

**Федеральное государственное автономное образовательное  
учреждение высшего образования «Национальный исследовательский  
университет ИТМО»**

**Факультет программной инженерии и компьютерной техники**

**Нейротехнологии и аффективные вычисления**

**Проектная работа**

**Выполнили:**

**Кобик Никита Алексеевич**

**Маликов Глеб Игоревич**

**Группа № Р3324**

**Преподаватель:**

**Русак Алена Викторовна**

**Санкт-Петербург**

**2025**

## Оглавление

<i>Обоснование выбранной темы проекта .....</i>	<i>3</i>
<i>Общая область интересов .....</i>	<i>4</i>
<i>Постановка проблемы и гипотезы .....</i>	<i>5</i>
<i>Цель, задачи и роли участников .....</i>	<i>6</i>
<i>Теория эмоций .....</i>	<i>9</i>
<i>Основные переменные .....</i>	<i>11</i>
<i>Особенности .....</i>	<i>12</i>
<i>Разработка .....</i>	<i>13</i>
<i>Модели распознавания эмоций .....</i>	<i>13</i>
<i>Telegram-бот .....</i>	<i>17</i>
<i>Результаты .....</i>	<i>20</i>
<i>Заключение .....</i>	<i>23</i>
<i>Источники .....</i>	<i>24</i>
<i>Приложения .....</i>	<i>26</i>

## **Обоснование выбранной темы проекта**

Современное общество всё активнее взаимодействует с интеллектуальными системами, и способность машин понимать эмоции человека становится одним из ключевых аспектов эффективной коммуникации. Эмоции играют важнейшую роль в человеческом поведении, принятии решений и социальной адаптации, поэтому их автоматическое распознавание открывает широкие горизонты в области образования, медицины, маркетинга и цифровых сервисов.

Однако большинство существующих решений по распознаванию эмоций фокусируются на одностороннем анализе — определении состояния пользователя. Наш проект расширяет этот подход: мы не просто распознаем эмоции человека, но и создаем обратную эмоциональную связь через визуальный образ — стикер с животным, выражающим аналогичную эмоцию. Такой подход делает взаимодействие более человечным и вовлекающим.

Особенность проекта также в универсальности использования: пользователь может отправить как фото, так и видео. Мы извлекаем лицо, анализируем его, классифицируем эмоцию, а затем находим соответствующий стикер, используя обученную модель, способную распознавать эмоции животных по изображениям. Это делает проект уникальным среди аналогов.

## **Общая область интересов**

Проект лежит на пересечении нескольких областей знаний: машинного обучения, компьютерного зрения, психологии эмоций и разработки чат-ботов.

В нашем случае применяется классическая архитектура сверточных нейронных сетей (CNN) для анализа изображений и классификации эмоций. Такие модели хорошо зарекомендовали себя в задачах facial expression recognition (FER). Мы также исследуем возможности обратной эмоциональной связи через визуальный отклик — стикер, что относится к области human-computer interaction (HCI).

Telegram-бот используется как универсальный интерфейс взаимодействия. Это удобный способ для пользователей протестировать и использовать систему без необходимости устанавливать сторонние приложения.

## **Постановка проблемы и гипотезы**

Существует множество решений, направленных на распознавание эмоций у человека. Однако крайне мало систем, которые бы не просто фиксировали эмоциональное состояние, но и отвечали бы на него. В частности, использование визуального, символического ответа в виде стикеров может усилить ощущение эмпатии со стороны системы.

**Проблема:** современные пользовательские интерфейсы не обеспечивают эмоциональной обратной связи, что снижает вовлеченность пользователя и ограничивает эмоциональное взаимодействие с ИИ.

**Гипотеза:** можно создать мультимодальную интеллектуальную систему, способную автоматически определять эмоции пользователя на изображениях (или видео) и отвечать ему стикером с животным, визуально выражающим такую же эмоцию. Такая система будет вызывать больше доверия и эмпатии у пользователей.

## **Цель, задачи и роли участников**

Основной целью проекта является автоматизацией проявления эмоции пользователей с помощью подбора стикеров с животными. Реализация данного решения позволит оптимизировать процесс выражения эмоций, сократить время на подбор визуальных средств коммуникации и улучшить качество цифрового общения. Для этого предлагается создание инструмента для анализа эмоционального состояния пользователей в мессенджере Телеграм.

### **Задачи:**

#### **1. Сбор и подготовка данных**

##### **1.1. Получение датасета животных (Маликов)**

- Выбрать и загрузить датасет.
- Провести первичную проверку целостности и корректности меток.

##### **1.2. Получение датасетов людей (Маликов)**

- Подобрать датасет содержащий размеченные изображения.
- Провести анализ качества данных, сверить формат и полноту разметки.

#### **2. Разработка и обучение моделей ИИ**

##### **2.1. Обучение модели для распознавания эмоций у людей (Маликов)**

- Спроектировать архитектуру модели, CNN или гибридные решения для работы с видео.
- Настроить гиперпараметры и определить стратегию обучения.
- Провести обучение модели на статических изображениях.
- Выполнить валидацию, оценку точности и оптимизацию модели.

##### **2.2. Обучение модели для распознавания эмоций у животных (Кобик)**

- Спроектировать специализированную архитектуру для анализа эмоций животных.
- Провести настройку гиперпараметров и выбрать алгоритмы оптимизации.

- Обучить модель на выбранном датасете животных.
- Провести тестирование и оптимизацию модели для повышения точности.

### **3. Подготовка базы стикеров**

#### **3.1. Выбор стикеров (Кобик)**

- Провести отбор изображений животных-стикеров, соответствующих различным эмоциональным состояниям.

#### **3.2. Разметка стикеров (Кобик)**

- Пропустить стикеры через модель для автоматической разметки эмоциональных характеристик.
- Проверить корректность разметки и при необходимости откорректировать метки.

#### **3.3. Сохранение стикеров в базе данных (Кобик)**

- Разработать структуру БД для хранения изображений стикеров и их разметки.
- Импортировать стикеры с соответствующими метками в базу данных.

### **4. Клиентская часть**

#### **4.1. Telegram-бот (Кобик)**

- Реализовать механизм проверки поступающих сообщений на наличие фото или видео.
- Разработать логику отправки полученного медиа на сервер для анализа.
- Организовать приём ответа от бэкенда с выбранным стикером.
- Обеспечить корректную отправку стикера пользователю через API Telegram.

### **5. Разработка серверной части**

#### **5.1. Обработка видео (Маликов)**

- Если входной медиафайл — видео, реализовать модуль разделения видео на отдельные кадры.

## 5.2. Предобработка изображений и выделение лиц (Маликов)

- Реализовать алгоритм детекции лиц на изображениях или кадрах видео.
- Адаптировать изображения для совместимости с входным форматом обученных моделей.

## 5.3. Запуск моделей и получение результатов (Кобик)

- Интегрировать вызов обученной модели для распознавания эмоций у людей в Telegram-бот.
- Обеспечить обработку как одиночных изображений, так и последовательностей кадров, в случае видео.
- Если видео, агрегировать результаты с нескольких кадров и возвращать случайный стикер из списка найденных эмоций.

## 5.4. Отправка результатов обратно в Telegram-бот (Кобик)

- Организовать передачу результатов, с выбранным стикером, от серверной части к Telegram-боту.



## Теория эмоций

В основе нашего проекта лежит теория базовых эмоций Пола Экмана. Согласно этой теории, существует универсальный набор эмоций, выражаемых через мимику: радость, грусть, гнев, отвращение, страх, удивление и нейтральное состояние. Именно на эти категории мы ориентировались при формировании датасетов и обучении моделей

Почему выбор пал именно на эту классификацию?

Экман показал, что ряд эмоциональных выражений универсален для людей разных культур и даже проявляется у врождённо слепых людей. К шести опорным критериям универсальности относятся:

1. узнаваемость выражения людьми из разных культур;
2. связь мимики с определёнными физиологическими реакциями;
3. стереотипность мимики у слепорождённых;
4. схожесть проявлений у приматов;
5. способность эмоции быстро возникать и так же быстро угасать;
6. наличие отличительных, легко описываемых выражений лица.

Эти критерии удобны для машинного анализа: количество классов ограничено, выражения дискретны, а набор визуальных индикаторов (подъём бровей, форма рта, расширение глаз) хорошо формализуется. Выбранные семь категорий представляют собой «минимально достаточный» базис, позволяющий достичь практической точности на реалистичном объёме данных

Для животных, изображенных на стикерах, использовалась та же классификация — как результат ручной разметки. Мы исходили из визуальной экспрессии: открытые пасти, уши, глаза, позы — всё это интерпретировалось аналогично человеческим эмоциям

## Ограничения теории

Мы осознаём, что эмоциональный опыт человека богаче, чем семь базовых состояний: существуют смешанные и культурно-специфичные эмоции («ностальгия», «застенчивость»). Однако модульная архитектура проекта позволяет при появлении размеченных корпусов расширить набор классов без перестройки всей системы.

## **Перенос классификации на стикеры с животными**

Ту же шкалу эмоций мы применили к изображениям животных, потому что:

- пользователю проще считывать параллели между своей мимикой и «зеркальным» отображением в виде животного;
- единое пространство меток упрощает сопоставление;
- визуальная экспрессия (положение ушей, раскрытая пасть, направление взгляда) надёжно коррелирует с аналогичными человеческими эмоциями.

Такой выбор поддерживает когнитивную согласованность интерфейса: бот реагирует «по-человечески», но через дружелюбный образ

## Основные переменные

Входное изображение: влияет качество, освещение, ракурс.

Наличие лица: критично для модели человека. Используется MTCNN.

Разрешение и масштаб: все изображения приводятся к 48x48 (для модели).

Баланс классов: важно обеспечить равномерное распределение эмоций в датасете.

Эмоция на стикере: зависит от визуальной выразительности — изображения подбирались вручную.

Эти переменные требуют как качественной предобработки, так и архитектурной устойчивости модели.

## Особенности

Интерфейс взаимодействия пользователя с системой лежит на Telegram-боте, принимающем изображение (или видео), далее система извлекает из полученного материала лицо, определяет эмоцию и подбирает соответствующий стикер

Система состоит из следующих компонентов:

- Две независимые модели CNN: для человека и для животных (стикеров);
- Детектор лиц (MTCNN);
- Механизм предобработки (перевод в градации серого, подбор разрешения, разбиение на кадры в случае видео);
- Telegram-бот с журналированием действий;
- База данных для хранения данных пользователей и размеченных стикеров.

Основной упор был сделан на универсальность и модульность архитектуры, что позволит в будущем дообучать модели, заменять или добавлять другие интерфейсы взаимодействия (например, веб-приложение).

Дополнительно был сделан упор на уменьшение времени простоя сервиса при заполнении или обновлении базы стикеров.

## Разработка

### Модели распознавания эмоций

Для распознавания эмоций людей были соединены датасеты FER-2013 и AffectNet, так как оба датасета уже размечены в одинаковой 7-классной схеме и не требуют особых предобработок кроме конвертации в размер 48x48 в grayscale в случае с изображениями из AffectNet. Таким образом размер датасета для тренировки модели составляет 43 258 изображений, а для валидации 20 384 изображений. Из-за высокой разницы в количестве изображений между классами потребовалось использование категориальной Focal-Loss( $\alpha = 0.25, \gamma = 2$ ), который устраняет крайний дисбаланс классов.

Эмоция	Изображения
Радость	9555
Грусть	7921
Нейтральное состояние	7723
Страх	5609
Гнев	5495
Удивление	5290
Отвращение	1665

*Таблица 1 - Количество изображений для каждой эмоции людей в датасете для тренировки*

Во время работы бота, в качестве входных данных принимаются изображения размером 48x48 в оттенках серого. Предобработка включала нормализацию и изменение размера. Для лиц — выделение через MTCNN и конвертация в grayscale.

Для распознавания эмоций животных был использован датасет Pets Facial Expression Dataset с добавлением 98 изображений, полученных из интернета и размеченных вручную. При предобработке, все изображения были изменены до размеров 224x224, при этом были оставлены цветными.

Для модели распознавания эмоций людей использовалась следующая архитектура:

- 2 слоя Conv2D 3x3 с BatchNorm

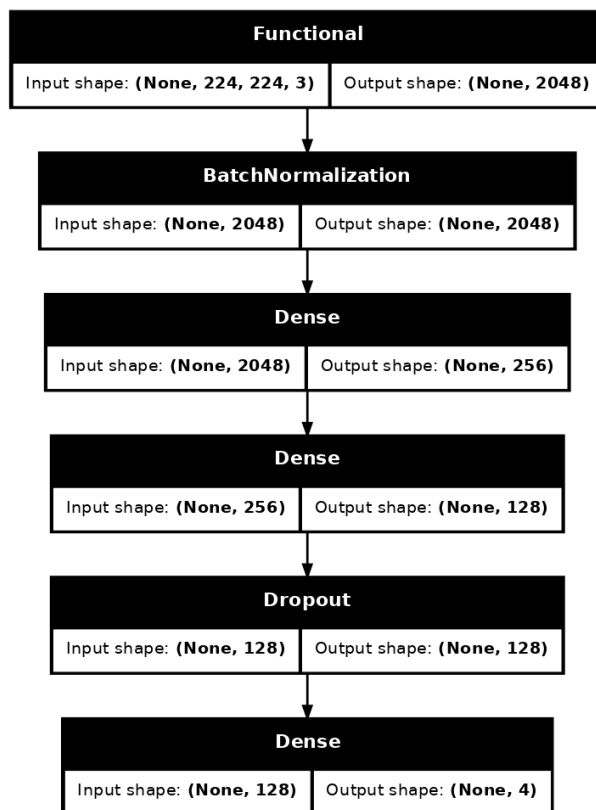
- Блок sep\_res\_block с 128 фильтрами и se\_block
- Блок sep\_res\_block с 256 фильтрами и se\_block
- Блок sep\_res\_block с 512 фильтрами и se\_block
- Блок sep\_res\_block с 728 фильтрами и se\_block
- GlobalAveragePooling2D
- Softmax-слой на выходе.

Данная модель включает в себя: Depthwise-Separable Convolution, позволяющий уменьшать количество операций в несколько раз без заметной потери точности, и Residual блоки + Squeeze-and-Excitation, которые ускоряют сходимость и дают адаптивное внимание к областям глаз/рта.

Для модели распознавания эмоций животных использовалась следующая архитектура на Keras:

- Модель EfficientNetB5
- BatchNormalization
- Dense 256 ReLU
- Dense 128 с Regularizer L2
- Dropout 0.45
- Softmax-слой на выходе.

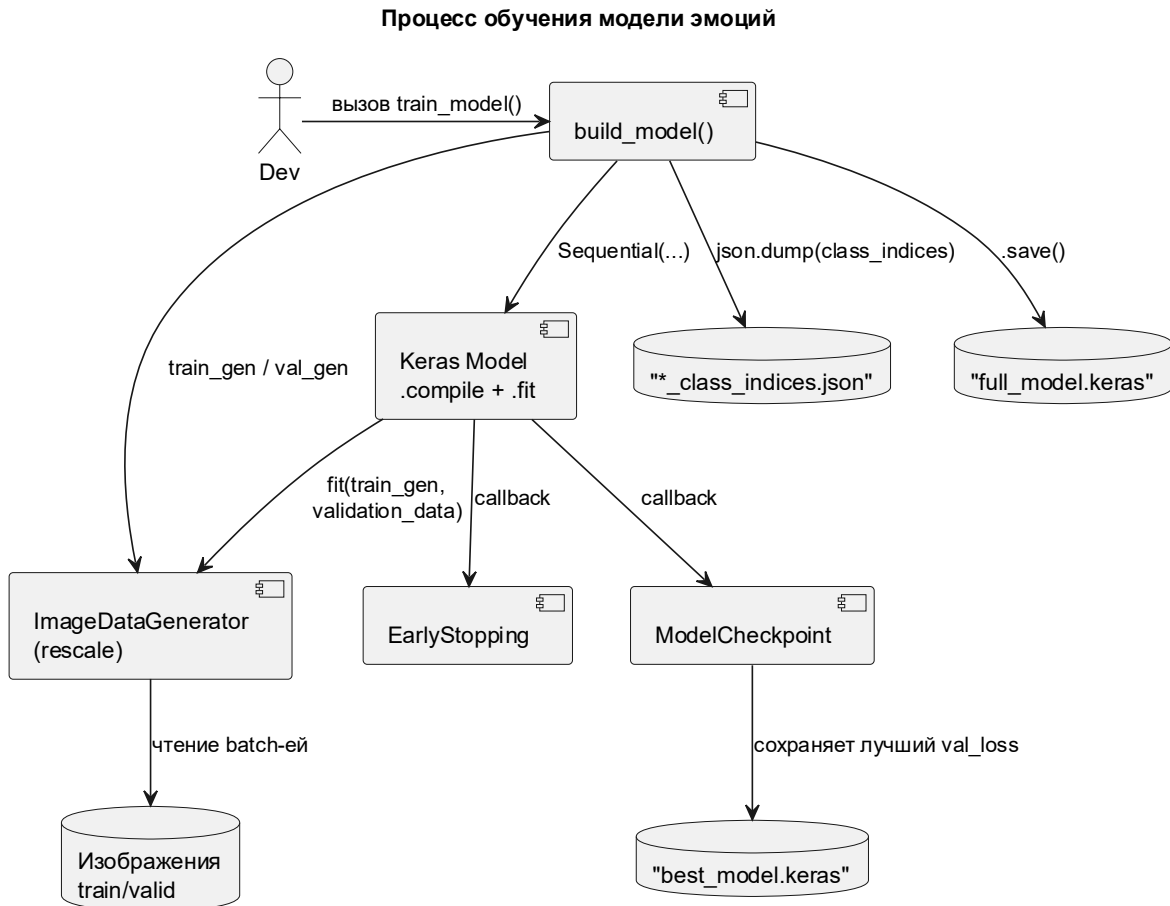
Решение использования EfficientNetB5 как база для модели было принято из-за малого размера датасета. При попытке тренировки собственной модели с нуля точность предсказания эмоций никогда не превышала 50% процентов, либо модель начинала всегда предсказывать одну и ту же эмоцию.



*Рисунок 1 - Схема слоев моделей анализа эмоции*

В процессе обучения использовались следующие механизмы:

- Ранняя остановка (EarlyStopping);
- Сохранение лучшей модели (ModelCheckpoint);
- Использование метрик
  - Accuracy;
  - Precision;
  - Recall.



*Рисунок 2 - Диаграмма процесса обучения модели эмоций*

Гиперпараметры каждой модели как например, размер батча 16, скорость обучения  $\approx 1 \times 10^{-3}$ , оптимизаторы AdamW/Adamax с регуляризацией веса и обрезкой градиента, а также стратегии ReduceLROnPlateau и EarlyStopping, были подобраны эмпирически: мы пробовали несколько комбинаций learning rate, batch size и методов регуляризации на валидационной выборке и выбирали те, которые давали наилучшее соотношение скорости сходимости и обобщающей способности, не преследуя автоматической оптимизации (Grid/Random Search), а опираясь на практический опыт и особенности каждого датасета.

Разметка классов сохраняется в файл формата \*.json и используется при инференсе.

Модуль face.py реализует извлечение лиц с помощью MTCNN и их подготовку для модели. Для видео используется cv2.VideoCapture, разбиение на кадры и выбор



наиболее четких с лицами. Анализируется каждый кадр до тех пор, пока не будет найдена уверенная эмоция.

## Telegram-бот

Telegram-бот реализован с использованием библиотеки aiogram. Он представляет собой асинхронную FSM-архитектуру, поддерживающую обработку фото, видео, а также систему модерации и добавления стикеров администраторами.

Бот обрабатывает входящие сообщения следующих типов:

- **Фотографии:** система сохраняет файл, извлекает лицо, производит предобработку, определяет эмоцию и отправляет стикер с такой же эмоцией, выбранный случайным образом из базы.
- **Видео:** система сохраняет видео, извлекает кадры, анализирует каждый аналогично процессу для фото выше, агрегирует эмоции и отвечает пользователю стикерами по каждой обнаруженной эмоции.

Ключевые моменты реализации:

- Все изображения сохраняются во временную директорию и удаляются после обработки.
- Используются функции `get_human_emotion()` для распознавания на лицах.
- База стикеров содержит привязку `sticker_id` и эмоции. Администратор может добавлять новые стикеры, отправляя их в чат. Бот извлекает кадры из стикера, анализирует эмоцию с помощью `get_animal_emotion()`, после чего предлагает администратору подтвердить или вручную задать правильную эмоцию.
- Стикеры сохраняются в базу данных и используются ботом при генерации отклика.

Таким образом, Telegram-бот выполняет не только роль клиента, но и инструмент сбора и разметки данных, а также систему обратной эмоциональной связи для пользователя.

Ниже показаны диаграммы последовательностей взаимодействия пользователей и администраторов с ботом.

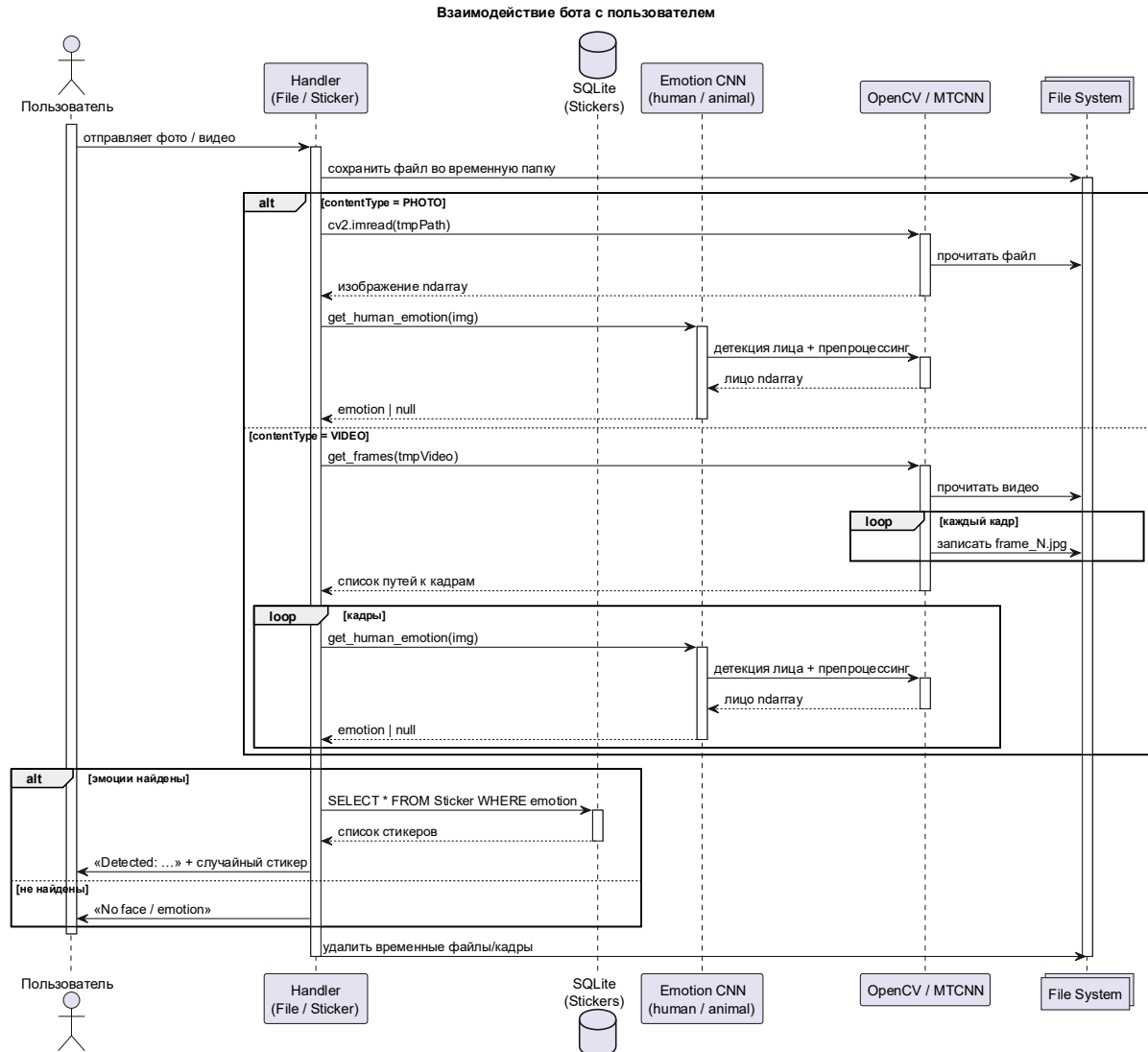
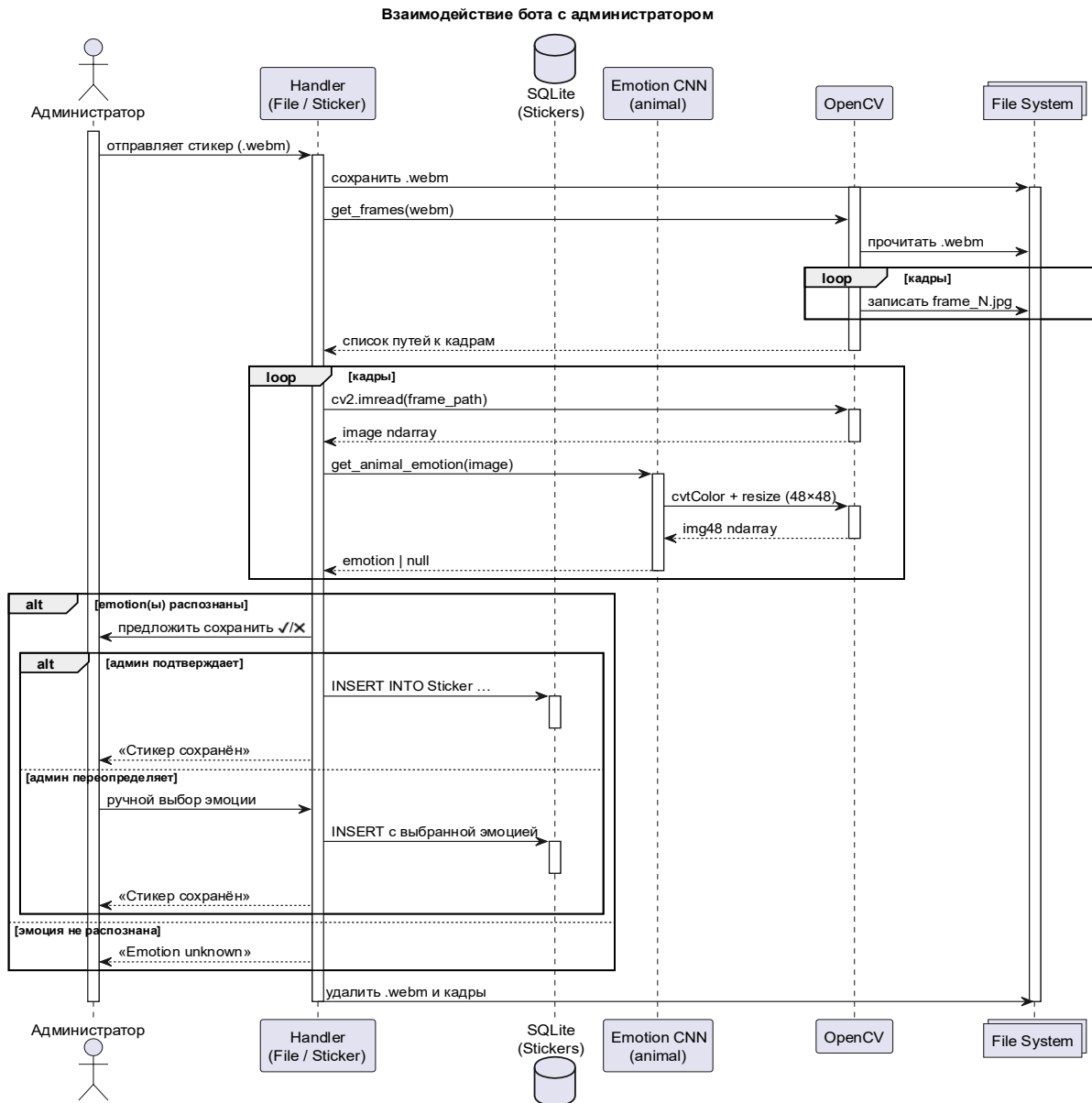


Рисунок 3 - Диаграмма последовательности взаимодействия бота с пользователем



*Рисунок 4 - Диаграмма последовательности взаимодействия бота с администратором*

## Результаты

В результате выполнения проекта были достигнуты следующие результаты:

- Обучены и протестированы две модели CNN для распознавания эмоций по изображениям людей и животных;
- Реализован детектор лиц и их предварительная обработка для подачи в модель;
- Обработаны и интегрированы возможности анализа видео и стикеров через разбиение на кадры;
- Реализована удобная система добавления стикеров и их модерации;
- Создан полнофункциональный Telegram-бот, взаимодействующий с пользователем и реагирующий визуально на его эмоциональное состояние;
- Подтверждена гипотеза о возможности эффективной визуальной эмоциональной обратной связи в ИИ-системах.

Ниже показаны результаты моделей распознавания в виде матриц ошибок

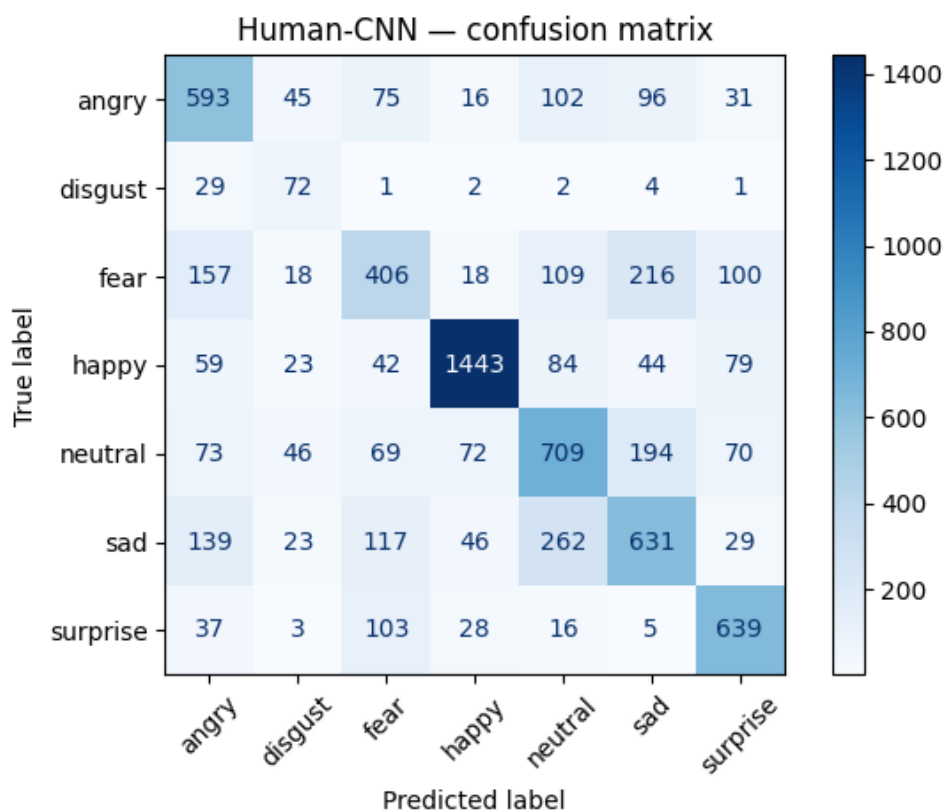


Рисунок 5 - Матрица ошибок модели людей

	precision	recall	f1-score	support
angry	0.55	0.62	0.58	958
disgust	0.31	0.65	0.42	111
fear	0.50	0.40	0.44	1024
happy	0.89	0.81	0.85	1774
neutral	0.55	0.58	0.56	1233
sad	0.53	0.51	0.52	1247
surprise	0.67	0.77	0.72	831
accuracy			0.63	7178
macro avg	0.57	0.62	0.58	7178
weighted avg	0.63	0.63	0.63	7178

Рисунок 6 - Результаты модели для людей

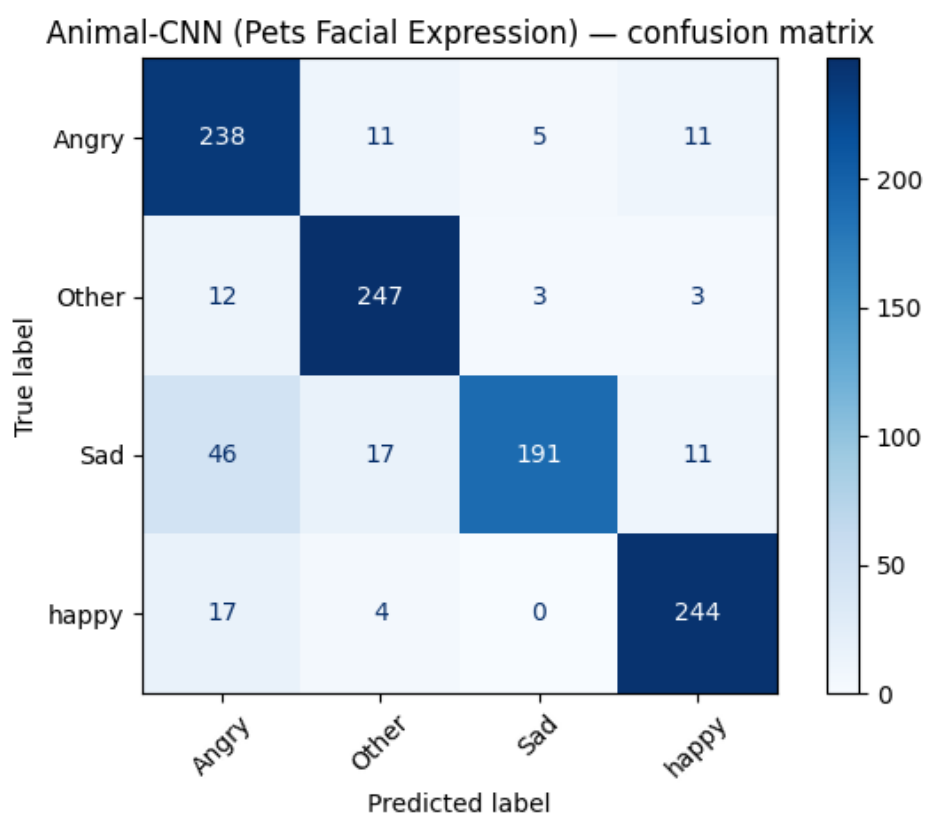


Рисунок 7 - Матрица ошибок модели животных

Animal-CNN (Pets Facial Expression) – confusion matrix				
	precision	recall	f1-score	support
Angry	0.76	0.90	0.82	265
Other	0.89	0.93	0.91	265
Sad	0.96	0.72	0.82	265
happy	0.91	0.92	0.91	265
accuracy			0.87	1060
macro avg	0.88	0.87	0.87	1060
weighted avg	0.88	0.87	0.87	1060

*Рисунок 8 - Результаты модели для животных*

## Заключение

Проект Telegram-бота для распознавания эмоций по изображениям человека и визуального отражения этих эмоций через стикеры животных представляет собой успешное сочетание компьютерного зрения, машинного обучения и разработки человеко-ориентированных интерфейсов. Мы добились стабильной и надежной работы системы, которая может быть использована в образовательных, развлекательных и исследовательских целях.

Система может быть расширена за счёт:

- добавления мультиязычности;
- внедрения ResNet, MobileNet или EfficientNet моделей для повышения точности;
- использования embedding-пространств (например, CLIP) для автоматической разметки и подбора стикеров;
- создания полноценного пользовательского приложения на Android/Web.

## Источники

Ли Б., Лима Д. Facial expression recognition via ResNet-50 // International Journal of Cognitive Computing in Engineering. – 2021. – Т. 2. – С. 57–64. – ISSN 2666-3074. – DOI: 10.1016/j.ijcse.2021.02.002. – Режим доступа: <https://www.sciencedirect.com/science/article/pii/S2666307421000073> (Дата обращения: 08.03.2025).

Мао Я., Лиу Я. Pet dog facial expression recognition based on convolutional neural network and improved whale optimization algorithm // Scientific Reports. – 2023. – Т. 13, № 1. – С. 3314. – ISSN 2045-2322. – DOI: 10.1038/s41598-023-30442-0. – Режим доступа: <https://doi.org/10.1038/s41598-023-30442-0> (Дата обращения: 08.03.2025).

Сумон Р.И., Али Х., Аккер С., Уддин С.М.И., Мозумдер М.А.И., Ким Х.-Ч. A Deep Learning-Based Approach for Precise Emotion Recognition in Domestic Animals Using EfficientNetB5 Architecture // Eng. – 2025. – Т. 6, № 1. – Ст. 9. – ISSN 2673-4117. – DOI: 10.3390/eng6010009. – Режим доступа: <https://www.mdpi.com/2673-4117/6/1/9> (Дата обращения: 08.03.2025).

Хаиреддин, Ю., Чен, З. Facial Emotion Recognition: State of the Art Performance on FER2013 // arXiv:2105.03588 [электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2105.03588> (Дата обращения: 08.03.2025).

FER-2013 (Facial Expression Recognition) : датасет изображений лиц с эмоциональными метками [Электронный ресурс] / Р.-L. Carrier, A. Courville, M. Mirza, I. Goodfellow [и др.]. – 2013. – Загл. с экрана. – URL: <https://paperswithcode.com/dataset/fer2013> ; <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data> (дата обращения: 22.05.2025).

Tanwar A. Pet's Facial Expression Image Dataset : [Электронный ресурс]. – Kaggle, 2023. – Загл. с экрана. – DOI: 10.34740/KAGGLE/DS/3546787. – URL:



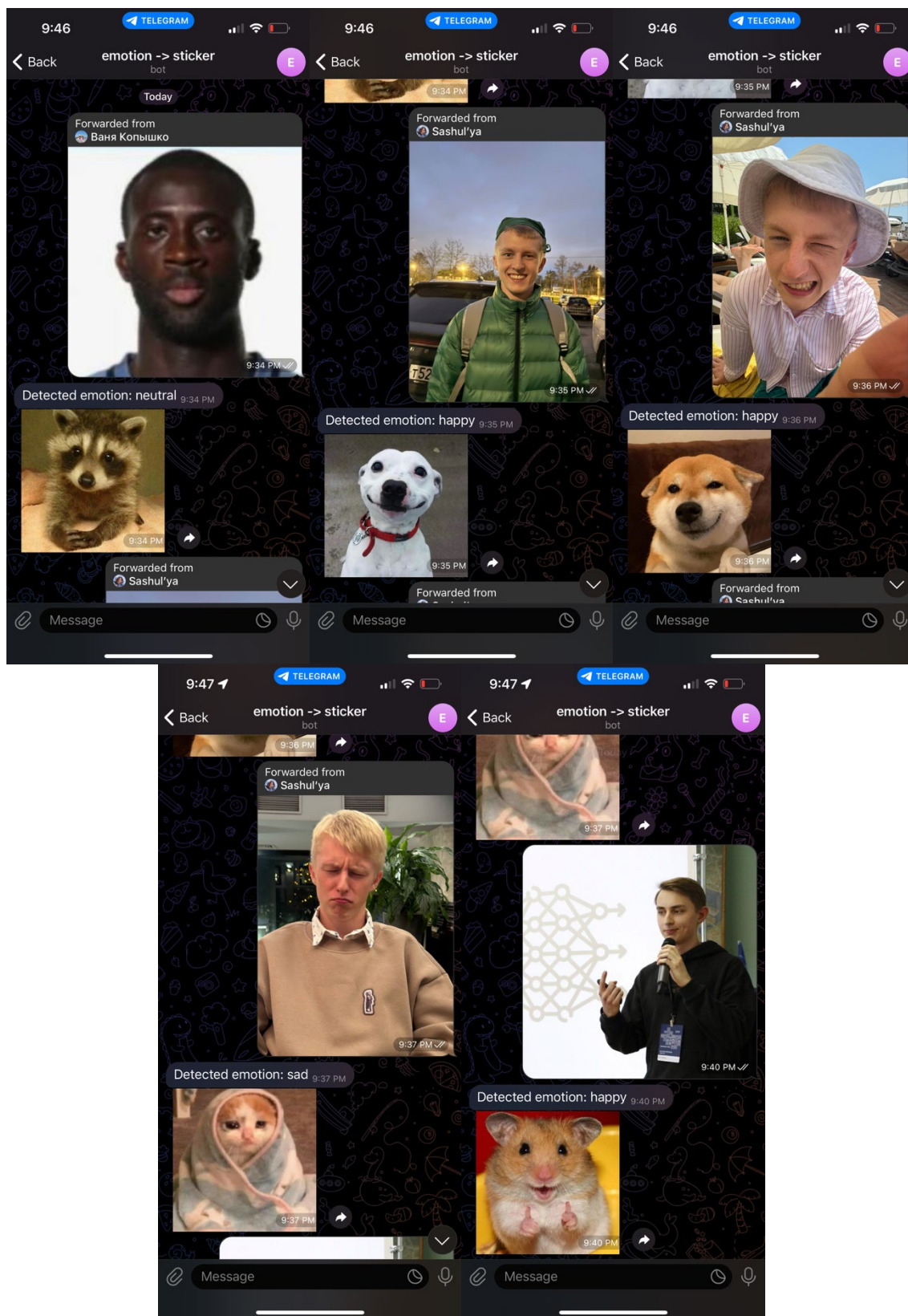
<https://www.kaggle.com/datasets/anshtanwar/pets-facial-expression-dataset> (дата обращения: 22.05.2025).

Lin T.-Y., Goyal P., Girshick R. [и др.] Focal Loss for Dense Object Detection [Электронный ресурс] / T.-Y. Lin, P. Goyal, R. Girshick и др. – arXiv:1708.02002, 2018. – Режим доступа: <https://arxiv.org/pdf/1708.02002> (дата обращения: 25.05.2025) – Загл. с экрана.

Хсиун Chieh Fei (Jeffee), Аертс Joerie. Pet Facial Expression Recognition [Электронный ресурс]. – Режим доступа: [https://github.com/jeffeehsiung/Pet\\_Facial\\_Expression\\_Recognition](https://github.com/jeffeehsiung/Pet_Facial_Expression_Recognition) (дата обращения: 26.05.2025).

## Приложения

Снимки экрана с работой бота (пользователь)



## Снимки экрана с работой бота (администратор)

