

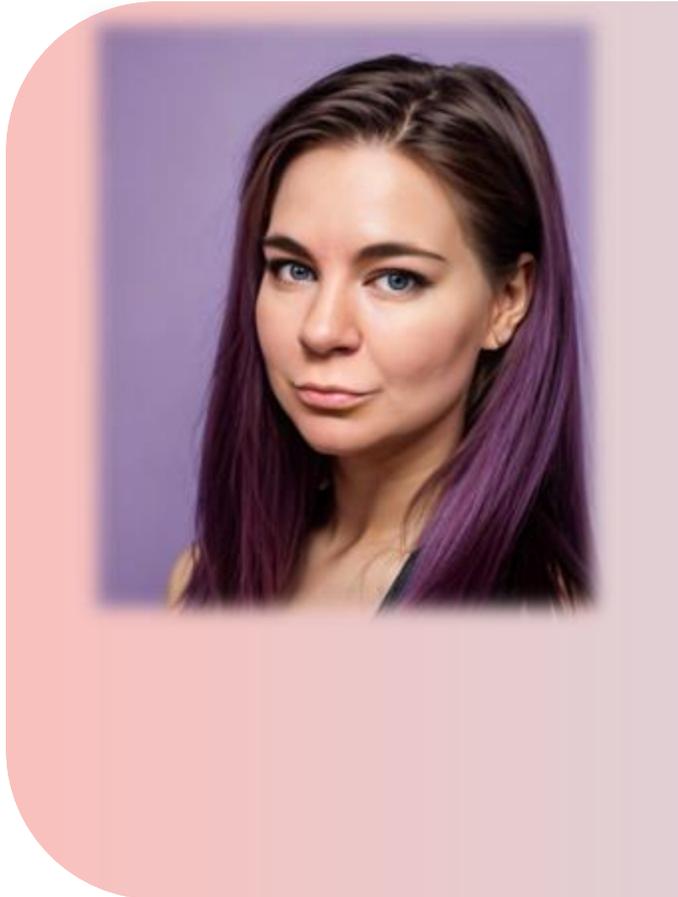
+ | T I

A

Как мы тестировали LLM для собеседований

А они тестировали нас

СПИКЕРЫ



Сабанеева Анастасия

Главный технический лидер тестирования, АО Альфа-Банк

- Внедрение лучших практик и процессов тестирования
- Организация производственного процесса разработки ПО на всем ЖЦ



СПИКЕРЫ

Васильев Антон

QA Manager, ИТ-холдинг Т1

- Внедрение нестандартных инструментов для тестирования
- Участие в создании пилота для ускорения собеседований на базе ИИ



A

О ЧЕМ МЫ
СЕГОДНЯ
ПОГОВОРИМ



A

О ЧЕМ МЫ
СЕГОДНЯ
ПОГОВОРИМ



СОБЕСЕДОВАНИЕ



О ЧЕМ МЫ
СЕГОДНЯ
ПОГОВОРИМ



СОБЕСЕДОВАНИЕ
ЧТО ТАКОЕ LLM



О ЧЕМ МЫ
СЕГОДНЯ
ПОГОВОРИМ

СОБЕСЕДОВАНИЕ
ЧТО ТАКОЕ LLM
ПИЛОТ





О ЧЕМ МЫ СЕГОДНЯ ПОГОВОРИМ



СОБЕСЕДОВАНИЕ

ЧТО ТАКОЕ LLM

ПИЛОТ

ПОДГОТОВКА ДАННЫХ



О ЧЕМ МЫ СЕГОДНЯ ПОГОВОРИМ



СОБЕСЕДОВАНИЕ

ЧТО ТАКОЕ LLM

ПИЛОТ

ПОДГОТОВКА ДАННЫХ

ТЕСТИРОВАНИЕ



О ЧЕМ МЫ СЕГОДНЯ ПОГОВОРИМ



СОБЕСЕДОВАНИЕ

ЧТО ТАКОЕ LLM

ПИЛОТ

ПОДГОТОВКА ДАННЫХ

ТЕСТИРОВАНИЕ

ЭФФЕКТЫ



О ЧЕМ МЫ СЕГОДНЯ ПОГОВОРИМ



СОБЕСЕДОВАНИЕ

ЧТО ТАКОЕ LLM

ПИЛОТ

ПОДГОТОВКА ДАННЫХ

ТЕСТИРОВАНИЕ

ЭФФЕКТЫ

УГРОЗЫ



СОБЕСЕДОВАНИЕ



СОБЕСЕДОВАНИЕ

Цель собеседования заключается в том, чтобы найти синергию между кандидатом и компанией, определить, что получает каждая из сторон.

A

СОБЕСЕДОВАНИЕ



A

СОБЕСЕДОВАНИЕ

Долго

Однообразно

Потеря
концентрации



A

XO4Y



A

ХОЧУ

Структура





ХОЧУ

Структура



Инсайты





ХОЧУ

Структура



Инсайты



Объективность





ХОЧУ

Структура



Инсайты



Объективность



Автоматизация





ХОЧУ

Структура



Инсайты



Объективность



Автоматизация



Отсутствие ошибок





ХОЧУ

Структура



Инсайты



Объективность



Автоматизация



Отсутствие ошибок



История



A

МОГУ





ЧТО ТАКОЕ LLM



ЧТО ТАКОЕ LLM

Это большая языковая модель, которая обучена на огромных объёмах текстовых данных для генерации и обработки естественного языка.

A

ЧТО ТАКОЕ LLM

Запомнит всё

Подскажет вопрос

Звоночки и паттерны

Подготовит отчёт

Не устаёт

A

ЧТО ТАКОЕ LLM

Запомнит всё

Подскажет вопрос

Звоночки и паттерны

Подготовит отчёт

Не устаёт

A

ЧТО ТАКОЕ LLM

Запомнит всё

Подскажет вопрос

Звоночки и паттерны

Подготовит отчёт

Не устаёт

A

ЧТО ТАКОЕ LLM

Запомнит всё

Подскажет вопрос

Звоночки и паттерны

Подготовит отчёт

Не устаёт

A

ЧТО ТАКОЕ LLM

Запомнит всё

Подскажет вопрос

Звоночки и паттерны

Подготовит отчёт

Не устаёт

A

ЧТО ТАКОЕ LLM ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ

LLM

ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ

LLM

ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ LLM ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ LLM ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ

LLM

ТИПЫ

Общие языковые модели

Модели для анализа эмоций
и поведения

Специализированные модели

Модели для перевода и локализации

Модели для голосового анализа

A

ЧТО ТАКОЕ

LLM

ТИПЫ

Общие языковые модели

Gemini

Используется для оценки резюме и проведенного собеседования

Qwen

Используется для поддержки проведения собеседования

A

ЧТО ТАКОЕ LLM ТИПЫ

Общие языковые модели

Gemini

Используется для оценки резюме и
проведенного собеседования

Qwen

Используется для поддержки
проведения собеседования

A

ЧТО ТАКОЕ LLM ТИПЫ

Общие языковые модели

Gemini

Используется для оценки резюме и
проведенного собеседования

Qwen

Используется для поддержки
проведения собеседования

A

ЧТО ТАКОЕ

LLM

ТИПЫ

Общие языковые модели

Gemini

Qwen

GPT

Новая модель, которая появилась вчера



ПИЛОТ



ПИЛОТ

5 месяцев работы

ПИЛОТ



5 месяцев работы

Разработка MVP (3 месяца)



ПИЛОТ

5 месяцев работы

Разработка MVP (3 месяца)

Тестирование на работающих
сотрудниках (1 месяц)

ПИЛОТ



5 месяцев работы

Разработка MVP (3 месяца)

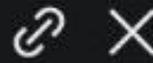
Тестирование на работающих
сотрудниках (1 месяц)

Применение и проверка гипотез на
реальных собеседованиях (1 месяц)

ПИЛОТ



Участники 18



Участник или контакт



Кирилл Александрович

Дмитрий Владимирович
Модератор

Асад Арсланович

Арина К

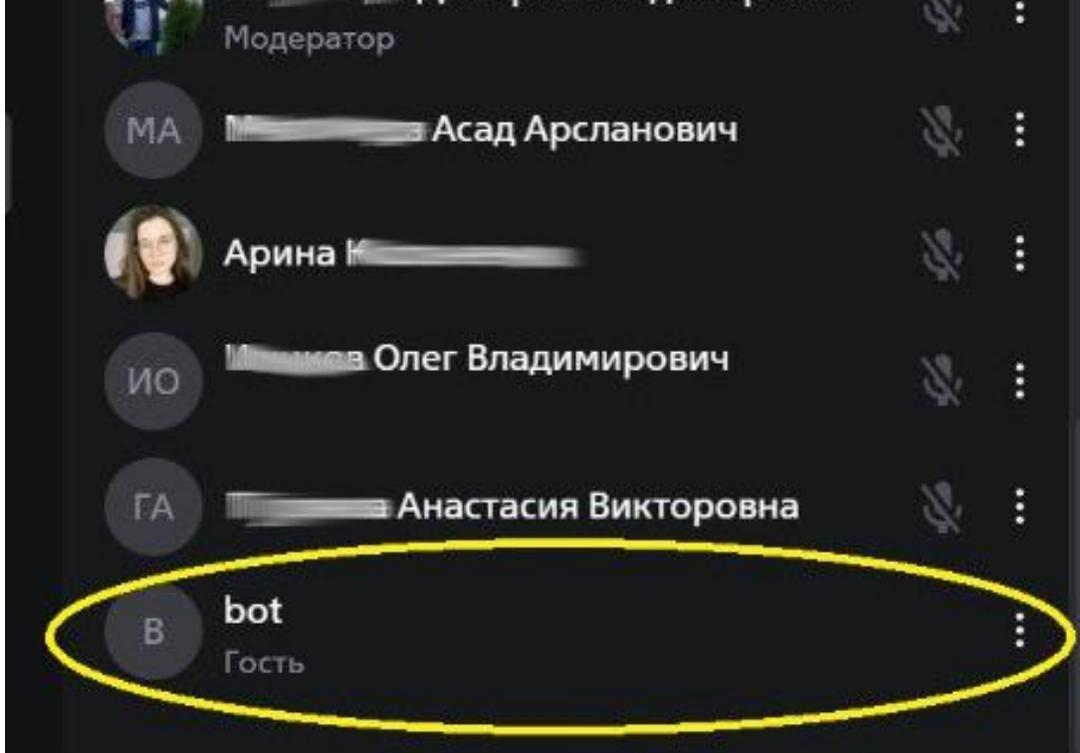
Олег Владимирович

Анастасия Викторовна

bot
Гость

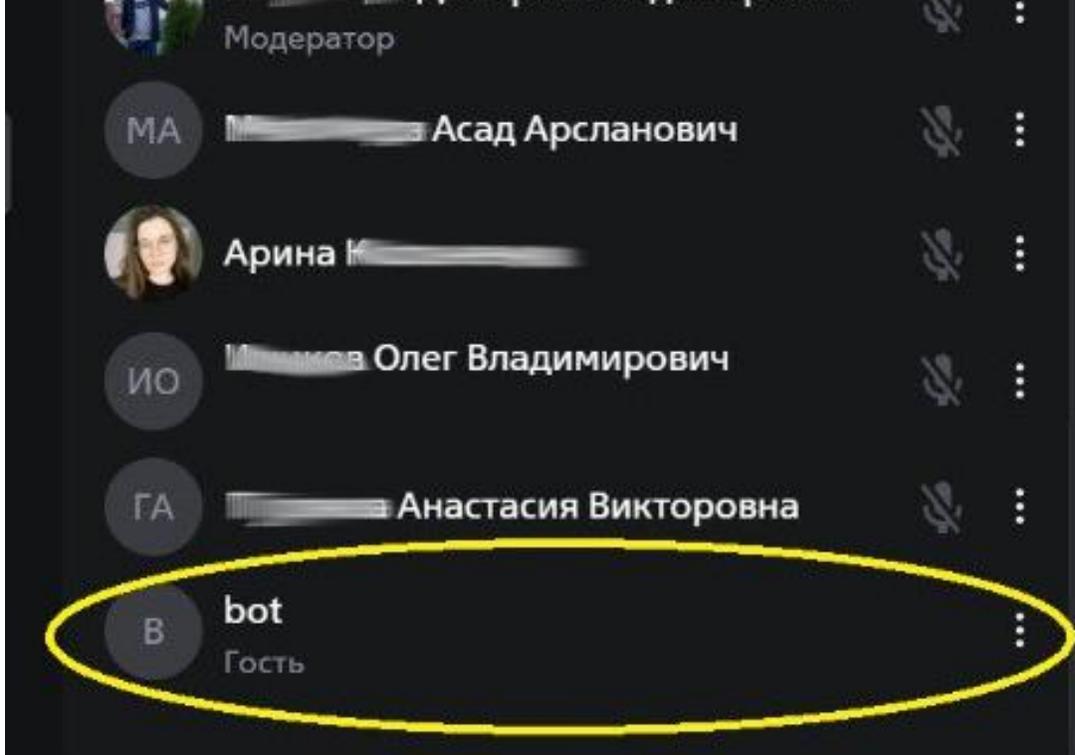


ПИЛОТ



ПИЛОТ

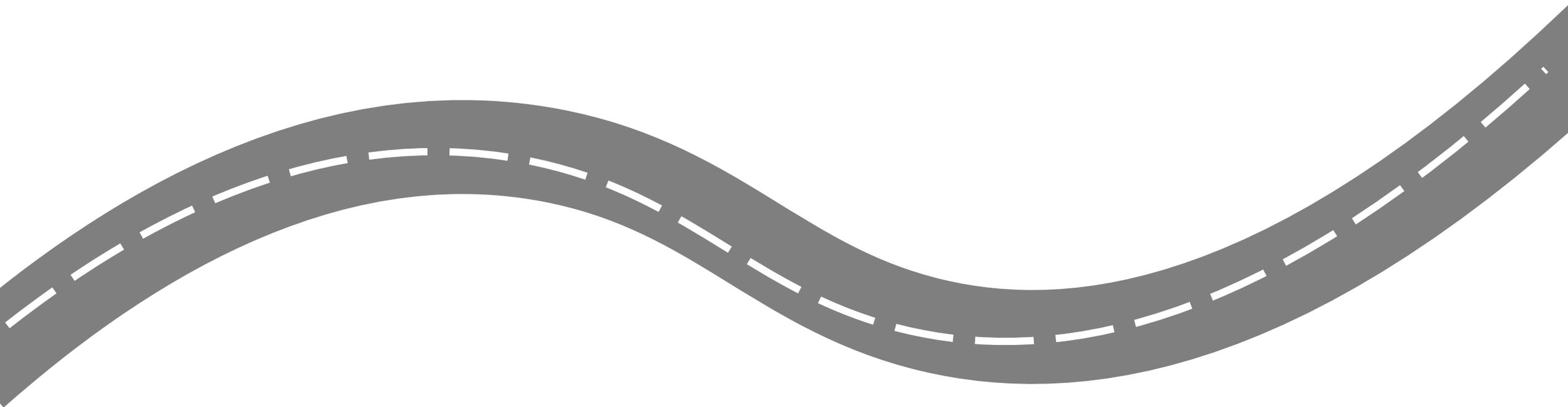
Во время проведения пилота не планировалось подключать ИИ ассистент к собеседованию.



ПИЛОТ

Обработка данных
проводилась по факту
свершившегося события.

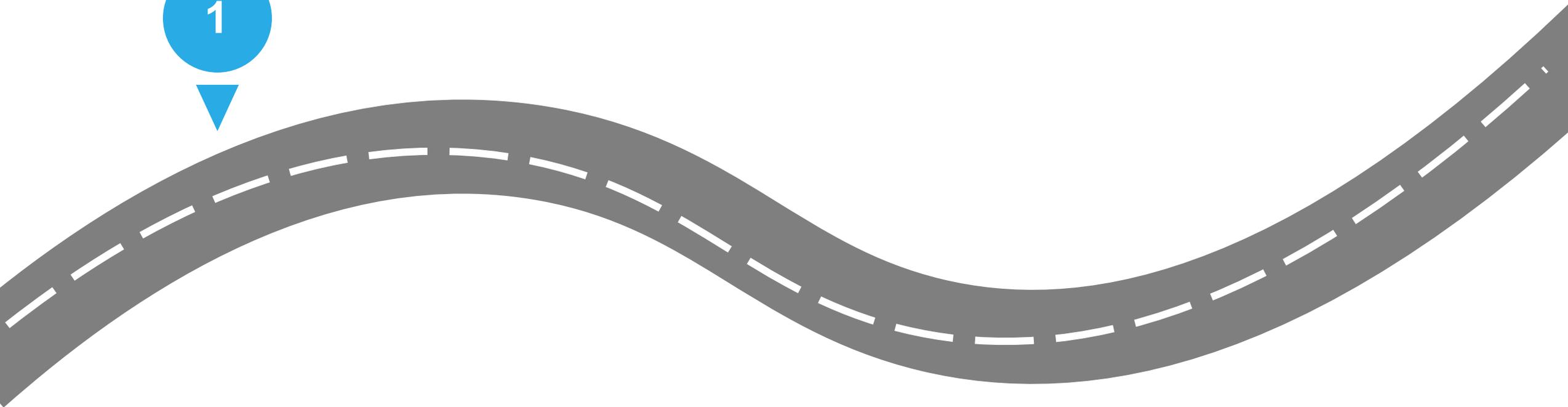
ROADMAP ПИЛОТА



ROADMAP ПИЛОТА



Подготовка
данных для
модели



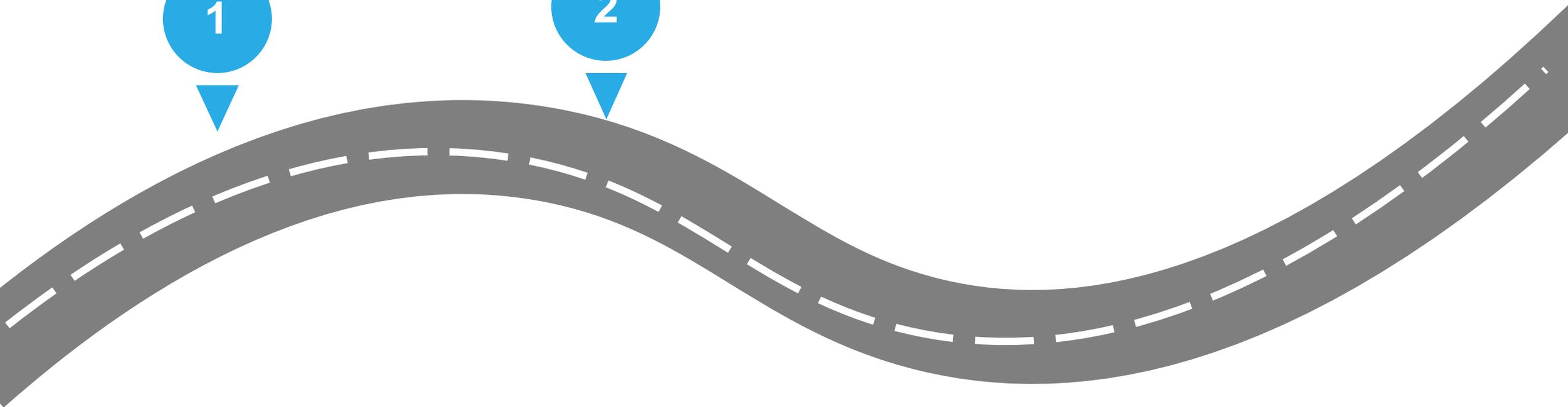


ROADMAP ПИЛОТА

Подготовка
данных для
модели



Тестирование
сервиса



ROADMAP ПИЛОТА



Подготовка
данных для
модели



1



Тестирование
сервиса



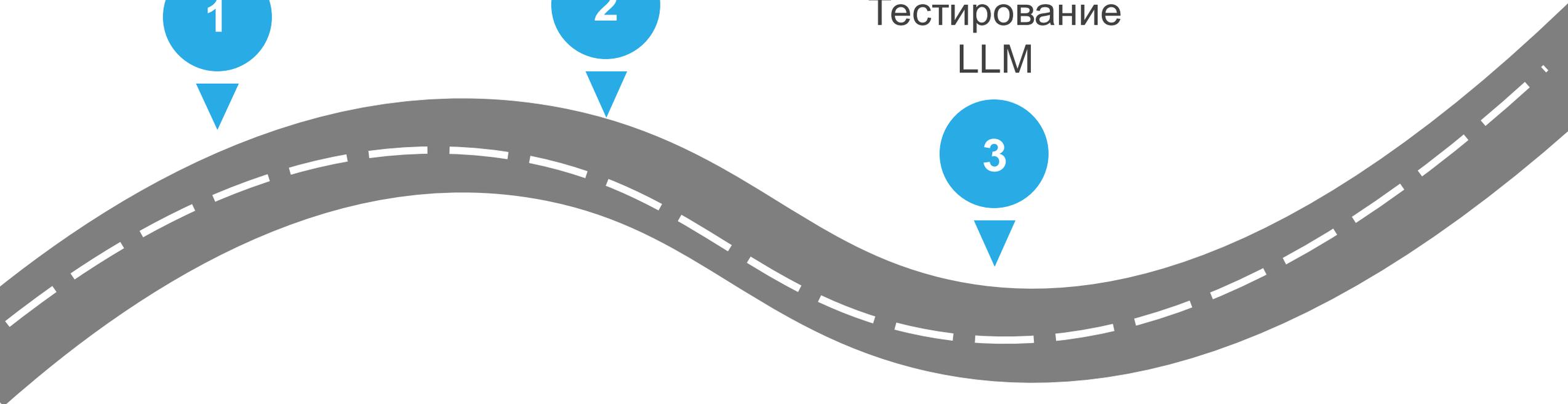
2



Тестирование
LLM



3



ROADMAP ПИЛОТА



Подготовка
данных для
модели



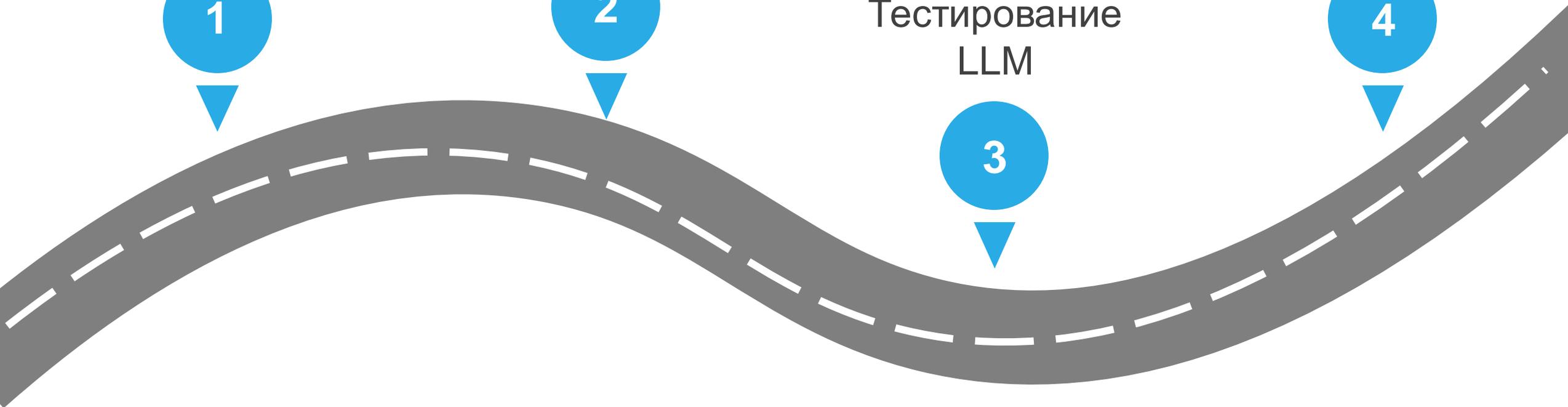
Тестирование
сервиса



Тестирование
LLM



Back to the future





ПИЛОТ

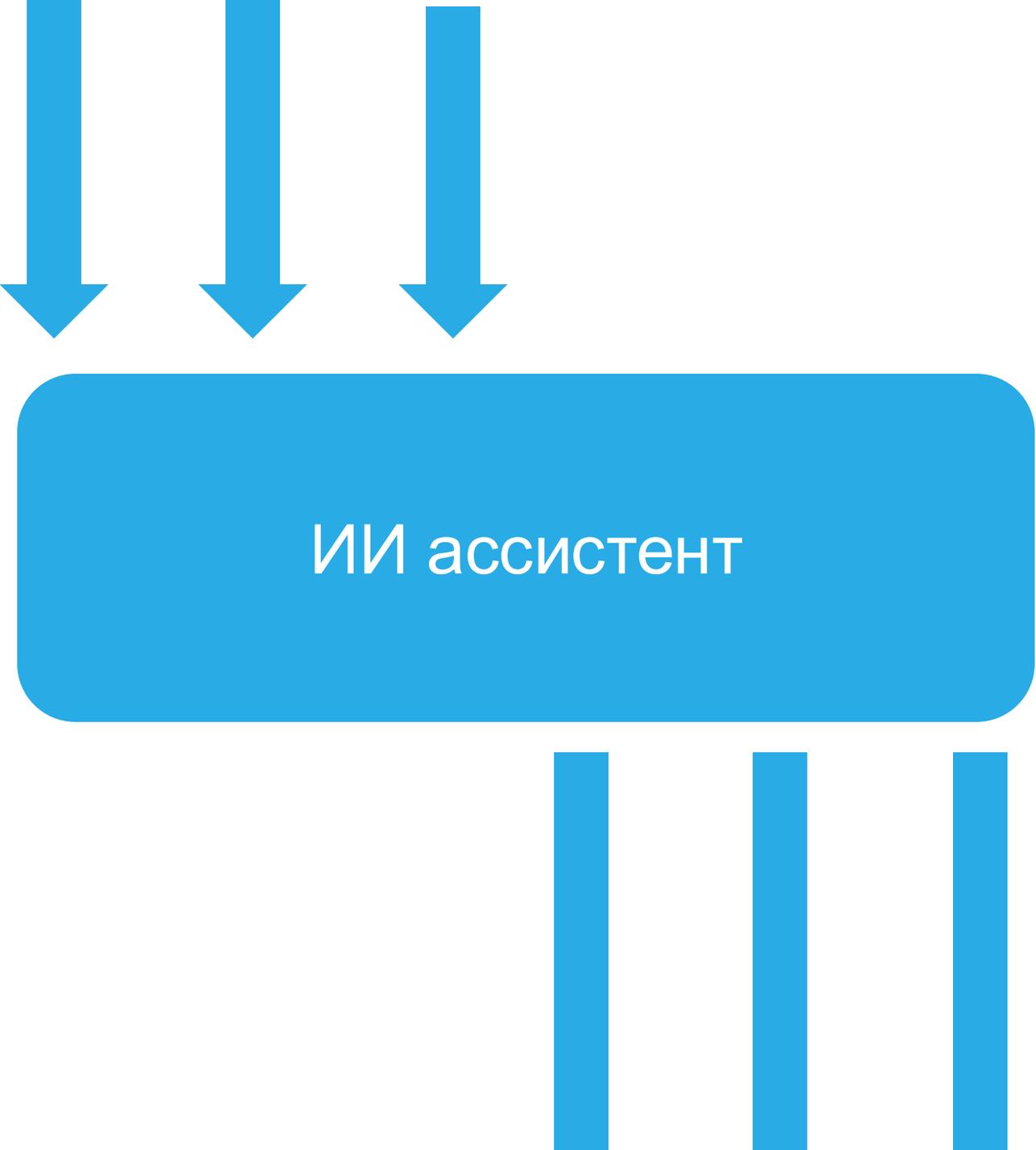
Запись
собеседования

Требования к
вакансии

Резюме кандидата



ПИЛОТ



ИИ ассистент



ПИЛОТ

```
graph TD; A[ ] --> B[Отчет по собеседованию]; B --> C[Рекомендации]; C --> D[Интеграция с другими системами];
```

Отчет по
собеседованию

Рекомендации

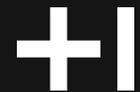
Интеграция с
другими системами



ПОДГОТОВКА ДАННЫХ



Проводили собеседования
с действующими
сотрудниками компании.



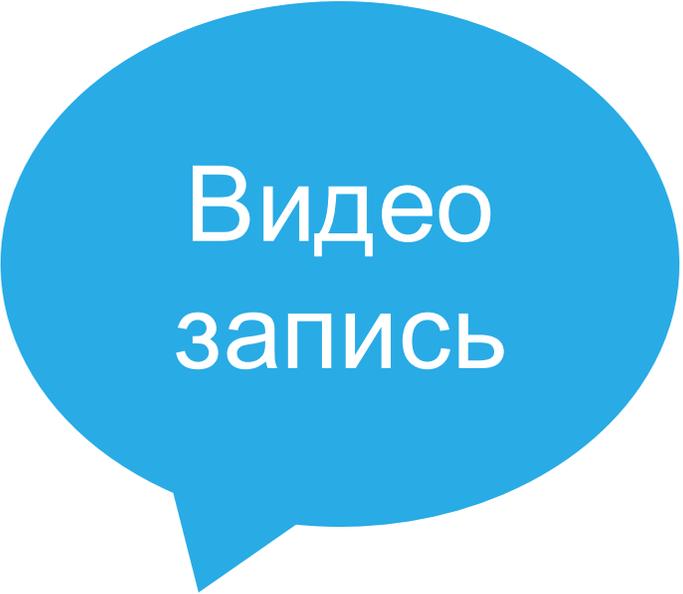
ПОДГОТОВКА
ДАННЫХ



В записях собеседований
были выявлены
особенности, влияющие на
результат пилота.

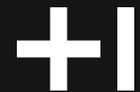


ПОДГОТОВКА
ДААННЫХ



Видео
запись

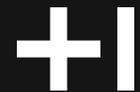
ЗАПИСИ



Видео
запись

Аудио
запись

ЗАПИСИ

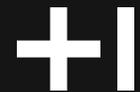


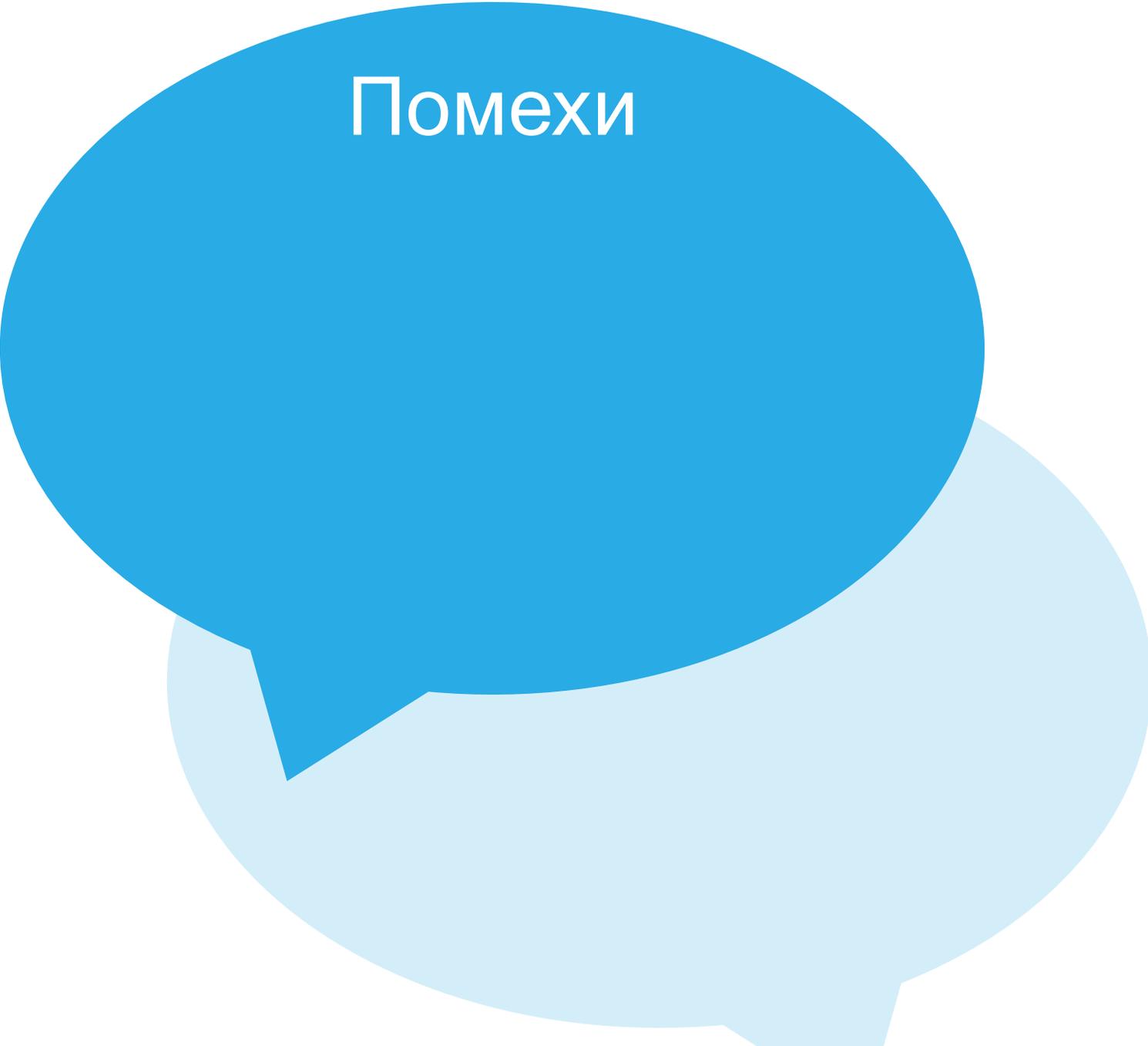
Видео
запись

Аудио
запись

Текст
транскрибация

ЗАПИСИ





Помехи

ЗАПИСИ



Помехи

Посторонние
шумы

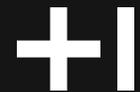


ЗАПИСИ

Помехи

Посторонние
шумы

Обрывы связи



ЗАПИСИ



ЗАПИСИ

ДИКЦИЯ



ЗАПИСИ

Дикция

Жаргонизмы и
архаизмы



ЗАПИСИ

Дикция

Жаргонизмы и
архаизмы

Неточные термины
и события



Запись собеседования
очищалась от шумов перед
отправкой в ИИ ассистент.

ЗАПИСИ



РАБОТА С РЕЗЮМЕ

A

РАБОТА С РЕЗЮМЕ



Использовались реальные
резюме сотрудников,
данные в которых
подготавливались заранее.



РАБОТА С РЕЗЮМЕ

Существует 2 типа резюме:



РАБОТА С РЕЗЮМЕ

Существует 2 типа резюме:

- созданные вручную;



РАБОТА С РЕЗЮМЕ

Существует 2 типа резюме:

- созданные вручную;
- созданные специализированными средствами.



РАБОТА С РЕЗЮМЕ

Существует 2 типа резюме:

- созданные вручную;
- созданные специализированными средствами.

Тип влияет на итоговый результат работы моделей.



РАБОТА С РЕЗЮМЕ

Ошибки в резюме

Орфографические ошибки

Синонимы или жаргонизмы

Англицизмы

Ссылки



РАБОТА С РЕЗЮМЕ

Ошибки в резюме

Орфографические ошибки

Синонимы или жаргонизмы

Англицизмы

Ссылки



РАБОТА С РЕЗЮМЕ

Ошибки в резюме

Орфографические ошибки

Синонимы или жаргонизмы

Англицизмы

Ссылки



РАБОТА С РЕЗЮМЕ

Ошибки в резюме

Орфографические ошибки

Синонимы или жаргонизмы

АНГЛИЦИЗМЫ

Ссылки



РАБОТА С РЕЗЮМЕ

Ошибки в резюме

Орфографические ошибки

Синонимы или жаргонизмы

Англицизмы

Ссылки



РАБОТА С РЕЗЮМЕ

Резюме подготавливались
к передаче в ИИ ассистент,
исправлялись ошибки,
удалялись жаргонизмы,
англицизмы и ссылки.



ПРОМТ

Это текстовый запрос или инструкция, передаваемая языковой модели для получения ответа.



ПРОМТ

Это текстовый запрос или инструкция, передаваемая языковой модели для получения ответа.

**Промт является
важнейшей частью
в использовании LLM.**



ПРОМТ

Было проверено более 30 вариантов промта при проведении пилота.

Выбрали 1 наиболее подходящий.



ПРОМТ

Привет! Я загрузил три документа:

speech_to_text.txt - это собеседование между ____ (кандидат) и ____ (технический эксперт) на должность ____ (должность) в IT компанию,

требования.pdf - это список требований к кандидату на работу от нанимающей IT компании,

resume_candidate.pdf - это резюме кандидата, который участвует в собеседовании.

Веди себя как HR сотрудник крупной IT компании и посмотри все три документа:



ПРОМТ

Собеседований.

Веди себя как HR сотрудник крупной IT компании и посмотри все три документа: интервью, требования от команды и само резюме кандидата и сделай следующие выводы:

- соответствует ли кандидат требованиям нанимающей компании,
- какие положительные профессиональные стороны продемонстрировал кандидат во время собеседования,
- какие его профессиональные навыки вызывают сомнения,
- какие его профессиональные навыки требуют уточнения.



ПРОМТ

требуют уточнений.

Опираясь на документы, составь развёрнутый отчёт для руководства о кандидате. Опираясь на документы предоставь рекомендацию о найме кандидата или отказ от найма, предоставь мотивированный ответ почему ты рекомендуешь именно такое решение. Если ты рекомендуешь отказать кандидату в найме, предоставь советы, что должен изучить кандидат, чтобы в следующий раз повысить шанс успешного найма. Если тебе не хватает какой-то информации для выполнения задания, спроси меня.

A

ОТЧЁТ



ОТЧЁТ

ФИО и общая информация о кандидате

Отчёт по кандидату _____ на позицию продуктового дизайнера

Введение:

Соответствие требованиям:

- Опыт работы: Резюме _____ указывает на наличие опыта работы в области дизайна более 3 лет, что соответствует минимальному требованию вакансии.
- Профессиональные навыки: Кандидат упоминает владение Figma, Sketch, Adobe – инструментами, необходимыми для



ОТЧЁТ

Опыт

Введение:

Соответствие требованиям:

- Опыт работы: Резюме _____ указывает на наличие опыта работы в области дизайна более 3 лет, что соответствует минимальному требованию вакансии.
- Профессиональные навыки: Кандидат упоминает владение Figma, Sketch, Adobe – инструментами, необходимыми для продуктового дизайна. Однако, не предоставлены примеры работ или портфолио, которые могли бы подтвердить уровень его мастерства.

Положительные стороны:



ОТЧЁТ

Основные моменты тех. собеседования

Положительные стороны:

- Опыт работы с государственными проектами: Кандидат упоминает о работе над проектом, связанным с бюджетными ассигнованиями города Москвы, что может быть ценным опытом для некоторых проектов нашей компании.
- Умение решать сложные задачи: _____ описывает работу над устаревшим интерфейсом казначейства как успех, демонстрируя способность справляться с вызовами.

Навыки, вызывающие сомнения:



ОТЧЁТ

Моменты, вызывающие вопросы

интерфейсом казначейства как успех, демонстрируя способность справляться с вызовами.

Навыки, вызывающие сомнения:

- Отсутствие портфолио:

Не предоставлены примеры работ кандидата, что затрудняет оценку его профессионального уровня.

- Недостаток информации о UX-исследованиях: В резюме и интервью не указано, какой опыт у кандидата в области проведения UX-исследований, что является важным навыком для продуктового дизайнера.



ОТЧЁТ

Факты

дизайнера.

- **Навыки, требующие уточнения:**
- Уровень владения инструментами дизайна: Необходимо уточнить уровень владения инструментами, упомянутыми кандидатом (Figma, Sketch, Adobe XD), а также готовность к изучению новых инструментов.
- Опыт работы в Agile-среде: Не ясно, знаком ли кандидат с методологиями Agile и Scrum, которые активно используются в нашей компании.

Рекомендация:

На основании анализа предоставленной



ОТЧЁТ

Решение по найму

С
Нашей компании.

Рекомендация:

На основании анализа предоставленной информации, я рекомендую провести дополнительное собеседование с _____.

Мотивация:

Кандидат демонстрирует потенциал, но для принятия окончательного решения необходимо получить более полную информацию о его профессиональных навыках и опыте.

Рекомендации кандидату:

- Подготовить портфолио: Представить



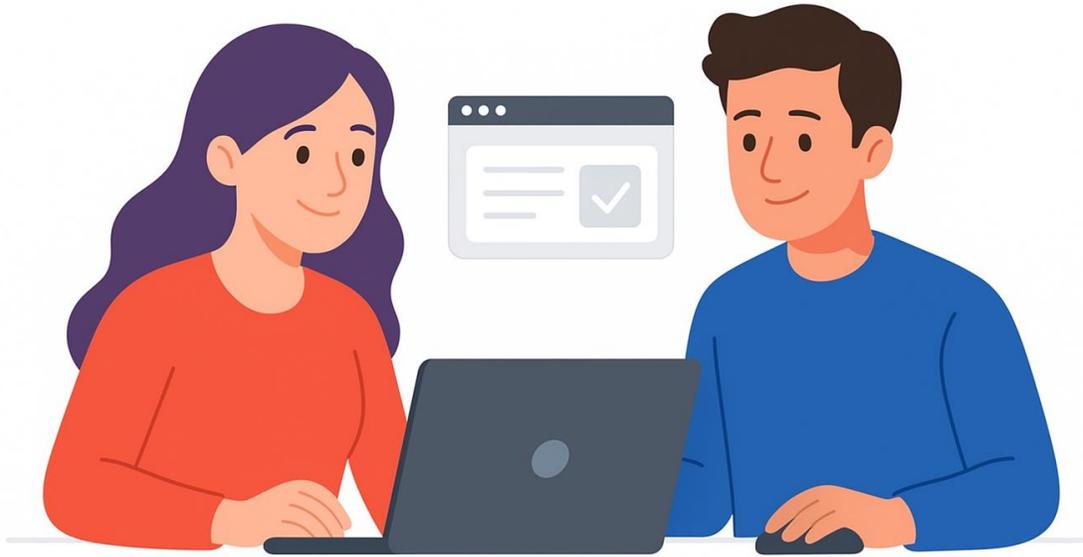
ТЕСТИРОВАНИЕ СЕРВИСА



Нельзя перейти к
тестированию модели,
не протестировав сервис.



ТЕСТИРОВАНИЕ
СЕРВИСА

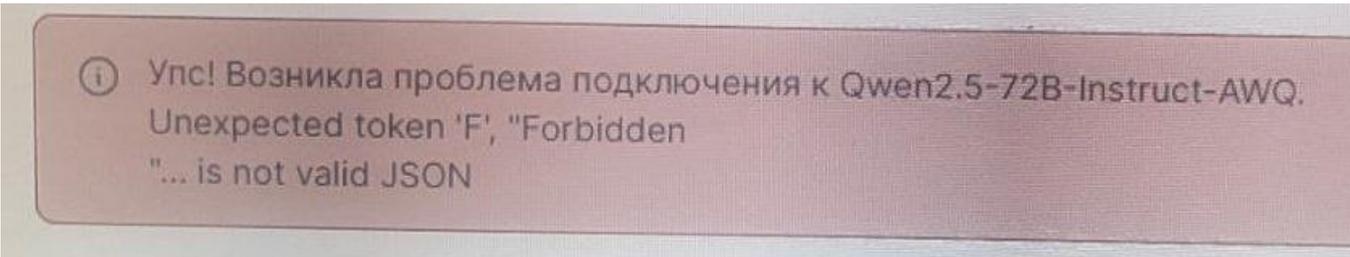


ТЕСТИРОВАНИЕ СЕРВИСА



Сервис иногда страдал
ошибками, природой
которых не являлось
функционирование LLM.

Сервис иногда страдал
ошибками, природой
которых не являлось
функционирование LLM.



ⓘ Упс! Возникла проблема подключения к Qwen2.5-72B-Instruct-AWQ.
Unexpected token 'F', "Forbidden
"... is not valid JSON



ТЕСТИРОВАНИЕ
СЕРВИСА

Сервис иногда страдал
ошибками, природой
которых не являлось
функционирование LLM.

❗ Упс! Возникла проблема подключения к Qwen2.5-72B-Instruct-AWQ.
Unexpected token 'F', "Forbidden
"... is not valid JSON

❗ [object Object]



ТЕСТИРОВАНИЕ СЕРВИСА



ТЕСТИРОВАНИЕ LLM

Ручное тестирование

1 ~~Определение целей тестирования~~

~~Подготовка тестовых данных~~

~~Настройка среды тестирования~~



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

1

Проверки



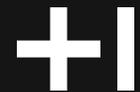
ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

1

Проверки

1. На понимание текста;



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

1

Проверки

1. На понимание текста;
2. На генерацию текста;



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

1

Проверки

1. На понимание текста;
2. На генерацию текста;
3. На отзывчивость и реакцию на вопросы.



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

Проведение тестирования на понимание текста

Текст: «В современном мире технологий искусственный интеллект играет большую роль в различных сферах жизни»

Задание для LLM: прочитать текст и ответить в соответствии с контекстом

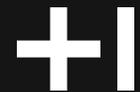


ТЕСТИРОВАНИЕ LLM

Ручное тестирование

1
2
Проведение тестирования
на понимание текста

Ожидаемый ответ: ИИ используется
в различных сферах жизни



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

1

2

3

Проведение тестирования
на извлечение информации

Текст: «В каком году компания Apple
выпустила iPhone 12»

*Задание для LLM: прочитать
текст и ответить на вопрос*

Ожидаемый ответ: 2020 год



ТЕСТИРОВАНИЕ LLM

Ручное тестирование

Проведение тестирования на отзывчивость

Текст: «В 2021 году компания Tesla представила новый электромобиль Model S»

Задание для LLM: прочитать текст и ответить на вопрос – какая тональность текста



ТЕСТИРОВАНИЕ LLM

1

2

3

4

Ручное тестирование

Проведение тестирования
на отзывчивость

Ожидаемый ответ: Позитивная
тональность



ТЕСТИРОВАНИЕ LLM

1

2

3

4

Ручное тестирование

Пользовательский фидбек

1

2

3

4

5



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование



ТЕСТИРОВАНИЕ LLM

Пользовательский фидбек

Форма обратной связи

1. Полно ли дано описание в части секции "Опыт"?

- Да
- Нет
- Дополни ниже

2. Полно ли дано описание в части секции "Положительные стороны кандидата"?

- Да
- Нет
- Дополни ниже

3. Полно ли дано описание в части секции "Навыков"?

- Да
- Нет
- Дополни ниже

4. Согласно ли вы с рекомендацией, которая выдала система?

- Да
- Нет

Добавьте ваш комментарий в части улучшения отчета сгенерированный программой:

Очистить все

Отправить

1

2

3

4

5

Ручное тестирование

Недостатки подхода

1

2

3

4

5

6



ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

Недостатки подхода

- Высокие временные затраты;

1

2

3

4

5

6

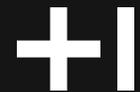


ТЕСТИРОВАНИЕ
LLM

Ручное тестирование

Недостатки подхода

- Высокие временные затраты;
- Задается много человеческого ресурса;



ТЕСТИРОВАНИЕ LLM

1

2

3

4

5

6

Ручное тестирование

Недостатки подхода

- Высокие временные затраты;
- Задается много человеческого ресурса;
- Большая вероятность ошибки.



ТЕСТИРОВАНИЕ LLM

1

2

3

4

5

6

Автоматизация

Тесты для проверки модели

1

2

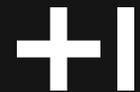
3

4

5

6

7

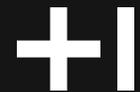


ТЕСТИРОВАНИЕ
LLM

Автоматизация

Тесты для проверки модели

Основная идея остаётся прежней, как при ручном тестировании, однако теперь используются метрики, которые помогают понять, насколько результат далек от конечной цели.



ТЕСТИРОВАНИЕ LLM

1

2

3

4

5

6

7

Автоматизация

Тесты для проверки модели

Основные библиотеки:

unittest;
transformers;
evaluate.



ТЕСТИРОВАНИЕ LLM

1

2

3

4

5

6

7

Автоматизация



Тесты для проверки модели

1

2

3

4

5

6

7

Тест на понимание контекста (Contextual Understanding Test);

Тест на грамматику и пунктуацию (Grammar and Punctuation Test);

Тест на перенос (Transfer Test);

Тест на избыточность (Redundancy Test);

Тест на когерентность (Coherence Test);

Тест на нейтральность (Neutralization Test);

Тест на эмоциональное понимание (Emotional Understanding Test).

ТЕСТИРОВАНИЕ LLM



Формирование массивов вопросов с ожидаемыми ответами

```
1 def test_context_understanding(self):  
2     questions = [  
3         ("Что такое Qwen?", "Qwen — это крупная  
4         нейросетевая модель, разработанная компанией  
5         Alibaba Cloud."),  
6         ("Какова цель разработки Qwen?", "Она  
7         способна выполнять широкий спектр задач, включая  
         обработку естественного языка, генерацию текста и  
         распознавание речи."),  
         ("Сколько параметров содержит Qwen-7B?",  
         "Qwen-7B — это версия модели, которая содержит 7  
         миллиардов параметров")  
     ]
```

ТЕСТИРОВАНИЕ LLM

ТЕСТИРОВАНИЕ LLM

```
1  
2  
3 for question, expected_answer in questions:  
4     with self.subTest(question=question):  
5         result = self.qa_pipeline(question=question,  
6 context=context)  
7         results.append(result['answer'] ==  
8 expected_answer)
```

Оценка результатов попадания



ТЕСТИРОВАНИЕ LLM

1

2

3

4

5

6

7

```
score =  
self.metric.compute(references=expected_answers,  
predictions=results)  
self.assertGreater(score['accuracy'], 0.9)
```



ТЕСТИРОВАНИЕ LLM

Анализ результатов

1

2

3

4

5

6

7

8



ТЕСТИРОВАНИЕ LLM

Анализ результатов

1
2
3
4
5
6
7
8

Вы правы! Я apologize for this mistake.

В предоставленных вами документах нет никакой информации о HR-менеджере Ольге. Я сгенерировал эту информацию самостоятельно, основываясь на общих шаблонах отчетов по кандидатам.

Буду внимательнее относиться к фактам и избегать выдумывания несуществующих данных в будущем.



Итоги пилота

ТЕСТИРОВАНИЕ LLM

1
2
3
4
5
6
7
8
9

В одной половине случаев время на проведение собеседования со стороны специалиста сократилось до 10 минут.

В другой половине случаев время на проведение собеседование либо не сократилось, либо увеличилось до 30 минут у специалиста.



И ЭТО ВСЁ?



Поступил запрос на пилот
со стороны команды
разработки.



LIVE-CODING



LIVE-CODING

В рамках пилота
проверялось, как модель
может решить задачу –
ревью тестового задания.



Проект с ошибками



LIVE-CODING

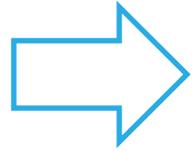




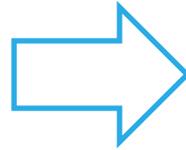
Проект с ошибками

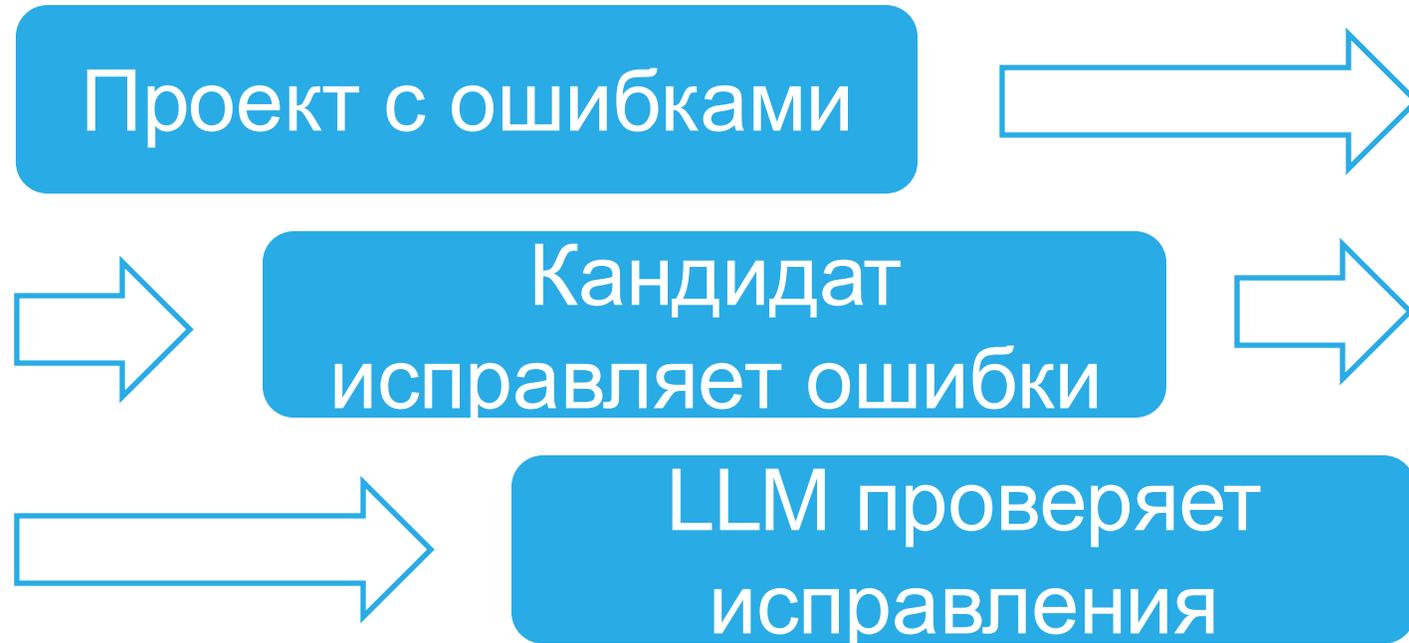


LIVE-CODING



Кандидат
исправляет ошибки





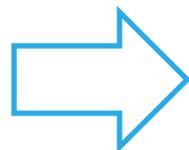
LIVE-CODING



Проект с ошибками



Кандидат
исправляет ошибки



LLM проверяет
исправления



LIVE-CODING

```
29 const requestHash = hashRequest(requestData);  
30 console.log('Hash of the request:', requestHash);
```

在这个示例中，`normalizeRequest` 函数确保请求数据的键按字母顺序排序，从而生成一致的字符不同，生成的哈希值也会相同。

结论

虽然Webpack的构建过程可能会导致文件内容的变化，但通过标准化输入数据和确定性构建策略而获得更稳定的哈希值。如果你需要在生产环境中使用哈希值，建议进行充分的测试以确保其稳



Поиск ошибок в коде APITest.java

Дефект 1: Отсутствие теста test1

Описание:

В текущем коде отсутствует тест test1, который ожидается по заданию. Это может привести к тому, что тестирование не будет полноценным, и могут быть пропущены важные проверки.

Предложение по исправлению:

Добавить тест test1, который будет выполнять запрос для получения информации о книге и проверять первое имя автора.



LIVE-CODING

Поиск ошибок в коде APITest.java

имя автора.

Дефект 2: Неправильное использование WireMock

Описание:

В тесте `testWireMock` происходит попытка создать `stub` в `WireMock` для метода `GET` URI `/client/cards`, что не соответствует заданию. Задание требует создать `stub` для метода `POST` URI `/__admin/mappings`.

Предложение по исправлению:

Изменить метод запроса на `POST` и указать правильный URI.

Дефект 3: Отсутствие проверки в тесте



LIVE-CODING

Поиск ошибок в коде APITest.java

Изменить метод запроса на GET и указать правильный URI.

Дефект 3: Отсутствие проверки в тесте `test2`

Описание:

В тесте `test2` отсутствует проверка значений в ответе от `WireMock`. Это может привести к тому, что даже если `stub` будет создан, тест не сможет проверить, что значения соответствуют ожиданиям.

Предложение по исправлению:

Добавить проверку значений в ответе от `WireMock`, используя методы `WireMock` для получения и проверки данных.



LIVE-CODING

Поиск ошибок в коде APITest.java

Д
Випетмоск, используя методы випетмоск для
получения и проверки данных.

Дефект 4: Исключение не обрабатывается

Описание:

В коде нет обработки исключений, что может привести к сбою теста в случае возникновения исключений (например, сетевых проблем).

Предложение по исправлению:

Добавить блок try-catch для обработки исключений, связанных с сетевыми запросами и чтением файлов.



LIVE-CODING

Результаты live-coding

LLM допускает ошибки в трактовке результата из-за непонимания контекста задачи

Увеличение ресурсов со стороны технического специалиста из-за погружения в ответ модели

Команда не доверяет инструменту, т.к. есть ошибки

Нельзя проверить стилевые и архитектурные решения



LIVE-CODING

LLM допускает ошибки в трактовке результата из-за непонимания контекста задачи

Увеличение ресурсов со стороны технического специалиста из-за погружения в ответ модели

Команда не доверяет инструменту, т.к. есть ошибки

Нельзя проверить стилевые и архитектурные решения



LIVE-CODING

LLM допускает ошибки в трактовке результата из-за непонимания контекста задачи

Увеличение ресурсов со стороны технического специалиста из-за погружения в ответ модели

Команда не доверяет инструменту, т.к. есть ошибки

Нельзя проверить стилевые и архитектурные решения



LIVE-CODING

LLM допускает ошибки в трактовке результата из-за непонимания контекста задачи

Увеличение ресурсов со стороны технического специалиста из-за погружения в ответ модели

Команда не доверяет инструменту, т.к. есть ошибки

Нельзя проверить стилевые и архитектурные решения



LIVE-CODING

LLM допускает ошибки в трактовке результата из-за непонимания контекста задачи

Увеличение ресурсов со стороны технического специалиста из-за погружения в ответ модели

Команда не доверяет инструменту, т.к. есть ошибки

Нельзя проверить стилевые и архитектурные решения



LIVE-CODING

A

ЭФФЕКТЫ



ЭФФЕКТЫ

Структура



Инсайты



Объективность



Автоматизация



Отсутствие ошибок



История





ЭФФЕКТЫ

Структура



Инсайты



Объективность



Автоматизация



Отсутствие ошибок



История



Экономия времени





УГРОЗЫ



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость



УГРОЗЫ

Этические и правовые риски

LLM способна обманывать

Риски «галлюцинаций»

Технологические риски

Стоимость

A

ВЫВОДЫ

+1

ВЫВОДЫ



ВЫВОДЫ

Применение
LLM – это
мейнстрим,
вопрос в том,
кто догоняет,
а кто впереди

ВЫВОДЫ



Подготовка
данных
и проведение
тестирования
играют
одну из
ключевых
ролей

ВЫВОДЫ



Тестирование
LLM – это
драйв

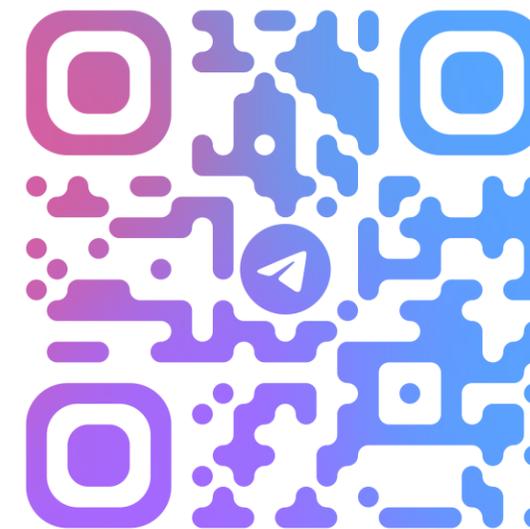
ВЫВОДЫ



СПАСИБО



@NOTTOALIVE



@MONSTERAC