Министерство науки и высшего образования Российской Федерации
ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»
Кафедра «школы бакалавриата (школа)»

Оценка работы 100 из 100
Руководитель от УрФУ: Домашних Иван Алексеевич,
старший преподаватель департамента математики,
механики и компьютерных наук

Тема задания на практику

План-проспект ВКР по теме исследование методов улучшения качества больших языковых моделей для прикладных задач

ОТЧЕТ

Вид практики Производственная практика
Тип практики Производственная практика, Научно-исследовательская работа

Руководитель практики от предприятия (организации):

Студент: Плисковский Лавр Юрьевич

Направление подготовки:
02.03.02 Фундаментальная информатика и информационные технологии

Группа МЕН-420810

Екатеринбург 2025

1. ВВЕДЕНИЕ

Научно-исследовательская работа является важнейшим этапом подготовки будущего специалиста, позволяя применить теоретические знания для решения реальных практических задач в области искусственного интеллекта. Это уникальная возможность погрузиться в современные методы машинного обучения, познакомиться с передовыми стандартами и внести вклад в развитие больших языковых моделей (LLM).

Работа над дипломным проектом ведется под руководством научного руководителя, который, являясь экспертом в области обработки естественного языка (NLP), формулирует актуальные для индустрии научные задачи и обеспечивает поддержку на всех этапах исследования.

Тема дипломной работы сфокусирована на повышении качества и надежности больших языковых моделей при их использовании в узкоспециализированных прикладных задачах. Проблема носит не теоретический, а практический характер: из-за своей обобщенной природы базовые LLM часто генерируют "галлюцинации", нерелевантную или фактически неверную информацию, что напрямую влияет на эффективность и безопасность их применения в бизнесе, медицине, юриспруденции и других областях.

Выбор темы обусловлен её критической важностью для успешного внедрения LLM в реальные рабочие процессы, а также личным профессиональным интересом к архитектурам и методам, позволяющим повысить точность, управляемость и достоверность генеративных моделей.

2. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

Основной задачей на начальном этапе являлось проведение первичного исследования по теме будущей выпускной квалификационной работы (ВКР) на основе анализа ограничений существующих LLM. Это включало в себя:

Анализ проблемы и ограничений: Изучение причин и последствий "галлюцинаций", нехватки доменных знаний и других сбоев LLM для формализации конкретных сценариев, приводящих к генерации некачественных ответов.

Изучение предметной области: Исследование существующих методов, подходов и best practices в области дообучения (fine-tuning), Retrieval-Augmented Generation (RAG), и других техник адаптации LLM.

Планирование работы: Формулировка четкой цели, конкретных задач, ожидаемых результатов и выбор методологии для дальнейшего исследования. Итогом этой работы стал детальный план-проспект будущей ВКР, нацеленный на разработку и проверку методов повышения качества моделей.

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ ПО ТЕМЕ ВКР (ПЛАН-ПРОСПЕКТ)

В ходе предварительной работы был разработан и согласован с руководителем структурированный план-проспект дипломной работы, который включает следующие ключевые разделы:

3.1. Цель исследования

Целью работы является разработка и экспериментальная проверка комплекса методов для улучшения качества ответов больших языковых моделей в прикладных задачах, направленных на повышение их точности, релевантности и фактической достоверности.

3.2. Актуальность темы

Актуальность исследования подтверждается многочисленными примерами, которые наглядно демонстрируют системную проблему базовых ("general-purpose") LLM:

Проблема: Модели, обученные на общих данных из интернета, не обладают достаточной экспертизой в узких областях и склонны генерировать правдоподобные, но ложные утверждения ("галлюцинации").

Последствия: Это приводит к риску дезинформации, снижению доверия к AI-системам и невозможности их полноценного использования в критически важных задачах, где цена ошибки высока (например, в медицинской диагностике, юридических консультациях, финансовом анализе).

Конкретные примеры:

Медицина: LLM может предложить корректное по форме, но неверное по сути лечение, основываясь на устаревших или нерелевантных данных.

Юриспруденция: Модель может ссылаться на несуществующие законы или судебные прецеденты, вводя пользователя в заблуждение.

Таким образом, актуальность заключается в необходимости разработки методов, которые позволяют "заземлить" LLM на фактические знания и адаптировать их к специфике конкретной прикладной области.

3.3. Методы решения

Для достижения цели планируется выполнить ряд последовательных задач:

Анализ и аудит: Детальный анализ архитектуры базовых LLM и их производительности на целевых прикладных задачах с фокусом на выявлении типичных ошибок и "галлюцинаций".

Разработка методологии: Создание комплексного подхода, сочетающего дообучение (fine-tuning) на специализированных датасетах и интеграцию с базой знаний через механизм Retrieval-Augmented Generation (RAG).

Проектирование архитектуры: Разработка экспериментального стенда, включающего векторную базу данных для хранения доменных знаний и конвейер (pipeline) для их извлечения и подачи в контекст модели.

Эксперимент и апробация: Внедрение метода на тестовых данных и проверка его работоспособности в условиях, моделирующих реальные запросы пользователей в выбранной прикладной области.

Верификация эффективности: Оценка результатов по ключевым метрикам качества: точность (accuracy), полнота (recall), BLEU, ROUGE, а также метрики фактической согласованности.

3.4. Ожидаемые результаты

По итогам работы планируется получить:

Практический метод адаптации LLM, направленный на минимизацию "галлюцинаций" и повышение релевантности ответов.

Результаты сравнительного тестирования (базовая модель vs. адаптированная), демонстрирующие улучшение качества генерации в прикладных задачах.

Набор рекомендаций для эффективного применения техник fine-tuning и RAG.

Актуализированный набор метрик для оценки качества и надежности специализированных LLM.

3.5. Стек технологий

Для реализации проекта будет использован технологический стек: Python, фреймворки PyTorch или TensorFlow, библиотека Hugging Face Transformers, LangChain. Для реализации RAG-подхода

— векторные базы данных (например, FAISS, ChromaDB) и системы мониторинга экспериментов (например, Weights & Biases).

4. ЗАКЛЮЧЕНИЕ

В ходе предварительной работы поставленная задача была выполнена. Проведен анализ ограничений современных LLM, что позволило сформулировать проблему не в общих чертах, а в терминах конкретных рисков и последствий для прикладного использования. На основе этого изучена актуальная литература, сформулированы цель, задачи и ожидаемые результаты будущего исследования. Разработанный план-проспект служит надежной дорожной картой для выполнения выпускной квалификационной работы.

Основной сложностью на данном этапе является выбор оптимального баланса между дообучением (fine-tuning) и RAG, так как каждый из методов имеет свои преимущества и недостатки. В дальнейшей работе планируется углубиться в этот анализ, детально проработать архитектуру гибридного подхода и приступить к практической реализации на тестовом стенде, с особым вниманием к сценариям, где базовые модели демонстрируют наихудшие результаты.

5. СПИСОК ИСТОЧНИКОВ (ЛИТЕРАТУРЫ)

1. Vaswani, A., et al. Attention is All You Need. – NIPS, 2017.
2. Lewis, P., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. – NeurIPS, 2020.
3. Документация Hugging Face Transformers: https://huggingface.co/docs/transformers
4. Brown, T. B., et al. Language Models are Few-Shot Learners (GPT-3). – NeurIPS, 2020.
5. Статьи и блоги о fine-tuning, RAG и prompt engineering (Towards Data Science, distill.pub, блоги OpenAI и Google AI).
6. Обзоры и бенчмарки для оценки LLM (SuperGLUE, MMLU).
7. Документация по используемым технологиям: PyTorch, TensorFlow, LangChain, FAISS.