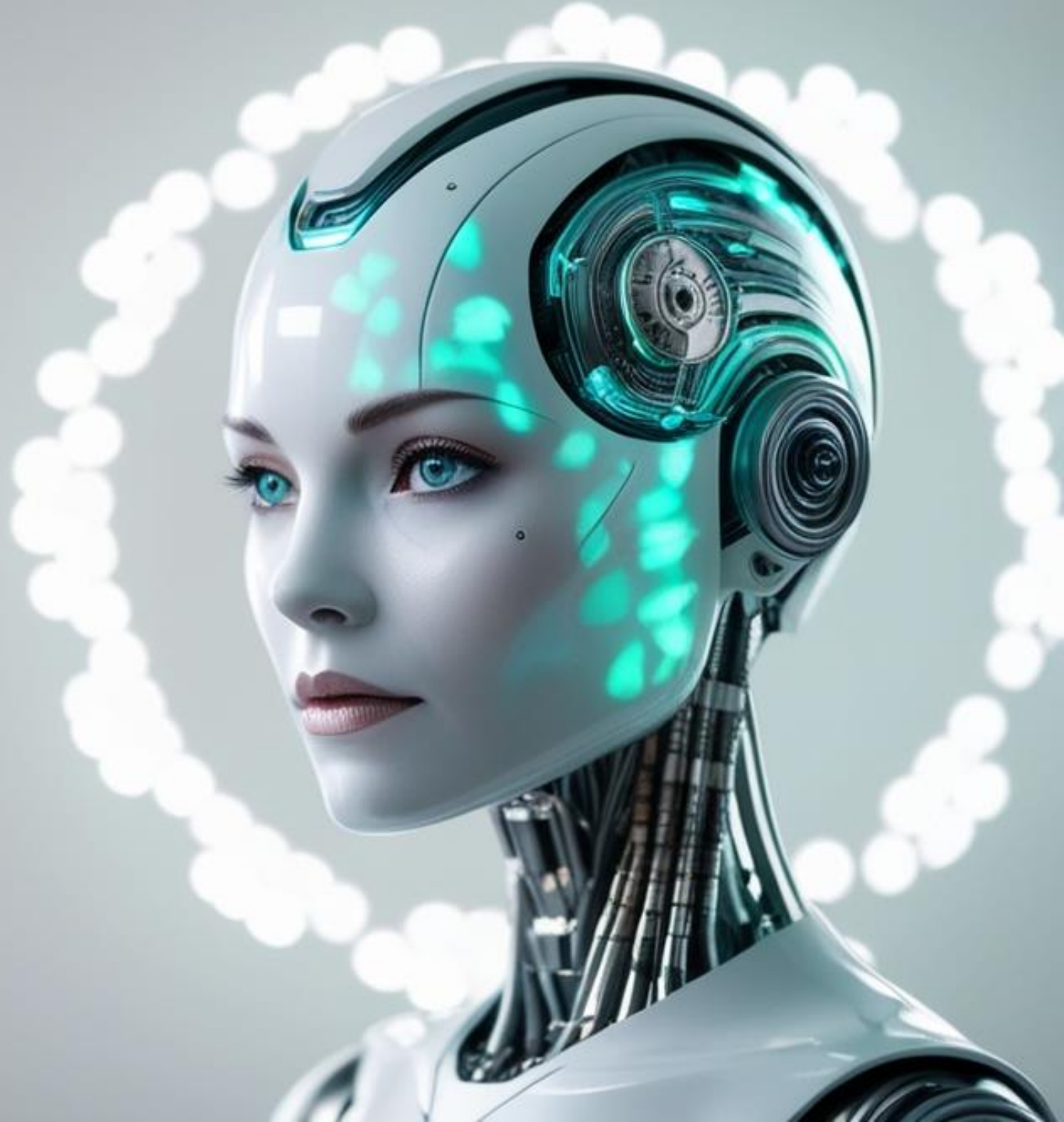
# AXIOM

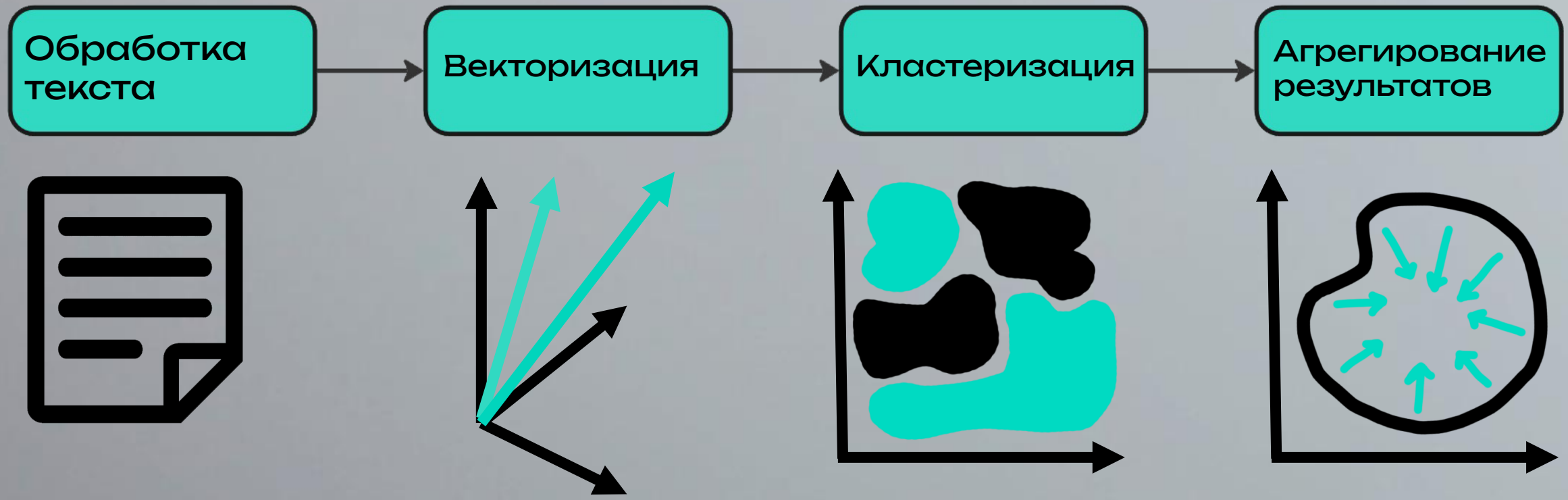
Nuclear IT Hack 2024



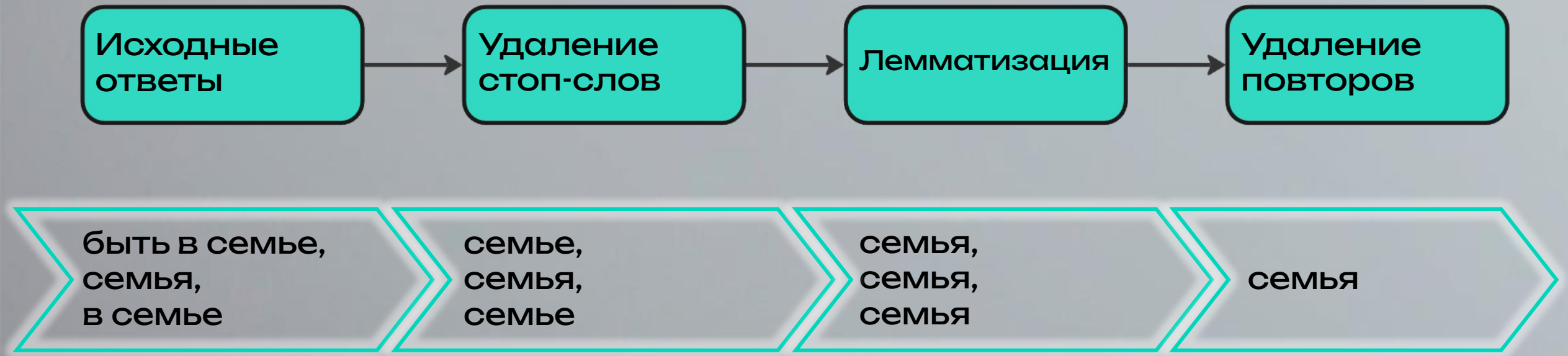
# Задача


Работа, семья, деньги, друзья, путешествия
...
...
...
Хорошие друзья, успех, карьера, эмоции, путешествия
свобода
друг
здоровье
эмоция
покой
амбиции
удовольствие
развитие
деньги
путешествие
признание
доверие
семья

# Решение


Обработка текста
Векторизация
Кластеризация
Агрегирование результатов

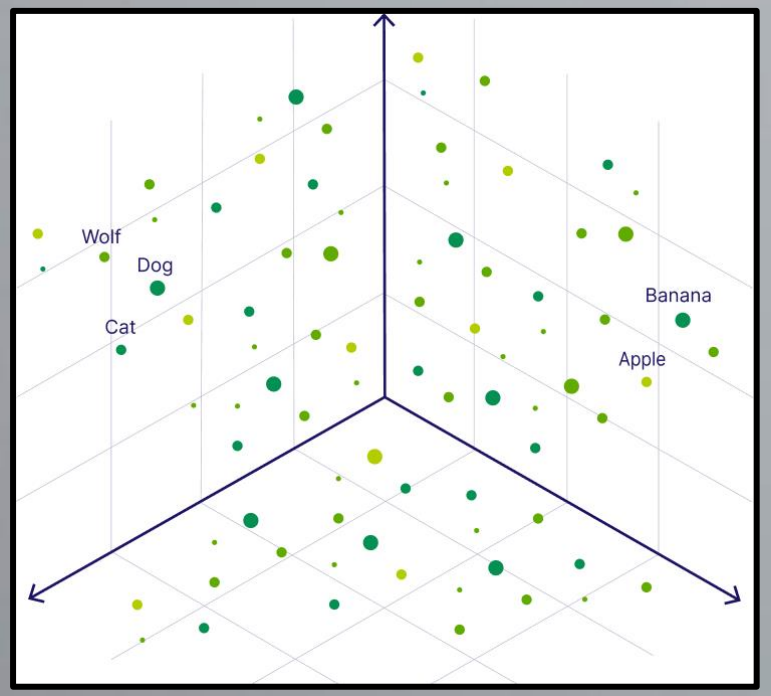
# Обработка текста

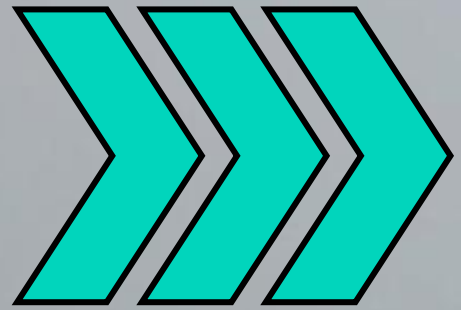


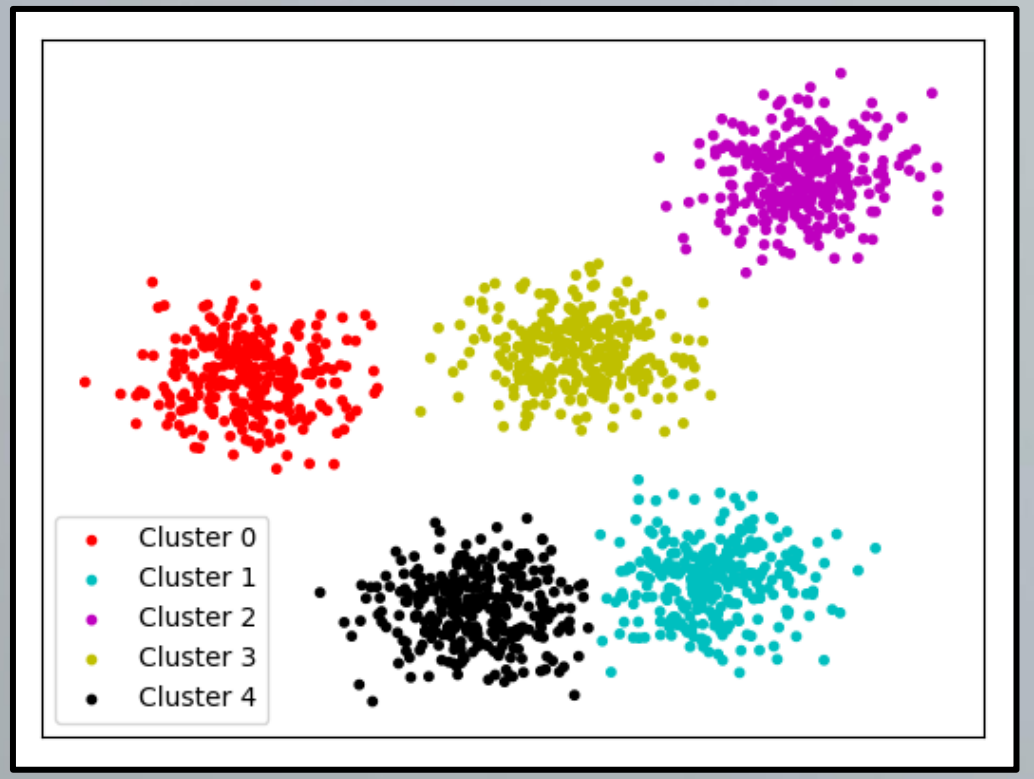
# Удаление синонимов

Векторизация текста:

- Word2vec
- Bert-like



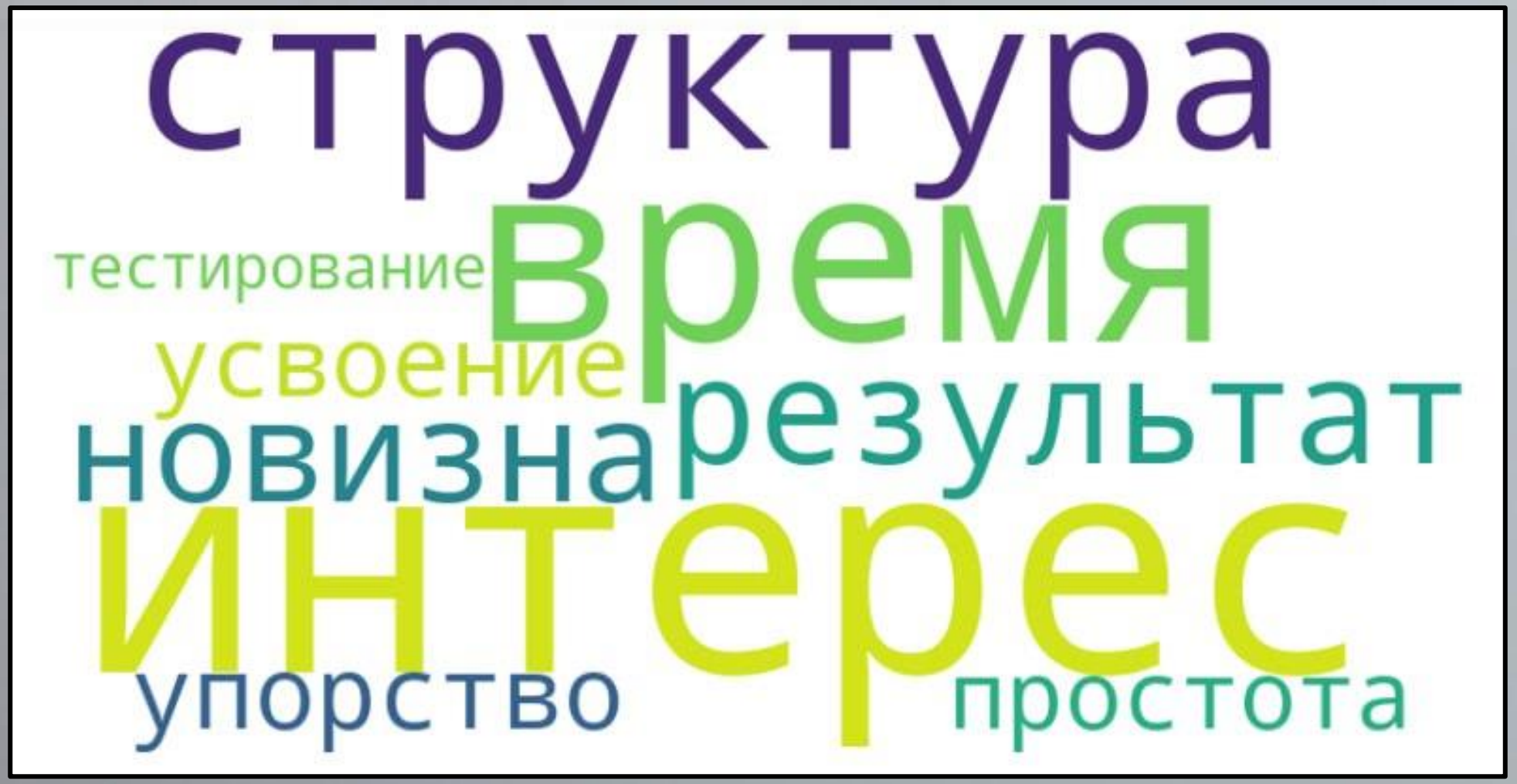





# Датасеты

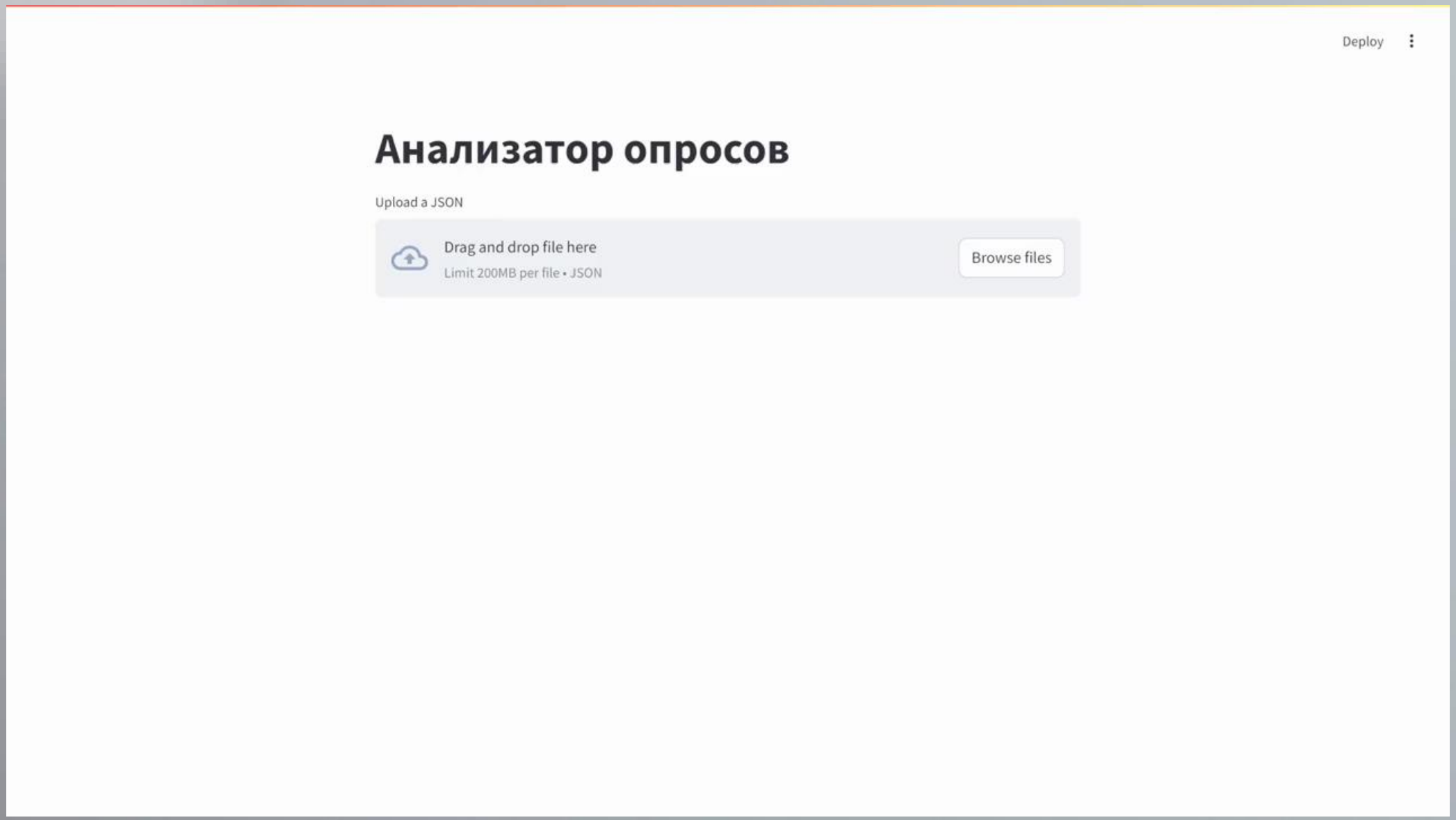
| Название | Кол-во ответов | Просторечия | Нецензурная лексика |
|---|---|---|---|
| Small | 100 | Нет | Нет |
| Medium | 500 | Да | Нет |
| Big | 2500 | Да | Нет |
| Large | 5000 | Да | Да |

# Примеры 1 - Medium



# Примеры 2 - Large

# Демо

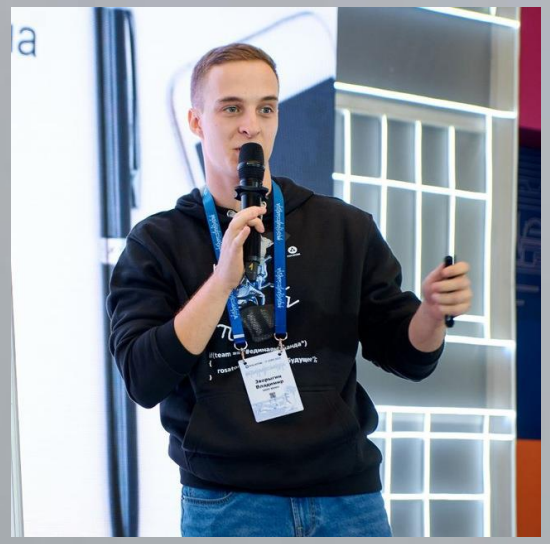

Deploy
Анализатор опросов
Upload a JSON
Drag and drop file here
Limit 200MB per file • JSON
Browse files

# Производительность

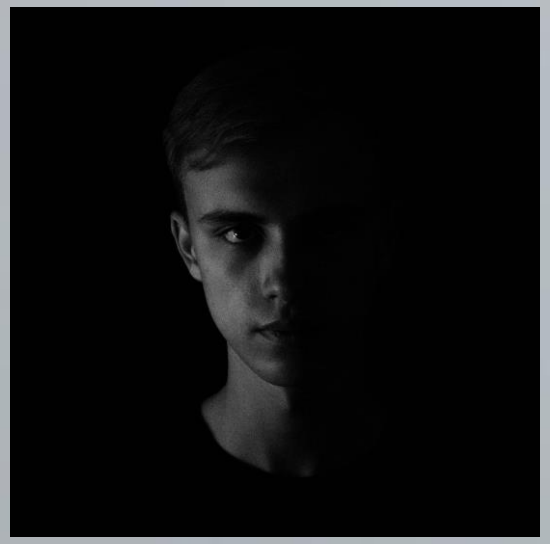
| | | Время работы (в сек) | | |
|---|---|---|---|---|
| Датасет | Кол-во ответов | Word2vec | BERT embeddings on *CPU* | BERT embeddings on *GPU* |
| Small | 100 | 1.7 | 15.7 | 5.7 |
| Medium | 500 | 1.8 | 70.6 | 14.7 |
| Big | 2500 | 3.9 | 371 | 60.6 |
| Large | 5000 | 6.9 | 723 | 119.6 |

# Команда

Андрей Миронов



Зворыгин Владимир

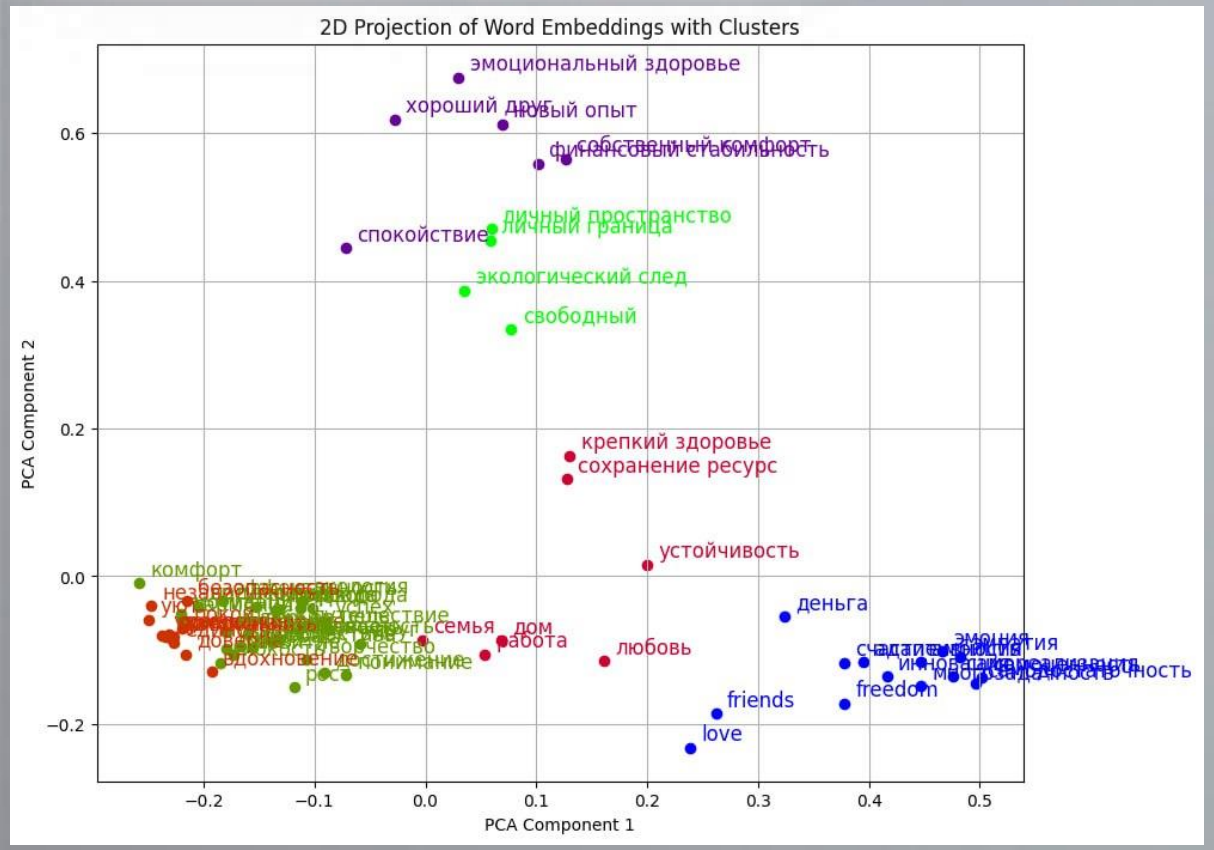


Диков Александр

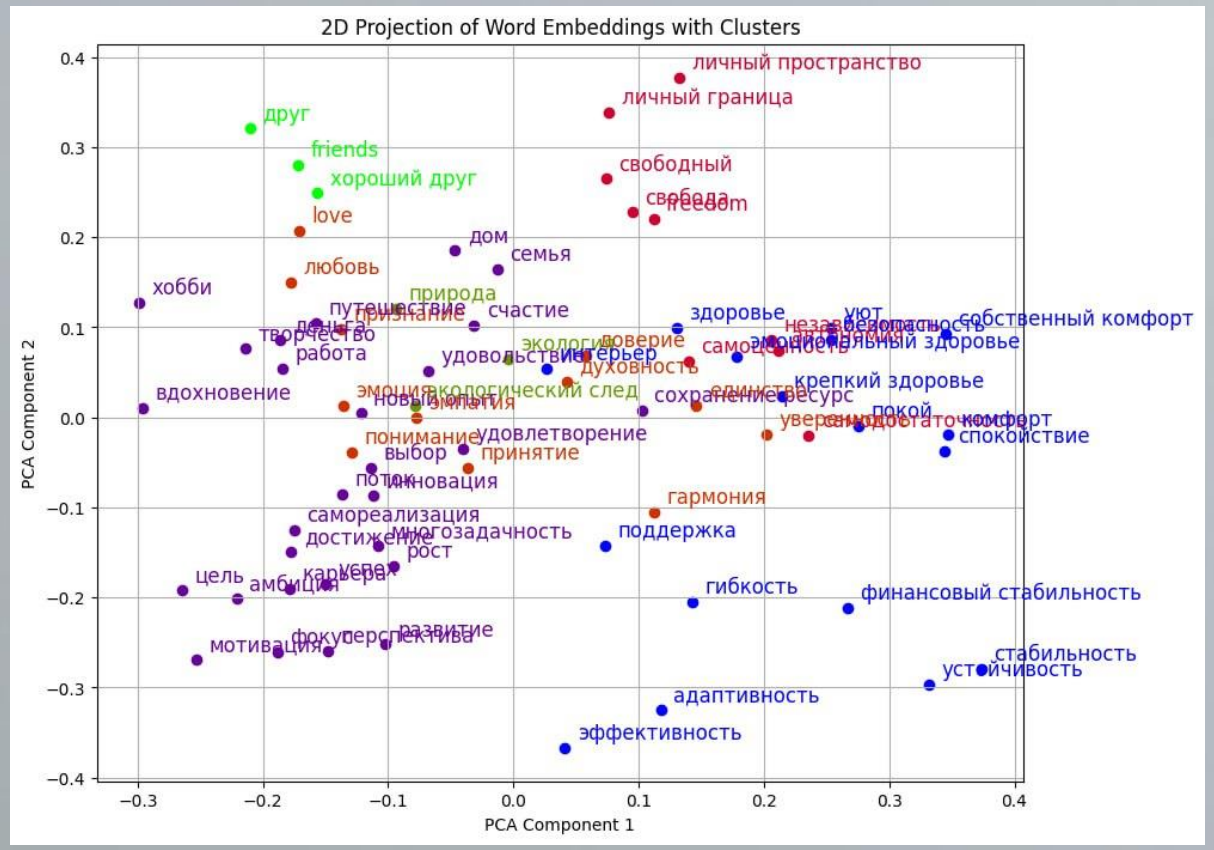
Алексей Садохин

# Приложения

## Датасет - small



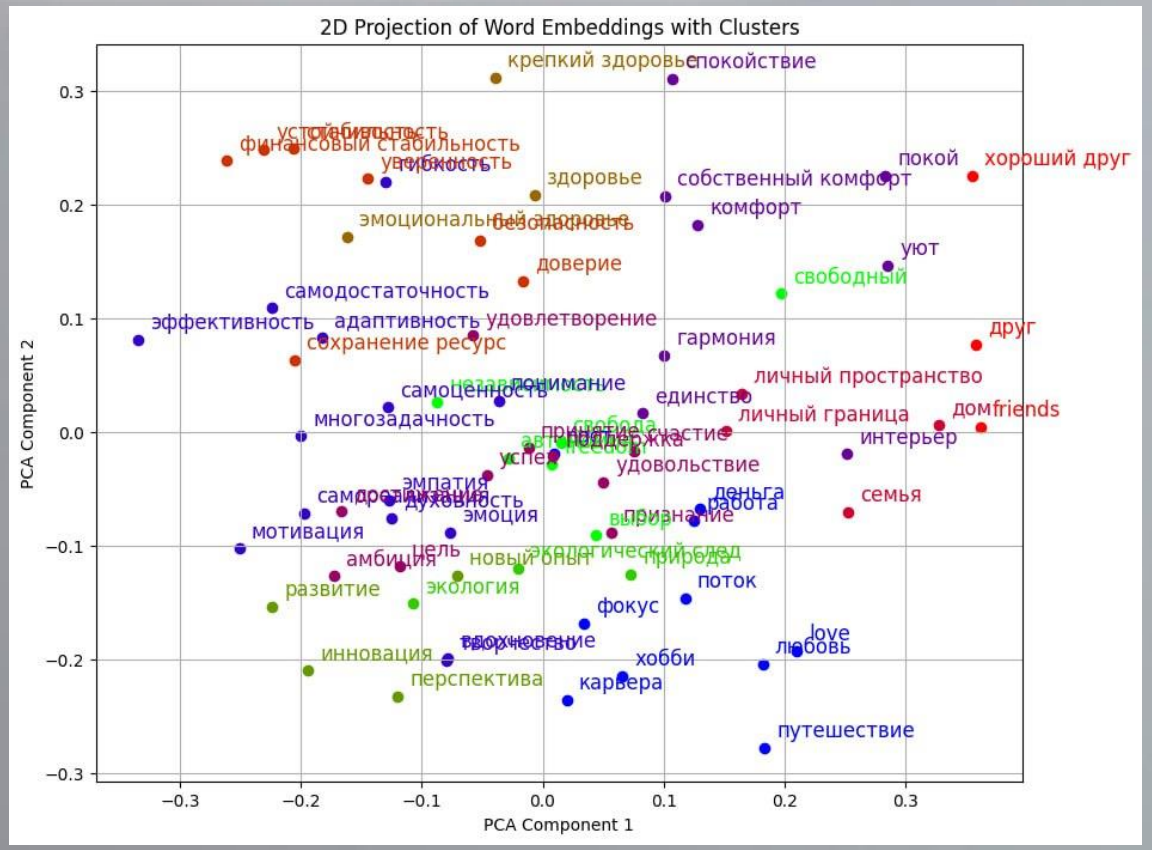

ruBert-large



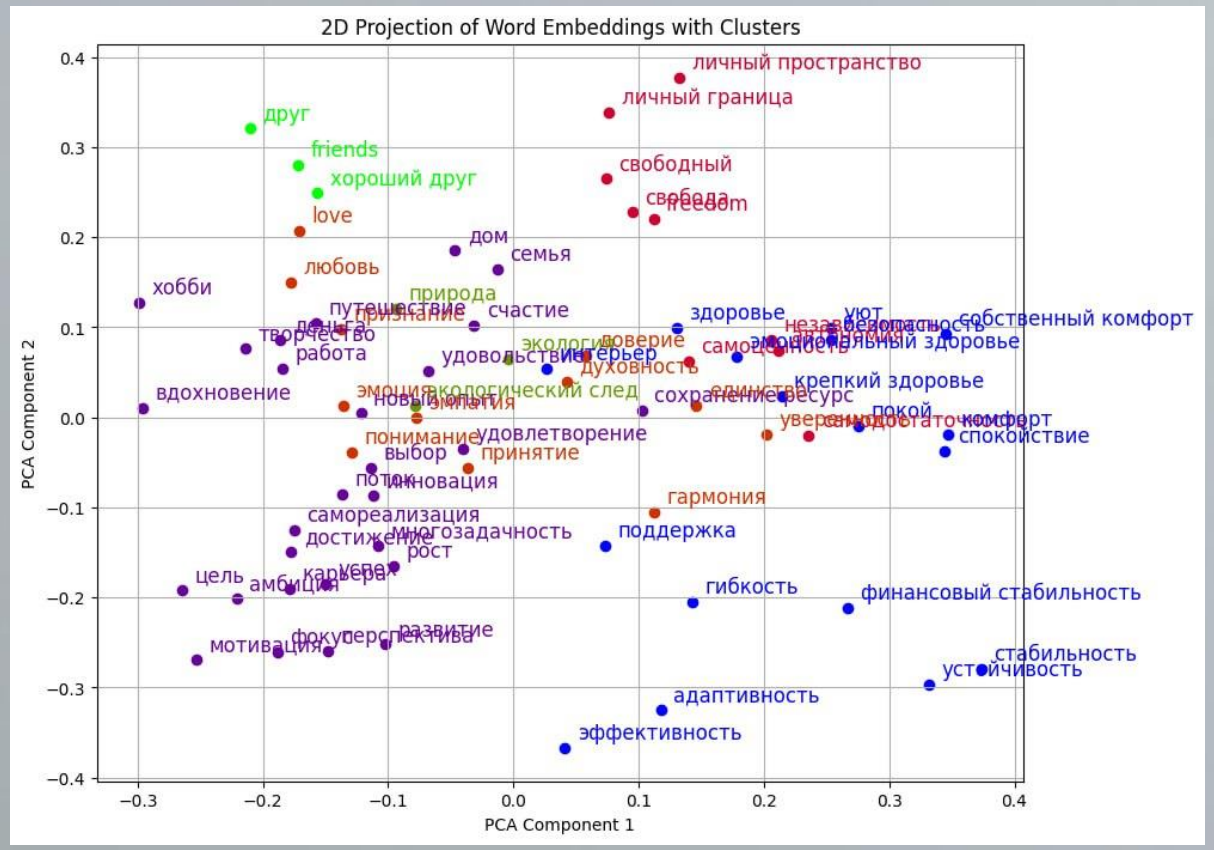

ru-en-RoSBERTa

# Приложения

## Датасет - small




ruElectra-large




ru-en-RoSBERTa