СРМ №2

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Срок выполнения по календарно-тематическому плану: 10-я неделя.

Цель работы

Самостоятельно исследовать возможности Generative AI и LLM, провести контролируемое сравнение их применения на прикладной задаче и определить целесообразность интеграции генеративного компонента в индивидуальный проект.

Связь с дисциплиной и индивидуальным семестровым проектом

Работа опирается на результаты ПЗ №1–9. Она подготавливает магистранта к ПЗ №11–13 по Generative AI, LLM и Prompt Engineering и определяет роль LLM в индивидуальном проекте.

Краткие теоретические сведения

Generative AI создаёт новый контент, а большие языковые модели работают с текстовым контекстом и могут решать широкий спектр языковых задач. Их оценка должна включать не только удобство ответа, но и фактическую корректность, устойчивость, соблюдение формата, конфиденциальность данных и риск галлюцинаций.

Методические указания

Работа выполняется преимущественно самостоятельно. Магистрант сам определяет конкретные методы, данные, параметры эксперимента и форму представления результатов в рамках заданных требований. Необходимо отделять обзор литературы от собственных экспериментальных результатов. Все сравниваемые варианты должны оцениваться в сопоставимых условиях. Каждый вывод должен опираться либо на источник, либо на полученный экспериментальный результат.

При использовании AI-сервисов запрещается передавать конфиденциальные, персональные или служебные данные без разрешения. Код и вычислительные эксперименты должны быть воспроизводимыми; для стохастических методов фиксируется seed/random_state, а данные, метрики и параметры описываются в отчёте.

Порядок выполнения

1. Уточнить исследовательский вопрос текущей работы.
2. Определить исходные данные и критерии оценки.
3. Изучить и сравнить подходящие методы.
4. Выполнить практический эксперимент или разработать прототип.
5. Проанализировать ошибки и ограничения.
6. Связать результат с индивидуальным семестровым проектом.
7. Оформить выводы и подготовиться к защите.

Задания

Задание 1. Выберите прикладную задачу из своего проекта, где потенциально полезны Generative AI или LLM.

Задание 2. Сформируйте тестовый набор не менее чем из 15–20 обезличенных примеров.

Задание 3. Определите минимум 5 критериев оценки до начала эксперимента.

Задание 4. Сравните минимум 2 модели, режима или способа использования LLM на одинаковых входных данных.

Задание 5. Зафиксируйте ошибки: неверные факты, пропуски, нарушение формата, нестабильные ответы.

Задание 6. Проанализируйте риски передачи данных внешним AI-сервисам.

Задание 7. Сделайте вывод: LLM является основным, вспомогательным или неприемлемым компонентом вашего проекта.

Пример выполнения

В проекте по фишингу LLM можно использовать не для окончательной классификации URL, а для объяснения уже рассчитанных признаков или анализа обезличенного текста сообщения. Тестовый набор содержит 20 сообщений с эталонной оценкой. Сравниваются корректность класса, полнота выделенных индикаторов, соблюдение JSON-формата и число неподтверждённых утверждений.

Индивидуальное задание

Разработайте собственную матрицу оценки LLM для темы проекта и приложите не менее 5 примеров, где модель ошиблась или дала спорный результат. Предложите механизм проверки ответа.

Рекомендуемая форма представления результатов

Раздел	**Что должно быть представлено**
Постановка	Цель, исследовательский вопрос, связь с индивидуальным проектом
Данные/материалы	Источник, структура, ограничения и допустимость использования
Методика	Алгоритмы, параметры, критерии и схема эксперимента
Результаты	Таблицы, метрики, схемы, примеры или журналы работы
Анализ	Сравнение, ошибки, ограничения, риски
Решение по проекту	Что включается в итоговую систему и почему

Требования к отчёту

7–9 страниц: постановка задачи; описание модели/режимов; тестовый набор; критерии; результаты; таблица ошибок; анализ конфиденциальности; решение об интеграции в проект.

Критерии оценивания

Работа оценивается в соответствии с критериями силлабуса: аккуратность и точность — 30 %, творчество и креативность — 30 %, полнота и зрелость — 40 %. Оригинальность учитывается коэффициентом 1,0; 0,5 или 0. Работа должна быть самостоятельной, а использование готовых решений без собственного анализа снижает коэффициент оригинальности.

Контрольные вопросы

1. Какой исследовательский вопрос решается в данной работе?

2. Почему выбранные методы соответствуют поставленной задаче?

3. Какие альтернативные подходы были рассмотрены?

4. Какие данные использованы и какие у них ограничения?

5. По каким критериям сравнивались варианты решения?

6. Какие ошибки или нестабильные случаи выявлены?

7. Какие угрозы достоверности эксперимента необходимо учитывать?

8. Какие вопросы конфиденциальности и безопасности возникают?

9. Как результаты работы изменяют индивидуальный семестровый проект?

10. Какой следующий этап исследования необходимо выполнить?

Список источников

1. Бессмертный И. А., Авдюшина А. Е., Кугаевских А. В., Королева Ю. А., Рущенко Н. Г. Системы искусственного интеллекта : методическое пособие. – Санкт-Петербург : Университет ИТМО, 2024. – 281 с.

2. Соснило А. И. Атлас искусственного интеллекта для бизнеса и власти : учебное пособие. – Санкт-Петербург : Университет ИТМО, 2022. – 98 с.

3. Сидоркина И. Г. Системы искусственного интеллекта : учебное пособие. – Москва : КНОРУС, 2018. – 248 с.

4. Баланов А. Н. Искусственный интеллект. Понимание, применение и перспективы : учебник для вузов. – Санкт-Петербург : Лань, 2024. – 312 с.

5. Калинин М. О., Крундышев В. М. Основы искусственного интеллекта. Безопасность искусственного интеллекта : учебное пособие. – Санкт-Петербург, 2024. – 220 с.

6. Рассел С., Норвиг П. Искусственный интеллект: современный подход. – 2-е изд. – Москва : Вильямс, 2018. – 230 с.

7. Андреянов Н. В., Евдокимова Т. С., Павлов А. Д. [и др.]. Методы искусственного интеллекта : учебное пособие. – Казань : КНИТУ-КАИ, 2024. – 392 с.