

# Анализ чатов Telegram с помощью LLM

\*AdminHelper\*

# Работали над проектом:



Желобанов Александр, БИСО-02-21, [daipagr](#)[rat](#)

Кашинцева Алина, БИСО-02-21, [itjustme888](#)

Одинцова Юлия, БИСО-02-21, [just-student-19](#)

Чирков Илья, БИСО-02-21, [All9wer](#)



# Содержание презентации

01

## Концепция

*«Что мы создаем?»*  
Расскажем об основной идее нашего продукта, какие проблемы решает.

02

## Техническая часть

*«Как мы это делаем?»*  
Расскажем о всех этапах работы над продуктом, что и как использовали.

03

## Результат

*«Что получили пользователи?»*  
Покажем итоговый продукт, продемонстрируем его полезность.

# Введение

## Проблема администраторов

Ручной мониторинг сотен сообщений ежедневно — трудоемко и неэффективно. Сложности в выявлении токсичных сообщений, спама и конфликтов в режиме реального времени.

Отсутствие аналитики по активности участников: сложно управлять вовлеченностью и выявлять ботов.

## Цель проекта

Создать автоматизированный инструмент для:

Анализа тональности сообщений (негатив/нейтрально/позитив).

Мониторинга активности участников и детекции подозрительного поведения.

# Почему это важно сейчас?



## Рост Telegram-сообществ:

- Крупные чаты (10k+ участников) требуют профессиональной модерации.
- Без автоматизации админы теряют контроль над контентом и аудиторией.

## Угрозы для репутации:

- Негативные сообщения и конфликты быстро эскалируют, если их не устранить.
- Спам и боты снижают качество обсуждений и доверие пользователей.

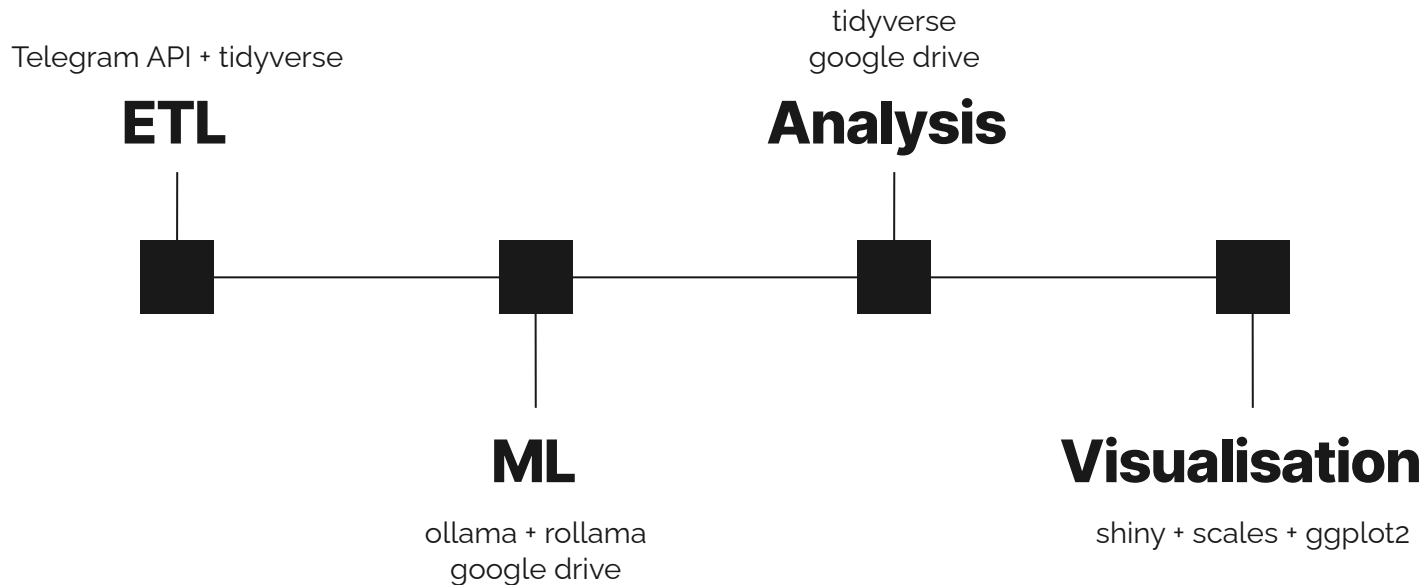
## Тренд на data-driven управление:

Современные админы нуждаются в аналитике, а не в ручном скроллинге чатов.

# Как будет выглядеть продукт для пользователя?

- Приложение для ввода параметров и начала работы
- Shiny-страница с аналитикой сообщений из чата

# Общий пайплайн работы



01

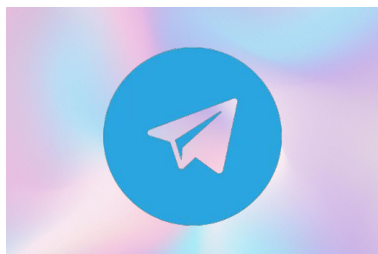
**ETL**

Extract & Transform & Load



# Стек технологий

- Telegram API
- Tidyverse
- Shiny app



# Как проходит работа?



- Создается API-ключ, через который будет происходить подключение
- Создается подключение через пользователя к чату, который он задал в качестве параметра
- Считывается история сообщений чата и список участников
- Данные преобразуются в удобный вид и сохраняются в таблицы
- Таблицы объединяются и фильтруются по типу сообщений
- В дальнейшем данные считываются раз в сутки, таблица с историей обновляется

# Что получаем в итоге?

## Таблица с историей сообщений

Таблица с общей историей сообщений, содержит в себе также данные о пользователе.  
Будет дополняться на этапе ML.

## Таблица со списком участников

Хранит в себе ID, username, ник пользователя и его статус. Будет использоваться для расчета метрик. Обновляется при каждом запуске.

~/xyq - Shiny

http://127.0.0.1:4225

Open in Browser

Publish

## Telegram Chat Parser

Настройки подключения

API ID

\*\*\*\*\*

API Hash

\*\*\*\*\*

Номер телефона

+79854487207

Пароль (если есть)

\*\*\*\*\*

Название чата

363 взвод

Настройки Google Drive

JSON ключ сервисного аккаунта

Browse...

No file selected

ID папки на Drive

1Xdj3qgYVN0ejUS-aPFJyaOdN17dPludj

Настройки парсера

Периодичность запуска (часы)

24

Лог выполнения

[2025-05-24 09:25:11] Настройки загружены из config.ini

#### API Hash

.....

#### Номер телефона

+79854487207

#### Пароль (если есть)

.....

#### Название чата

363 взвод

#### Настройки Google Drive

##### JSON ключ сервисного аккаунта

Browse...

No file selected

#### ID папки на Drive

1Xdj3qgYVN0ejUS-aPFJyaOdN17dPludj

#### Настройки парсера

##### Периодичность запуска (часы)

24

Сохранить настройки

Запуск

Остановить

| ID     | Дата            | id_Автора | Автор      | Username     | Тип   | Текст               |
|--------|-----------------|-----------|------------|--------------|-------|---------------------|
| 531341 | 2025-05-23 8:12 | 501047371 | Ilya       | @all9wer     | Текст | хм, хорошо щд напи  |
| 531338 | 2025-05-23 8:11 | 790062753 | Vare Папян | @papyan_vahe | Текст | В пятницу к 9 могу. |
| 531335 | 2025-05-23 8:10 | 501047371 | Ilya       | @all9wer     | Текст | я прост одним сооб  |
| 531333 | 2025-05-23 8:10 | 501047371 | Ilya       | @all9wer     | Текст | а мб не в среду мож |
| 531332 | 2025-05-23 8:09 | 790062753 | Vare Папян | @papyan_vahe | Текст | К часам 11-12 могу  |

| ID         | Имя              | Username      | Телефон     | Статус  |
|------------|------------------|---------------|-------------|---------|
| 1960910170 | Артем Десятников | @desart8104   | 79859140500 | Обычный |
| 5064521217 | Григорий Верещак | @Vgrom7       | 79260705719 | Обычный |
| 790062753  | Vare Папян       | @papyan_vahe  | 79205807888 | Обычный |
| 2022891075 | Борис Воловик    | @BorisSailman | 79151118661 | Обычный |
| 1404254364 | Паша Попов       | @St_Hati      |             | Обычный |

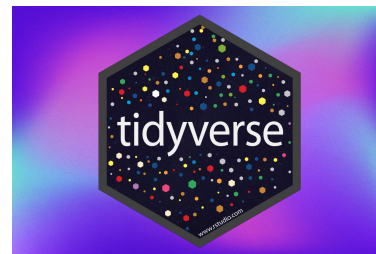
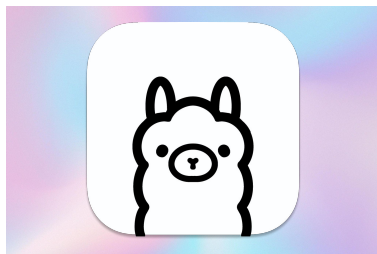
02

**ML**

Работа с LLM

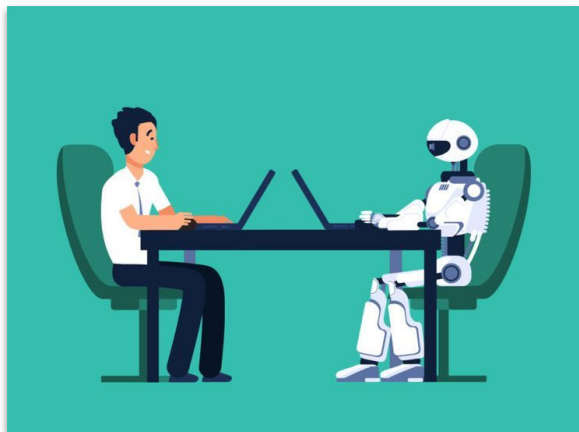
# Стек технологий

- Ollama + OxW/Saiga\_YandexGPT\_8B:q4\_K\_M
- Rollama
- Tidyverse
- Google Drive





# Как происходит работа?



- С помощью ollama локально устанавливается нужная модель.
- С помощью библиотеки rollama задаются параметры работы с моделью и промт
- Отправляется запрос к модели, получается ответ от нее в виде текста (на конкретное сообщение)
- Текст фильтруется и проверяется на вхождение нужной метки
- Метка записывается в таблицу
- Общая таблица отправляется на диск

# Что получаем в итоге?

## Обновление общего файла

Данные добавляются к уже существующей таблице, хранящей предыдущие обработанные данные.

## Начальный файл

Создается таблица с данными, включающими кроме истории сообщений также тональность и тему для каждого сообщения. Csv-файл загружается на Google диск.

| message_id | timestamp           | author_username | author_name  | text                | member_status | ID        | dialog_id | topic              | sentiment |
|------------|---------------------|-----------------|--------------|---------------------|---------------|-----------|-----------|--------------------|-----------|
| 2          | 2025-02-27 11:13:00 | @zaytsev1337    | илюха зайцев | Готово              | Обычный       | 365013927 | 0         | повседневность и к | neutral   |
| 5          | 2025-02-28 8:18:00  | @zaytsev1337    | илюха зайцев | Хочу сесть презу п  | Обычный       | 365013927 | 1         | другое             | negative  |
| 4          | 2025-02-28 8:18:00  | @zaytsev1337    | илюха зайцев | Есть какие-нибудь   | Обычный       | 365013927 | 1         | образование и карь | neutral   |
| 3          | 2025-02-28 8:18:00  | @zaytsev1337    | илюха зайцев | Че, народ, как успе | Обычный       | 365013927 | 1         | повседневность и к | neutral   |
| 8          | 2025-02-28 8:22:00  | @all9wer        | Илья         | код для бота почти  | Премиум       | 501047371 | 1         | другое             | neutral   |
| 7          | 2025-02-28 8:22:00  | @all9wer        | Илья         | это саня хотел      | Премиум       | 501047371 | 1         | повседневность и к | neutral   |
| 6          | 2025-02-28 8:22:00  | @all9wer        | Илья         | надо узнать про се  | Премиум       | 501047371 | 1         | технологии и соцсе | neutral   |

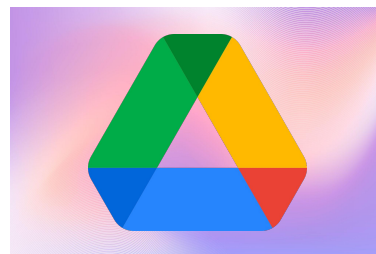
**03**

# **Analysis**

Работа с полученными данными

# Стек технологий

- Tidyverse
- Google Drive



# Как проходит работа?



Считывается обработанная таблица из предыдущего шага и таблица со списком пользователей.

На основе этих данных рассчитываются метрики, такие как:

- Средняя длина сообщений
- Статистика сообщений по периодам
- Количество активных пользователей
- Статистика активности

И другие.

# Что получаем в итоге?

## Обработанные данные

Теперь они готовы к использованию в различных графиках и для вывода на дашборд.

04

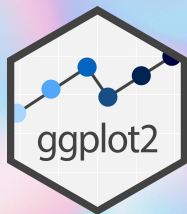
# Visualisation

Составление дашборда

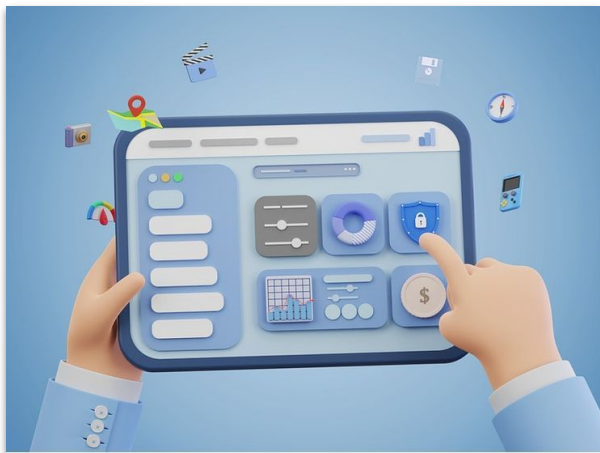


# Стек технологий

- Ggplot/scales
- Shiny



# Как проходит работа?



На основе обработанных данных с предыдущего шага

- Создаются функции для вызовов графиков
- Задается общая канва веб-страницы, куда размещаются графики
- Задается серверная часть
- Собирается приложение

# Что получаем в итоге?

## Веб-страница

На ней располагаются метрики, таблицы и графики. Страница кликабельная и дает возможность точно посмотреть данные конкретной точки.



**Демонстрация решения**

## Анализ сообщений

Участников в чате

5

Активных участников

5 (100%)

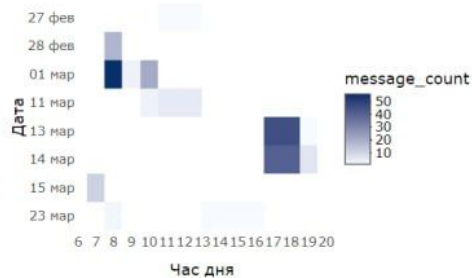
Всего сообщений

215

Чат существует дней

86

### Активность по часам



### Статистика участников

| Пользователь | Количество сообщений | Средняя длина      |
|--------------|----------------------|--------------------|
| @all9wer     | 48                   | 28.166666666666667 |
| @bedyaz      | 14                   | 54.14285714285715  |
| @diepaigrat  | 11                   | 24.18181818181818  |
| @lila_s_19   | 28                   | 41.25              |
| @zaytsev1337 | 114                  | 34.41228070175438  |

### Основные темы

| Тема                      | Количество сообщений |
|---------------------------|----------------------|
| другое                    | 76                   |
| образование и карьера     | 60                   |
| повседневность и юмор     | 27                   |
| технологии и соцсети      | 24                   |
| отношения                 | 19                   |
| ментальное здоровье       | 7                    |
| путешествия и приключения | 1                    |

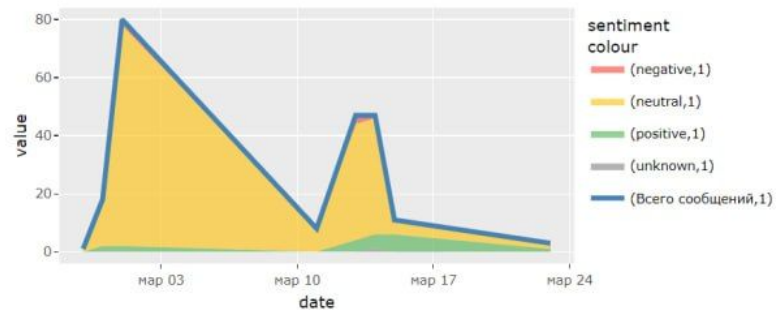
### Тональность по датам



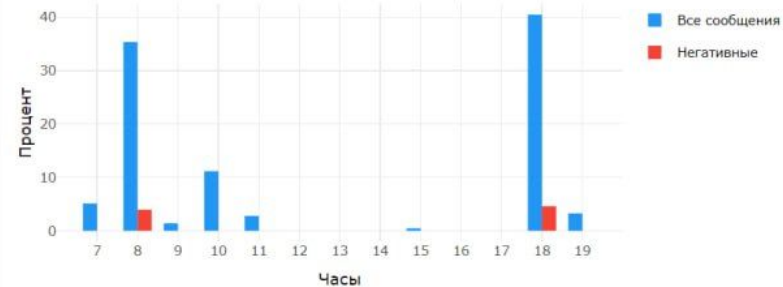
### Тональность по времени



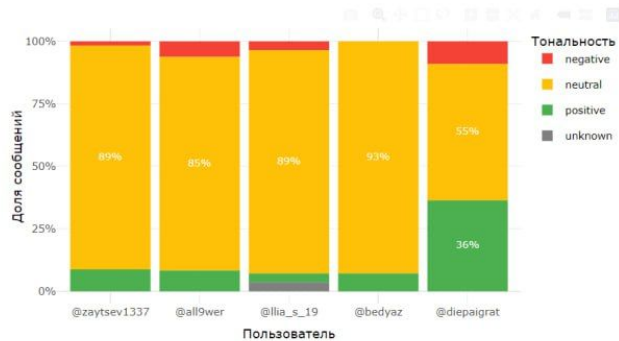
## Тональность по датам



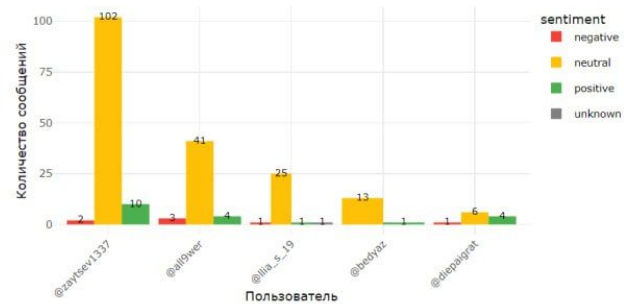
## Тональность по времени



## Распределение тональности по пользователям



## Количество сообщений по тональности





# Практическая польза

## Кейсы использования:

- Модерация: Автоматическое выявление токсичных сообщений.
- Анализ аудитории: Понимание активности участников для улучшения вовлеченности.
- Отчеты: Генерация статистики для владельцев чатов.

## Преимущества:

- Сокращение времени на ручную проверку на 60-70%.
- Проактивное управление конфликтами.

Вызовы, с которыми мы столкнулись в  
процессе работы

и какие нашли решения



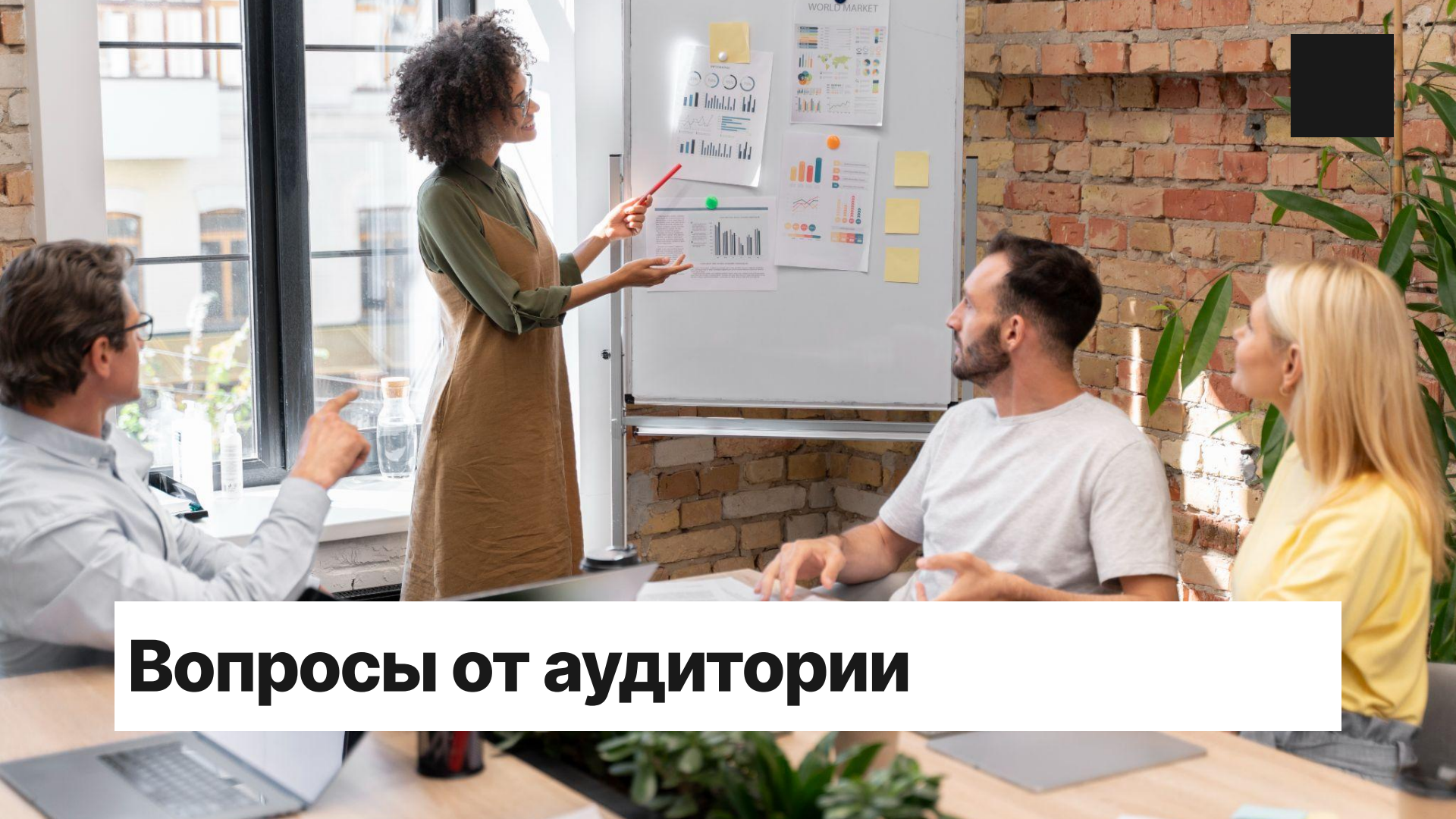


# Планы развития



## Что будет дальше?

- Анализ сообщений в режиме реального времени.
- Интеграция автоматических действий: бан по правилам, шаблонные ответы.
- Расширение анализа: детекция изображений/стикеров с запрещенным контентом.
- Упрощение onboarding: шаблоны дашбордов под разные типы сообществ.



**Вопросы от аудитории**

# Спасибо за внимание!



Проект размещается по адресу:  
<https://github.com/daipaigrat/ZHOKACH>