

Тестовое задание на позицию ML-инженер

Задание:

Есть транскрипт встречи, необходимо сгенерировать для него followup (буллет-поинты). Followup -- протокол встречи, в нем должно содержаться самая важная информация (итоги, договоренности, ключевые решения), сгруппированная по темам. Например:

Обновление продукта:

- новая страница подписки
- улучшено саммари в отчете
- повышена производительность продукта

Конференции и встречи:

- в среду встреча с X в 20:00
- с конференции Y пришло 2 новых клиента

....

Важно, чтобы итоговый followup содержал все важные моменты со встречи и не содержал “лишнего” (ненужных итогов и воды).

Использовать можно любые модели (ChatGPT, YandexGPT, open-source модели) и любые подходы (zero-shot, few-shot), экспериментировать с промптами для улучшения качества результата.

Важный момент: большинство моделей не могут вместить весь текст транскрипта целиком в контекстное окно (не получится получить followup по всему транскрипту целиком), поэтому можно разбивать текст на части, прогонять их и потом объединять между собой определенным образом (можно также через отдельный промпт просить у модели объединить). Это нужно для того, чтобы объединить под одну тему все итоги и результаты из разных частей транскрипта. Например тема “Обновление продукта” могла обсуждаться в первые 10 минут встречи и в самом конце.

Будет оцениваться:

- подход кандидата к решению
- креатив в решении задачи
- качество полученного результата (сжатость информации, связанность и отсутствие воды)
- процесс подбора промптов и обоснование выбора используемых моделей
- интерес к подобного рода задачам

Транскрипт, по которому нужно построить followup:

<https://drive.google.com/file/d/1FHQWj8bYMzrl6hJWfTAseXYjOvdJ68mG/view?usp=sharing>