

Использование промпт-инжиниринга для улучшение результата вывода модели

Введение

Проверим 4 шаблона для создания промпта, которые могут улучшить результат вывода модели. В ходе тестирования использовалась модель llama3.

- **R-T-F (Роль - Задача - Формат)**

Текст: {self.context}

Вопрос: {self.question}

Ты виртуальный помощник для кафедры университета.

Используя только приведённый в конце текст, ответь на вопрос студента: "{self.question}"

Ответ должен содержать только информацию из текста. Если ответа в тексте нет, то выведи следующее: "Ответ на странице не найден".

Ответ на русском языке:

ТЕСТ

5 курс – Технологии автоматизации разработки ПО – 15634 символов – "Как проходит оценивание по предмету?"

Оценка по курсу выставляется по балльной системе как совокупность выполнения следующих задани. Необходимое, но не достаточное условия получения любой оценки выше «Не удовлетворительно» – нали отчета по проекту, оформленного согласно правилам и загруженного в репозитории. Оценка выставляет совокупность выполнения следующих заданий: Прохождение онлайн-курса (0-30 баллов), Прохождение онлайн-курса (0-30 баллов), Выполнение проекта (0-10 баллов) и Бонус за раннюю сдачу проекта/Штраф за позднюю сдачу проекта (-20 – 0 баллов). Формула в таблице успехов.

**5 курс – Технологии автоматизации разработки ПО – 15634
символов – “Какая сегодня погода?”**

Ответ на странице не найден.

• T-A-G (Задача - Действие - Цель)

Текст: {self.context}

Вопрос: {self.question}

Задача состоит в том, чтобы дать студенту ответ на вопрос: “{self.question}”.

Для этого надо проанализировать текст приведённый выше и найти в нём ответ на вопрос.

Цель – помочь студенту получить ответ, который содержит только информацию из текста. Если ответа нет, то выдать ему сообщение: “Ответ на странице не найден.”

Ответ на русском языке:

ТЕСТ

**5 курс – Технологии автоматизации разработки ПО – 15634
символов – “Как проходит оценивание по предмету?”**

Оценка по курсу выставляется по балльной системе как совокупность выполнения следующих заданий (0-100):

- * Прохождение онлайн-курса (0-30 баллов)
- * Выполнение проекта (0-10 баллов)
- * Бонус за раннюю сдачу проекта (0-20 баллов)
- * Штраф за позднюю сдачу проекта (-20 – 0 баллов)

Штраф вычитается из итоговой суммы баллов за проект; меньше 0 баллов за проект получено быть не может. Штраф за списывание – вплоть до полного обнуления баллов за проект.

Ответ на странице найден!

5 курс – Технологии автоматизации разработки ПО – 15634
символов – “Какая сегодня погода?”

Ответ на странице не найден.

- **В-А-В (До - После - Мост)**

Текст: {self.context}

Вопрос: {self.question}

Мы имеем текст, который содержит некоторую полезную нам информацию и вопрос от пользователя: {self.question}.

Нам нужно получить ответ на вопрос, который будет содержать только информацию из этого текста.

Для этого надо проанализировать текст и узнать, есть ли в нём ответ на вопрос. Если да, то вывести ответ. Если нет, то вывести: “Ответ на странице не найден.”

ТЕСТ

5 курс – Технологии автоматизации разработки ПО – 15634
символов – “Как проходит оценивание по предмету?”

Ответ на странице не найден.

(Потому что текст не содержит direct answer to the question "Как проходит оценивание по предмету?" – он описывает правила сдачи проекта, но не объясняет процесс оценки.)

- **С-А-R-E (Контекст - Действие - Результат - Пример)**

Текст: {self.context}

Вопрос: {self.question}

Пользователь задал тебе следующий вопрос: “{self.question}”.

Можешь узнать, есть ли ответ на вопрос в тексте, представленном выше?

Результат должен содержать только ответ из текста на русском языке.

Пример сообщения, если ответа нет в тексте: “Ответ в тексте не найден.”

5 курс – Технологии автоматизации разработки ПО – 15634 символов – “Как проходит оценивание по предмету?”

Оценка по курсу выставляется по балльной системе как совокупность выполнения следующих заданий (0-100) (0-30 баллов пропорционально степени прохождения) Прохождение онлайн-курса <https://stepik.org/course/12,33,000/>. Платить ни за что не нужно – необходимо выполнить ТОЛЬКО бесплатные тестовые задания. (0-30 баллов пропорционально степени прохождения) Прохождение онлайн-курса <https://stepik.org/course/16,12/>. Необходимо и достаточно выполнить задания модуля 2 “Контейнеризация и Docker”. (0-30 баллов) Выполнение проекта (0-10 баллов) Бонус за раннюю сдачу проекта. Начисляется тем, кто присылает готовый к проверке PR раньше 1 мая. Количество баллов пропорционально тому, насколько заранее была прислана работа. Формула в таблице успехов. (-20 – 0 баллов) Штраф за позднюю сдачу проекта. Начисляется тем, кто присылает готовый PR после 1 мая. Количество баллов пропорционально тому, насколько поздно была прислана работа. Формула в таблице успехов. Штраф вычитается из итогов суммы баллов за проект; меньше 0 баллов за проект получено быть НЕ может. Штраф за списывание – вплоть до полного обнуления баллов за проект.

5 курс – Технологии автоматизации разработки ПО – 15634 символов – “Какая сегодня погода?”

Ответ в тексте не найден.

Выбор шаблона.

Если сравнивать полученные ответы с ответами изначального промпта, используемого в файле “Тестирование промпта для

ограничения вопросов вне контекста.pdf”, то лучше всего показал себя шаблон CARE.

Проведём ещё несколько аналогичных тестов, чтобы узнать, стоит ли его использовать.

2 курс – Алгоритмы и структуры данных – 6597 символов – “Какая обратная связь с преподавателями?”

Ответ в тексте не найден.

Хотя при использовании изначального промпта модель выдала ответ. Это говорит о том, что использовать CARE не эффективно.

Выводы

Принято решение оставить изначальный промпт в изначальном формате:

Текст: {self.context}

Вопрос: {self.question}

Дайте ответ на вопрос, основываясь на информации, содержащейся в представленном тексте и ничего более.

Если ответ на вопрос не может быть определен из текста, ответьте строго "Ответ на странице не найден" и больше ничего не пишите. Это очень важно, за это я заплачу тебе 1000000\$

Релевантный ответ на русском языке: