

Министерство науки и высшего образования Российской Федерации
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ**

«Национальный исследовательский университет ИТМО»



Факультет программной инженерии и компьютерной техники
09.03.04 - Нейротехнологии и программирование

Выпускная квалификационная работа
**«Telegram-бот для стилизации изображений с использованием
диффузионных моделей»**

Выполнил: Ле Нгок Тхиен - Группа: Р342111
Руководитель: к.т.н., доцент ФПИиКТ
Штенников Дмитрий Геннадьевич

Санкт-Петербург

2024

Актуальность темы



Растущий спрос: рост использования социальных сетей подчеркивает необходимость создания цифрового контента.



Технологические инновации: диффузионные модели позволяют создавать креативные и высококачественные изображения.



Проблема доступности: высокие требования к оборудованию и технические навыки ограничивают доступ пользователей.

Цель: автоматизация стилизации пользовательских изображений.



Задачи:

1. Проведение анализа предметной области и обзора существующих аналогов. Выбор наиболее подходящих методов для решения задачи.
2. Реализация модели диффузии и метода PhotoMaker с использованием библиотеки PyTorch и пакета Diffusers.
3. Разработка веб-приложения с пользовательским интерфейсом в виде бота в Telegram, используя API Telegram Bot и базу данных на SQLite.
4. Проведение тестирования и оптимизации производительности приложения.
5. Выгрузка исходного кода приложения в открытый репозиторий.

Диффузионные модели

Таблица 1: Список заметных диффузионных моделей



Название модели	Дата выпуска	Разработчик	Использование	Открытый исходный код
DALL-E	Январь 2021	OpenAI	Веб-интерфейс / API (облачный)	Нет
DALL-E 2	Апрель 2022	OpenAI	Веб-интерфейс / API (облачный)	Нет
Imagen	Май 2022	Google	Веб-интерфейс, API (По приглашениям)	Нет
Midjourney	Июль 2022	Midjourney, Inc.	Бот в Discord (облачный)	Нет
Stable Diffusion	Август 2022	Stability AI	Локальное или облачное использование	Да

Методы персонализации

Таблица 2: Методы персонализации в диффузионных моделях



Метод	Тип метода	Количество изображений при выводе	Время (секунда)
DreamBooth	С настройкой	Множественное количество	1284
Textual Inversion	С настройкой	Множественное количество	2400
FastComposer	Без настройки	1	8
IP-Adapter	Без настройки	1	12
PhotoMaker	Без настройки	1 или несколько фотографий	10

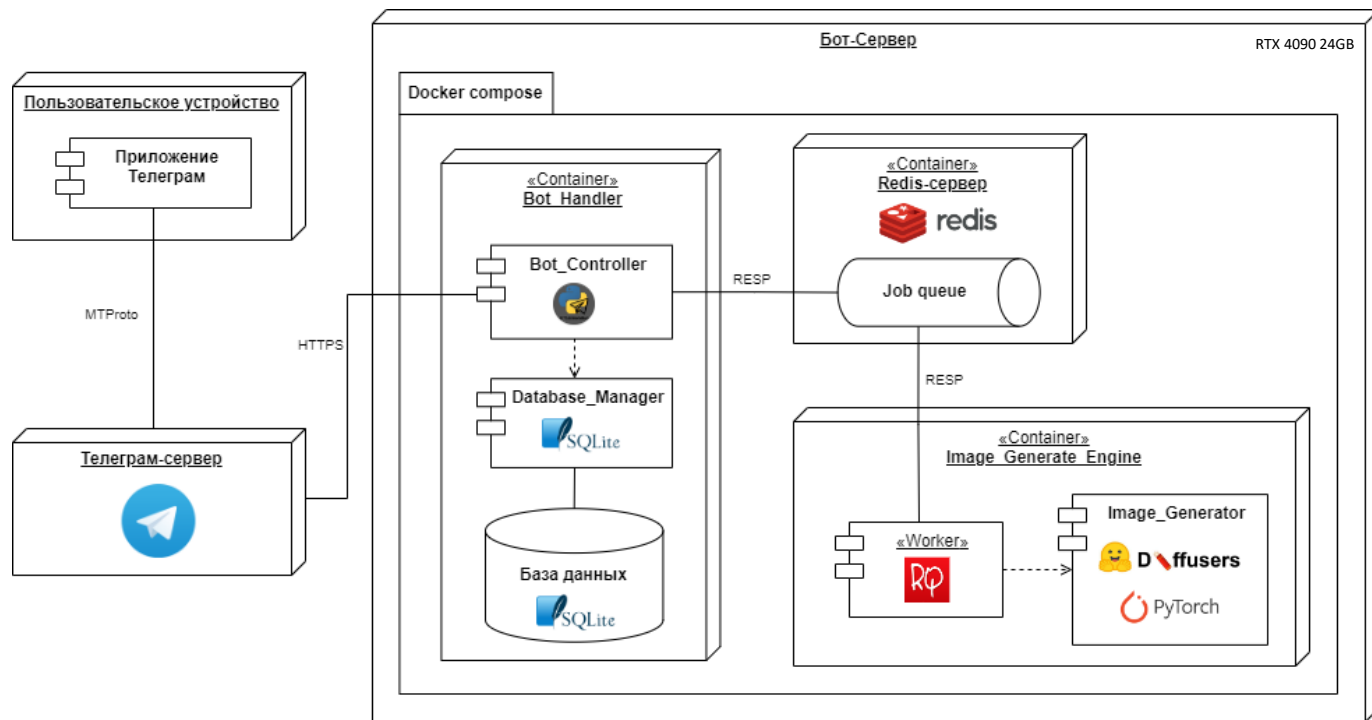


Рисунок 1: Высокоуровневая архитектура системы Telegram-бота

Проектирование

Bot_Handler

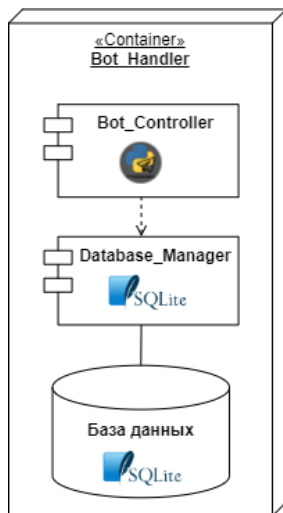


Таблица 3: Список команд Telegram-бота

	Команда	Описание
Команды взаимодействия с пользователем	/start	Запустить бота
	/help	Показать список команд
	/menu	Руководство по созданию запросов
Команды управления фотографиями пользователя	/upload	Добавить новое фото в коллекцию
	/show_uploads	Посмотреть галерею всех загруженных фото
	/delete	Удалить конкретное фото из коллекции
Команда создания изображения	/generate	Создать изображение на основе творческого запроса
Команды рекомендации запросов	/popular_prompts	Исследовать наиболее часто используемые запросы
	/top_rated_prompts	Просмотр запросов с самым высоким рейтингом

Проектирование

Bot_Handler

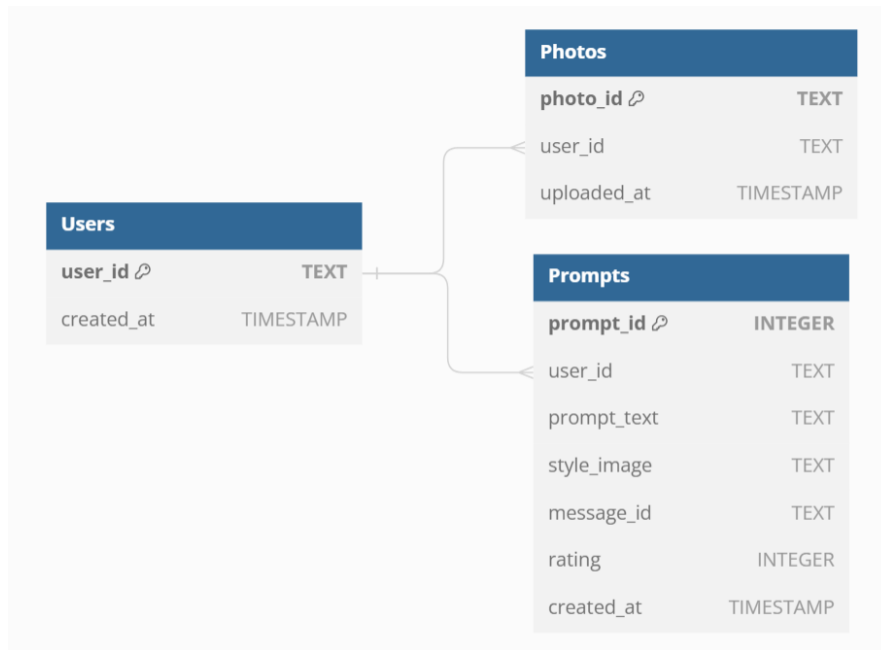
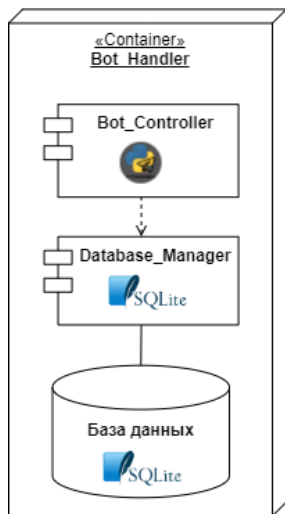


Рисунок 2: Схемы базы данных Телеграм-бота

Redis-сервер



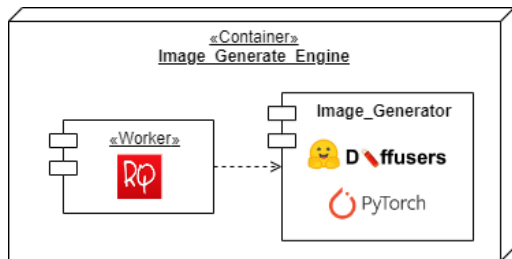
Высокие требования к вычислительным ресурсам: каждый запрос на создание изображения требует значительных вычислительных ресурсов.

- Сервер развертывания настроен на обработку одного запроса за раз.

Redis как управление очередью задач: использует FIFO для управления очередями задач.

Проектирование

Image_Generate_Engine



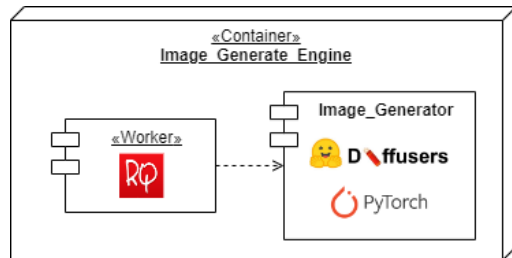
RQ-worker: (библиотека Redis Queue) управления заданиями из сервера Redis.

Image_Generator:

- модель **Stable Diffusion XL (Diffusers)**
- метод **PhotoMaker (Pytorch)**
- используйте весовые модели, доступные на **Hugging Face Hub**.




Проектирование

Image_Generate_Engine

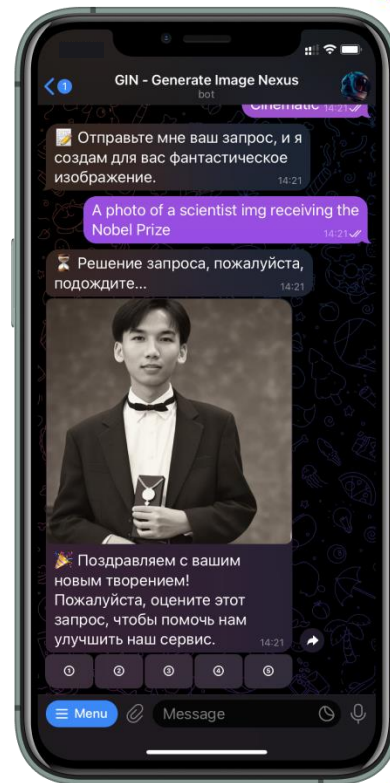
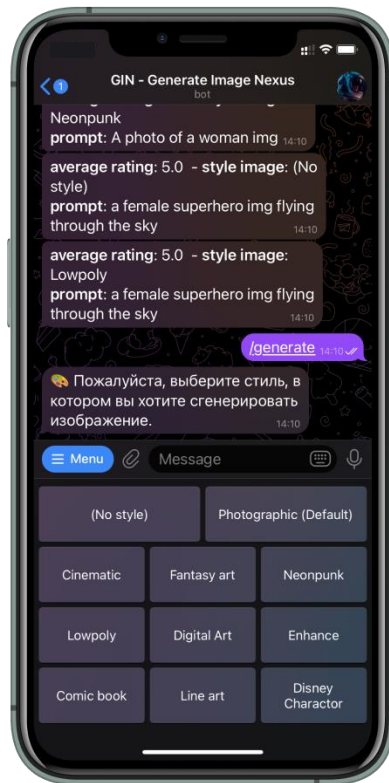
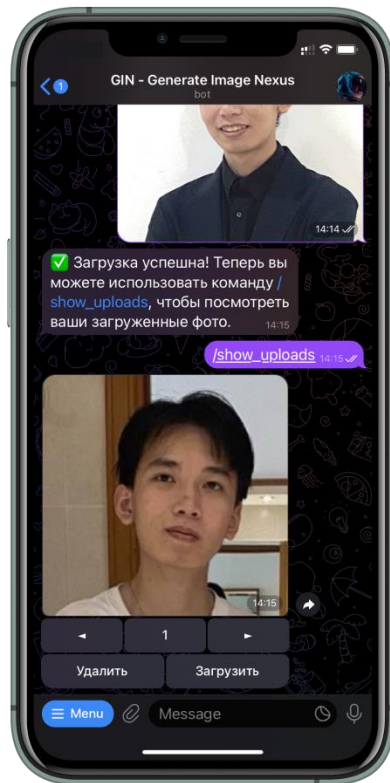
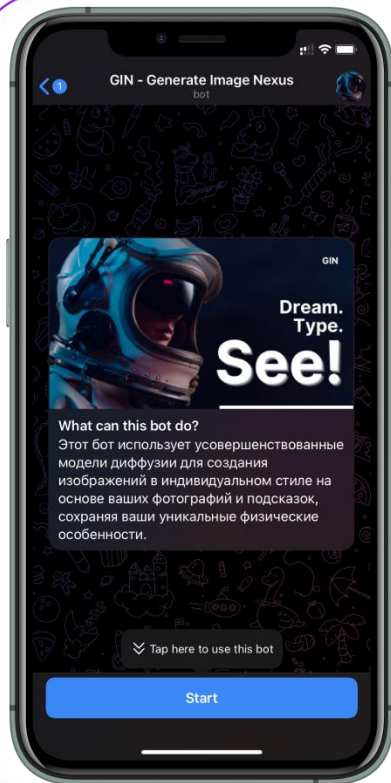


Функция выбора стиля

Бот предоставляет выбор predetermined стилей для упрощения процесса.

ЗАПРОС	A photo of a man as a Russian astronaut	A man riding a futuristic motorbike	A man wearing a doctoral cap
СТИЛЬ	Photographic	Neonpunk	Line art
ИЗМЕНЕННЫЙ ЗАПРОС	realistic photograph A photo of a man as a Russian astronaut. 35mm photograph, film, professional, 4k, highly detailed	neonpunk style A man riding a futuristic motorbike . cyberpunk, vaporwave, neon, vibes, vibrant, crisp, detailed, dark purple shadows, high contrast	line art drawing A man wearing a doctoral cap . professional, sleek, modern, minimalist, graphic, line art, vector graphics
РЕЗУЛЬТАТ			

Реализация



Мониторинг



собрать показатели
производительности
из очереди заданий



визуализировать
собранные показатели
производительности

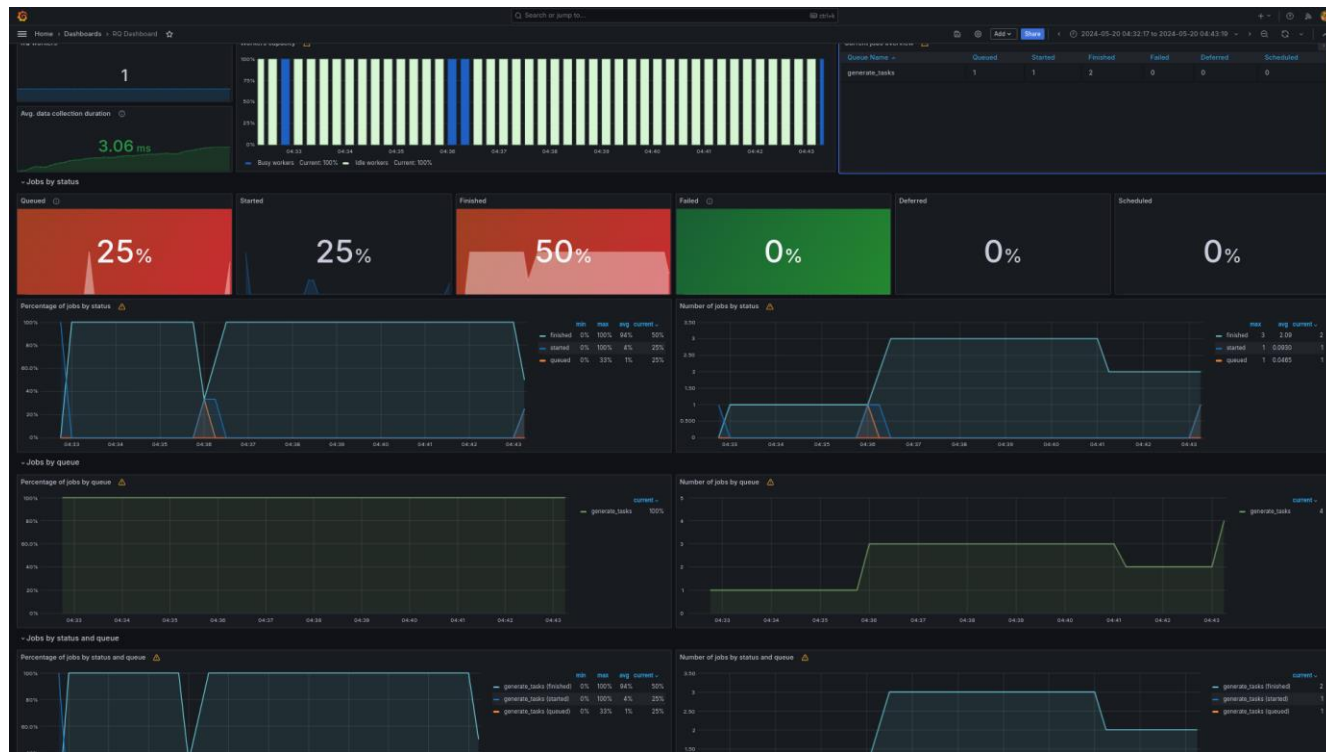


Рисунок 3: Интерфейс мониторинга Grafana

В результате выполнения ВКР достигнуты следующие результаты:



- Проведен подробный анализ **6 диффузионных моделей** и **5 методов персонализации** изображений.
- Интеграция модели **Stable Diffusion XL** и метода **PhotoMaker** обеспечила высокое качество результата и персонализацию изображений.
- Приложение оснащено удобным интерфейсом, включающим **9 команд** для улучшения взаимодействия с пользователем. Приложение объединяет **10 дополнительных предопределенных опций стиля**.
- Система спроектирована и разработана для одновременной обработки нескольких запросов, обеспечивая стабильную производительность и масштабируемость. Среднее время обработки запроса составляет **40 секунд**.
- В систему включен **компонент мониторинга**, обеспечивающий оптимальную производительность и оперативное выявление любых потенциальных проблем.
- Проект опубликован в открытом репозитории GitHub.

Таким образом, все задачи данной работы выполнены и цель достигнута.