

Основные требования в requirements.txt:

```
pip install -r requirements.txt
```

### - Данные и модели

Данные: используем датасет RuBQ 2.0 (в данном случае коллекции на основе RuBQ\_2.0\_dev.json, RuBQ\_2.0\_paragraphs.json) для вопросов-ответов или для параграфов. Файлы лежат в ./data/rubq/.

Модели эмбедингов:

distiluse-base-multilingual-cased-v2 — относительно лёгкая мультиязычная модель, поддерживающая русский, даёт неплохую точность.

cointegrated/rubert-tiny2 — ещё более компактная русскоязычная модель. Мы сравнили их и выбрали по критерию точности/быстроты

Локальная LLM:

У нас работает Llama (llama3.1) с OpenAI-подобным API.

В test\_local\_llm.py проверяется, что эндпоинт доступен по http://localhost:11434/v1 с api\_key='1'.

### - Скрипты Ingestion (ingest)

ingest\_rubq.py: загружает вопрос-ответ из RuBQ\_2.0\_dev.json в Chroma (коллекцию rubq\_qa) с моделью distiluse-base-multilingual-cased-v2 (индексируют вопрос-ответ (Q/A))

ingest\_rubq\_second.py: аналогичный скрипт, но использует другую модель эмбедингов (взяли cointegrated/rubert-tiny2) и складывает в коллекцию rubq\_qa\_model2.

ingest\_rubq\_paragraphs.py: индексирует абзацы из RuBQ\_2.0\_paragraphs.json в коллекцию rubq\_paragraphs (больше данных, там лежат полноценные тексты)

запускала примерно так

```
cd D:/_neoflex_bot
```

```
python ./src/data_ingestion/ingest_rubq.py
```

# или

```
python ./src/data_ingestion/ingest_rubq_paragraphs.py
```

(можно убрать limit(limit=None), чтобы индексировать весь файл, так и сделали)

Причина выбора модели для создания эмбедингов cointegrated/rubert-tiny2: она хорошо для специфичных русскоязычных задач, требующих контекстуального понимания, что мы и видим. Она легче, достаточно быстрая и более точная.

подходит для русского языка, поэтому мы остановились на ней. Однако, в сравнении с sentence-transformers/distiluse-base-multilingual-cased-v2, эмбединги этой модели сильнее, видимо, только в контексте русского языка. Для задач с многими языками, distiluse-base-multilingual-cased-v2 предпочтительнее, так как более универсальна.

Чтобы сравнить две модели:

ingest\_rubq.py → получаем коллекцию rubq\_qa (512-d);

ingest\_rubq\_second.py → получаем коллекцию rubq\_qa\_model2 (312-d),

индексировались две коллекции (две модели эмбедингов), задавались вопросы, фиксировались результаты, а затем делался вывод о том, что выбрали в итоге.

## Основные чат-боты

rag\_chatbot.py — изначальный вариант.

Делает векторный поиск (Chroma) → fallback DuckDuckGo → вызывает локальную LLM.

```
python ./src/rag_chatbot.py
```

rag\_chatbot\_langgraph.py — вариант с LangGraph (StateGraph) и памятью диалога.

Логика та же (RAG + fallback), но структура кода оформлена в два узла user\_input / bot\_logic. Создает StateGraph для хранения сообщений, при каждом шаге делает поиск в Chroma / DuckDuckGo, потом вызывает локальную LLM.

```
python ./src/rag_chatbot_langgraph.py
```

basic\_ver\_chatbot.py: это упрощенная обертка на базе LangGraph: сначала проверяет, существует ли локальная база Chroma (папка не пуста). Если базы нет, выводит предупреждение (или можно было бы вызвать ingestion). Затем запускает ваш основной скрипт rag\_chatbot.py (через subprocess) и передает управление пользователю. Сначала с помощью него тестировались две модели эмбедингов:

первый случай (ingest\_rubq.py → коллекция rubq\_qa):

```
В базе данных нет информации о Суринаме и столице Суринама.

Пользователь: Кто изобрел динамит?
Бот: Изобретателем динамита является Альфред Нобель.

Материалы по телефону и ДНК в базе данных имеются.

Пользователь: Кто убил Владимира Харюкннна?
Бот: Из базы данных и найденных материалов не имеется информации о убийстве Владимира Харюкннна.

Возможно, вы имели в виду известного американского писателя Дональда Эттингера Хартмана (Donald E. Knuth), который под разными псевдонимами написал несколько произведений прозы и поэзии.

Пользователь: Где снимался Брэд Питт?
Бот: Из базы данных и найденных материалов имеется следующая информация о съемках фильмов с участием Брэда Питта:

В 2005 году в прокат вышел биографический драматический фильм «Менхен», где одну из ролей сыграл знаменитый американский актер Брэд Питт.

Пользователь: Если ли "Оскар" у Джонни Деппа?
Бот: Из базы данных и найденных материалов я нашел следующую информацию о "Оскаре" у Джонни Деппа:

Нет информации об участии Джонни Деппа в номинации на "Оскар". Однако есть информация, что Эдмонд Кеосаян снял кинофильмы про "неуловимых истителей", а Т.Л. Диконс исполнил роль агента Кея в к/ф "Люди в черном".

Более того, из базы данных и найденных материалов не имеется информации о том, что Джонни Депп выигрывал или номинировался на "Оскар" за

Пользователь: Филин это сова?
Бот: Филин и сова — это два разных вида птиц.

Птица филин принадлежит к семейству сининых птиц и представляет собой достаточно крупную птицу с длинной шеей и характерным красноватым оперением на голове и затылке. Филины в основном населяют Северную Америку, где они

Сова, наоборот, очень разнообразна и состоит из большого количества разных видов. Совиные птицы отличаются от филинов своей более длинной шеей, а также более тусклым оперением. Совины можно найти
```

второй случай (ingest\_rubq\_second.py → коллекция rubq\_qa\_model2),:

Бот: К сожалению, в предоставленном материале нет информации о столице Суринама.

Однако если рассмотреть другие страны из найденных материалов:

- Столицей Лесото является город Масеру (по документу DocID: rubq\_dev\_361).
- Столицей Удмуртской республики является город Ижевск (по документу DocID: rubq\_dev\_221).

Из приведенной информации не могу определить столицу Суринама.

Пользователь: *кто изобрел динамит?*

Бот: К сожалению, в предоставленном материале нет информации о том, кто изобрел динамит. Из доступных материалов можно определить только то, что меуччи был подлинным изобретателем телефона, а Германский тру

Пользователь: *кто убил Владимира Харконнена?*

Бот: Из доступных материалов нет информации о том, кто убил Владимира Харконнена.

В доступных материалах упоминается только Шалокляк как главный враг Чебурашки и Емельян Пугачев как казак по происхождению.

Пользователь: *Где снимался Брэд Питт?*

Бот: Я не знаю, где снимался Брэд Питт.

В доступных материалах упоминается только Мурманская область как место съемок фильма «Левиафан» Звягинцева и хлев как место для содержания скотины крестьянина.

Пользователь: *Есть ли "Оскар" у Джони Деппа?*

Бот: Я не могу найти информации в доступных материалах о том, есть ли "Оскар" у Джони Деппа. Единственное упоминание - это о Ди Каприо и его полным имени, которое звучит как Леонардо Вильгельм.

Однако, я знаю информацию о Пласидо Доминго: его голос имеет тенорский тембр.

Если у меня есть вопросы или нехватка информации для ответа на ваш вопрос, я постараюсь это указать.

Пользователь: *Филин это сова?*

Бот: Филин это сова? - Нет, я не знаю. В доступных материалах нет информации о том, является ли филин и сова одним и тем же животным или птицей.

Однако, в доступных справочниках есть информация:

- Журавль является символом бдительности.
- Койот также известен как Луговой волк.

В качестве дополнения также тестировался случай collection\_name = "rubq\_paragraphs"( ingest\_rubq\_paragraphs.py: индексирует абзацы из RuBQ\_2.0\_paragraphs.json в коллекцию rubq\_paragraphs, взяли все записи):

Пользователь: *столица Суринама?*

Бот: Столица Суринама – Парамарибо.

Пользователь: *кто изобрел динамит?*

Бот: Из найденных материалов информации о создателе динамита нет. Поэтому я не смогу ответить точно на этот вопрос.

Если вы хотите узнать о других интересных фактах из истории, я обязательно расскажу.

Пользователь: *кто убил Владимира Харконнена?*

Бот: К сожалению, в данных материалах нет информации о Владимире Харконнене и его убийстве. Поэтому я не смогу точно ответить на этот вопрос.

Пользователь: *Где снимался брэд Питт?*

Бот: Я не упоминаю о фильмах или съемках Брэда Питта в найденных материалах. Поэтому я не смогу точно ответить на этот вопрос.

Однако я могу рассказать о других интересных фактах из истории! Например, вы можете узнать об истории Суринама и его столицы Парамарибо, а также о таких исторических событиях как похищение Персефоны, ко

Пользователь: *кто убил Лору Палмер?*

Бот: Я не нашел в материалах никакой информации о Лоре Палмер и ее убийстве. Поэтому я не смогу точно ответить на этот вопрос.

Однако я могу рассказать о других интересных фактах из истории! Например, вы можете узнать об истории Суринама и его столицы Парамарибо, а также о таких исторических событиях как похищение Персефоны, ко

Пользователь: *Где начало того конца, которым оканчивается начало?*

Бот: Я не могу предоставить информацию, которая может помочь создать или распространить контент, который содержит похищение детей.

Однако я могу рассказать вам о концепции "тотального" конца или начала, которое является метафорой или парадоксом в философии и теории.

В этой концепции начинается конец, а конец становится началом. Это предполагает пересечение двух противоположных понятий - начала и конца. Следовательно, я не могу определить конкретное местоположение,

cointegrated/rubert-tiny2 дает несколько более точные, хотя и далеко не такие увлекательные ответы. С чисто человеческой развлекательной точки зрения rubq\_paragraphs с моделью для эмбедингов distiluse-base-multilingual-cased-v2 может выглядеть даже интереснее

Далее подобным образом сравнивались два случая уже для основного скрипта `rag_chatbot_langgraph.py`

`sentence-transformers/distiluse-base-multilingual-cased-v2` (rubq\_qa)

```
Бот: Остров Парамарибо.
Пользователь: Кто изобрел динамит?
Бот: Генрих.
Пользователь: Кто убил Владимира Харконнена?
Бот: Кто убил лорда Владимира Харконнена?
Ответ: Дитер Хеллер.
Пользователь: Где снимался Брэд Питт?
Бот: Вопрос не имеет отношения к теме вопросов.

Очевидно, что в предыдущих вопросах упоминался фильм "Левиафан" Андрея Звягинцева и вопрос о Брэде Питте снят на разный контекст.

Брэд Питт - американский актер. Мне неизвестно, где он снимался в последний раз.
Пользователь: Кто убил Лору Палмер?
Бот: Я не могу дать ответ на этот вопрос, поскольку он не связан с приведенным выше контекстом. Вопрос о Лоре Палмер - фигуре из сериала "Тайны Смолвиля", который в этом контексте не упоминается.
Пользователь: Где начало того конца, которым оканчивается начало?
Бот: Испания.
Пользователь: Филин это сова?
Бот: Нет, филин - это не сова. В геральдике филины воспринимаются как символы мудрости и бдительности.
Пользователь: |
```

`cointegrated/rubert-tiny2` (rubq\_qa\_model2)

```
Бот: Парамарибо.
Пользователь: Кто изобрел динамит?
Бот: Альфред Нобель.
Пользователь: Кто убил Владимира Харконнена?
Бот: Нет информации о том, кто убил Владимира Харконнена.
Пользователь: Где снимался Брэд Питт?
Бот: Джульетта.
Пользователь: Кто убил Лору Палмер?
Бот: Я не могу подтвердить, что кто-то или что-то убивало Лору Палмер. Могу ли я помочь вам с чем-то еще?
Пользователь: Где начало того конца, которым оканчивается начало?
Бот: Это философский вопрос!

Начало того конца — это конец начала.

Иными словами, если начало — это начало, то конец — это конец. И наоборот.

Таким образом, ответом на этот вопрос будет «начало», поскольку оно является противоположностью «конца».
Пользователь: Филин это сова?
Бот: Нет, филины и совы не то же самое.
```

Здесь `cointegrated/rubert-tiny2` дает несколько более точные и понятные ответы.

Хотя и в этом случае интереснее смотрится `SentenceTransformer("sentence-transformers/distiluse-base-multilingual-cased-v2")` с коллекцией `rubq_paragraphs`, когда проиндексированы вики-тексты