

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования

Российский государственный гидрометеорологический университет
(РГГМУ)

Институт информационных систем и геотехнологий

Направление подготовки: 09.03.03 «Прикладная информатика»

Профиль подготовки: «Прикладные информационные системы и
геотехнологии»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(БАКАЛАВРСКАЯ РАБОТА)**

На тему: «Разработка приложения интеллектуального ассистента на базе технологий глубокого обучения.»

Научный руководитель,
к.т.н

_____Петров Я.А.

Исполнитель,
студент группы ПИ-Б20-2-2

_____Попов В.Н.

Санкт-Петербург 2024 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Предпроектный анализ	5
1.1 Анализ предметной области	5
1.2 Сравнительный анализ	6
1.3 Системный анализ	7
2 Проектирование информационной системы	8
2.1 Концептуальное проектирование	8
2.2 Функциональное проектирование	8
2.3 Диаграмма последовательности	8
2.4 Диаграмма развертывания	8
3 Разработка системы	9
3.1 Выбор средств разработки	9
3.2 Что-то про структуру приложения	9
3.3 Что-то про извлечение сущности из фраз пользователя и их метчинг с сущностями из базы знаний	9
3.4 Что-то про метод ближайших соседей и обогащение ответа	9
3.5 Что-то про генерацию ответа	9
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	10

ВВЕДЕНИЕ

За последние десятилетия произошел огромный скачок в развитии информационных технологий: от создания первой электронной вычислительной машины, до сложных генеративных нейронных сетей (GAN). Сейчас компании проводят множественные исследования для выявления возможностей таких технологий и необходимости дальнейших инвестиций в данную отрасль. Одним из представителей GAN стали большие языковые модели (LLM). В настоящее время лидерами в данной отрасли стали: OpenAI, которые придумали реализовать интерфейс для взаимодействия с нейронной сетью в виде чата; ПАО Сбербанк, реализовавшие отечественную LLM в условиях изоляции и ограниченных ресурсов; Meta (признана в РФ экстремистской организацией и запрещена), разработавшие малую языковую модель (SLM) LLaMA, ставшая прорывом для энтузиастов, у которых нет таких ресурсов для реализации LLM, как у больших игроков рынка.

В данной выпускной квалификационной работе будет разработана платформа, позволяющая взаимодействовать с ресурсами предприятия, использующая LLM/SLM для генерации релевантных ответов.

Актуальность данной темы обусловлена возможностью оптимизации многих процессов, увеличении производительности сотрудников, повышении лояльности клиентов. В контексте высшего учебного заведения (ВУЗ), использование технологий подобного рода повышает конкурентоспособность ВУЗа, что наряду с предыдущими пунктами является положительной метрикой.

Объект исследования — большие языковые модели (LLM).

Предмет исследования — является применение языковых моделей для бизнеса.

Цель работы — проектирование и разработка приложения, которое позволяет взаимодействовать с структурой предприятий посредством приложения для коммуникации.

В качестве интерфейса для пользователя была выбрана оболочка в виде чат-бота. Чат-боты давно вошли в жизнь большинства населения. Это подтверждается информацией аналитической компании «eMarketer», согласно которой, чат-ботами пользуются более 1,4 млрд. человек на планете.

Для выполнения поставленной цели были поставлены следующие зада-

чи:

- Выполнить анализ предметной области;
- Провести сравнительный анализ информационных систем;
- Изучить сроки реализации проекта;
- Смоделировать схему бизнес-процессов;
- Составить описание документов бизнес-процессов;
- Сформировать перечень требований к ИС;
- Исследовать подходы SWOT
- Описать сценарии вариантов использования;
- Визуализировать описанные сценарии вариантов использования;
- Создать модель диаграммы компонентов;
- Создать модель диаграммы развертывания;
- Реализовать бизнес-логику ассистента и перенести его в интерфейс

бота;

В работе будет рассматриваться РГГМУ (далее Университет), но применяться бот сможет не только в конкретном учебном заведении, а для любых предприятий.

Во время разработки ассистента использовалась методология Agile. Она позволила работать в удобном темпе и формировать требования во время разработки.

В ходе выполнения практической части выпускной квалификационной работы были использованы:

Python 3.13, LangChain, FAISS, HuggingFace.

1 Предпроектный анализ

1.1 Анализ предметной области

Индустрия информационных технологий является одной из наиболее динамичных и быстроразвивающихся отраслей, где каждый год появляются новые тенденции и совершенствуются технологии, которые позволяют улучшить пользовательский опыт и приносить большую выгоду бизнесу. Одной из ключевых технологий стала технология трансформер, предложенная в статье “Attention is all you need”.

Одной из ключевых особенностей трансформеров является их способность обрабатывать большие объемы текста без потери информации для выполнения задач таких как машинный перевод, обработка естественного языка, когнитивный анализ текста и генерирование текста. Трансформер состоит из блоков кодировщиков и декодеров, которые обрабатывают входные данные и генерируют выходные данные. Большое количество параметров сети позволяет ей улучшить качество работы по сравнению с другими моделями. В последнее время наблюдается тренд на внедрение больших языковых моделей в различные отрасли бизнеса, например системы автоматических ответов на вопросы, чат-боты, умные помощники.

Преимуществом таких решений является быстрый поиск информации и выдача её в удобоваримом виде, когда без использования таких ассистентов на поиск необходимой информации может потребоваться достаточный промежуток времени. В данной дипломной работе предлагается разработать информационную систему-помощника в виде чат-бота который будет представлять собой полезный инструмент как для студентов, так и для сотрудников ВУЗа. Основная идея информационной системы состоит в том, чтобы получить универсальный инструмент для взаимодействия со всей структурой университета. В рамках чат-бота пользователь сможет получить всю необходимую информацию, например информацию о заселении в общежития, списке необходимых документов для поступления т.п.

В общем и целом, интеграция технологии больших языковых моделей является актуальной и перспективной темой для дипломной работы, которая позволит изучить основы построения архитектуры приложения, интеграции технологий в предприятия, основы работы с нейросетями и машинным обучением, а также тестирования решений, где нет очевидных метрик для

измерения результата.

1.2 Сравнительный анализ

На данный момент прямых конкурентов у моего решения нет, но я не отрицаю того, что в настоящий момент может разрабатываться схожее решение. Из схожих решений можно отметить следующие решения:

Боты от университетов. Такие решения не имеют возможности масштабирования, имеют ограниченный пул вопрос/ответ и привязанны к какой-то определенной платформе.

Virtual Spirits. Эта зарубежная компания специализируется на создании на создании чат-ботов для различных предприятий. Из преимуществ имеется возможность настройки внешнего вида бота.

Таблица 1.1 — Сравнительный анализ

Информационная система	Удобный сбор информации	Возможность найти информацию не касающуюся учебы	Необходимость аутентификации
Virtual Spirits	-	-	+
ИС от ВУЗ	-	-	+
Моя ИС	+	+	-

Опираясь на проведенный анализ можно подвести некоторый итог: В итоговой системе не будет системы авторизации, так как мне кажется, что вся информация должна быть в открытом доступе для всех возможных пользователей ИС.

Под удобным сбором информации подразумевается интуитивно понятный процесс заполнения базы знаний, который может осуществляться как вручную, так и при помощи API, парсинга или других методов получения информации.

Возможность получать информацию не связанную с обучением мне кажется одним из ключевых преимуществ моей информационной системы: для абитуриентов может быть важно получить информацию как о возможном расписании, так и о поступлении в ВУЦ, получении БСК или же информации о истории университета.

1.3 Системный анализ

Исследование проектируемой ИС проводилось в виде нескольких типов анализа: SWOT, VCM, BPR и ISA.

SWOT-анализ — метод стратегического планирования, суть которого заключается в выявлении факторов внутренней и внешней среды организации и разделении их на четыре категории: Strengths, Weaknesses, Opportunities, Threats. Сильные и слабые стороны представляют факторы внутренней среды объекта анализа. В свою очередь возможность и угрозы представляют собой внешнюю среду объекта анализа.

Таблица 1.2 — SWOT анализ

	Положительное влияние	Негативное влияние
Внутренняя среда	— Актуальность — Простота использования	— Нейросетевые галлюцинации — Необходимость тщательно прорабатывать интеграцию во избежание проблем с безопасностью
Внешняя среда	— Упрощение навигации по ресурсам ВУЗа — Получение поддержки от государства	— Регуляции со стороны государства — Бюрократия с какой-либо стороны

Из положительных аспектов можно выделить простоту конечного использования и улучшение взаимодействие с предприятием.

Из отрицательных факторов стоит выделить следующие аспекты: нейросетевые галлюцинации, проблемы с безопасностью и бюрократия.

Что про документацию [1]

- 2 Проектирование информационной системы**
- 2.1 Концептуальное проектирование**
- 2.2 Функциональное проектирование**
- 2.3 Диаграмма последовательности**
- 2.4 Диаграмма развертывания**

3 Разработка системы

3.1 Выбор средств разработки

3.2 Что-то про структуру приложения

3.3 Что-то про извлечение сущности из фраз пользователя и их метчинг с сущностями из базы знаний

3.4 Что-то про метод ближайших соседей и обогащение ответа

3.5 Что-то про генерацию ответа

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 *Ganegedara T.* Word2Vec: как работать с векторными представлениями слов [Электронный ресурс] / пер. с англ. С. Рунов. — 2018. — Режим доступа: <https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obuchenija/>.