложных рыночных условиях и эффективно реагировать на новые вызовы. В качестве помощника аналитика может выступать «Цифровой ассистент аналитика».

«Цифровой ассистент аналитика» представляет собой API (Application Programming Interface, дословно – «интерфейс программирования приложения») на

основе HTTP запросов (Hypertext Transfer Protocol – «протокол передачи гипертекста»), который позволяет получить необходимую информацию в ответ на текстовый запрос пользователя, введенный на простом языке. Для обработки запроса цифровой ассистент реализует паттерн проектирования цепочка обязанностей и путем использования нескольких блоков-преобразователей, а также API внешних систем, получает ответ на изначальный запрос пользователя с приемлемой точностью.

Основная технология сервиса – это генеративные предобученные трансформеры, особый вид нейронных сетей, основная задача которых состоит в выполнении заданной текстовой инструкции.

Подобный цифровой ассистент может оказать существенную помощь аналитику в проведении так называемого разведывательного анализа данных, используя при этом только текстовый запрос, для составления которого не обязательно обладать компетенциями в области анализа данных. А применение API позволяет использовать не только веб-интерфейс, но и интерфейс мобильного приложения.

Запросы к базам данных являются неотъемлемой частью работы любого аналитика. Базы данных хранят огромное количество информации, которая может быть использована для анализа и принятия решений. В этой связи умение формулировать и выполнять запросы к базам данных становится критически важным навыком. Тем не менее на практике часто возникает ситуация, когда аналитики не владеют языком SQL на должном уровне. SQL (Structured Query Language) – это стандартный язык для работы с базами данных, который позволяет извлекать необходимую информацию, манипулировать данными и управлять структурой базы.

Недостаточное владение SQL может приводить к ряду проблем. Во-первых, это замедляет процесс анализа данных, так как аналитики вынуждены обращаться за помощью к IT-специалистам или разработчикам, что может затягивать сроки выполнения задач и снижать оперативность принятия решений. Во-вторых, отсутствие навыков работы с SQL может ограничивать аналитиков в их возможностях глубже исследовать данные и строить более сложные и точные модели анализа. Это в свою очередь может негативно сказаться на качестве выводов и рекомендаций, предоставляемых бизнесу.

Одним из перспективных направлений в решении этой проблемы является использование моделей обработки естественного языка (Natural Language Processing, NLP), таких как большие языковые модели (Large Language Models, LLM). Эти модели обладают способностью понимать и генерировать человеческий язык, что позволяет применять их для преобразования текстовых запросов в SQL-запросы.

Тематика курсовых и дипломных работ может быть ориентирована на разработку приложений, которые принимают запросы аналитиков в текстовой форме и преобразуют их в SQL-запросы с использованием LLM. Основная цель приложений – сокращение разрыва между аналитиками и базами данных, предоставляя им возможность эффективно взаимодействовать с данными без необходимости глубоких знаний SQL. Это позволит аналитикам сосредоточиться на интерпретации данных и принятии решений, а не на технических аспектах их извлечения.

При выполнении курсовых и дипломных работ по данной тематике студенты будут овладевать знаниями и приобретать умения и навыки в области разработки искусственных нейронных сетей, в том числе такого перспективного направления как генеративные предобученные трансформеры.