

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Филологический факультет

УТВЕРЖДЕНО:

Декан

И.В. Тубалова

Оценочные материалы по дисциплине

Промпт инжиниринг

по направлению подготовки

45.04.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) подготовки:

Компьютерная и когнитивная лингвистика

Форма обучения

Очная

Квалификация

Магистр

Год приема

2024

СОГЛАСОВАНО:

Руководитель ОП

З.И. Резанова

Председатель УМК

Ю.А. Тихомирова

Томск – 2025

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий и актуальные концепции в области лингвистики.

ПК-4 Способен разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.2 Решает профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий

ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта

ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта

ИПК-4.3 Обеспечивает выполнение проекта в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта, в соответствии с установленными целями, сроками и затратами

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

– контрольная работа;

В качестве промежуточного текущего контроля каждый студент в обязательном порядке выполняет контрольную работу по основным принципам и техникам промпт-инжиниринга. Контрольная работа выполняется письменно и прикрепляется в файле в системе Moodle, Затем происходит краткий устный доклад и обсуждение результатов на практических занятиях.

Задание № 1 для Контрольной работы (ОПК-1, ПК-4, ИОПК-1.2):

Сравнительный анализ эффективности разных промпт-стратегий в решении одной и той же задачи. Спроектируйте 4 промпта: Zero-shot, Few-shot (с 2–3 примерами), Chain-of-Thought, Structured (JSON/XML/Markdown)

Протестируйте каждый промпт на нескольких запросах — сохраните ответы.

Оцените по критериям: Точность (factuality), Полнота (coverage), Структурированность (structure compliance). Наличие галлюцинаций

Проведите сравнительный анализ: Какой тип промпта лучше для вашей задачи и почему? Какие ошибки характерны для каждого типа? Как можно гибридизировать подходы? Как влияет структурирование на воспроизводимость?

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Зачет в третьем семестре проводится в форме проверки, презентации и защиты творческой аналитической контрольной работы по применению промпт-инжиниринга к исследовательскому или прикладному проекту студента по теме ВКР.

Для зачета можно выбрать одно любое задание или синтезировать любые два задания. Выбранное задание нужно обязательно связать с темой своего курсового проекта или ВКР.

Отчет в письменной форме и краткий устный доклад-презентация в итоге.

Задания для итоговой контрольной проектной работы (для выбора):

ЗАДАНИЕ 2. (ИОПК-1.2, ИПК-4.1, ИПК-4.2, ИПК-4.3):

Построить систему Retrieval-Augmented Generation, где LLM генерирует ответы, опираясь на внешние источники, и сравнить её с “чистой” генерацией — с оценкой снижения галлюцинаций (насколько RAG снизил галлюцинации). Используйте для оценки метрики и методики: Faithfulness → насколько В ближе к ретривированному контексту); Hallucination Rate → сколько ложных утверждений в А vs В; FEVER (Fact Extraction and VERification) → подтверждается ли каждое утверждение в ответе ретривированным контекстом.

ЗАДАНИЕ 3. (ОПК-1, ПК-4, ИОПК-1.2, ИПК-4.1, ИПК-4.2):

1. Построение саморефлексирующей системы анализа текста: Self-Consistency + ReAct + Meta-Prompting. Цель: Создать многоуровневую систему, где LLM: - генерирует несколько ответов (self-consistency), - рассуждает и действует (ReAct), - получает инструкции от “мета-промпта” (meta-prompting).

Этапы:

- Выберите задачу лингво-когнитивного анализа текста.

- Спроектируйте 2 системы:

Система А (линейная): один промпт → один ответ.

Система В (рефлексирующая):

Этап 1: Generate 5 вариантов ответа (self-consistency)

Этап 2: Примени ReAct: “Рассмотри каждый вариант. Какие доказательства в тексте его поддерживают? Какие противоречат?”

Этап 3: Meta-prompt: “Ты — _укажите роль_. Выбери лучший вариант и перепиши его, устранив противоречия.”

Оцените обе системы по:

Точности (сверка с вашей экспертной разметкой), Глубине анализа, Устойчивости к галлюцинациям,

Вычислительной сложности (токены, этапы)

Рефлексия: Преимущества и недостатки усложнения? Где “ломается” ReAct? Как улучшить meta-prompt для вашей области?

ЗАДАНИЕ 4. (ПК-4, ИОПК-1.2, ИПК-4.1, ИПК-4.2, ИПК-4.3):

Управление контекстным окном: сравнение стратегий при работе с длинными текстами или многошаговыми задачами.

Цель: Исследовать, как разные стратегии управления контекстом (разбиение на подзадачи, использование delimiters, token budgeting, summarization) влияют на качество анализа длинных или сложных материалов.

Этапы:

- Выберите длинный текст и сложную задачу (например, лингво-когнитивный анализ главы книги, сравнение 3 статей, разбор корпуса диалогов).

- Спроектируйте 4 подхода:

A: Загрузить всё в один промпт (если позволяет контекстное окно)

B: Разбить на подзадачи (Split complex tasks) — каждый абзац/статью отдельно → потом агрегировать

C: Использовать delimiters + summarization: отдельный анализ для отдельных фрагментов текстов. На основе summaries сделай общий вывод.

D: Token budgeting + iterative refinement: На первом проходе — только ключевые тезисы (макс. 200 токенов). На втором — углубление в спорные места.

Оцените результаты по: Полноте охвата, Глубине анализа, Наличию “потерь” при разбиении, Эффективности использования токенов.

Рефлексия: Какой подход лучше для вашей задачи? Где теряется контекст? Можно ли автоматизировать разбиение?

ЗАДАНИЕ 5. (ИПК-4.1, ИПК-4.2, ИПК-4.3):

Гиперпараметры как инструмент управления смыслом: экспериментальное исследование влияния temperature, top_p, presence_penalty на интерпретацию текста”

Цель: Эмпирически проверить, как гиперпараметры генерации влияют на толкование, креативность и точность в рамках вашей исследовательской задачи — и построить “профиль гиперпараметров” для разных типов задач.

Этапы:

1. Выберите задачу лингво-когнитивной интерпретации текста.

2. Зафиксируйте один промпт — максимально чёткий и структурированный.

3. Варьируйте гиперпараметры в 5–7 комбинациях:

- Базовая: temp=0.7, top_p=0.9, presence_penalty=0.0

- Креативная: temp=1.2, top_p=0.95, presence_penalty=0.5

- Консервативная: temp=0.3, top_p=0.5, presence_penalty=1.0

- “Анти-галлюцинация”: temp=0.5, top_k=20, frequency_penalty=0.8

- “Максимальная осторожность”: temp=0.1, stop=["возможно", "предположительно"]

4. Для каждой комбинации сгенерируйте по несколько ответов.

5. Постройте “карту влияния гиперпараметров” — например:

→ Высокая температура → больше креатива, но выше риск галлюцинаций.

→ Низкий top_p → стабильнее, но беднее семантика.

Рефлексия: Какие настройки оптимальны для вашей задачи? Можно ли “подписывать” гиперпараметры под тип задачи (анализ vs. генерация vs. рефлексия)?

ЗАДАНИЕ 6. (ИОПК-1.2, ИПК-4.1, ИПК-4.2, ИПК-4.3):

Спроектируйте и протестируйте минималистичный оптимизированный LLM-пайплайн для одной из классических прикладных NLP задач.

Используйте материалы и данные, связанные с вашим исследовательским проектом.

Классифицируйте ошибки LLM (hallucination, truncation, misalignment, format drift и др.). Зафиксируйте “слепые зоны” модели.

ЗАДАНИЕ 7. (ПК-4, ИОПК-1.2, ИПК-4.1, ИПК-4.2, ИПК-4.3):

Коллаборатория ИИ-агентов или AI-ассистентов для создания системы сбора и обработки данных.

Используй любой из доступных сервисов: www.chatwoot.com, rasa.com, www.llamaindex.ai, www.crewai.com, chatdev.ai, www.langchain.com, www.langflow.org,

flowiseai.com, www.vtiger.com, www.sim.ai или найдите другой интересующий репозиторий на GitHub.

Используйте текстовый материал по своему проекту в качестве "сырья" для ИИ-агентов.

Придумайте 3–5 "агентов" — каждому дайте:

Подробно прописанную роль (например, не просто "анализируй", "а ты — лингвист-семиотик, специализируешься на вербальных маркерах идентичности"),

Задачу (промпт по прикладному NLP),

Ограничения, Формат вывода.

Организируйте поток данных между ИИ-агентами — как эстафету:

Агент 1 собирает или извлекает примеры (термины, фразы, интенции и т.п.) → передает их Агенту 2.

Агент 2 классифицирует их по категориям (фреймам, доменам и т.п.) → передает их Агенту 3.

Агент 3 сравнивает с теорией X (сравнивает с теоретической базой или глоссарием — как "верификатор".) → передал Агенту 4.

Агент 4 пишет вывод (обобщает, выявляет паттерны, строит гипотезы — как "исследователь") → Агент 5 его проверяет и улучшает вывод (запоминает ошибки других агентов, рефлексиирует, критикует, редактирует).

ЗАДАНИЕ 8. (ОПК-1, ПК-4, ИОПК-1.2, ИПК-4.1):

На основе тематического тезауруса или онтологии, связанных с вашим исследовательским проектом, составьте систему промптов для проверки и повышения SEO-эффективности тематических сайтов, апробируйте и оцените результативность такого промпт-инжиниринга.

ЗАДАНИЕ 9. (ОПК-1, ПК-4, ИОПК-1.2, ИПК-4.1):

На основе тематического тезауруса или онтологии, связанных с вашим исследовательским проектом, составьте систему промптов для sentiment-анализа и intent-анализа. Создайте на сервисе www.coze.com или любым другим способом чат-бот, который будет учитывать выявленные тематические интенции в диалоге с пользователем.

ЗАДАНИЕ 10. (ИОПК-1.2, ИПК-4.1, ИПК-4.2, ИПК-4.3):

Сравнение эффективности разных методов промпт-инжиниринга для структурирования и очистки данных, а также разных форматов линеаризации таблиц на одной задаче, связанной с вашим исследовательским проектом.

- Протестируйте 2–3 формата линеаризации таблиц (CSV, Markdown, JSONL, TSV) на стабильность парсинга LLM.

- Зафиксируйте и классифицируйте ошибки модели (hallucination, format drift, entity misalignment, truncation).

- Определите оптимальный pipeline: метод + формат → минимальная ошибка, максимальная воспроизводимость.

ЗАДАНИЕ 11. (ИПК-4.1, ИПК-4.2, ИПК-4.3):

Создание MVP веб-интерфейса с LLM-ассистентом: структурирование текста или обработка структурированных данных.

Разработайте минимально жизнеспособный веб-интерфейс (MVP), который через браузер позволяет пользователю загружать неструктурированный текст (или

структурированные данные) и получать их обработанную/структурированную версию. Реализация — на любом языке (JS/Python/etc), финал — рабочая HTML-страница.

Требования:

Используйте LLM как co-pilot для генерации, рефакторинга и дебаггинг кода. Ведите журнал промптов: каждый шаг разработки (планирование, генерация, отладка) — через явные, версионизируемые промпты.

Примените минимум 3 разные методики промпт-инжиниринга. Рефлексия:

По итогу - какой подход к промптингу оказался наиболее эффективен для быстрой итерации и разработки MVP?

ЗАДАНИЕ 12. (ИПК-4.1, ИПК-4.2, ИПК-4.3):

Сравнительный анализ метрик качества на собственном корпусе текстов.

Этапы выполнения:

— Сформируйте небольшой корпус (5–10 пар “входной промпт → эталонный ответ”), релевантный вашей исследовательской теме.

— Эталон может быть написан вами вручную или взят из авторитетного источника (например, учебник, статья, официальная документация).

— Для каждого промпта сгенерируйте по 2–3 варианта ответа, меняя формулировку промпта (разные стили, уровни детализации, инструкции). Используйте любую публичную LLM

Применение метрик:

— Рассчитайте вручную или с помощью открытых инструментов (например, bert-score через Hugging Face Spaces, онлайн-калькуляторы BLEU/ROUGE):

BLEU-4 (на основе n-грамм), ROUGE-L (через онлайн-инструменты или Python-скрипты), BERTScore (через Hugging Face Inference API или Spaces), субъективную оценку (Human Evaluation) по 2 критериям: Информативность (Faithfulness), Следование инструкции (Instruction Following) — по шкале 1–5.

— Постройте таблицу/график, где по осям: автоматические метрики vs. человеческие оценки.

— Какая метрика лучше коррелирует с субъективной оценкой?

— Где метрики «ложно срабатывают» и почему?

ЗАДАНИЕ 13. (ИПК-4.1, ИПК-4.2, ИПК-4.3):

Классификация и оценка бенчмарков: построение персонализированного “мини-бенчмарка”

Цель:

На основе классификации бенчмарков создать и протестировать собственный мини-бенчмарк, релевантный вашей исследовательской теме, и оценить на нём поведение LLM.

Этапы выполнения:

Создание мини-бенчмарка (5–7 задач)

— На основе вашей исследовательской темы создайте:

- 2 задачи с множественным выбором (MC)
- 2 генеративные задачи (открытые)
- 1 задачу на “следование инструкции”
- 1 задачу на “верность фактам” (Faithfulness)

— Формализуйте критерии оценки для каждой задачи.

— Протестируйте 2–3 публичных LLM.

— Зафиксируйте:

- Точность на MC-задачах
- Качество генерации (по вашей шкале Faithfulness/Consistency/Instruction Following)

— Попробуйте “обмануть” модель — дать провокационный или двусмысленный промпт.

Рефлексия и выводы

- Сравните модели: на каких задачах и какая лучше? Почему?
- Как ваш мини-бенчмарк соотносится с известными крупными бенчмарками?
- Какие ограничения выявились? Что нужно добавить/изменить?

Работа может быть зачтена, если она соответствует следующим критериям:

Работа демонстрирует наличие у магистранта знаний о промпт-инжинирингу, основных и специальных техниках, умение адаптировать приемы, стратегии и технологии промпт-инжиниринга к предметной области своего исследовательского или прикладного проекта. Магистрант должен проявить способность позиционировать свою работу в контексте современных теории и тенденций в области искусственного интеллекта, LLM и промпт-инжиниринга.

Продолжительность презентации и защиты в форме ответов на вопросы 15-20 минут на одного студента.

Устная презентация работы и письменный отчет к ней должна отвечать следующим критериям:

1. успешное удержание внимания на речи и презентации докладчика;
2. адекватное оформление презентационных материалов по времени, дизайну и структуре;
3. разделение поданной информации на главную и второстепенную;
4. умение адаптировать материалы к специфике задач контрольной работы;
5. умение корректно (в рамках научной дискуссии) отвечать на поставленные вопросы аудитории и сделанные замечания.

Результаты зачета определяются оценками «зачтено», «не зачтено».

Работа должна демонстрировать умение чётко формулировать исследовательскую задачу, описывать этапы проектирования и тестирования промптов, а также аргументированно интерпретировать полученные результаты. Особое внимание уделяется владению научным стилем, логике изложения, корректному использованию терминологии и способности критически оценивать собственные методы и выводы.

При оценке проекта основное внимание уделяется:

- пониманию принципов построения, тестирования и анализа промптов;
- качеству оформления отчёта (структура, воспроизводимость эксперимента, интерпретация метрик);
- способности обосновать выбор стратегий, гиперпараметров или архитектур LLM-пайплайнов.

Критерии оценки:

Зачтено — проект оригинален и полностью соответствует поставленной задаче; отчёт структурирован, логичен, содержит самостоятельно подобранные и корректно проанализированные примеры; выявлены закономерности, предложены обоснованные выводы. Допущено не более одной незначительной ошибки.

Зачтено — проект выполнен, но примеры заимствованы, не оригинальны или интерпретированы поверхностно, структура отчёта недостаточно чёткая; примеры заимствованы из учебных материалов; имеются отдельные пробелы в объяснении методов или результатов, устранимые при уточнении. Допущено до двух ошибок.

Зачтено — работа неполная, слабо структурирована; студент затрудняется связать выбранные техники промпт-инжиниринга с целями исследования; допущены существенные ошибки в интерпретации ответов модели или метрик. Допущено до трёх ошибок.

Не зачтено — работа не выполнена или содержит фундаментальные ошибки в понимании базовых принципов промпт-инжиниринга; отсутствует логика эксперимента,

интерпретация результатов невозможна; студент не может привести или объяснить примеры. Допущено более четырёх существенных ошибок.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Тесты для Контроля остаточных знаний:

Компетенция: ОПК-1

Способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий, и актуальные концепции в области лингвистики.

1. Какой лингвистический принцип лежит в основе метода Chain-of-Thought prompting?

- А) Принцип экономии (принцип Оккама)
- В) Концепция дискурсивной когерентности
- С) Теория речевых актов
- Д) Принцип последовательного логического рассуждения как формы текстовой когерентности

когерентности

- Правильный ответ: D

2. Какой из перечисленных аспектов семантики наиболее уязвим при галлюцинациях LLM?

- А) Синтагматические отношения
- В) Денотативное значение
- С) Прагматическая нагрузка
- Д) Фонологическая структура

- Правильный ответ: B

3. Какой лингвистический подход обосновывает использование structured prompting (например, JSON)?

- А) Генеративная грамматика
- В) Функционально-коммуникативная теория
- С) Теория текстовой структуры и форматирования как прагматического маркера
- Д) Структурализм

- Правильный ответ: C

4. При анализе интенгов в диалоге с LLM-ассистентом наиболее релевантной является:

- А) Теория метафоры Лакоффа
- В) Классификация иллокутивных сил по Серлю
- С) Фонология высказывания
- Д) Морфологический разбор

- Правильный ответ: B

5. Какой лингвистический феномен может привести к «format drift» при генерации структурированных данных?

- А) Полисемия
- В) Нарушение прагматических норм форматирования
- С) Синтаксическая неоднозначность
- Д) Лексическая сочетаемость

- Правильный ответ: B

Компетенция: ПК-4

Способен разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

1. Какой из перечисленных проектов НАИБОЛЕЕ соответствует прикладной задаче в компьютерной лингвистике с использованием LLM?
A) Анализ поэтического размера в стихах Пушкина
B) Создание MVP-интерфейса для автоматического структурирования научных аннотаций
C) Сравнение диалектных форм в русских говорах
D) Реконструкция праславянской лексики
- Правильный ответ: B
2. Какой инструмент НЕ относится к техническим средствам реализации LLM-пайплайна?
A) LangChain
B) spaCy
C) Hugging Face Transformers
D) LaTeX
- Правильный ответ: D
3. Для реализации RAG-системы в задаче анализа юридических текстов НЕОБХОДИМО использовать:
A) Векторную базу данных (например, FAISS или Chroma)
B) Систему машинного перевода
C) Корпус древнерусских текстов
D) Фонетический анализатор
- Правильный ответ: A
4. Какой из перечисленных подходов позволяет интегрировать тематический тезаурус в LLM-систему?
A) Fine-tuning на корпусе с аннотациями
B) Использование мета-промптов с включённой онтологией
C) Изменение архитектуры внимания
D) Увеличение batch size
- Правильный ответ: B
5. Какой из проектов напрямую связан с SEO-оптимизацией через промпт-инжиниринг?
A) Генерация мета-описаний с учётом ключевых запросов и семантического ядра
B) Кластеризация пользовательских отзывов
C) Транскрибирование аудиозаписей
D) Построение деривационных гнёзд
- Правильный ответ: A

Компетенция: ИОПК-1.2

Решает профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий.

1. Какой лингвистический принцип объясняет необходимость использования delimiters в промптах для сложных задач?
A) Принцип языковой экономии
B) Теория информационной структуры высказывания (тема/рема)
C) Принцип маркированности
D) Теория речевой деятельности
- Правильный ответ: B

2. При sentiment-анализе с использованием LLM наиболее важной является:
- A) Теория эмоциональной окраски лексики
 - B) Фонологическая теория
 - C) Морфологическая парадигма
 - D) Синтаксическая валентность
- Правильный ответ: A
3. Какой лингвистический феномен затрудняет корректную интерпретацию промптов с высокой температурой?
- A) Полисемия и контекстуальная неоднозначность
 - B) Алломорфия
 - C) Синкопа
 - D) Акцентуация
- Правильный ответ: A
4. Какой теоретический подход лежит в основе self-consistency prompting?
- A) Теория вероятностного языкового моделирования
 - B) Концепция консенсуса в дискурсе
 - C) Принцип максимальной энтропии
 - D) Теория интертекстуальности
- Правильный ответ: B
5. При анализе «слепых зон» LLM в задачах классификации интенгов важно учитывать:
- A) Теорию прагматических импликатур
 - B) Фонетические законы
 - C) Грамматические категории времени
 - D) Лексикографические нормы
- Правильный ответ: A

Компетенция: ИПК-4.1

Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

1. Какая формулировка цели проекта НАИБОЛЕЕ корректна для задачи по снижению галлюцинаций в LLM?
- A) «Изучить историю развития языковых моделей»
 - B) «Разработать и протестировать RAG-систему для повышения фактологической достоверности ответов в домене медицины»
 - C) «Написать эссе о проблемах ИИ»
 - D) «Сравнить два стиля оформления промптов»
- Правильный ответ: B
2. Какое обоснование применения ИИ-технологий является НАИБОЛЕЕ уместным для проекта по структурированию неформатированных текстов?
- A) «LLM позволяют автоматизировать извлечение структурированной информации без ручной разметки»
 - B) «ИИ моден и популярен»
 - C) «Студенты должны уметь пользоваться ChatGPT»
 - D) «Это требует преподаватель»
- Правильный ответ: A

3. Какая цель проекта соответствует задаче управления длинным контекстом?

А) «Создать бота для общения»

В) «Сравнить эффективность стратегий разбиения и суммаризации при анализе многостраничных документов»

С) «Перевести текст на английский»

Д) «Изучить влияние шрифтов на читаемость»

- Правильный ответ: В

4. Какое утверждение НАИЛУЧШИМ образом обосновывает использование гиперпараметров в исследовании интерпретации текста?

А) «Гиперпараметры позволяют управлять балансом между креативностью и точностью генерации»

В) «Модель работает быстрее при высокой температуре»

С) «Это стандартная практика»

Д) «Так делают в YouTube»

- Правильный ответ: А

5. Какая формулировка цели проекта отражает применение промпт-инжиниринга в SEO?

А) «Повысить видимость сайта за счёт генерации контента, соответствующего семантическому ядру и требованиям поисковых систем»

В) «Написать 100 статей»

С) «Научиться пользоваться Google»

Д) «Улучшить дизайн сайта»

- Правильный ответ: А

Компетенция: ИПК-4.2

Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта.

1. Какой этап НЕ входит в программу действий для реализации self-reflexive LLM-системы?

А) Генерация 5 вариантов ответа

В) Применение ReAct-рассуждений

С) Обучение новой языковой модели с нуля

Д) Применение мета-промпта для выбора итогового ответа

- Правильный ответ: С

2. Какой порядок действий корректен для тестирования влияния гиперпараметров на интерпретацию текста?

А) Выбор задачи → фиксация промпта → варьирование temperature/top_p → сбор ответов → анализ

В) Сбор данных → обучение модели → настройка гиперпараметров → тестирование

С) Написание отчёта → генерация ответов → выбор гиперпараметров

Д) Выбор модели → изменение архитектуры → тестирование

- Правильный ответ: А

3. Какой компонент НЕ требуется для реализации RAG-системы?

А) Эмбединг-модель

- В) Векторное хранилище
- С) Система fine-tuning'а модели
- Д) Механизм ретривала
- Правильный ответ: С

4. Какой шаг является ключевым при создании MVP с LLM-ассистом для структурирования текста?

- А) Версионирование промптов для каждого этапа разработки
- В) Покраска интерфейса в синий цвет
- С) Использование только одной LLM
- Д) Отказ от тестирования
- Правильный ответ: А

5. Какой подход к управлению контекстом НАИБОЛЕЕ рационален при анализе трёх научных статей?

- А) Загрузить всё в один промпт
- В) Разбить на подзадачи → проанализировать каждую → агрегировать

Выводы

- С) Удалить все примеры из промпта
- Д) Использовать только temperature = 0
- Правильный ответ: В

Компетенция: ИПК-4.3

Обеспечивает выполнение проекта в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта, в соответствии с установленными целями, сроками и затратами.

1. Какой показатель НАИБОЛЕЕ релевантен для оценки эффективности использования ресурсов при тестировании промптов?

- А) Количество использованных токенов
- В) Красивое оформление отчёта
- С) Число упоминаний ChatGPT
- Д) Длина промпта в символах
- Правильный ответ: А

2. Какой из перечисленных факторов может нарушить соблюдение сроков в проекте по созданию LLM-пайплайна?

- А) Отсутствие версионирования промптов и воспроизводимости экспериментов
- В) Использование открытых моделей
- С) Наличие чёткой цели
- Д) Применение метрик качества
- Правильный ответ: А

3. Какой подход обеспечивает соответствие проекта по RAG установленной цели снижения галлюцинаций?

- А) Оценка Faithfulness и Hallucination Rate на контрольном наборе
- В) Генерация как можно большего числа ответов
- С) Использование только одного источника
- Д) Отказ от внешнего контекста
- Правильный ответ: А

4. Что из перечисленного свидетельствует о контроле над «затратами» в проекте с LLM?
- A) Минимизация количества итераций за счёт продуманного промпт-дизайна
 - B) Использование самой дорогой модели всегда
 - C) Генерация 1000 ответов на каждый промпт
 - D) Отсутствие логирования запросов
- Правильный ответ: A
5. Какой элемент проекта по структурированию таблиц обеспечивает соответствие цели «максимальной воспроизводимости»?
- A) Фиксация формата линеаризации и версии модели
 - B) Случайный выбор промптов
 - C) Отказ от метрик
 - D) Использование только ручной проверки
- Правильный ответ: A

Информация о разработчиках

Бочаров Алексей Владимирович, канд. ист. наук, доцент, кафедра общей, компьютерной и когнитивной лингвистики (по совместительству), доцент