**Бюджетное Учреждение Высшего Образования**

**Ханты-Мансийского автономного округа – Югры**

**«СУРГУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

Политехнический институт

Кафедра автоматизированных систем обработки информации и управления

**Курсовой проект по дисциплине «Проектирование и эксплуатация ИЭС»**

Тема курсовой работы:

**«Проектирование Системы распознавания эмоций человека на основе искусственной нейронной сети»**

Выполнил: студент группы 606-12

Демьянцев Виталий Владиславович

Проверил: Доцент кафедры АСОИУ, к. т. н.

Гавриленко Тарас Владимирович

Сургут, 2025 г.

**СОДЕРЖАНИЕ**

[1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ 7](#_Toc195530195)

[1.1. Эмоции человека и их роль в HR 7](#_Toc195530196)

[1.2. Технологи мультимодального анализа 8](#_Toc195530197)

[1.3. Искусственные нейронные сети 9](#_Toc195530198)

[2. ОБЗОР АНАЛОГОВ 10](#_Toc195530199)

[2.1. Существующие решения 10](#_Toc195530200)

[2.2. Сравнительный анализ аналогов 12](#_Toc195530201)

[2.3. Выводы по обзору аналогов 13](#_Toc195530202)

[3. ТЕХНИЧЕСКОЕ ЗАДАНИЕ 14](#_Toc195530203)

[3.1. Общие сведения 14](#_Toc195530204)

[3.2. Назначение и цели разработки 14](#_Toc195530205)

[3.3. Требования к системе 16](#_Toc195530206)

[3.4. Основные пользователи 19](#_Toc195530207)

[4. ПРОЕКТИРОВАНИЕ СИСТЕМЫ 20](#_Toc195530208)

[4.1 BPMN-диаграмма 20](#_Toc195530209)

[4.2 IDEF0-диаграмма 22](#_Toc195530210)

[4.3 DFD-диаграмма 24](#_Toc195530211)

[4.4 Схема интерфейса 26](#_Toc195530212)

[5. ВИДЫ ОБЕСПЕЧЕНИЯ 29](#_Toc195530213)

[5.1 Лингвистическое обеспечение 29](#_Toc195530214)

[5.2 Математическое обеспечение 30](#_Toc195530215)

[5.3 Алгоритмическое обеспечение 31](#_Toc195530216)

[5.4 Техническое обеспечение 32](#_Toc195530217)

[5.5 Программное обеспечение 33](#_Toc195530218)

[5.6 Правовое обеспечение 34](#_Toc195530219)

[5.7 Эргономическое обеспечение 35](#_Toc195530220)

[5.8 Информационное обеспечение 35](#_Toc195530221)

[5.8.1 . ER-диаграмма базы данных 35](#_Toc195530222)

[5.8.2 Логическая модель базы данных 38](#_Toc195530223)

[ЗАКЛЮЧЕНИЕ 40](#_Toc195530224)

ВВЕДЕНИЕ

В современном мире технологии искусственного интеллекта (ИИ) играют ключевую роль в совершенствовании процессов взаимодействия человека и машины. Одной из перспективных областей применения ИИ является распознавание эмоций человека, которое позволяет создавать более интуитивные и персонализированные системы. Такие технологии находят применение в различных сферах, включая медицину, образование, маркетинг, кибербезопасность и управление персоналом (HR). В контексте HR распознавание эмоций может использоваться для оценки эмоционального состояния сотрудников, повышения эффективности коммуникации, предотвращения профессионального выгорания и оптимизации процессов подбора персонала.

Эмоции человека проявляются через множество каналов, включая выражения лица, интонации голоса, жесты и физиологические показатели. Традиционные методы анализа эмоций, основанные на субъективных оценках или анкетировании, имеют ограничения, связанные с низкой точностью, трудоёмкостью и невозможностью оперативного мониторинга. Современные технологии, такие как компьютерное зрение, обработка аудиосигналов и искусственные нейронные сети (ИНС), открывают новые возможности для автоматизированного распознавания эмоций. Мультимодальные системы, объединяющие анализ видеопотока (выражения лица) и аудиопотока (голосовые характеристики), обеспечивают более точное и комплексное определение эмоционального состояния.

Данная работа посвящена разработке мультимодальной системы распознавания эмоций человека на основе искусственной нейронной сети, ориентированной на применение в области управления персоналом. Система направлена на анализ выражений лица и голосовых характеристик сотрудников для классификации их эмоционального состояния, что позволит повысить эффективность HR-процессов, включая оценку вовлечённости, диагностику стресса и поддержку принятия управленческих решений. Особое внимание уделяется интеграции данных из видеопотока и аудиопотока, что делает задачу междисциплинарной и требующей комбинированного подхода.

**Актуальность темы** обусловлена следующими факторами:

1. **Рост значимости эмоционального интеллекта в HR**: Современные подходы к управлению персоналом подчеркивают важность учёта эмоционального состояния сотрудников для повышения производительности и снижения текучести кадров.
2. **Ограниченность традиционных методов**: Субъективные оценки и анкетирование не обеспечивают высокой точности и оперативности, что затрудняет своевременное выявление проблем.
3. **Развитие технологий ИИ**: Прогресс в области компьютерного зрения, обработки аудиосигналов и глубокого обучения создаёт основу для разработки автоматизированных систем анализа эмоций.
4. **Потребность в мультимодальном анализе**: Комплексный подход, объединяющий видео- и аудиоданные, позволяет учитывать разнообразные аспекты эмоционального состояния, повышая точность классификации.
5. **Экономическая эффективность**: Автоматизация анализа эмоций способствует снижению затрат на HR-процессы, таких как подбор персонала и мониторинг вовлечённости, за счёт оперативного предоставления данных для принятия решений.

**Объект исследования**: процесс распознавания эмоций человека на основе анализа выражений лица и голосовых характеристик.

**Предмет исследования**: мультимодальная система распознавания эмоций на основе искусственной нейронной сети, ориентированная на применение в области управления персоналом.

**Цель исследования**: разработка и исследование мультимодальной системы распознавания эмоций человека, обеспечивающей точную классификацию эмоционального состояния на основе анализа видеопотока и аудиопотока с использованием искусственных нейронных сетей для применения в HR-процессах.

**Задачи исследования**:

1. Исследование предметной области, включая анализ существующих методов и технологий распознавания эмоций.
2. Изучение подходов к проектированию мультимодальных систем, интегрирующих видео- и аудиоданные.
3. Проектирование системы: разработка архитектуры нейронной сети, алгоритмов обработки данных и интерфейса взаимодействия.
4. Анализ эффективности применения свёрточных нейронных сетей (CNN) для обработки видеопотока и рекуррентных нейронных сетей (RNN) для анализа аудиопотока.
5. Разработка программного обеспечения для интеграции данных из видеопотока и аудиопотока.
6. Тестирование системы в смоделированных HR-сценариях с использованием стандартных метрик точности, полноты и F1-меры.
7. Формирование рекомендаций по внедрению системы в HR-процессы предприятий.

**Научная новизна** заключается в разработке мультимодальной системы, интегрирующей анализ видеопотока и аудиопотока для распознавания эмоций, с акцентом на её адаптацию к задачам управления персоналом, что ранее не рассматривалось в аналогичных исследованиях с такой степенью детализации.

**Практическая значимость** работы состоит в создании инструмента, который может быть использован HR-специалистами для повышения эффективности оценки эмоционального состояния сотрудников, оптимизации подбора персонала и предотвращения профессионального выгорания.

Выполнение указанных задач позволит достичь поставленной цели и создать систему, способную повысить качество управления персоналом за счёт автоматизированного анализа эмоций.

# АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

При разработке мультимодальной системы распознавания эмоций человека, ориентированной на применение в области управления персоналом, необходимо тщательно исследовать предметную область. Это позволяет выявить ключевые аспекты, определяющие требования к системе, включая особенности эмоционального состояния, технологии обработки данных и роль искусственных нейронных сетей. В данном разделе рассматриваются основные элементы предметной области: эмоции человека, мультимодальные технологии анализа и архитектуры нейронных сетей.

## Эмоции человека и их роль в HR

Эмоции представляют собой комплексные психофизиологические состояния, которые проявляются через выражения лица, голосовые интонации, жесты и физиологические показатели. В контексте HR эмоции играют важную роль, поскольку влияют на мотивацию, вовлечённость и производительность сотрудников. Основные категории эмоций, такие как радость, грусть, гнев, страх, удивление и отвращение, могут быть дополнены нейтральным состоянием и более сложными эмоциями, такими как стресс или усталость.

В HR-процессах распознавание эмоций может использоваться для:

* **Оценки вовлечённости**: выявление эмоционального состояния во время рабочих встреч или собеседований.
* **Диагностики стресса**: мониторинг признаков профессионального выгорания.
* **Подбора персонала**: анализ реакции кандидатов на вопросы для оценки их соответствия корпоративной культуре.
* **Обучения и развития**: адаптация программ обучения с учётом эмоциональных реакций сотрудников.

Основной задачей является точная классификация эмоций в реальном времени, что требует обработки данных из нескольких источников (видео- и аудиопотоков) для повышения надёжности

## Технологи мультимодального анализа

Для распознавания эмоций используются различные технологии, которые можно разделить на два основных направления: анализ видеопотока и аудиопотока.

* **Компьютерное зрение**: Камеры фиксируют выражения лица, которые анализируются с помощью алгоритмов обработки изображений. Основные признаки включают положение уголков рта, движение бровей и морщины на лбу. Такие системы способны классифицировать эмоции на основе визуальных данных в реальном времени.
* **Обработка аудиосигналов**: Микрофоны записывают голосовые характеристики, такие как высота тона, темп речи и интенсивность. Эти параметры позволяют выявить эмоциональные состояния, такие как гнев (повышенная громкость) или грусть (замедленная речь).
* **Мультимодальный анализ**: Комбинирование видео- и аудиоданных повышает точность распознавания за счёт учёта взаимодополняющих признаков. Например, улыбка в сочетании с радостным тоном голоса подтверждает положительную эмоцию.

Каждая технология имеет свои особенности. Компьютерное зрение эффективно для статического анализа выражений лица, но может быть ограничено условиями освещения. Обработка аудиосигналов устойчива к визуальным помехам, но зависит от качества звука. Мультимодальный подход позволяет минимизировать эти ограничения, обеспечивая более надёжные результаты.

## Искусственные нейронные сети

Искусственные нейронные сети являются основным инструментом для обработки мультимодальных данных. В контексте данной работы ИНС применяются для:

* **Анализа видеопотока**: Свёрточные нейронные сети (CNN) используются для извлечения признаков из изображений лица, таких как контуры губ или положение глаз.
* **Анализа аудиопотока**: Рекуррентные нейронные сети (RNN) или их модификации (LSTM, GRU) обрабатывают временные характеристики аудиосигналов, выявляя изменения интонаций.
* **Интеграции данных**: Полносвязные слои или трансформеры объединяют признаки из видео- и аудиопотоков для финальной классификации эмоций.

Наиболее перспективные архитектуры включают:

* **CNN**: ResNet, VGGNet, MobileNet для анализа изображений.
* **RNN**: LSTM, GRU для обработки временных данных.
* **Трансформеры**: применяются для мультимодального анализа благодаря способности учитывать взаимосвязи между различными типами данных.

Проведённый анализ предметной области показывает, что мультимодальный подход с использованием ИНС позволяет создать систему, способную эффективно решать задачу распознавания эмоций в контексте HR. Это обеспечивает основу для дальнейшего проектирования и разработки.

# ОБЗОР АНАЛОГОВ

В области распознавания эмоций человека существует ряд решений, использующих технологии компьютерного зрения, обработки аудиосигналов и искусственного интеллекта. Такие системы применяются в маркетинге, медицине, образовании и HR для анализа эмоционального состояния. В данном разделе рассматриваются основные аналоги, их функциональные возможности, преимущества и недостатки, а также проводится сравнительный анализ для определения перспективного подхода к разработке системы.

## Существующие решения

1. **Системы на основе компьютерного зрения**

Примером является библиотека Affectiva, которая использует свёрточные нейронные сети для анализа выражений лица. Такие системы фиксируют видеопоток и классифицируют эмоции на основе визуальных признаков, таких как улыбка или нахмуренные брови.

**Преимущества**:

* Высокая точность распознавания при хорошем освещении.
* Возможность анализа в реальном времени.
* Гибкость настройки под конкретные категории эмоций.

**Недостатки**:

* Зависимость от качества видеопотока и угла обзора.
* Ограниченная эффективность при обработке сложных или неоднозначных выражений.
* Отсутствие учёта аудиоданных.

1. **Системы на основе анализа аудиосигналов**

Примером служит платформа Beyond Verbal, которая анализирует голосовые характеристики, такие как высота тона и темп речи, для определения эмоционального состояния. Такие решения часто применяются в call-центрах.

**Преимущества**:

* Устойчивость к визуальным помехам.
* Низкие требования к оборудованию (достаточно микрофона).
* Возможность анализа в условиях ограниченной видимости.

**Недостатки**:

* Ограниченная точность при низком качестве звука.
* Невозможность учёта визуальных признаков.
* Сложности с распознаванием эмоций при нейтральном тоне.

1. **Мультимодальные системы**

Примером является система Emotient, которая комбинирует анализ выражений лица и голосовых данных с использованием CNN и RNN. Такие решения применяются в маркетинге и HR для комплексного анализа эмоционального состояния.

**Преимущества**:

* Высокая точность за счёт комбинирования данных.
* Адаптивность к различным сценариям (например, собеседования или рабочие встречи).
* Возможность интеграции с другими системами управления.

**Недостатки**:

* Высокая вычислительная сложность.
* Необходимость больших размеченных датасетов для обучения.
* Сложность настройки для специфических задач HR.

1. **Системы на основе традиционных методов**

Примером являются решения, использующие алгоритмы SVM или KNN для классификации эмоций на основе заранее выделенных признаков (например, геометрия лица или спектральные характеристики голоса).

**Преимущества**:

* Низкие вычислительные требования.
* Простота реализации для базовых задач.
* Независимость от больших датасетов.

**Недостатки**:

* Низкая точность по сравнению с нейронными сетями.
* Трудоёмкость предварительной обработки данных.
* Ограниченная адаптивность к сложным сценариям.

## Сравнительный анализ аналогов

Для оценки существующих решений ниже приведена таблица, в которой сравниваются ключевые характеристики систем по основным критериям.

Таблица 1. Сравнение аналогов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Функциональная возможность | Компьютерное зрение | Анализ аудиосигналов | Мультимодальные системы | Традиционные методы |
| Точность классификации | Высокая | Средняя | Высокая | Низкая |
| Учёт видеопотока | Да | Нет | Да | Ограниченно |
| Учёт аудиопотока | Нет | Да | Да | Ограниченно |
| Адаптивность к HR-сценариям | Средняя | Низкая | Высокая | Низкая |
| Вычислительная сложность | Высокая | Средняя | Высокая | Низкая |
| Требования к данным | Высокие | Средние | Высокие | Средние |
| Стоимость внедрения | Высокая | Средняя | Высокая | Низкая |

* **Точность классификации**: Мультимодальные системы обеспечивают наивысшую точность благодаря комбинированию данных, тогда как традиционные методы уступают из-за ограничений в обработке сложных признаков.
* **Учёт видео- и аудиопотока**: Только мультимодальные системы способны полноценно анализировать оба типа данных.
* **Адаптивность к HR-сценариям**: Мультимодальные системы лучше подходят для задач HR благодаря возможности учёта контекста и комплексного анализа.
* **Вычислительная сложность**: Компьютерное зрение и мультимодальные системы требуют мощных вычислительных ресурсов, в отличие от аудиоанализа и традиционных методов.
* **Стоимость**: Традиционные методы являются наиболее экономичными, тогда как мультимодальные системы требуют значительных вложений.

## Выводы по обзору аналогов

Анализ существующих решений показывает, что ни одна из технологий в отдельности не решает задачу комплексного распознавания эмоций с учётом требований HR. Системы на основе компьютерного зрения эффективны для анализа выражений лица, но игнорируют голосовые характеристики. Системы анализа аудиосигналов устойчивы к визуальным помехам, но ограничены в распознавании сложных эмоций. Традиционные методы просты в реализации, но уступают по точности. Мультимодальные системы демонстрируют наибольший потенциал, объединяя преимущества видео- и аудиоанализа, однако их внедрение связано с высокой сложностью и стоимостью.

Для разработки системы в рамках данной работы целесообразно ориентироваться на мультимодальный подход, интегрирующий свёрточные нейронные сети (CNN) для обработки видеопотока и рекуррентные нейронные сети (RNN) для анализа аудиопотока. Это позволит достичь высокой точности, надёжности и функциональности, необходимых для применения в HR-процессах.

# ТЕХНИЧЕСКОЕ ЗАДАНИЕ

## Общие сведения

* + 1. Наименование темы (проекта)

Полное наименование: «Разработка мультимодальной системы распознавания эмоций человека на основе искусственной нейронной сети для применения в процессах управления персоналом».

* + 1. Основания для разработки
* Дипломная работа в рамках Сургутского государственного университета.
* Письменное задание научного руководителя от «01» марта 2025 г.
* Цель разработки — создание системы, обеспечивающей точное распознавание эмоционального состояния человека на основе анализа видеопотока (выражения лица) и аудиопотока (голосовые характеристики) с использованием искусственных нейронных сетей для поддержки HR-процессов, таких как оценка вовлечённости, диагностика стресса и оптимизация подбора персонала.
  + 1. Исполнители
* Студент: Демьянцев Виталий Владиславович, группа 606-12.
* Научный руководитель: Горбунов Дмитрий Владимирович, Старший преподаватель кафедры АСОИУ.
  + 1. Дата начала и дата окончания работ
* Начало работ: «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2025 г.
* Окончание работ: «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2025 г.

## Назначение и цели разработки

* + 1. Назначение

Разрабатываемая мультимодальная система предназначена для автоматического распознавания эмоций человека на основе анализа видеопотока (выражения лица) и аудиопотока (голосовые характеристики) в реальном времени. Система ориентирована на применение в области управления персоналом (HR) для решения задач, связанных с оценкой эмоционального состояния сотрудников, диагностикой стресса, анализом вовлечённости и поддержкой подбора персонала. Решение направлено на повышение эффективности HR-процессов, улучшение корпоративной культуры и снижение рисков профессионального выгорания.

* + 1. Цели и задачи

**Главная цель**: разработка и исследование мультимодальной системы распознавания эмоций человека, обеспечивающей точную классификацию эмоционального состояния на основе интеграции видео- и аудиоданных с использованием искусственных нейронных сетей для применения в HR-процессах.

**Основные задачи**:

1. Исследование предметной области и анализ существующих методов распознавания эмоций.
2. Определение требований к системе и выбор подходящих инструментов для обработки видео- и аудиопотоков.
3. Проектирование архитектуры системы, включая интеграцию свёрточных и рекуррентных нейронных сетей.
4. Разработка программного обеспечения для обработки видеопотока (анализ выражений лица) и аудиопотока (анализ голосовых характеристик).
5. Реализация алгоритмов классификации эмоций на основе мультимодальных данных.
6. Тестирование системы в смоделированных HR-сценариях (например, имитация собеседований или рабочих встреч) с оценкой точности, полноты и F1-меры.
7. Формирование отчёта с результатами тестирования и рекомендациями по внедрению системы в HR-процессы.

## Требования к системе

* + 1. Функциональные требования

1. **Обработка видеопотока**:
   * Захват и анализ видеопотока с камеры для распознавания выражений лица.
   * Классификация эмоций (радость, грусть, гнев, страх, удивление, нейтральное состояние и др.) на основе визуальных признаков.
2. **Обработка аудиопотока**:
   * Захват и анализ аудиопотока с микрофона для извлечения голосовых характеристик (высота тона, темп речи, интенсивность).
   * Классификация эмоций на основе аудиоданных с учётом интонаций.
3. **Мультимодальная интеграция**:

Объединение данных из видео- и аудиопотоков для повышения точности классификации.

* + Генерация единого результата классификации эмоций с указанием вероятности каждой категории.

1. **Классификация эмоций**:
   * Поддержка не менее 7 категорий эмоций (радость, грусть, гнев, страх, удивление, отвращение, нейтральное состояние).
   * Возможность расширения для учёта дополнительных состояний (например, стресс, усталость).
2. **Визуализация и уведомления**:
   * Отображение текущего эмоционального состояния в реальном времени на пользовательском интерфейсе.
   * Генерация уведомлений при выявлении критических состояний (например, высокий уровень стресса).
3. **Хранение данных**:
   * Ведение базы данных с результатами анализа (время, эмоция, вероятность).
   * Возможность экспорта отчётов в форматах PDF и CSV для HR-специалистов.
4. **Интеграция с HR-процессами**:
   * Поддержка сценариев, таких как анализ эмоций во время собеседований, рабочих встреч или мониторинга состояния сотрудников.
   * Возможность настройки под конкретные HR-задачи (например, оценка вовлечённости).
     1. Нефункциональные требования

* Производительность:
* Время обработки одного кадра видеопотока — не более 100 мс.
* Время обработки аудиофрагмента (5 секунд) — не более 200 мс.
* Поддержка анализа в реальном времени (не менее 15 кадров в секунду для видео).
* Надёжность:
* Устойчивость к помехам (например, низкое освещение для видео, шум для аудио).
* Обеспечение корректной работы при частичной потере данных (например, отсутствие аудиопотока).
* Масштабируемость:
* Возможность обработки данных с нескольких камер и микрофонов одновременно.
* Поддержка добавления новых категорий эмоций без изменения архитектуры.
* Интерфейс:
* Интуитивно понятный интерфейс для HR-специалистов с минимальными требованиями к обучению.
* Поддержка русского и английского языков.
* Конфиденциальность:
* Обеспечение защиты персональных данных сотрудников в соответствии с Федеральным законом №152-ФЗ «О персональных данных».
* Шифрование хранимых данных и ограничение доступа к результатам анализа.
* Точность:
* Точность классификации эмоций — не менее 85% на стандартных датасетах (например, FER-2013 для видео, RAVDESS для аудио).
* F1-мера для мультимодальной классификации — не менее 0.80.
  + 1. Требования к программному обеспечению
* **Языки программирования**: Python (для нейронных сетей и обработки данных).
* **Среда разработки**: PyCharm или Jupyter Notebook.
* **Библиотеки**:
* PyTorch или TensorFlow (реализация нейронных сетей).
* OpenCV (обработка видеопотока).
* Librosa (обработка аудиопотока).
* Pandas, NumPy (анализ данных).
* Flask или Django (разработка веб-интерфейса, при необходимости).
* **СУБД**: SQLite или PostgreSQL для хранения результатов анализа и логов.
* **Датасеты**:
* Для видеопотока: FER-2013, AffectNet.
* Для аудиопотока: RAVDESS, TESS.
  + 1. Требования к аппаратному обеспечению
* **Камера**: Разрешение не менее 720p, частота кадров — 30 fps.
* **Микрофон**: Частота дискретизации — не менее 16 кГц, подавление шума.
* Вычислительная **платформа**:
* Процессор: 4 ядра (Intel i5 или выше) или GPU (например, NVIDIA GTX 1660 для ускорения нейронных сетей).
* Оперативная память: 16 ГБ.
* Накопитель: SSD 256 ГБ.
* **Операционная система**: Windows 10/11 или Linux (Ubuntu 20.04 или выше).

## Основные пользователи

* **HR-специалисты**: Персонал, использующий систему для оценки эмоционального состояния сотрудников, анализа вовлечённости и поддержки подбора персонала.
* **Руководители подразделений**: Лица, принимающие решения на основе данных об эмоциональном климате в коллективе.
* **Технические специалисты**: Инженеры и разработчики, ответственные за настройку, обслуживание и обновление системы.

# ПРОЕКТИРОВАНИЕ СИСТЕМЫ

Для проектирования мультимодальной системы распознавания эмоций человека необходимо построить BPMN-диаграмму для описания бизнес-процесса, DFD-диаграмму для отображения потоков данных, IDEF0-диаграмму для представления общей структуры системы, ER-диаграмму для описания структуры базы данных, а также схему интерфейса для определения концепции взаимодействия с пользователем.

## BPMN-диаграмма

Первым этапом в проектировании системы является построение BPMN диаграммы. BPMN-диаграмма отражает детальное описание бизнес-процессов. Диаграмма представлена ниже (рис. 1).

Далее представлена таблица обозначений нотации BPMN:

Таблица 2. Обозначения в нотации BPMN

| Объект | Описание |
| --- | --- |
|  | Начало работы системы |
|  | Конец работы системы |
|  | Процесс |
|  | Хранилище данных |
|  | Комментарий к процессу |
|  | Обозначение глобального процесса |
|  | Документ/отчет |
|  | Условное разветвление бизнес-процесса |

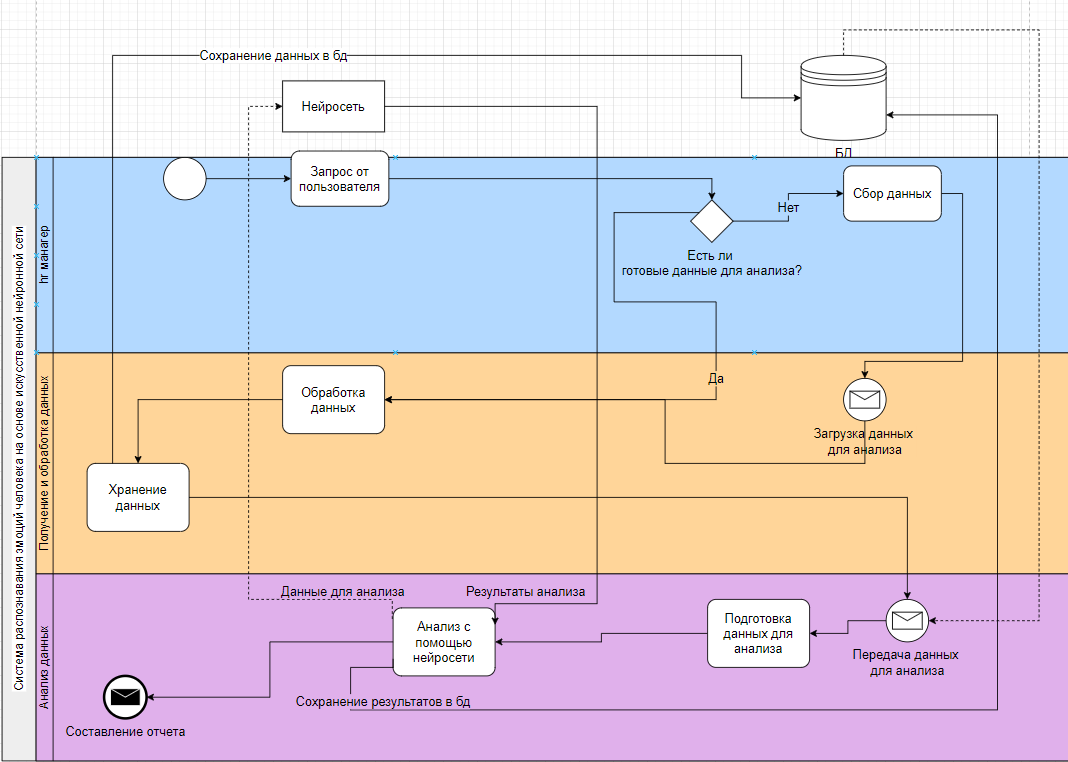


Рисунок 1. BPMN диаграмма системы

Глобально бизнес-процесс системы можно разделить на три модуля: взаимодействие с HR-менеджером, получение и обработка данных, анализ данных.

1. **Взаимодействие с HR-менеджером**:
   * Процесс начинается с запроса HR-менеджера на анализ эмоционального состояния кандидата или сотрудника. HR-менеджер инициирует систему, проверяя наличие готовых данных для анализа (например, ранее сохранённые видео- и аудиозаписи).
   * Если данные отсутствуют, система переходит к сбору новых данных. Если данные есть, они загружаются из базы данных для обработки.
2. **Получение и обработка данных**:
   * При необходимости сбора новых данных система захватывает видеопоток (изображения лица) и аудиопоток (голос) с помощью камеры и микрофона. Данные проходят предварительную обработку: нормализация видео, фильтрация шума в аудио.
   * Обработанные данные сохраняются в базу данных и передаются в модуль анализа. Если используются готовые данные, они загружаются из базы данных и также передаются на анализ.
3. **Анализ данных**:
   * Данные обрабатываются нейронной сетью. Видеопоток анализируется для распознавания выражений лица, а аудиопоток — для определения голосовых характеристик (например, высоты тона, темпа речи).
   * Результаты анализа (классифицированные эмоции с вероятностями) сохраняются в базу данных.
   * На основе результатов формируется отчёт, содержащий информацию об эмоциональном состоянии (например, процентное распределение эмоций за сеанс).

Диаграмма отражает последовательность процессов, взаимодействие с базой данных и нейросетью, а также условные развилки, обеспечивая полное представление о работе системы.

## IDEF0-диаграмма

Следующим этапом проектирования является построение IDEF0-диаграммы, которая предоставляет общее описание структуры системы на верхнем уровне. Диаграмма представлена ниже (рис. 2).

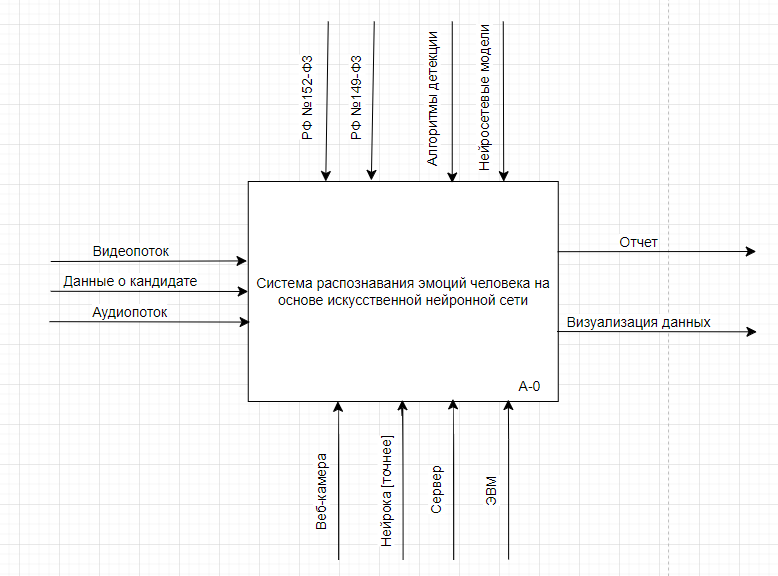


Рисунок 2. IDEF0диаграмма системы

IDEF0-диаграмма описывает систему как единый процесс «Распознавание эмоций человека» с входами, выходами, управляющими факторами и механизмами:

* **Входы**:
  + Видеопоток (изображения лица с камеры).
  + Аудиопоток (голосовые данные с микрофона).
* **Выходы**:
  + Классифицированные эмоции (например, «Радость, 90%»).
  + Отчёты об эмоциональном состоянии.
* **Управляющие факторы**:
  + Параметры анализа (категории эмоций, сценарий HR, например, собеседование).
  + Модели нейронных сетей (CNN для видео, RNN для аудио).
* **Механизмы**:
  + Аппаратное обеспечение (камера, микрофон, сервер).
  + Программное обеспечение (PyTorch, OpenCV, Librosa).
  + Пользователь (HR-менеджер).

Диаграмма демонстрирует, как входные данные преобразуются в результаты анализа с использованием заданных механизмов и под управлением параметров, заданных пользователем.

## DFD-диаграмма

На следующем этапе проектирования построена DFD-диаграмма, которая отображает потоки данных между компонентами системы и хранилищами. Диаграмма представлена ниже (рис. 3).

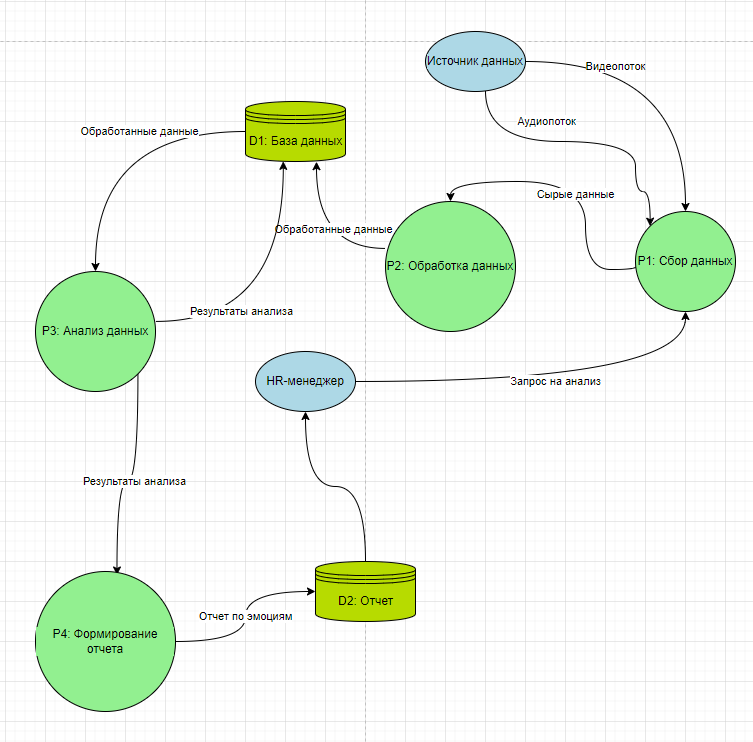


Рисунок 3. DFD диаграмма системы

DFD-диаграмма включает следующие компоненты:

* **Внешние сущности**:
  + **HR-менеджер**: Инициирует запрос на анализ и получает отчёт.
  + **Источник данных**: Камера и микрофон, предоставляющие видео- и аудиопотоки.
* **Процессы**:
  + **P1: Сбор данных**: Получение видео- и аудиопотоков, их предварительная обработка (например, сжатие или фильтрация).
  + **P2: Обработка данных**: Извлечение признаков из видео (например, ключевые точки лица) и аудио (например, MFCC-коэффициенты).
  + **P3: Анализ данных**: Классификация эмоций с использованием нейронной сети на основе обработанных данных.
  + **P4: Формирование отчёта**: Создание отчёта с результатами анализа (например, статистика эмоций).
* **Хранилища данных**:
  + **D1: База данных**: Хранит сырые данные, обработанные данные, результаты анализа.
  + **D2: Отчёт**: Содержит итоговый отчёт, доступный HR-менеджеру.
* **Потоки данных**:
  + Видео- и аудиопотоки от источника к процессу сбора.
  + Сырые данные от сбора к обработке.
  + Обработанные данные от обработки к базе данных и анализу.
  + Результаты анализа от анализа к базе данных и формированию отчёта.
  + Отчёт от формирования к HR-менеджеру.

Диаграмма наглядно иллюстрирует, как данные движутся от ввода (камера, микрофон) через обработку и анализ к выводу (отчёт для HR-менеджера).

## Схема интерфейса

Для определения концепции взаимодействия с пользователем разработан прототип интерфейса системы в WPF. Схема представлена ниже (рис. 4).

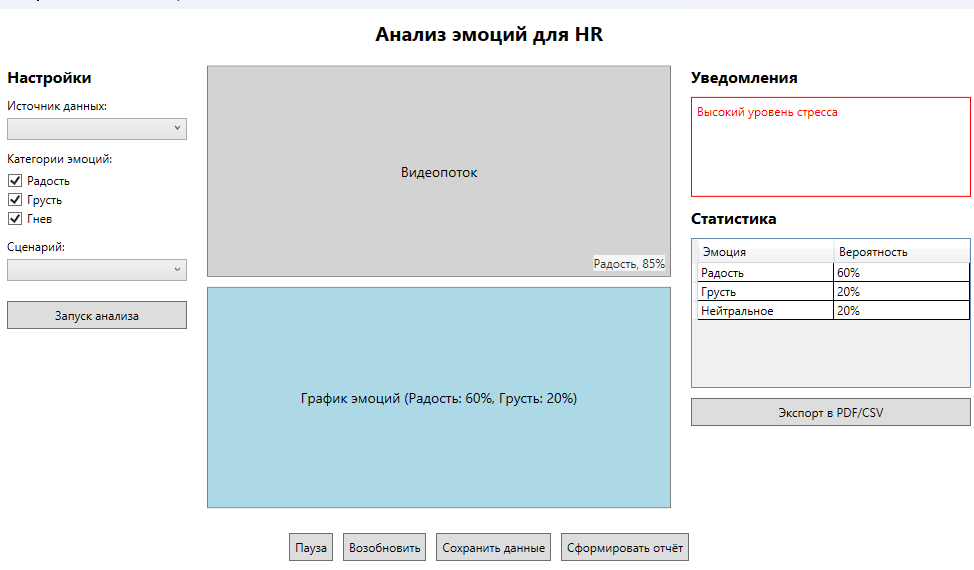


Рисунок 4. Схема интерфейса системы

Интерфейс системы спроектирован для удобства HR-менеджеров и включает следующие элементы:

* **Панель настроек**:
  + Выбор источника данных (камера, микрофон, загруженные файлы).
  + Настройка параметров анализа: выбор категорий эмоций (радость, грусть, гнев и т.д.), сценарий использования (например, собеседование, мониторинг).
  + Кнопка «Запуск анализа».
* **Область отображения видеопотока**:
  + Показывает текущий видеопоток с лица кандидата или сотрудника.
  + Накладывает метки с текущей эмоцией (например, «Радость, 85%») и вероятностью.
* **График эмоций**:
  + Отображает динамику эмоционального состояния в реальном времени (например, изменение эмоций за последние 5 минут).
  + Поддерживает переключение между категориями (радость, стресс и т.д.).
* **Панель уведомлений**:
  + Выводит предупреждения о критических состояниях (например, «Высокий уровень стресса»).
  + Позволяет настроить пороговые значения для уведомлений.
* **Область статистики**:
  + Показывает сводку по эмоциям за сеанс (например, «Радость: 60%, Грусть: 20%, Нейтральное: 20%»).
  + Поддерживает экспорт статистики в PDF или CSV.
* **Кнопки управления**:
  + «Пауза/Возобновить анализ».
  + «Сохранить данные» (видео, аудио, результаты).
  + «Сформировать отчёт» (генерация документа с результатами).

Интерфейс разработан с учётом принципов интуитивной навигации: элементы управления сгруппированы логически, а визуализация данных (графики, метки) упрощает интерпретацию результатов. Поддерживается русский язык, а структура интерфейса минимизирует время обучения HR-менеджера.

# ВИДЫ ОБЕСПЕЧЕНИЯ

Разработка мультимодальной системы распознавания эмоций человека требует комплексного подхода к обеспечению её функционирования. В данном разделе описаны лингвистическое, математическое, алгоритмическое, техническое, программное, правовое, эргономическое и информационное обеспечение системы.

## Лингвистическое обеспечение

В качестве основного языка программирования для реализации системы выбран **C#**, который удобен для создания приложений с графическим интерфейсом благодаря поддержке .NET Framework и интеграции с Visual Studio. Среда разработки — **Visual Studio 2022**, обеспечивающая удобные инструменты для написания, отладки и тестирования программ на C#.

Для обработки данных с камер и микрофонов, а также реализации нейросетевых алгоритмов используется язык программирования **Python**. Python применяется для анализа видеопотока (распознавание выражений лица), аудиопотока (определение голосовых характеристик) и интеграции с библиотеками машинного обучения, такими как TensorFlow и PyTorch.

Для управления базой данных, в которой хранятся данные о сеансах, кандидатах, эмоциях и отчётах, используется язык запросов **SQL** в СУБД **PostgreSQL**. Это обеспечивает эффективное хранение и быстрый доступ к структурированным данным.

Ниже приведены некоторые профессиональные термины, используемые в проекте:

Таблица 3. Термины и их определения

|  |  |
| --- | --- |
| **Термин** | **Определение** |
| Компьютерное зрение | Технология анализа видеопотока для распознавания объектов, в данном случае — выражений лица. |
| MFCC | Мел-частотные кепстральные коэффициенты, используемые для анализа аудиопотока (голосовых характеристик). |
| Нейронная сеть | Математическая модель, применяемая для классификации эмоций на основе видео- и аудиоданных. |
| Сеанс | Единичный процесс мониторинга эмоций кандидата или сотрудника в заданное время. |

Эти термины отражают ключевые аспекты работы системы и используются в документации и коде.

## Математическое обеспечение

Для реализации системы применяются математические методы и алгоритмы, обеспечивающие обработку данных и классификацию эмоций.

Обработка видеопотока:

Для анализа видеопотока используется сверточная нейронная сеть (CNN), которая обучается на датасете изображений лиц (например, FER2013). Классификация эмоций (радость, грусть, гнев и т.д.) выполняется с помощью функции активации **Softmax**:

где — вероятность принадлежности объекта к классу , — выходной сигнал нейрона для класса , n — количество классов (например, «шлем есть», «шлема нет»).

Для оценки качества модели используется функция потерь **Cross-Entropy Loss**:

Где N — количество примеров в обучающей выборке, — истинная метка класса, — предсказанная вероятность.

Обработка аудиопотока:

Для анализа аудиопотока применяются мел-частотные кепстральные коэффициенты (MFCC), которые извлекаются из голосовых данных. Классификация выполняется рекуррентной нейронной сетью (RNN) с функцией активации Softmax. Точность модели также оценивается через Cross-Entropy Loss.

Метрики качества:

Для контроля обучения используются метрики:

* **Accuracy**: доля правильно классифицированных эмоций,
* **Loss**: значение функции потерь, отражающее расхождение между предсказаниями и истинными значениями.

Эти методы обеспечивают точную классификацию эмоций и их интеграцию для мультимодального анализа.

## Алгоритмическое обеспечение

Алгоритмы системы разделены на этапы обработки данных и анализа эмоций:

1. **Алгоритм сбора данных**:
   * Захват видеопотока с камеры (разрешение 1080p) и аудиопотока с микрофона.
   * Предварительная обработка: нормализация яркости видео, фильтрация шума в аудио.
   * Сохранение данных в базе данных (видео, аудио, метаданные сеанса).
2. **Алгоритм обработки данных**:
   * Извлечение признаков из видео: определение ключевых точек лица (например, уголки рта, брови) с помощью OpenCV.
   * Извлечение MFCC из аудио с использованием библиотеки Librosa.
   * Подготовка данных для передачи в нейронную сеть (нормализация, преобразование в тензоры).
3. **Алгоритм анализа эмоций**:
   * Анализ видеопотока с помощью CNN для классификации выражений лица.
   * Анализ аудиопотока с помощью RNN для классификации голосовых характеристик.
   * Интеграция результатов: объединение вероятностей эмоций с использованием взвешенного среднего (например, 60% видео + 40% аудио).
   * Сохранение результатов (эмоция, вероятность) в базе данных.
4. **Алгоритм формирования отчёта**:
   * Сбор данных о сеансе (время, эмоции, вероятности).
   * Генерация статистики (например, процентное распределение эмоций).
   * Экспорт отчёта в формате CSV или PDF.

Эти алгоритмы обеспечивают последовательную обработку данных от ввода до вывода результатов.

## Техническое обеспечение

Требования к операционной системе

* **Windows 10/11** (64-разрядные версии) для настольного приложения.
* Возможна адаптация под Linux для серверной части (опционально).

Минимальные требования

* **Процессор**: 4 ядра (Intel Core i3 или AMD Ryzen 3, начиная с 2015 года).
* **Оперативная память**: 8 ГБ.
* **Видеокарта**: NVIDIA GT 1030 или встроенная графика Intel HD Graphics 620.
* **Свободное место на диске**: 50 ГБ (рекомендуется SSD).

Рекомендуемые требования

* **Процессор**: 6 ядер (Intel Core i5 или AMD Ryzen 5, начиная с 2018 года).
* **Оперативная память**: 16 ГБ.
* **Видеокарта**: NVIDIA GTX 1060 6 ГБ или выше (для ускорения нейросетевых вычислений).
* **Свободное место на диске**: 100 ГБ (SSD).

Аппаратное обеспечение

* **Камеры**: Разрешение 1080p, частота кадров 30 fps, поддержка инфракрасного режима для работы в условиях слабого освещения.
* **Микрофоны**: Частота дискретизации 44.1 кГц, направленные или всенаправленные в зависимости от сценария (собеседование, мониторинг).
* **Сервер**: Для обработки данных в реальном времени (опционально, если анализ выполняется локально).

Среда разработки

* **Visual Studio 2022** для C#.
* **PyCharm или Jupyter Notebook** для Python.

Версия браузера

* Для веб-интерфейса (если планируется): последние версии Chrome, Firefox, Edge (поддержка HTML5 и WebRTC).

## Программное обеспечение

Для реализации системы используются следующие программные средства:

* **Visual Studio 2022**: разработка на C# с использованием .NET Framework 4.7.2 или .NET 6.
* **Python 3.9+**: реализация нейросетевых алгоритмов с библиотеками:
  + **TensorFlow** или **PyTorch** для обучения и inference нейронных сетей.
  + **OpenCV** для обработки видеопотока.
  + **Librosa** для анализа аудиопотока.
* **PostgreSQL 15**: управление базой данных с поддержкой расширений для хранения больших данных (видео, аудио).

Требования к коду

* Код должен быть модульным, с чётким разделением на слои (сбор данных, обработка, анализ, интерфейс).
* Использование объектно-ориентированного подхода в C# и Python.
* Комментарии к коду на русском языке для упрощения поддержки.

Сборка приложения доступна для Windows (32- и 64-разрядные архитектуры), с возможностью портирования на другие платформы.

## Правовое обеспечение

Законодательство в сфере персональных данных

Система обрабатывает изображения лиц и голосовые данные, что подпадает под регулирование законодательства о защите персональных данных. В России основным нормативным актом является **Федеральный закон №152-ФЗ "О персональных данных"**, который устанавливает правила сбора, обработки, хранения и использования данных, позволяющих идентифицировать личность.

Для соответствия требованиям закона:

* **Согласие пользователя**: Перед началом сеанса пользователю (кандидату или сотруднику) отображается всплывающее окно с текстом согласия на обработку данных. Подтверждение фиксируется в базе данных.
* **Меры безопасности**: Данные шифруются с использованием AES-256, доступ ограничен ролями (например, только HR-менеджер), проводятся регулярные проверки на уязвимости.
* **Прозрачность**: Пользователям предоставляется информация о целях сбора данных (анализ эмоций), сроках хранения (например, 1 год) и способах удаления.

Авторские права и лицензии

В системе используются открытые датасеты и предобученные модели:

* **FER2013**: датасет изображений лиц для обучения CNN, распространяется под лицензией Creative Commons (CC BY-SA 4.0).
* **LibriSpeech**: датасет аудиозаписей для анализа голоса, доступен под лицензией CC BY 4.0.
* **Предобученные модели**: модели из TensorFlow Hub или PyTorch (например, VGG-Face), распространяемые под Apache License 2.0.

Все сторонние библиотеки (OpenCV, Librosa) используются в соответствии с их лицензиями (BSD, MIT). Код системы является оригинальным, за исключением адаптации открытых решений, что не нарушает авторских прав.

## Эргономическое обеспечение

Для удобства использования системы применяются следующие решения:

* **Мониторы**: Дисплей с диагональю от 24 дюймов, разрешением 1920x1080, частотой обновления 60 Гц для комфортного просмотра видеопотока и графиков.
* **Клавиатура и мышь**: Эргономичные устройства с минимальным уровнем шума для длительной работы HR-менеджера.
* **Камеры**: Регулируемые крепления для точного позиционирования, угол обзора 90°.
* **Интерфейс**: Минималистичный дизайн с крупными шрифтами (не менее 14 pt), контрастными цветами (например, тёмный фон, светлый текст) и интуитивным расположением элементов управления.

Эти меры снижают нагрузку на пользователя и повышают эффективность работы с системой.

## Информационное обеспечение

### . ER-диаграмма базы данных

ER-диаграмма описывает информационную модель мультимодальной системы распознавания эмоций, включая сущности, их атрибуты и связи. Она обеспечивает структурированное хранение данных, необходимых для работы системы. Диаграмма представлена ниже (рис. 5).

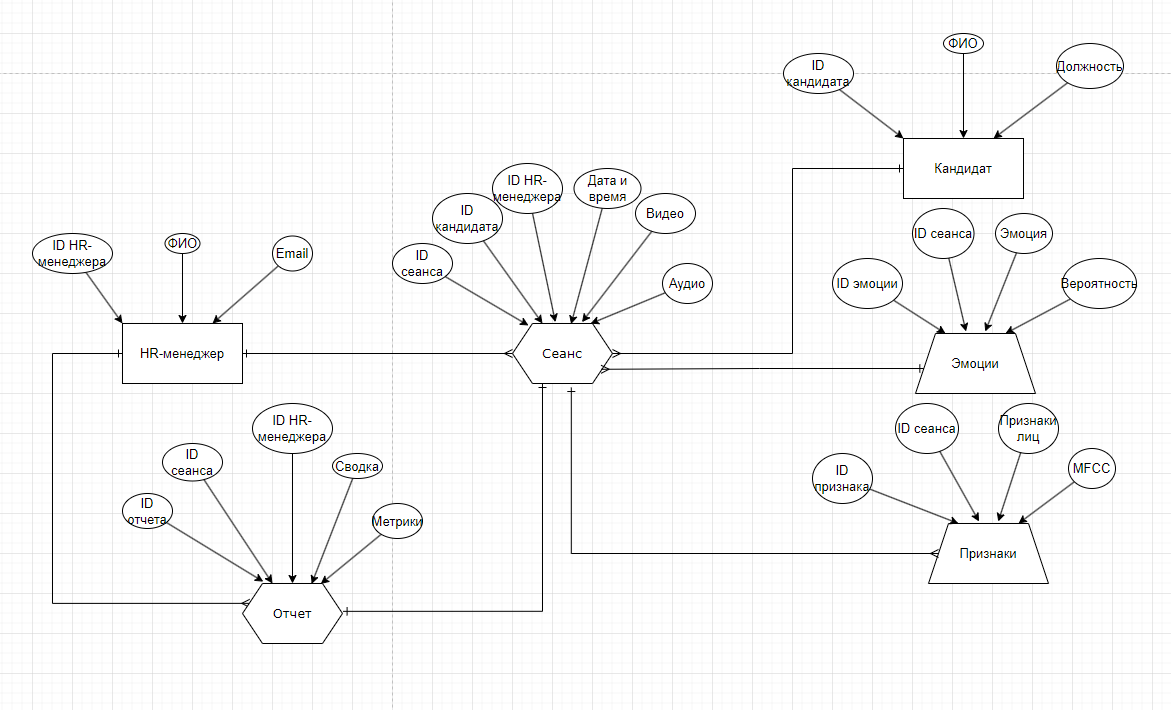


Рисунок 5. ER-диаграмма базы данных

**Основные** **сущности**:

* HR-менеджер:

Таблица 4. Описание сущности "HR-менеджер"

|  |  |  |
| --- | --- | --- |
| № | Наименование | Назначение |
| 1 | ID HR-менеджера | Уникальный идентификатор HR-менеджера |
| 2 | ФИО | Полное имя HR-менеджера |
| 3 | Email | Электронная почта для уведомлений |

* Сеанс:

Таблица 5. Описание сущности "Сеанс"

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | ID сеанса | Уникальный идентификатор сеанса |
| 2 | ID кандидата | Ссылка на кандидата |
| 3 | ID HR-менеджера | Ссылка на HR-менеджера |
| 4 | Дата и время | Временная метка начала сеанса |
| 5 | Видео | Путь к видеофайлу или данные в формате BLOB |
| 6 | Аудио | Путь к аудиофайлу или данные в формате BLOB |

* Кандидат:

**Таблица 6. Описание сущности "Кандидат"**

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | ID кандидата | Уникальный идентификатор кандидата |
| 2 | ФИО | Полное имя кандидата |
| 3 | Должность | Занимаемая или желаемая должность |

**Эмоции**:

* **Таблица 7. Описание сущности "Эмоции"**

|  |  |  |
| --- | --- | --- |
| № | Наименование | Назначение |
| 1 | ID эмоции | Уникальный идентификатор эмоции |
| 2 | ID сеанса | Ссылка на сеанс |
| 3 | Эмоция | Тип эмоции (радость, грусть и т.д.) |
| 4 | Вероятность | Вероятность эмоции (0–1) |

* Отчёт:

**Таблица 8. Описание сущности "Отчёт"**

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | ID отчёта | Уникальный идентификатор отчёта |
| 2 | ID сеанса | Ссылка на сеанс |
| 3 | ID HR-менеджера | Ссылка на HR-менеджера |
| 4 | Метрики | Статистика эмоций (например, % распределения) |
| 5 | Сводка | Итоговое описание состояния |

* Признаки:

**Таблица 9. Описание сущности "Признаки"**

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | ID признака | Уникальный идентификатор признака |
| 2 | ID сеанса | Ссылка на сеанс |
| 3 | Признаки лиц | Координаты ключевых точек лица (JSON или массив) |
| 4 | MFCC | Мел-частотные кепстральные коэффициенты аудио (массив) |

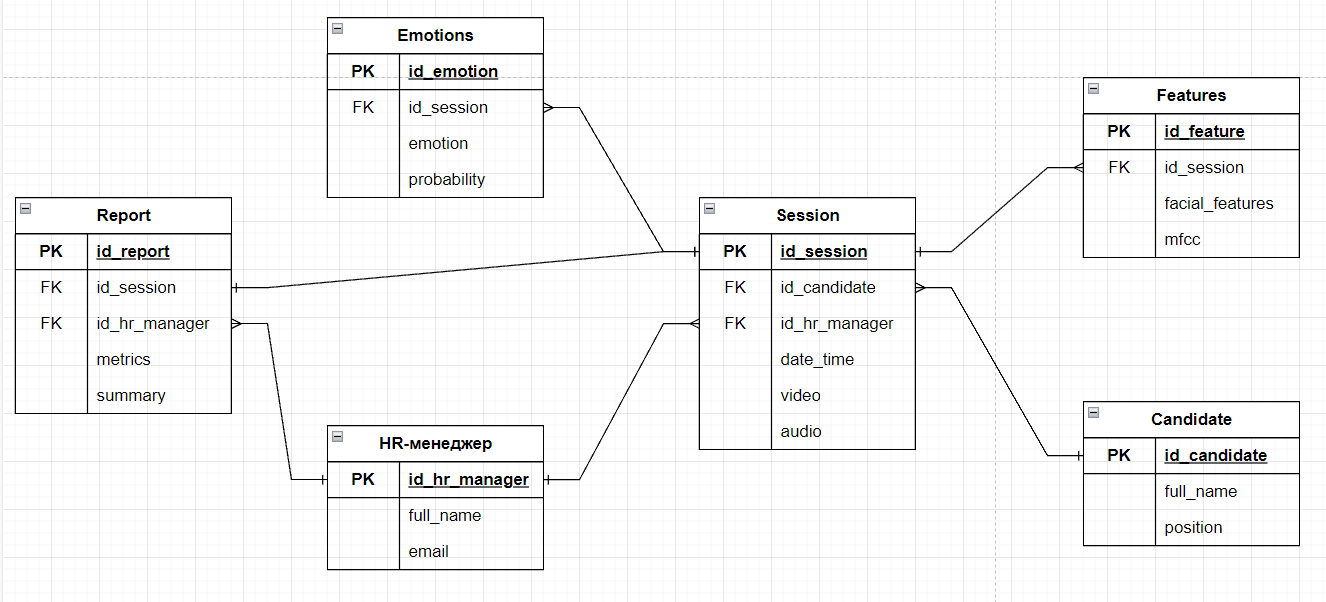
**Связи:**

* **HR-менеджер → Сеанс: связь "один-ко-многим" (1:M).**
* **Кандидат → Сеанс: связь "один-ко-многим" (1:M).**
* **Сеанс → Эмоции: связь "один-ко-многим" (1:M).**
* **Сеанс → Признаки: связь "один-ко-многим" (1:M).**
* **Сеанс → Отчёт: связь "один-к-одному" (1:1).**
* **HR-менеджер → Отчёт: связь "один-ко-многим" (1:M).**

Логическая модель обеспечивает структурированное хранение данных о сеансах, эмоциях и отчётах, поддерживая быстрый доступ и обработку информации.

### Логическая модель базы данных

Логическая модель базы данных представляет собой детализированную структуру таблиц системы, включая их поля, типы данных и связи. Она основана на ER-диаграмме и предназначена для реализации в СУБД PostgreSQL. Логическая модель представлена на рисунке ниже (рис. 6).



**Рисунок 6. Логическая модель базы данных**

**Описание таблиц**:

1. **HR\_Manager** 
   * **id\_hr\_manager (PK)**
   * **full\_name**
   * **email**
2. **Candidate** 
   * **id\_candidate (PK)**
   * **full\_name**
   * **position**
3. **Session** 
   * **id\_session (PK)**
   * **id\_candidate (FK)**
   * **id\_hr\_manager (FK)**
   * **date\_time**
   * **video**
   * **audio**
4. **Emotions** 
   * **id\_emotion (PK)**
   * **id\_session (FK)**
   * **emotion**
   * **probability**
5. **Features** 
   * **id\_feature (PK)**
   * **id\_session (FK)**
   * **facial\_features**
   * **mfcc**
6. **Report** 
   * **id\_report (PK)**
   * **id\_session (FK)**
   * **id\_hr\_manager (FK)**
   * **metrics**
   * **summary**

**Связи**

* **HR\_Manager → Session: 1:M (FK: id\_hr\_manager)**
* **Candidate → Session: 1:M (FK: id\_candidate)**
* **Session → Emotions: 1:M (FK: id\_session)**
* **Session → Features: 1:M (FK: id\_session)**
* **Session → Report: 1:1 (FK: id\_session)**
* **HR\_Manager → Report: 1:M (FK: id\_hr\_manager)**

# ЗАКЛЮЧЕНИЕ

В рамках данной работы была разработана мультимодальная система распознавания эмоций человека, предназначенная для использования в HR-процессах. Эта задача обладает высокой актуальностью в условиях роста интереса к автоматизации оценки персонала и повышения объективности анализа эмоционального состояния кандидатов и сотрудников. Применение современных технологий, таких как компьютерное зрение, анализ аудиопотока и нейронные сети, позволило создать решение, способное улучшить процессы подбора персонала, мониторинга вовлечённости и оценки стресса, минимизируя субъективность традиционных методов.

В ходе выполнения работы были решены следующие задачи:

1. Проведён анализ предметной области, в рамках которого изучены методы распознавания эмоций, мультимодальные подходы и их применение в HR. Определены ключевые аспекты, такие как необходимость интеграции видео- и аудиоданных для повышения точности анализа.
2. Выполнен обзор аналогов, который показал, что существующие решения (системы анализа лиц, голосовые анализаторы) имеют ограничения в точности и универсальности. На основе анализа был обоснован выбор мультимодального подхода, объединяющего обработку видеопотока и аудиопотока.
3. Разработано техническое задание, включающее функциональные и нефункциональные требования к системе, а также этапы её проектирования и реализации. Определены основные пользователи (HR-менеджеры) и требования к программному и аппаратному обеспечению.
4. Спроектирована система с использованием BPMN-диаграммы, описывающей бизнес-процессы сбора данных, их обработки, анализа эмоций и формирования отчётов. Диаграмма отражает взаимодействие между HR-менеджером, системой и модулем нейросетевого анализа.
5. Описаны виды обеспечения системы, включая лингвистическое, математическое, алгоритмическое, техническое, программное, правовое, эргономическое и информационное обеспечение.

Разработанная система позволяет в реальном времени анализировать эмоциональное состояние человека на основе видеопотока (выражения лица) и аудиопотока (голосовые характеристики), предоставляя HR-менеджерам детализированные отчёты. Применение сверточных нейронных сетей (CNN) для анализа видео и рекуррентных нейронных сетей (RNN) для аудио обеспечивает высокую точность классификации эмоций (радость, грусть, гнев и т.д.). Интеграция результатов в мультимодальной модели повышает надёжность анализа, а удобный интерфейс и уведомления позволяют оперативно интерпретировать данные для принятия решений.

Однако система имеет потенциал для дальнейшего улучшения. В перспективе возможно:

* Интеграция с корпоративными HR-платформами (например, SAP SuccessFactors или Workday) для автоматизации процессов оценки персонала.
* Расширение функционала для анализа групповых взаимодействий (например, на совещаниях) с использованием нескольких камер и микрофонов.
* Применение адаптивных моделей машинного обучения для повышения точности в условиях шумной среды или нестандартных выражений эмоций.
* Разработка мобильного приложения для HR-менеджеров, упрощающего доступ к отчётам и уведомлениям в реальном времени.

Таким образом, разработанная система представляет собой эффективный инструмент для повышения качества HR-процессов. Её внедрение может улучшить объективность оценки кандидатов и сотрудников, снизить влияние человеческого фактора на анализ эмоционального состояния и повысить эффективность управления персоналом в организациях.