

Дататон 2023



TimeTravelTech

О КУНИСЬ В ИСКУССТВО



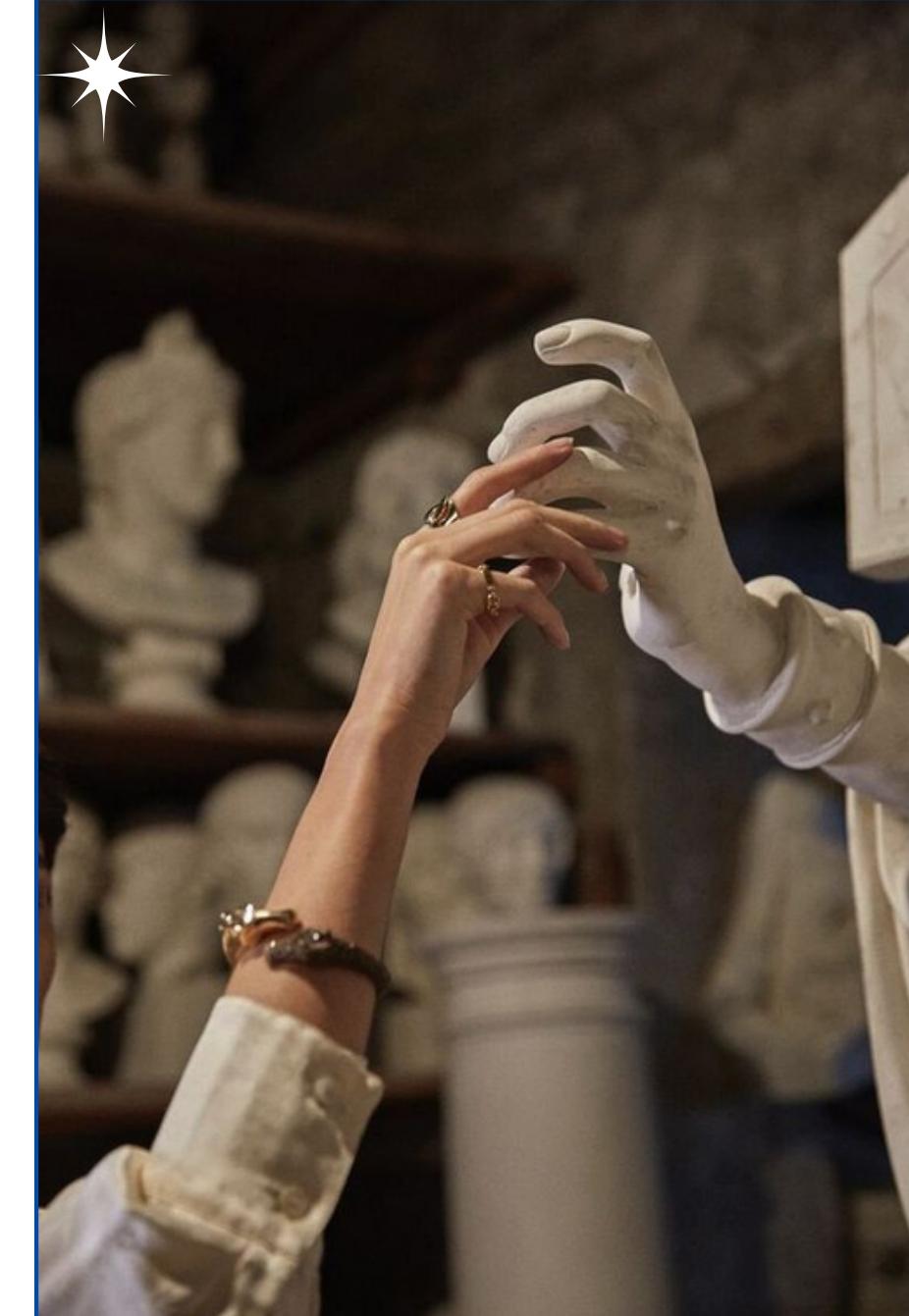
Митягина Дарья & Дяминова Эльвира
«Науки о данных»

ПЛАН

01. Идея
02. Пайплайн
03. СV модуль. Датасет для модели по классификации изображений
04. СV модуль. Модель для классификации картин
05. Podcast модуль. Использованные предобученные модели
06. Image generation модуль
07. Платформа и демонстрация
08. Дальнейшие шаги

ИДЕЯ

Мы верим, что в мире, насыщенном технологиями и экранами, наступил момент остановиться и взглянуть вокруг. Мы убеждены, что оторваться от своего телефона и обратить внимание на искусство вокруг нас — это ключ к настоящему вдохновению и пониманию красоты. TimeTravelTech стремится сделать этот шаг к искусству легким и доступным, предлагая краткие и яркие моменты встречи с творчеством, которые вписываются в любой график. Давайте вместе откроем для себя новые грани красоты вне мира электронных экранов!



ПАЙПЛАЙН

Фотография картины ->

СV модуль ->

Распознанное имя автора ->

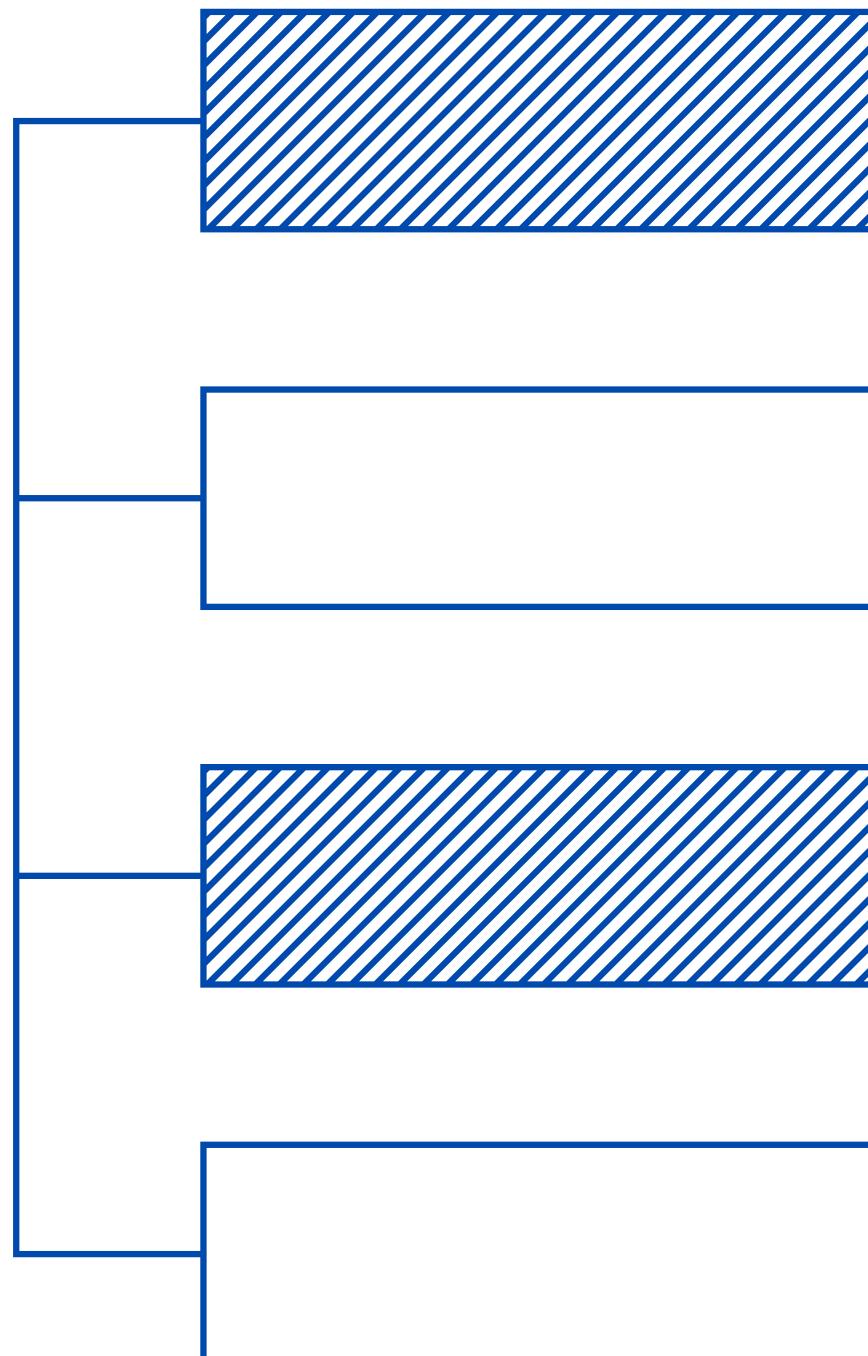
Генерация текста ->

Преобразования текста в речь ->

Эффекты на аудио ->

Подкаст ->

Генерация изображения



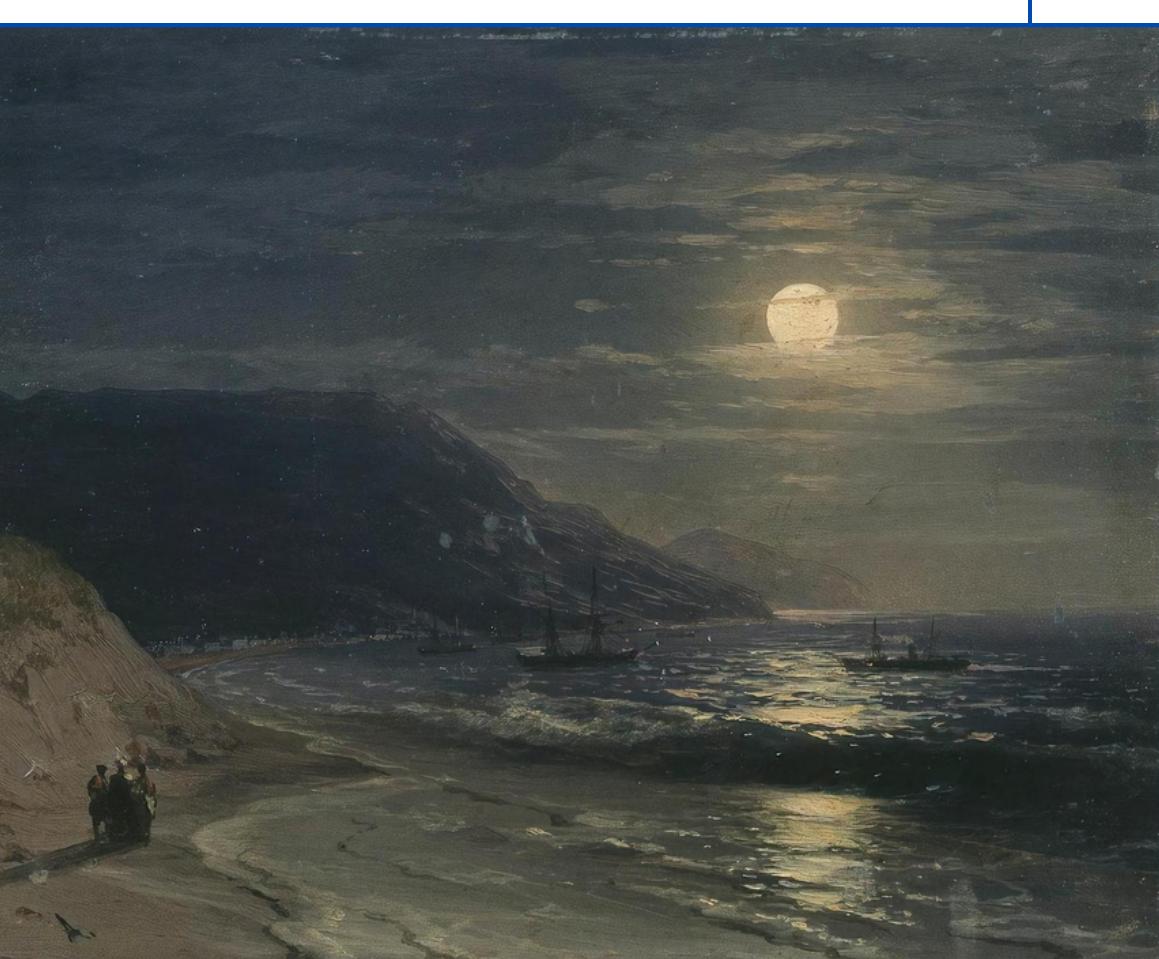
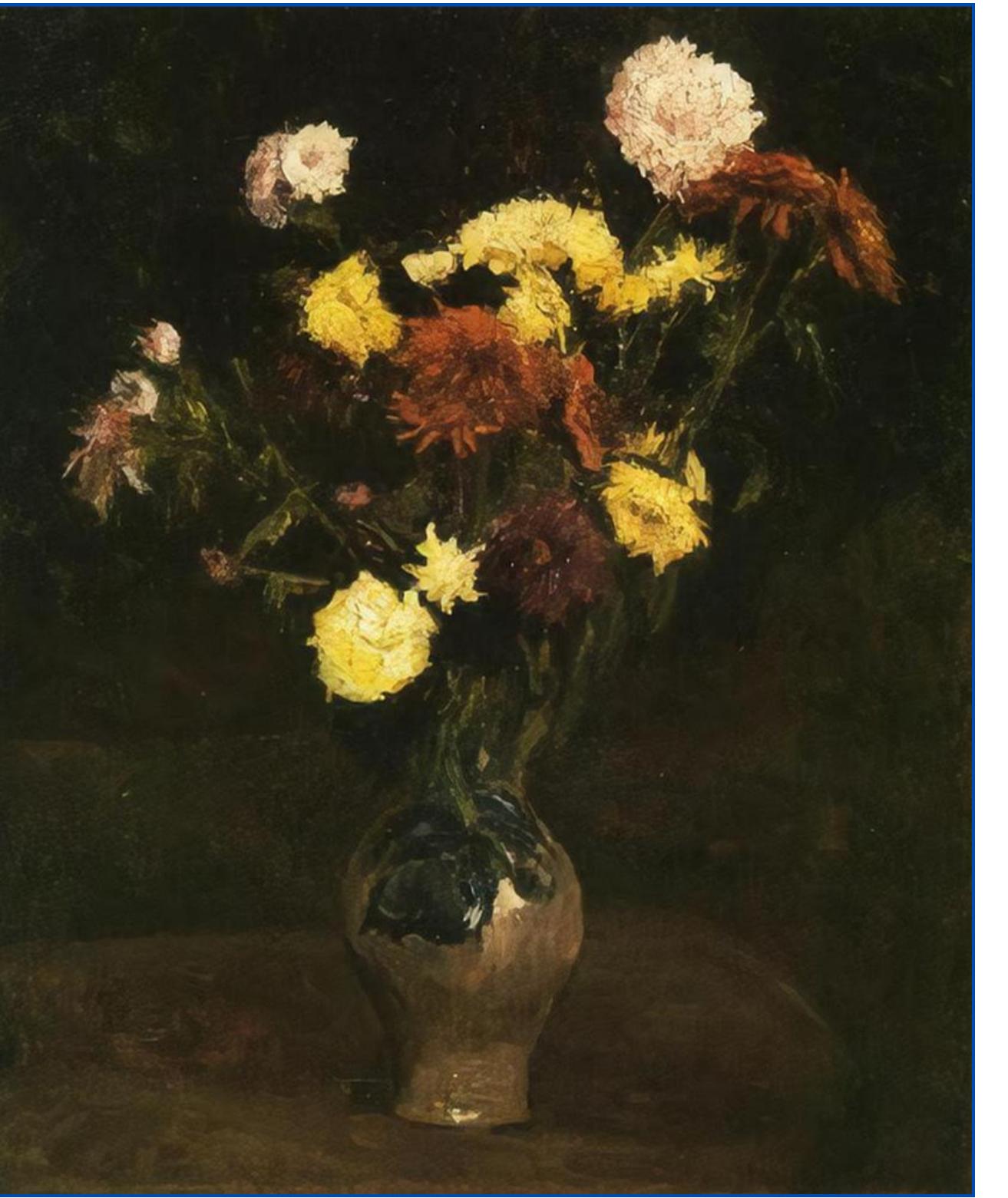
СВ МОДУЛЬ. ДАТАСЕТ ДЛЯ МОДЕЛИ ПО КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Оригинальный датасет «Wikiart» — это коллекция художественных изображений, полученных из веб-сайта www.wikiart.org с информацией о произведениях искусства. В нем содержится 80020 уникальных изображений от 1119 художников в 27 стилях.

Для этой работы исходный набор данных был сокращен.

Предобработка данных:

- были отобраны 5 художников — camille pissarro, claude monet, edgar degas, ivan aivazovsky, vincent van gogh
- для каждого из художников оставлен только один главный стиль в его творчестве.



МОДЕЛЬ ДЛЯ КЛАССИФИКАЦИИ КАРТИН ПО ХУДОЖНИКАМ

inception_v3 — модель, которая была выбрана для этой работы.

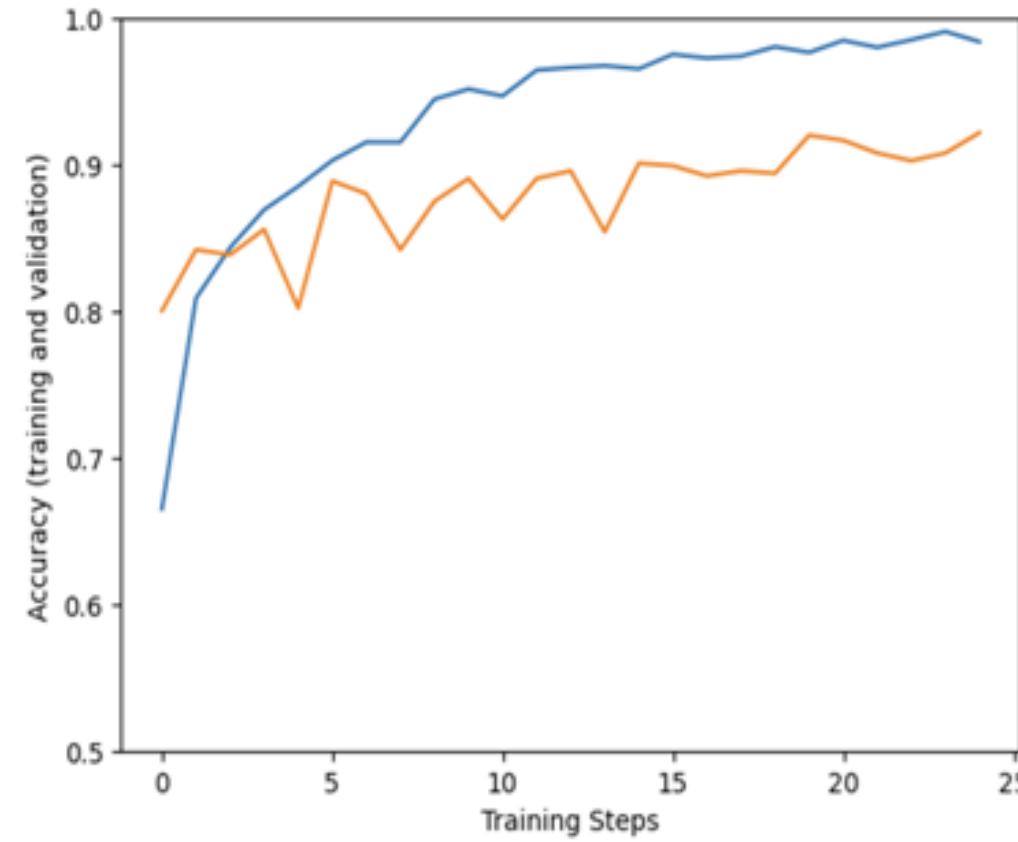
В итоге **inception_v3** была обучена на 25 эпохах.

В процессе обучения также происходила аугментация данных: **RandomTranslation** (смещение внутри заданного размера) и **RandomRotation** (угол поворота). Такие варианты аугментации были выбраны, чтобы сымитировать фотографии, которые люди могут делать в музеях. Для обучения модели использовался **TensorFlow** (2.10.0).

Результаты на последних эпохах:

```
Epoch 23/25
145/145 [=====] - 183s 1s/step - loss: 0.6976 - accuracy: 0.9853 - val_loss: 0.8510 - val_accuracy: 0.9028
Epoch 24/25
145/145 [=====] - 201s 1s/step - loss: 0.6886 - accuracy: 0.9909 - val_loss: 0.8618 - val_accuracy: 0.9080
Epoch 25/25
145/145 [=====] - 199s 1s/step - loss: 0.6891 - accuracy: 0.9840 - val_loss: 0.8174 - val_accuracy: 0.9219
```

РЕЗУЛЬТАТЫ МОДЕЛИ



classes: ('camille pissarro', 'claude monet', 'edgar degas', 'ivan aivazovsky', 'vincent van gogh')

precision: [0.84, 0.92, 0.86, 0.98, 0.94]

recall: [0.83, 0.93, 0.86, 0.98, 0.94]

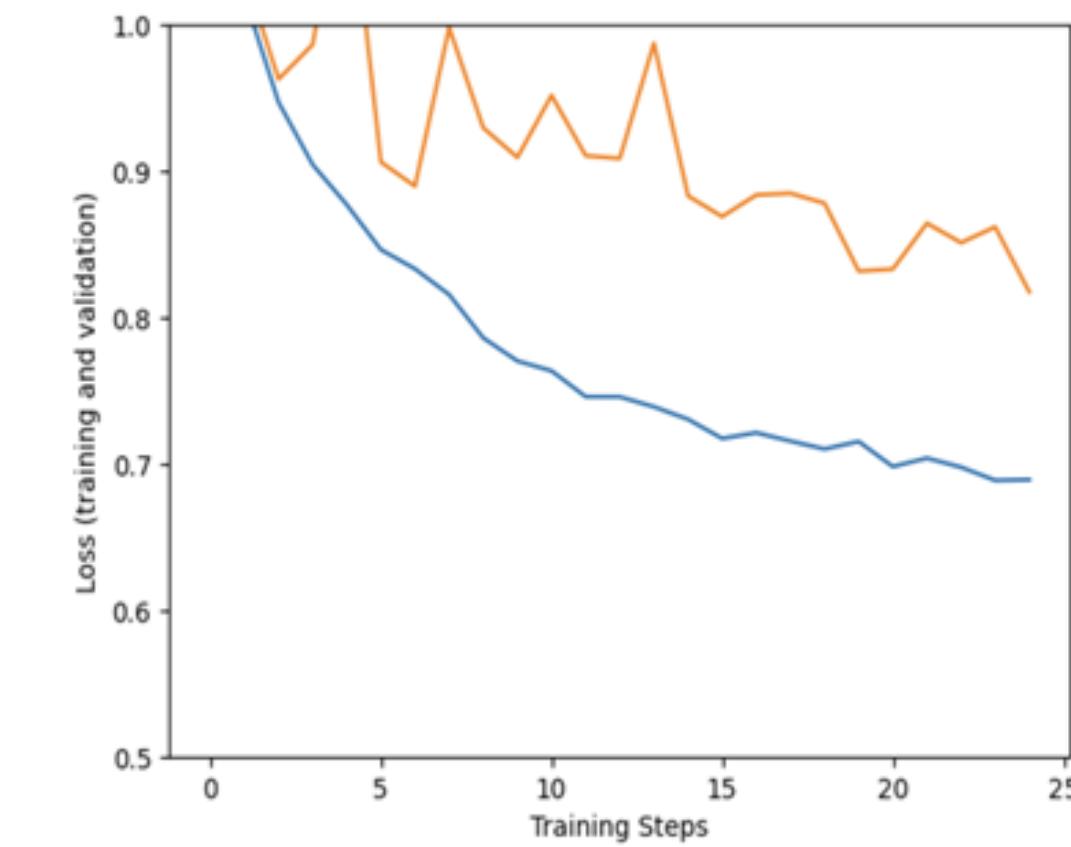
fscore: [0.84, 0.93, 0.86, 0.98, 0.94]

mean precision: 0.91

mean recall: 0.91

mean fscore: 0.91

В качестве метрики был выбран метод precision_recall_fscore_support из библиотеки sklearn (sklearn.metrics) для вычисления precision, recall, F-measure для каждого класса.



PODCAST МОДУЛЬ. ИСПОЛЬЗОВАННЫЕ ПРЕДОБУЧЕННЫЕ МОДЕЛИ

- Text-davinci-003 - это крупная модель GPT-3 с 175 миллиардами параметров. Она обучена на огромном объеме данных и умеет генерировать текст, отвечать на вопросы, выполнять переводы и другие задачи обработки языка. Модель способна учиться на широком контексте
- Text-to-speech (TTS). Модель предлагает на выбор шесть предустановленных голосов и два варианта модели: tts-1 и tts-1-hd. TTS оптимизирован для случаев использования в реальном времени, а tts-1-hd оптимизирован по качеству.

IMAGE GENERATION МОДУЛЬ

DALL·E представляет собой версию [GPT-3](#) с 12 миллиардами параметров, обученную генерировать изображения из текстовых описаний на датасете из пар текст-изображение.

Конкретно в нашем прототипе данная модель используется для создания картин в стиле того или иного художника.

Леонардо Да Винчи



Иван Айвазовский



Винсент Ван Гог

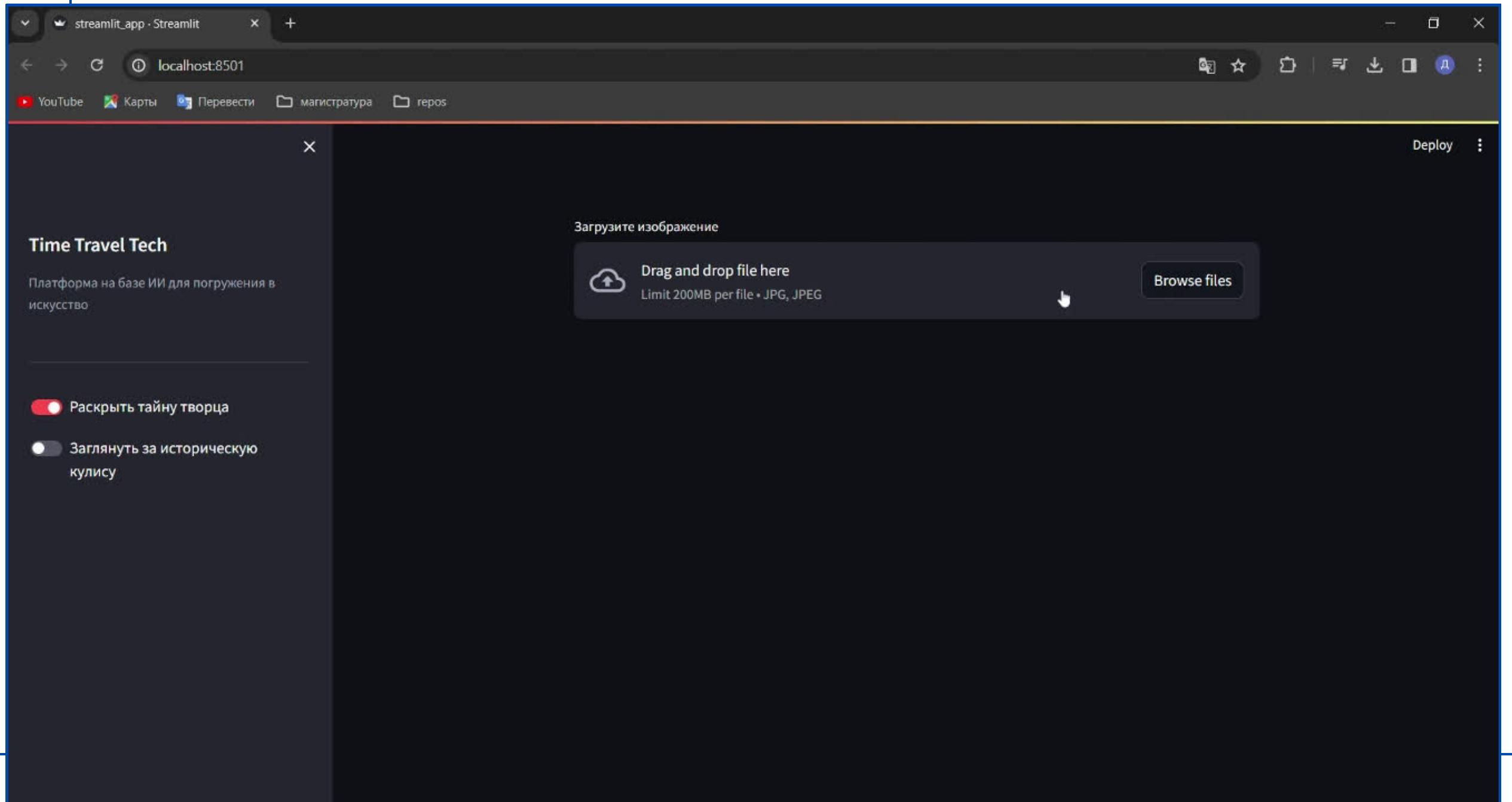


Камиль Писсарро



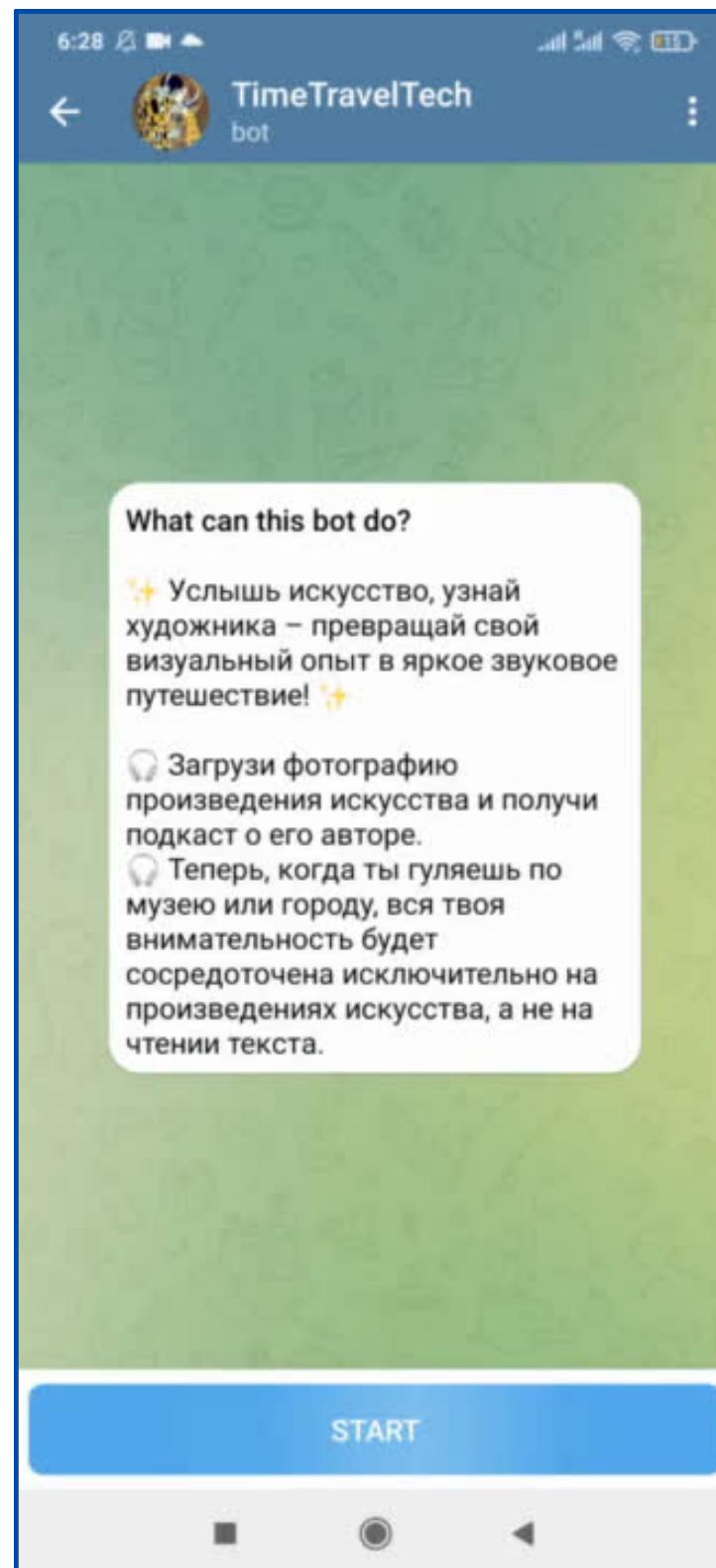
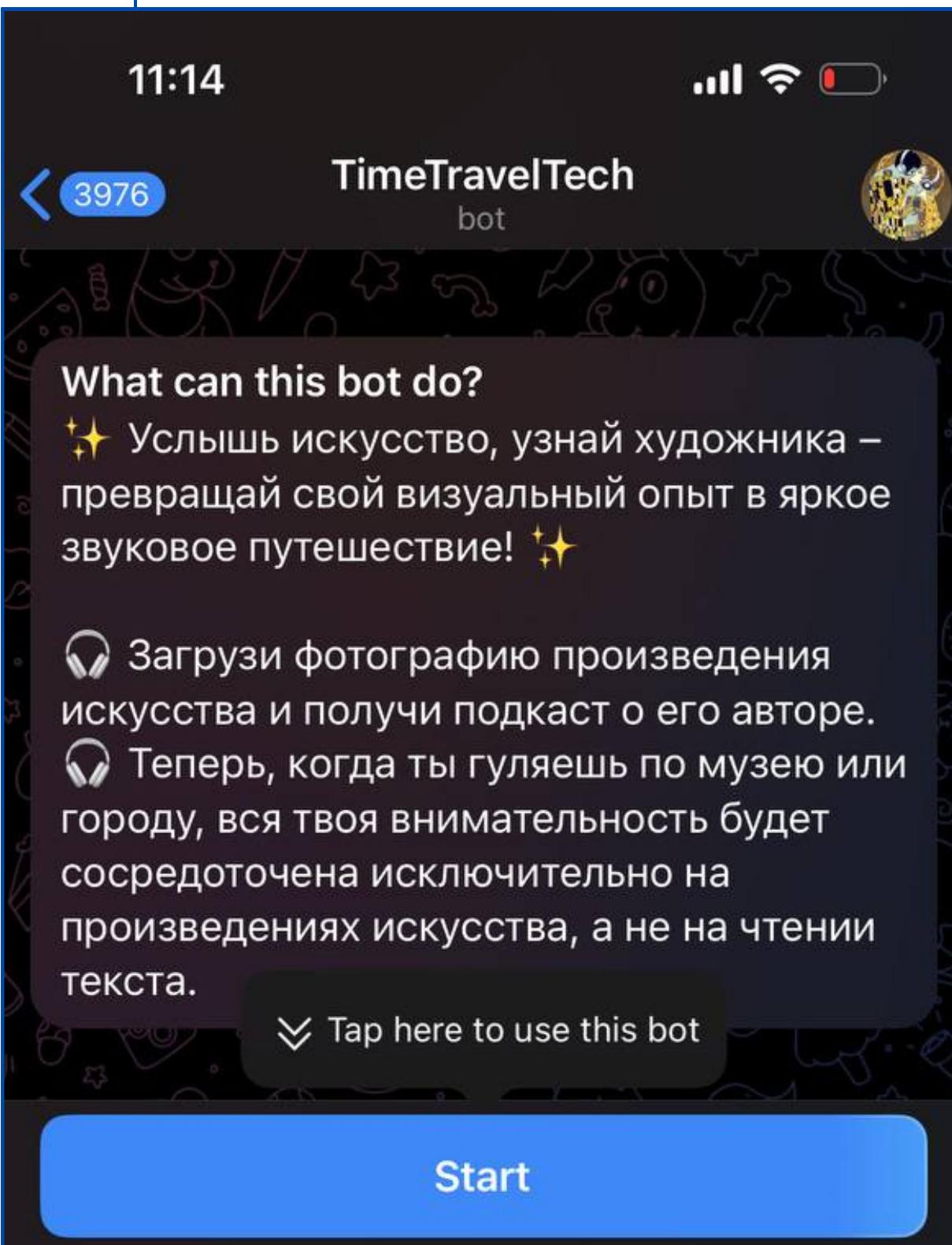
ПЛАТФОРМА И ДЕМОНСТРАЦИЯ

Для визуальной составляющей было разработано локальное веб-приложение на основе Streamlit для агрегирования результатов таких модулей, как классификация картин по художникам,



создание историй об их жизни с последующем генерацией подкаста.

Для инференса модели по классификации картин был создан локальный веб-сервер на Flask.



МОБИЛЬНАЯ ВЕРСИЯ

Был создан телеграм чат бот, полностью повторяющий пайплайн. Загрузи фото любой картины, и наш интеллектуальный чат-бот определит автора, создаст захватывающий подкаст и расскажет тебе удивительные истории об искусстве. Исследуй, учись и погружайся в мир художественной гармонии.

ДАЛЬНЕЙШИЕ ШАГИ

01. При получении больших ресурсов планируется проведение экспериментов:

- с большим количеством эпох на уже выбранной модели;
- с более сложными архитектурами нейронных сетей для задачи классификации;
- использование полного датасета с 80020 изображениями и 1119 художниками;
- создание модели для классификации не только художника, но и его стиля.

02. Планируются эксперименты с другими LL моделями

03. Расширить доступ к прототипу