

# RELATÓRIO TÉCNICO

## Tech Challenge - Fase 4

Análise Automatizada de Vídeo com Visão Computacional e  
Inteligência Artificial

FIAP - IADT

**Autores:**

Diego Silva Prado - RM364919

Francisco Ferreira de Araújo - RM361133

Ricardo Almeida da Rocha - RM362655

**Professor:** Carlos Expedito Rodrigues Aragão

Link para o github: <https://github.com/araujofran/tech-challenge-fase4-video-analysis>

Link para o vídeo: <https://youtu.be/6ImdLmI43GE>

Data de Entrega: 09 de Janeiro de 2026

## Sumário

Autores:.....	1
<b>1. INTRODUÇÃO .....</b>	<b>3</b>
<b>2. CONTEXTUALIZAÇÃO DO PROBLEMA .....</b>	<b>3</b>
<b>3. OBJETIVOS .....</b>	<b>4</b>
3.1 Objetivo Geral.....	4
3.2 Objetivos Específicos.....	4
<b>4. VISÃO GERAL DO FLUXO COMPUTACIONAL .....</b>	<b>4</b>
Fluxo geral da aplicação: .....	4
<b>5. ARQUITETURA DA SOLUÇÃO .....</b>	<b>5</b>
5.1 Detecção Facial (DNN) .....	5
5.2 Rastreamento Facial .....	5
5.3 Análise de Expressões Emocionais .....	5
5.4 Detecção de Atividades e Anomalias .....	5
<b>6. GERAÇÃO DO RELATÓRIO AUTOMÁTICO.....</b>	<b>6</b>
<b>7. DEMONSTRAÇÃO EM VÍDEO .....</b>	<b>6</b>
<b>8. TECNOLOGIAS UTILIZADAS .....</b>	<b>7</b>
<b>9. RESULTADOS OBTIDOS .....</b>	<b>7</b>
<b>10. INSIGHTS OBTIDOS .....</b>	<b>7</b>
<b>11. CONCLUSÃO FINAL .....</b>	<b>8</b>
<b>12. TECH CHALLENGE – RELATÓRIO FINAL.....</b>	<b>8</b>
<b>13. LINKS DE ENTREGA.....</b>	<b>9</b>

# 1. INTRODUÇÃO

O presente relatório descreve o desenvolvimento de uma aplicação de **análise automatizada de vídeo**, construída como entrega do **Tech Challenge – Fase 4 (IADT)**. O projeto integra técnicas modernas de **Visão Computacional**, **Aprendizado de Máquina** e **Inteligência Artificial**, aplicadas diretamente sobre dados visuais dinâmicos (vídeos).

O desafio proposto nesta fase tem como foco a análise de vídeos reais, exigindo a implementação de funcionalidades como **reconhecimento facial**, **análise de expressões emocionais**, **detecção de atividades e anomalias**, bem como a **geração automática de relatórios e demonstração em vídeo**.

A solução desenvolvida neste trabalho busca atender integralmente a esses requisitos, adotando uma abordagem modular, reproduzível e alinhada às boas práticas de engenharia de software e ciência de dados.

## 2. CONTEXTUALIZAÇÃO DO PROBLEMA

A análise manual de vídeos é um processo custoso, sujeito a vieses humanos e pouco escalável. Em contextos como monitoramento, segurança, análise comportamental e auditoria, torna-se essencial o uso de sistemas automatizados capazes de extrair informações relevantes de grandes volumes de dados audiovisuais.

Diante desse cenário, o desafio da Fase 4 consiste em projetar uma aplicação capaz de:

- Identificar rostos humanos em vídeos;
- Inferir estados emocionais a partir de expressões faciais;
- Analisar padrões de movimento e identificar comportamentos anômalos;
- Consolidar estatísticas e gerar relatórios automáticos;
- Demonstrar o funcionamento da aplicação por meio de um vídeo explicativo.

## 3. OBJETIVOS

### 3.1 Objetivo Geral

Desenvolver uma aplicação em Python para **análise automatizada de vídeo**, utilizando técnicas de Visão Computacional e Inteligência Artificial, capaz de extrair informações faciais, emocionais e comportamentais de forma integrada.

### 3.2 Objetivos Específicos

- Implementar detecção e rastreamento facial em vídeos;
- Realizar análise de expressões emocionais frame a frame;
- Classificar atividades com base em padrões de movimento;
- Identificar e contabilizar anomalias comportamentais;
- Gerar relatórios estatísticos automáticos;
- Produzir um vídeo demonstrativo narrado, conforme exigido pelo desafio.

## 4. VISÃO GERAL DO FLUXO COMPUTACIONAL

A aplicação foi estruturada seguindo um **pipeline sequencial e modular**, permitindo clareza no fluxo de dados e facilidade de manutenção.

### Fluxo geral da aplicação:

1. Carregamento do vídeo de entrada
2. Leitura frame a frame
3. Detecção e rastreamento de rostos
4. Inferência de emoções por rosto
5. Análise de movimento e classificação de atividades
6. Registro estatístico dos eventos
7. Geração do vídeo processado com overlays
8. Consolidação dos dados em relatório final

## 9. Produção de vídeo narrado (podcast técnico)

# 5. ARQUITETURA DA SOLUÇÃO

A arquitetura da solução é composta pelos seguintes módulos principais:

## 5.1 Detecção Facial (DNN)

A detecção de rostos é realizada utilizando um modelo de **Deep Neural Network (DNN)** pré-treinado, baseado em arquitetura SSD (Single Shot Detector). Esse modelo permite identificar múltiplos rostos em diferentes escalas dentro de cada frame do vídeo.

Os rostos detectados são delimitados visualmente por **bounding boxes**, servindo como base para as análises subsequentes.

## 5.2 Rastreamento Facial

Após a detecção inicial, os rostos são rastreados ao longo dos frames, permitindo manter consistência na análise emocional e na identificação de padrões de movimento associados a cada indivíduo.

## 5.3 Análise de Expressões Emocionais

Para cada rosto identificado, o sistema aplica um modelo de **Facial Expression Recognition (FER)**, responsável por inferir o estado emocional predominante. As emoções detectadas são contabilizadas ao longo de todo o vídeo, possibilitando a geração de estatísticas consolidadas.

## 5.4 Detecção de Atividades e Anomalias

A análise de atividades é baseada na variação espacial dos rostos entre frames consecutivos. Com base nessa variação, os movimentos são classificados em:

- Parado
- Movimento leve
- Movimento brusco

Movimentos bruscos são tratados como **anomalias**, conforme definido no escopo do desafio.

## 6. GERAÇÃO DO RELATÓRIO AUTOMÁTICO

Ao final do processamento do vídeo, a aplicação gera automaticamente um relatório estatístico contendo:

- Total de frames processados;
- Distribuição das emoções detectadas;
- Quantidade de atividades classificadas;
- Número total de anomalias identificadas.

Esse relatório é exportado em formato texto e serve como base para validação dos resultados e apresentação do projeto.

## 7. DEMONSTRAÇÃO EM VÍDEO

Como parte dos requisitos do Tech Challenge – Fase 4, foi produzido um **vídeo demonstrativo** contendo:

- Execução da aplicação;
- Visualização do reconhecimento facial;
- Exibição das emoções inferidas;
- Detecção de atividades e anomalias;
- Apresentação do relatório final.

Para facilitar a compreensão do fluxo da aplicação, o vídeo foi narrado em formato de **podcast técnico**, explicando cada etapa do processamento.

Quando a duração do podcast excedeu a do vídeo base, o vídeo foi automaticamente repetido até cobrir toda a narração, garantindo a exibição completa do conteúdo explicativo.

---

## 8. TECNOLOGIAS UTILIZADAS

- Python 3
  - OpenCV
  - MediaPipe
  - FER (Facial Expression Recognition)
  - NumPy
  - FFmpeg
  - IA Generativa (Text-to-Speech)
- 

## 9. RESULTADOS OBTIDOS

A aplicação processou integralmente o vídeo proposto, identificando com sucesso:

- Múltiplos rostos ao longo dos frames;
- Emoções predominantes ao longo do tempo;
- Atividades classificadas por intensidade de movimento;
- Anomalias comportamentais.

Os resultados foram consolidados automaticamente em relatório e apresentados de forma clara no vídeo final.

---

## 10. INSIGHTS OBTIDOS

- A integração de múltiplas técnicas de Visão Computacional em um único pipeline mostrou-se eficaz e robusta.
- A análise emocional baseada em expressões faciais permite extrair informações qualitativas relevantes de vídeos.
- A detecção de atividades por variação espacial é uma abordagem simples, porém eficiente, para identificação de comportamentos anômalos.
- A automação completa do pipeline reduz significativamente o esforço humano na análise de vídeos.

## 11. CONCLUSÃO FINAL

O projeto desenvolvido atende integralmente aos requisitos do **Tech Challenge – Fase 4**, demonstrando a aplicação prática de técnicas de Inteligência Artificial e Visão Computacional na análise automatizada de vídeos.

A solução apresentada é modular, extensível e alinhada às boas práticas de engenharia de software, além de fornecer uma base sólida para futuras expansões, como análise em tempo real, integração com bancos de dados ou uso de modelos mais avançados de aprendizado profundo.

## 12. TECH CHALLENGE – RELATÓRIO FINAL

Total de frames analisados: 3326

Número de anomalias detectadas: 95

Distribuição de Emoções:

- neutral: 1963

- sad: 253

- happy: 514

- surprise: 224

- angry: 65



Distribuição de Atividades:

- parado: 2218
- movimento leve: 706
- movimento brusco: 95

O sistema analisou o vídeo identificando rostos, inferindo emoções por meio de modelos de deep learning e classificando atividades. Movimentos que não seguiram o padrão geral foram classificados como anômalos.

## 13. LINKS DE ENTREGA

- **Repositório GitHub:** <https://github.com/araujofran/tech-challenge-fase4-video-analysis>
- **Vídeo Demonstrativo:** <https://youtu.be/6lmdLml43GE>