

УДК 004.89

DOI: [10.26102/2310-6018/2025.50.3.032](https://doi.org/10.26102/2310-6018/2025.50.3.032)

Анализ эмоций на видеоданных с использованием локальных и облачных решений искусственного интеллекта

Л.В. Агамиров^{1,2,3}, В.Л. Агамиров^{1,2}, В.А. Вестяк², Н.В. Тутова^{1✉}, С.А. Базунов⁴,
Ю.Н. Зеляник²

¹Московский технический университет связи и информатики, Москва,
Российская Федерация

²Московский авиационный институт, Москва, Российская Федерация

³Национальный исследовательский университет «МЭИ», Москва,
Российская Федерация

⁴Университет БРИКС, Москва, Российская Федерация

Резюме. Актуальность исследования обусловлена растущей потребностью в высокоточной и интерпретируемой системе распознавания эмоций на основе видеоданных, что имеет решающее значение для развития человеко-ориентированных технологий в образовании, медицине и системах взаимодействия человек–компьютер. В связи с этим статья направлена на выявление различий и перспектив применения локального решения DeepFace и облачной модели GPT-4o (OpenAI) для анализа коротких видеосюжетов с эмоциональными выражениями. Методологически исследование базируется на эмпирическом сравнительном анализе: использован метод скользящего среднего для сглаживания временных рядов эмоциональных оценок и оценки устойчивости и когнитивной интерпретируемости. Результаты показали, что DeepFace обеспечивает стабильную локальную обработку и высокую устойчивость к артефактам, в то время как GPT-4o демонстрирует способность к сложной семантической интерпретации и высокой чувствительности к контексту. Обоснована эффективность гибридного подхода, сочетающего вычислительную автономность и гибкость интерпретации. Таким образом, синергия локальных и облачных решений открывает перспективы для создания более точных, адаптивных и масштабируемых систем аффективного анализа. Материалы статьи имеют практическую ценность для специалистов в области аффективных вычислений, интерфейсного дизайна и когнитивных технологий.

Ключевые слова: аффективные вычисления, распознавание эмоций, анализ видеоданных, DeepFace, языковая модель GPT-4o, гибридная система анализа, семантический анализ текста, мультимодальное взаимодействие, интерпретируемость нейросетей, когнитивные технологии.

Для цитирования: Агамиров Л.В., Агамиров В.Л., Вестяк В.А., Тутова Н.В., Базунов С.А., Зеляник Ю.Н. Анализ эмоций на видеоданных с использованием локальных и облачных решений искусственного интеллекта. *Моделирование, оптимизация и информационные технологии*. 2025;13(3). URL: <https://moitvvt.ru/ru/journal/pdf?id=1982> DOI: 10.26102/2310-6018/2025.50.3.032

Emotion analysis on video data using on-premise and cloud-based artificial intelligence solutions

L.V. Agamirov^{1,2,3}, V.L. Agamirov^{1,2}, V.A. Vestyak², N.V. Toutova^{1✉}, S.A. Bazunov⁴,
Yu.N. Zelyanik²

¹Moscow Technical University of Communications and Informatics, Moscow,
the Russian Federation

²Moscow Aviation Institute, Moscow, the Russian Federation

³National Research University «MPEI», Moscow, the Russian Federation

⁴BRICS University, Moscow, the Russian Federation

Abstract. The relevance of the study is due to the growing need for a highly accurate and interpretable emotion recognition system based on video data, which is crucial for the development of human-centered technologies in education, medicine, and human-computer interaction systems. In this regard, the article aims to identify the differences and application prospects of the local DeepFace solution and the cloud-based GPT-4o (OpenAI) model for analyzing short video clips with emotional expressions. Methodologically, the study is based on empirical comparative analysis: a moving average method was used to smooth the time series of emotional assessments and to evaluate stability and cognitive interpretability. The results showed that DeepFace provides stable local processing and high resistance to artifacts, while GPT-4o demonstrates the ability for complex semantic interpretation and high sensitivity to context. The effectiveness of a hybrid approach combining computational autonomy and interpretative flexibility is substantiated. Thus, the synergy of local and cloud solutions opens up prospects for creating more accurate, adaptive, and scalable affective analysis systems. The materials of the article are of practical value to specialists in the fields of affective computing, interface design, and cognitive technologies.

Keywords: affective computing, emotion recognition, video data analysis, DeepFace, GPT-4o language model, hybrid analysis system, semantic text analysis, multimodal interaction, neural network interpretability, cognitive technologies.

For citation: Agamirov L.V., Agamirov V.L., Vestyak V.A., Toutova N.V., Bazunov S.A., Zelyanik Yu.N. Emotion analysis on video data using on-premise and cloud-based artificial intelligence solutions. *Modeling, Optimization and Information Technologies*. 2025;13(3). URL: <https://moitvvt.ru/journal/pdf?id=1982> DOI: 10.26102/2310-6018/2025.50.3.032

Введение

Современные технологии искусственного интеллекта всё глубже проникают в повседневную жизнь, в том числе в задачи анализа человеческих эмоций. Автоматическое определение эмоционального состояния по изображениям и видеопотоку открывает новые возможности в системах безопасности, маркетинге, телемедицине, образовании и человеко-компьютерном взаимодействии. Особенно востребовано распознавание по визуальным данным с камер наблюдения, веб-камер и мобильных устройств, где критичны минимальная задержка и простота интеграции в существующие программные продукты. Существующие локальные решения [1] (на примере DeepFace с моделью FER2013 mini-XCEPTION¹) обеспечивают автономность и высокую скорость обработки, но ограничены объёмом предобученных данных и требуют регулярного обновления моделей. Облачные сервисы (через OpenAI API) предлагают более точный мультимодальный анализ и гибкие форматы ответа, однако зависят от стабильного подключения и имеют переменные задержки и затраты. В этой работе представлен гибридный программный прототип, объединяющий локальный модуль на базе DeepFace² и облачный компонент OpenAI API: исследуются архитектурные подходы, методы интеграции, показатели точности и производительности, а также перспективы применения в реальном времени и при ограниченных ресурсах.

Материалы и методы

На современном этапе существует множество программных и аппаратных решений, направленных на распознавание эмоций по видео. Эти решения условно можно разделить на локальные (работающие на пользовательском устройстве) и облачные (использующие удалённые вычислительные ресурсы через API).

¹ FER-2013. Kaggle. URL: <https://www.kaggle.com/datasets/msambare/fer2013> (дата обращения: 23.05.2025).

² serengil. deepface. GitHub. URL: <https://github.com/serengil/deepface> (дата обращения: 14.06.2025).

Одним из наиболее известных инструментов для распознавания эмоций является библиотека DeepFace, включающая предобученные модели, способные классифицировать эмоции по выражению лица. В частности, в её составе реализована модель FER2013 mini-XCEPTION, обученная на открытом датасете, содержащем метки таких эмоций, как «злость», «страх», «отвращение», «грусть», «счастье», «удивление» и «нейтральное состояние».

К числу основных преимуществ DeepFace относится возможность автономной (оффлайн) работы, что позволяет использовать её в системах без подключения к сети. Кроме того, при наличии графических процессоров (GPU) обеспечивается высокая скорость обработки. Немаловажно и то, что библиотека отличается простотой интеграции с такими популярными инструментами, как OpenCV³ и другими средствами Python-экосистемы.

Помимо DeepFace, существуют и другие локальные решения. Так, проект OpenFace, распространяемый с открытым исходным кодом, ориентирован на трекинг лицевых выражений. Нейросеть EmotionNet разработана с прицелом на устойчивый анализ эмоций в условиях естественного освещения, что делает её пригодной для применения в неконтролируемых средах. Библиотека Dlib⁴ предоставляет инструменты для извлечения ключевых признаков лица (landmarks), которые могут использоваться в задачах классификации эмоционального состояния.

Тем не менее, локальные решения обладают и определёнными ограничениями. Прежде всего, они требуют использования предобученных моделей, что может накладывать ограничения на гибкость и адаптивность системы. Кроме того, по точности они зачастую уступают более современным архитектурам, построенным на основе трансформеров.

Интеграция с внешними интеллектуальными сервисами позволяет разработчикам быстро внедрять функциональность анализа эмоций без необходимости создания собственных моделей. Например, Azure Emotion API⁵, входящий в состав Microsoft Cognitive Services, предоставляет оценку по нескольким базовым эмоциям. Сервис Google Cloud Vision API ориентирован на универсальный анализ изображений и способен извлекать лицевые признаки, полезные для анализа выражений. Также стоит отметить OpenAI API, представляющий собой универсальный мультимодальный интерфейс: при помощи модели GPT-4o он способен анализировать изображения и предоставлять текстовое описание эмоционального состояния, что открывает возможности для гибкой интерпретации и диалога.

В частности, использование OpenAI API позволяет построить диалоговую систему, принимающую изображение на вход и возвращающую текстовое описание эмоции. Ключевыми преимуществами данного подхода являются актуальность и мощность модели, а также возможность задания конкретного формата ответа. Вместе с тем, облачные решения имеют и ряд недостатков. Во-первых, они требуют стабильного интернет-соединения. Во-вторых, стоимость использования таких сервисов может значительно увеличиваться при росте числа запросов. В-третьих, сохраняются потенциальные риски, связанные с конфиденциальностью передаваемых данных.

Распознавание эмоций представляет собой междисциплинарную область, объединяющую знания из психологии, нейробиологии и компьютерных наук. Основой подобных систем выступает модель эмоций, на которую обучаются алгоритмы

³ OpenCV – Open Computer Vision Library. URL: <https://opencv.org/> (дата обращения: 14.06.2025).

⁴ davisking, dlib. GitHub. URL: <https://github.com/davisking/dlib> (дата обращения: 14.06.2025).

⁵ Azure AI Vision Documentation. Microsoft Learn. URL: <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/> (дата обращения: 23.05.2025).

искусственного интеллекта. Одной из наиболее известных и широко применяемых является модель Пола Экмана [2], выделяющая шесть базовых эмоций: радость (*happy*), печаль (*sad*), гнев (*angry*), страх (*fear*), отвращение (*disgust*) и удивление (*surprise*). Позднее в перечень была добавлена нейтральная эмоция, необходимая для повышения точности классификации в реальных условиях. Эти эмоции считаются универсальными и проявляются схожим образом у представителей разных культур, что делает их удобными для компьютерного анализа.

На физиологическом уровне эмоции выражаются через микромимику и напряжение лицевых мышц, изменение выражения глаз, бровей и губ, а также особенности позы и наклона головы. В более продвинутых системах используется мультимодальный анализ, учитывающий все эти признаки. При обработке видеопотока особое значение приобретает анализ лицевых признаков. Он включает определение ключевых точек лица (*landmarks*), построение векторных представлений лица (*face embeddings*) и вычисление плотностей вероятностей различных эмоциональных состояний [3].

Современные системы распознавания эмоций основаны на методах машинного и глубокого обучения. Центральную роль играют сверточные нейронные сети (CNN) [4], как основной инструмент обработки изображений. Одним из примеров таких архитектур является Xception, применяемая в модели FER2013 mini-XCEPTION. Также активно внедряются трансформерные архитектуры и мультимодальные модели, такие как GPT-4o от OpenAI, которые объединяют текстовые и визуальные данные, обеспечивая более точную классификацию за счёт использования семантического контекста.

Существуют три основных подхода к классификации эмоций [5, 6]. Первый – это классификация по изображению, при которой система получает отдельный кадр с лицом и возвращает одну из эмоций. Второй подход — регрессия эмоций, при которой вместо одной категории система возвращает вероятностное распределение по всем возможным эмоциям. Третий метод – темпоральный анализ, в котором учитывается последовательность кадров, что позволяет сглаживать случайные колебания и кратковременные проявления эмоций, например, с применением оконного фильтра. Совмещение всех трёх подходов обеспечивает более устойчивую и достоверную оценку эмоционального состояния [7].

Для реализации локального подхода в данной работе была выбрана библиотека DeepFace, предоставляющая функциональность распознавания лиц и определения эмоций с использованием предобученных моделей. Такой подход позволяет выполнять анализ эмоций на стороне клиента, без передачи изображений в облачные сервисы, что особенно важно с точки зрения обеспечения конфиденциальности данных и автономной работы системы.

Реализация включает несколько ключевых компонентов. Сначала осуществляется захват видеопотока с камеры с помощью OpenCV (Листинг 1). Затем выполняется предобработка изображений, включая нормализацию, изменение размеров и цветовое преобразование [3]. Далее производится анализ эмоций с использованием DeepFace, а полученные результаты проходят этап сглаживания для повышения стабильности выводов (Листинги 2 и 3). После этого производится визуализация результата непосредственно на изображении с помощью PIL⁶, включая отрисовку текста с поддержкой кириллицы. Наконец, система обеспечивает управление пользовательским интерфейсом и корректное завершение работы.

⁶ Clark J.A. Python Imaging Library (Fork). PyPI. URL: <https://pypi.org/project/pillow/> (дата обращения: 14.06.2025).

Python

```
camera = cv2.VideoCapture(0)
frame = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
frame_pil = Image.fromarray(frame)
```

Листинг 1 – Захват и предварительная обработка кадра
 Listing 1 – Frame capture and preprocessing

Для определения эмоций используется метод `DeepFace.analyze`, который возвращает вероятностное распределение эмоций и координаты лица на изображении. Это позволяет не только классифицировать эмоцию, но и выделить соответствующую область (Листинг 2).

Python

```
results = DeepFace.analyze(frame, actions=['emotion'], enforce_detection=False)
```

Листинг 2 – Анализ эмоций с помощью DeepFace
 Listing 2 – Emotion analysis using DeepFace

Чтобы избежать резких скачков эмоций между кадрами (эффект «дёргания»), реализован класс сглаживания на основе скользящего окна, усредняющего вероятности эмоций по последним N кадрам (Листинг 3).

Python

```
class EmotionSmoother:
    def __init__(self, window_size=5):
        self.history = []

    def add_emotion(self, probs):
        self.history.append(probs)
        if len(self.history) > self.window_size:
            self.history.pop(0)

    def get_smoothed_emotion(self):
        if not self.history:
            return None
        avg_probs = {k: sum(d[k] for d in self.history)/len(self.history) for k in self.history[0]}
        return max(avg_probs, key=avg_probs.get)
```

Листинг 3 – Класс сглаживания эмоций
 Listing 3 – Emotion smoothing class

На кадре отображается прямоугольник вокруг лица и надпись с текущей (сглаженной) эмоцией. Выход из программы осуществляется нажатием клавиши 'q'.

Современные облачные сервисы предоставляют широкие возможности для анализа изображений и видео. В данной работе используется OpenAI API, обладающий возможностью обработки изображений и генерации текстовых ответов с помощью мультимодели GPT-4o, объединяющей визуальные и языковые возможности.

Облачная реализация состоит из нескольких этапов. Сначала осуществляется захват кадра с камеры. Затем изображение кодируется в формат base64. Полученный код включается в запрос к OpenAI API вместе с параметрами, определяющими формат и содержание ожидаемого ответа. После обработки сервер возвращает текстовое описание с указанием распознанной эмоции. Полученные данные визуализируются на

видеопотоке. Этот процесс повторяется каждые 0,5 секунды, обеспечивая обновление результатов в режиме, близком к реальному времени.

Для использования OpenAI API требуется получение персонального ключа (API_KEY), который указывается в настройках клиентской части системы (Листинг 4).

Python

```
from openai import OpenAI
client = OpenAI(api_key="ВАШ_API_КЛЮЧ")
```

Листинг 4 – Инициализация клиента OpenAI API
Listing 4 – OpenAI API client initialization

Запрос содержит:

- текстовую инструкцию (например, «Определи эмоцию на изображении, ответь одним словом»);
- изображение в base64;
- указание модели gpt-4o-mini или gpt-4o (Листинг 5).

Python

```
response = client.chat.completions.create(
    model="gpt-4o",
    messages=[
        {
            "role": "user",
            "content": [
                {"type": "text", "text": "Какие эмоции выражает человек на этом изображении? Ответ одним словом: счастлив, грустен, злой, нейтральный или удивленный."},
                {"type": "image_url", "image_url": {"url": f"data:image/jpeg;base64,{image_base64}"}}
            ],
        },
    ],
)
emotion = response.choices[0].message.content
```

Листинг 5 – Отправка изображения и получение эмоции через OpenAI API
Listing 5 – Image request and emotion extraction via OpenAI API

После получения эмоции от API, результат отображается на видео с использованием PIL. Текст рисуется на экране, обновляясь каждый 0,5 секунды.

Результаты

Локальные и облачные системы распознавания эмоций обладают различными характеристиками, и выбор между ними зависит от условий применения, доступных ресурсов и требований к системе (Таблица 1).

Таблица 1 – Сравнение характеристик локального и облачного решений
Table 1 – Comparison of local and cloud-based solutions

Параметр	DeepFace (локально)	GPT-4o (облако)
Средняя задержка, мс	7,7 ± 0,8 (p90 = 8,5)	1171 ± 389 (p90 = 1782)
Топ-1 accuracy, %	66,14	87,14
Macro-F1	0,525	0,875
Стоимость за 1000 кадров*	0 \$	≈ 0,48 \$

* Расчёт стоимости: тариф Vision модель GPT-4o (≈ 0,00048 \$ / изобр. ≤ 1 Мп) × 1000.

Для оценки точности обеих систем использовали датасет СК+ в репозитории Kaggle⁷ – 981 кадр-апекс 48×48 px, размеченный по семи базовым эмоциям. Метриками служили Top-1 accuracy (доля кадров, в которых алгоритм безошибочно определил эмоцию) и Macro-F1, уравнивающая вклад всех классов, чтобы проверить, насколько модель одинаково хорошо работает и с частыми, и с редкими эмоциями. Скорость характеризовалась средней задержкой, мс – временем от поступления кадра до получения результата. Кроме этих количественных показателей оценивались зависимость от сети, затраты на внедрение и эксплуатацию, гибкость архитектуры и масштабируемость, простота интеграции и уровень конфиденциальности пользовательских данных (Таблица 2). Все замеры проведены на ноутбуке Ryzen 5 5600H/ 16 ГБ RAM без дискретного GPU: DeepFace работал локально на CPU, а GPT-4o Vision вызывался через однопоточные HTTP-запросы при скорости канала 10 Мбит/с.

Таблица 2 – Оценка по ключевым критериям DeepFace и OpenAI API
Table 2 – Evaluation of DeepFace and OpenAI API by key criteria

Критерий	DeepFace (локально)	OpenAI API (облачно)
Зависимость от интернета	Отсутствует	Полная
Расходы	Бесплатно (открытый код)	Тарифицируется по запросам
Гибкость	Возможность дообучения, изменения модели	Зависимость от формата запроса и API
Простота интеграции	Требует настройки модели и среды	Прост в использовании, минимум кода
Конфиденциальность данных	Обработка локально, не требует передачи данных	Требует отправки изображений на удалённые сервера
Масштабируемость	Ограничена ресурсами устройства	Неограниченная (автоматически на стороне провайдера)
Обновление модели	Ручное (перекачать веса)	Автоматическое на стороне API

Локальное решение на базе DeepFace обладает рядом существенных преимуществ. Прежде всего, его можно использовать в оффлайн-режиме, что особенно важно при работе в условиях ограниченного доступа к интернету. При наличии графического процессора (GPU) достигается высокая скорость обработки. Кроме того, система отличается относительной простотой развёртывания и интеграции в существующие приложения.

Однако существуют и ограничения. Точность классификации эмоций у таких решений, как правило, ниже по сравнению с трансформерными моделями. Пользователю необходимо вручную загружать и инициализировать предобученные модели. Кроме того, DeepFace поддерживает ограниченное количество категорий эмоций, что снижает гибкость системы в более сложных сценариях [8].

Для повышения точности локальной модели можно рассмотреть два наиболее эффективных мероприятия. Во-первых, выполнить тонкую настройку (fine-tuning) на собственной выборке из ≈ 2–3 тыс. размеченных изображений с базовыми аугментациями (mixup, цветовые преобразования), что может увеличить Top-1 accuracy на 6–8 п.п. Во-вторых, заменить используемую в модуле *emotion* лёгкую трёхблочную CNN более мощным, но всё ещё компактным бэкбоном (например, ConvNeXt-T или EfficientNet-V2-S), переобучив его на том же наборе эмоций. Эти архитектуры дают дополнительно 3–4 п.п. прироста точности без заметного влияния на задержку обработки.

⁷ Датасеты СК+. Яндекс Диск. URL: <https://disk.yandex.ru/d/G3lvtYdkVN2V-Q> (дата обращения: 14.06.2025).

Облачное решение на базе OpenAI API, напротив, предоставляет доступ к мультимодели последнего поколения. Оно позволяет выполнять комплексный анализ изображений с учётом контекста и обладает высокой степенью универсальности. Преимущество заключается также в том, что модели регулярно обновляются, что обеспечивает актуальность и повышение качества распознавания.

Тем не менее, такие системы имеют и ограничения. Обработка одного запроса сопровождается задержкой порядка 1–2 секунд. Для функционирования необходим стабильный доступ к сети. Возникают потенциальные риски, связанные с передачей изображений в облачные хранилища. Наконец, использование API может повлечь за собой значительные расходы при росте объёмов запросов.

Обсуждение

Выбор между локальными и облачными подходами напрямую зависит от сценария применения. Локальные системы особенно эффективны в условиях ограниченного или отсутствующего интернет-доступа, а также в задачах, требующих минимальной задержки, таких как системы реального времени [9]. В то же время облачные API предпочтительны для масштабируемых решений, предполагающих высокую точность, мультимодальность и гибкость настройки.

Облачное решение на базе OpenAI API может быть применимо в различных прикладных сценариях. Оно особенно эффективно в чат-ботах с визуальным вводом, где требуется не только распознавание текста, но и обработка изображений. В CRM-системах такие модели способны оценивать эмоциональное состояние клиента на основе его фото, тем самым повышая адаптивность и эффективность коммуникации. На образовательных и консультационных онлайн-платформах мультимодальная модель GPT-4o позволяет учитывать эмоциональный отклик обучающегося или собеседника, адаптируя сценарии взаимодействия в реальном времени.

Анализ практических наблюдений показывает, что в задачах, где критически важна минимальная задержка (например, в системах безопасности, наибольшую ценность представляют локальные решения. В тех же случаях, где главным критерием выступают точность и универсальность, особенно при наличии стабильного интернет-соединения, более эффективными оказываются облачные сервисы. Всё большую популярность получает гибридный подход, при котором локальная предварительная обработка сочетается с облачным анализом как резервным или уточняющим инструментом. Это позволяет достичь высокой адаптивности и устойчивости при различных условиях эксплуатации.

Распознавание эмоций по видеоданным с использованием искусственного интеллекта продолжает оставаться активно развивающимся направлением, находящим применение в разнообразных отраслях. Несмотря на заметный прогресс в области моделей и методов, существует значительный потенциал для дальнейшего совершенствования систем. Одним из перспективных векторов развития является создание гибридных архитектур, сочетающих локальные и облачные ресурсы. Такой подход обеспечивает устойчивую работу даже при нестабильной связи, позволяет гибко подбирать модель под конкретные условия и задачи, а также оптимизирует стоимость эксплуатации и ресурсоёмкость.

Современные исследования всё чаще ориентированы на мультимодальный анализ, объединяющий визуальные, аудиальные и текстовые сигналы [10]. Такой синтез данных позволяет достоверно и всесторонне оценивать эмоциональное состояние пользователя. В системах следующего поколения планируется интеграция анализа

голоса, мимики, жестов и вербальной лексики, что существенно повысит чувствительность алгоритмов и снизит количество ложных классификаций.

Будущее подобных технологий — в персонализированном подходе. Модели, обучающиеся на индивидуальных данных конкретного пользователя, смогут учитывать уникальные особенности его мимики, культурные и этнические различия, возрастные и гендерные параметры. Это приведёт к более высокой точности даже при ограниченных объёмах входных данных.

Потенциальные области применения охватывают широкий спектр задач. В сфере образования такие системы позволяют адаптировать учебные материалы в зависимости от уровня вовлечённости студентов. В медицине и телемедицине они могут быть использованы для раннего выявления признаков депрессии или тревожности. В системах безопасности осуществляется идентификация подозрительного или агрессивного поведения. В маркетинге распознавание эмоций применяется для анализа реакции клиентов на продукты и рекламный контент. Наконец, в HR-сфере возможна оценка невербальной реакции кандидатов во время собеседований.

С развитием технологий особую актуальность приобретает вопрос этической и правовой регламентации. Ключевыми задачами являются защита персональных данных, запрет на несанкционированный сбор видеоинформации, а также обеспечение прозрачности алгоритмов и объяснимости принимаемых решений. Соблюдение этих принципов станет основой для формирования доверия пользователей и успешного внедрения интеллектуальных систем анализа эмоций в социальную и профессиональную практику.

Заключение

Рассмотрены два альтернативных подхода к распознаванию эмоций по видеоданным: локальный – с использованием библиотеки DeepFace, и облачный – с применением мультимодального API GPT-4o от OpenAI. Каждый из подходов реализован в рамках программного прототипа и протестирован по ряду ключевых характеристик: точность распознавания, скорость обработки, устойчивость к условиям съёмки, затраты на эксплуатацию и требования к вычислительной инфраструктуре. Локальная модель DeepFace продемонстрировала устойчивую работу в оффлайн-среде, обеспечив стабильность и автономность анализа, особенно при ограниченном интернет-доступе и строгих требованиях к конфиденциальности данных. Облачная модель GPT-4o показала более высокую точность и способность к сложной семантической интерпретации эмоциональных выражений, но потребовала постоянного сетевого соединения и учета стоимости каждого API-запроса. Проведённый сравнительный анализ с применением метода скользящего среднего подтвердил устойчивость локального подхода и высокую интерпретируемость облачного, что позволило обосновать целесообразность гибридного решения.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Serengil S.I., Ozpinar A. LightFace: A Hybrid Deep Face Recognition Framework. In: *2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 15–17 October 2020, Istanbul, Turkey*. IEEE; 2020. P. 1–5. <https://doi.org/10.1109/ASYU50717.2020.9259802>
2. Razzaq M.A., Hussain J., Bang J., et al. A Hybrid Multimodal Emotion Recognition Framework for UX Evaluation Using Generalized Mixture Functions. *Sensors*. 2023;23(9). <https://doi.org/10.3390/s23094373>
3. Горячкин Б.С., Китов М.А. Компьютерное зрение. *E-Scio*. 2020;(9):317–345.

4. Zhao X., Wang L., Zhang Yu., Han X., Deveci M., Parmar M. A Review of Convolutional Neural Networks in Computer Vision. *Artificial Intelligence Review*. 2024;57(4). <https://doi.org/10.1007/s10462-024-10721-6>
5. Kalateh S., Estrada-Jimenez L.A., Nikghadam-Hojjati S., Barata J. A Systematic Review on Multimodal Emotion Recognition: Building Blocks, Current State, Applications, and Challenges. *IEEE Access*. 2024;12:103976–104019. <https://doi.org/10.1109/ACCESS.2024.3430850>
6. Poria S., Majumder N., Hazarika D., Cambria E., Gelbukh A., Hussain A. Multimodal Sentiment Analysis: Addressing Key Issues and Setting Up the Baselines. *IEEE Intelligent Systems*. 2018;33(6):17–25. <https://doi.org/10.1109/MIS.2018.2882362>
7. Mujiyanto M., Setyanto A., Utami E., Kusri K. Facial Expression Recognition with Deep Learning and Attention Mechanisms: A Systematic Review. In: *2024 7th International Conference on Informatics and Computational Sciences (ICICoS), 17–18 July 2024, Semarang, Indonesia*. IEEE; 2024. P. 12–17. <https://doi.org/10.1109/ICICoS62600.2024.10636857>
8. Тимофеева О.П., Неимушев С.А., Неимушева Л.И., Тихонов И.А. Распознавание эмоций по изображению лица на основе глубоких нейронных сетей. *Труды НГТУ им. Р.Е. Алексева*. 2020;(1):16–24. https://doi.org/10.46960/1816-210X_2020_1_16
Timofeeva O.P., Neimushchev S.A., Neimushcheva L.I., Tikhonov I.A. Facial Emotion Recognition Using Deep Neural Networks. *Trudy NGTU im. R.E. Alekseeva*. 2020;(1):16–24. (In Russ.). https://doi.org/10.46960/1816-210X_2020_1_16
9. Pascual A.M., Valverde E.C., Kim J.-I., et al. Light-FER: A Lightweight Facial Emotion Recognition System on Edge Devices. *Sensors*. 2022;22(23). <https://doi.org/10.3390/s22239524>
10. Барабанщиков В.А., Суворова Е.В. Оценка эмоционального состояния человека по его видеоизображению. *Экспериментальная психология*. 2020;13(4):4–24. <https://doi.org/10.17759/exppsy.2020130401>
Barabanshikov V.A., Suvorova E.V. Human Emotional State Assessment Based on a Video Portrayal. *Experimental Psychology (Russia)*. 2020;13(4):4–24. (In Russ.). <https://doi.org/10.17759/exppsy.2020130401>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Агамиров Левон Владимирович, доктор технических наук, профессор, Национальный исследовательский университет «МЭИ», Московский технический университет связи и информатики, Московский авиационный институт, Москва, Российская Федерация.
e-mail: itno_agamirov@mail.ru
ORCID: [0009-0009-6909-9399](https://orcid.org/0009-0009-6909-9399)

Levon V. Agamirov, Doctor of Engineering Sciences, Professor, National Research University "MPEI", Moscow Technical University of Communications and Informatics, Moscow Aviation Institute, Moscow, the Russian Federation.

Агамиров Владимир Леонович, кандидат технических наук, доцент, Московский технический университет связи и информатики, Московский авиационный институт, Москва, Российская Федерация.
e-mail: avhere@yandex.ru
ORCID: [0000-0001-9181-7726](https://orcid.org/0000-0001-9181-7726)

Vladimir L. Agamirov, Candidate of Engineering Sciences, Associate Professor, Moscow Technical University of Communications and Informatics, Moscow Aviation Institute, Moscow, the Russian Federation.

Вестяк Владимир Анатольевич, доктор физико-математических наук, доцент,

Vladimir A. Vestyak, Doctor of Physical and Mathematical Sciences, Docent, Head of the

заведующий кафедрой «Прикладные программные средства и математические методы», Московский авиационный институт, Москва, Российская Федерация.

email: kaf311@yandex.ru

ORCID: [0009-0001-8496-9438](https://orcid.org/0009-0001-8496-9438)

Department of Applied software tools and mathematical methods, Moscow Aviation Institute, Moscow, the Russian Federation.

Тугова Наталья Владимировна, кандидат технических наук, доцент, заведующий кафедрой «Бизнес-информатика» Московского технического университета связи и информатики, Москва, Российская Федерация.

e-mail: e-natasha@mail.ru

ORCID: [0000-0002-2851-8472](https://orcid.org/0000-0002-2851-8472)

Natalia V. Toutova, Candidate of Engineering Sciences, Docent, Head of the Department of Business Informatics, Moscow Technical University of Communications and Informatics, Moscow, the Russian Federation.

Базунов Сергей Александрович, студент, Университет БРИКС, Москва, Российская Федерация.

e-mail: bazunovsergei01@gmail.com

ORCID: [0009-0007-8392-5402](https://orcid.org/0009-0007-8392-5402)

Sergei A. Bazunov, student, BRICS University, Moscow, the Russian Federation.

Зеляник Юлия Николаевна, студентка, Московский авиационный институт, Москва, Российская Федерация.

email: yuzelyanik@yandex.ru

ORCID: [0009-0008-7002-6704](https://orcid.org/0009-0008-7002-6704)

Yulia N. Zelyanik, student, Moscow Aviation Institute, Moscow, the Russian Federation.

Статья поступила в редакцию 02.06.2025; одобрена после рецензирования 29.07.2025; принята к публикации 06.08.2025.

The article was submitted 02.06.2025; approved after reviewing 29.07.2025; accepted for publication 06.08.2025.