

Рекомендательная система на основе ИИ

# Погнали, чувак!



**Хакатон МФТИ "Науки о данных"**  
**(DS2023)**

Раздел: «Умный городской гид»

27.12.2023

**Команда № 1 Data Sorcerers**

Роман Вяткин - капитан  
Даврон Ихматуллаев  
Дмитрий Косачев  
Ярослав Баймлер  
Вероника Заславская

# Постановка задачи и решение

**Задача дататона:** улучшить опыт от посещения новых мест с помощью ИТ.

**Гипотеза:** удовлетворенность посетителя зависит от соответствия места его ожиданиям!

**Решение:** удобный, релевантный подбор места и мероприятия с помощью **NLP** (векторизация, семантический анализ, частеречная разметка)



# ТГ-бот и рекомендательная система

Наш клиент: стартап Listing.events, входит в Топ 5 органической выдачи Яндекса по ключевым фразам



Санкт-Петербург

Сегодня в клубах Санкт-Петербурга - афиша вечеринок и тусовок

ВС 31 ДЕК

**NEW YEAR** SandboxMusicRoom

SandBoxMusicRoom

ВС 31 ДЕК

**НОВЫЙ ГОД** Base

Base

ВТ 2 ЯНВ

**DISTRICT** Ритмы

СБ 20 ЯНВ

**DARKNESS RISES** Factory 3



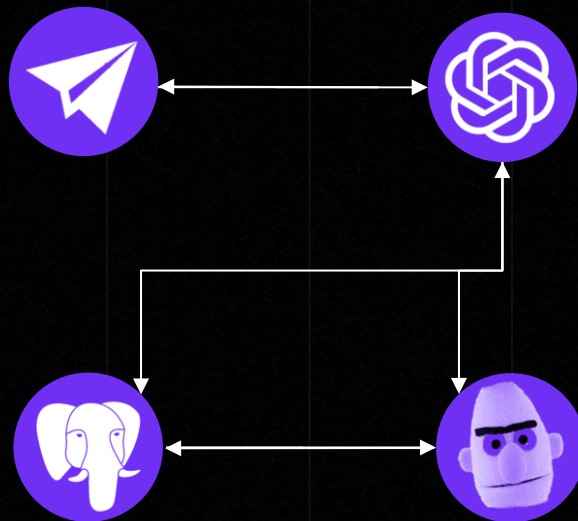


- Описание бизнес задачи и выбор инструментов DS для ее решения
- Анализ накопленных проектом listing.events данных
- Разработка рекомендательной системы и выбор оптимальный предобученных нейросетей
- Разработка чат-бота и подключения к нему GPT 3.5
- Тестирование и подсчет метрик

# Как это работает?

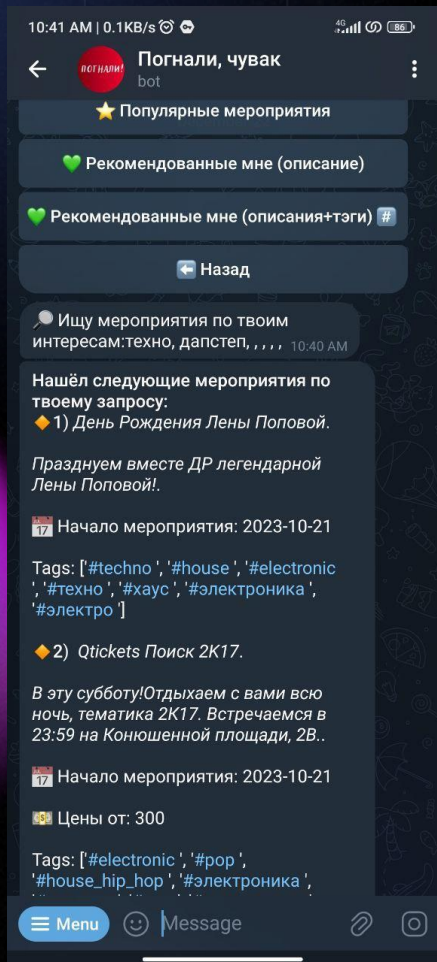
Пользовательский  
интерфейс через  
телеграмм бот

Postgresq хранит данные  
пользователей и  
мероприятия



Обработка  
пользовательских  
запросов на естественном  
языке через GPT 3.5

Модель all-MiniLM-L6-v2  
токенизирует запросы  
пользователей и  
параметры объектов



# Примененные технологии

- **Telegram API:** асинхронный фреймворк AiogramOpenAI API
- **Sentence-Transformers:** sentence-transformers/all-MiniLM-L6-v2 с Hugging Face.
- **NLP:** BERT (векторизация, семантический анализ, частеречная разметка)
- **ChatGPT Turbo 1.106**



# Итоговые метрики



**Для выбора наилучшей модели, а также объективной оценки качества были выбраны метрики:**

**Personalization** – 0,437

**Intra-list Similarity** - 0.849

Попробуйте сами! @pognali\_gpt\_bot

Репозиторий: [https://github.com/fitlemon/pognali\\_gpt\\_bot](https://github.com/fitlemon/pognali_gpt_bot)

### Планы по дальнейшей разработке:

- Коллаборативная рекомендательная система;
- Проведение кластерного анализа объектов;
- Доработка контентной рекомендательной системы с учетом кластеризации;
- Разработка API для обмена данными;
- Проведение А/Б тестов;
- Интеграция с билетными сервисами.

