INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS INHUMAS
BACHARELADO EM ENGENHARIA DE SOFTWARE

FELIPE ANTONIO ESTACIO DE MORAIS
LUIS FERNANDO RIBEIRO HONORATO
PEDRO GABRIEL CUNHA MENDES

FERRAMENTA PARA A CRIAÇÃO AUTOMATIZADA DE FRASES COM BASE NA ANÁLISE DE EMOÇÕES E RECONHECIMENTO DA EXPRESSÃO FACIAL

INHUMAS
24 de setembro de 2025

FELIPE ANTONIO ESTACIO DE MORAIS
LUIS FERNANDO RIBEIRO HONORATO
PEDRO GABRIEL CUNHA MENDES

FERRAMENTA PARA A CRIAÇÃO AUTOMATIZADA DE FRASES COM BASE NA ANÁLISE DE EMOÇÕES E RECONHECIMENTO DA EXPRESSÃO FACIAL

Trabalho de Conclusão de Curso, no formato de monografia, apresentado ao curso de Bacharelado em Engenharia de Software, do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Campus Inhumas, como requisito parcial à obtenção do título de Bacharel em Engenharia de Software.

Orientador: Ricardo Rodrigues Dias de Lima

Inhumas

24 de setembro de 2025

Dados Internacionais de Catalogação na Publicação (CIP)

Morais, Felipe Antonio Estacio de.

M828 Ferramenta para a criação automatizada de frases com base na análise de emoções e reconhecimento da expressão facial [Manuscrito]. / Felipe Antonio Estacio de Morais, Luis Fernando Ribeiro Honorato, Pedro Gabriel Cunha Mendes – Inhumas: IFG, 2025.

87 f.: il.

Inclui bibliografia.

Orientador: Prof. Ms. Ricardo Rodrigues D. de Lima

Trabalho de Conclusão de Curso em Bacharelado em Engenharia de Software – IFG/Câmpus Inhumas, 2025.

1. Inteligência artificial. 2. Reconhecimento facial. 3. Processamento de linguagem natural. 4. Honorato, Luis Fernando Ribeiro. 5. Mendes, Pedro Gabriel Cunha. 6. Lima, Ricardo Rodrigues de (orientador). I. Título.

CDD 303.483

Ficha catalográfica elaborada pela bibliotecária Maria Aparecida Rodrigues de Souza - CRB/1-1497

Instituto Federal de Educação, Ciência e Tecnologia de Goiás

Campus Inhumas - Biblioteca Atena

24 de setembro de 2025

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinalada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

[ ] Tese

[ ] Dissertação

[ ] Monografia – Especialização

[**X**] TCC – Graduação

[ ] Produto Técnico e Educacional - Tipo:

[ ] Artigo Científico

[ ] Capítulo de Livro

[ ] Livro

[ ] Trabalho Apresentado em Evento

Nome Completo dos Autores: FELIPE ANTONIO ESTACIO DE MORAIS; LUIS FERNANDO RIBEIRO HONORATO; PEDRO GABRIEL CUNHA MENDE

Matrículas: 20211030180209; 20211030180314; 20211030180292

Título do Trabalho: FERRAMENTA PARA A CRIAÇÃO AUTOMATIZADA DE FRASES COM BASE NA ANÁLISE DE EMOÇÕES E RECONHECIMENTO DA EXPRESSÃO FACIAL

Autorização - Marque uma das opções

[**X**] Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto).
[ ] Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data de 24 de setembro de 2025 (Embargo).
[ ] Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção 2 ou 3, marque a justificativa:

[ ] O documento está sujeito a registro de patente.

[ ] O documento poderá vir a ser publicado como livro, capítulo de livro ou artigo.

[ ] Outra justificativa:

Documento assinado digitalmente
FELIPE ANTONIO ESTACIO DE MORAIS
Data: 24/09/2025 17:50:41-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor

Documento assinado digitalmente
LUIS FERNANDO RIBEIRO HONORATO
Data: 24/09/2025 18:01:03-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor

Documento assinado digitalmente
PEDRO GABRIEL CUNHA MENDES
Data: 24/09/2025 18:04:51-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O(A) referido(a) autor(a) declara que:

i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;

ii. obteve autorização de quaisquer materiais inclusos no documento de que não detém os direitos de autoria, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cumpre direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;

iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Inhumas/Goiás
Local

24 de setembro de 2025
Data

Documento assinado digitalmente
gov.br
FELIPE ANTONIO ESTACIO DE MORAIS
Data: 24/09/2025 17:52:24-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Documento assinado digitalmente
gov.br
LUIS FERNANDO RIBEIRO HONORATO
Data: 24/09/2025 18:01:55-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Documento assinado digitalmente
gov.br
PEDRO GABRIEL CUNHA MENDES
Data: 24/09/2025 18:03:23-0300
Verifique em https://validar.iti.gov.br

Assinatura do Autor e/ou Detentor dos Direitos Autorais

ATA DE DEFESA DE MONOGRAFIA

Na presente data realizou-se a sessão pública de defesa da Monografia intitulada **Ferramenta para a criação automatizada de frases com base na análise de emoções e reconhecimento da expressão facial**, sob orientação de Ricardo Rodrigues Dias de Lima , apresentada pelo aluno **Felipe Antonio Estacio de Morais (20211030180209)** do Curso **Bacharelado em Engenharia de Software (Câmpus Inhumas)**. Os trabalhos foram iniciados às **20:00** do dia **26/08/2025** pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

- **Ricardo Rodrigues Dias de Lima** (Presidente)
- **Rodrigo Candido Borges** (Examinador Interno)
- **Victor Hugo Lazaro Lopes** (Examinador Interno)

A banca examinadora, tendo terminado a apresentação do conteúdo da Monografia, passou à argüição do candidato. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelo aluno, tendo sido atribuído o seguinte resultado:

[X] Aprovado [ ] Reprovado Nota : 8,3

Observação / Apreciações:

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu **Ricardo Rodrigues Dias de Lima** lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Documento assinado eletronicamente por:

- **Rodrigo Candido Borges** (002.592.511-35), em **01/09/2025 14:57:14** com chave **177d5f12875d11f0a98d005056a537a4**.
- **Victor Hugo Lazaro Lopes** (940.053.971-15), em **30/08/2025 11:05:52** com chave **705232ac85aa11f0b648005056a537a4**.
- **Ricardo Rodrigues Dias de Lima** (991.163.241-53), em **30/08/2025 10:51:36** com chave **720a59e685a811f08bdc005056a537a4**.

Este documento foi emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QrCode ao lado ou acesse https://suap.ifg.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Ata de Projeto Final

Data da Emissão: 01/09/2025

Código de Autenticação: db6e9e

ATA DE DEFESA DE MONOGRAFIA

Na presente data realizou-se a sessão pública de defesa da Monografia intitulada **Ferramenta para a criação automatizada de frases com base na análise de emoções e reconhecimento da expressão facial**, sob orientação de Ricardo Rodrigues Dias de Lima , apresentada pelo aluno **Luis Fernando Ribeiro Honorato (20211030180314)** do Curso **Bacharelado em Engenharia de Software (Câmpus Inhumas)**. Os trabalhos foram iniciados às **20:00** do dia **26/08/2025** pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

- **Ricardo Rodrigues Dias de Lima** (Presidente)
- **Rodrigo Candido Borges** (Examinador Interno)
- **Victor Hugo Lazaro Lopes** (Examinador Interno)

A banca examinadora, tendo terminado a apresentação do conteúdo da Monografia, passou à argüição do candidato. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelo aluno, tendo sido atribuído o seguinte resultado:

[X] Aprovado [ ] Reprovado Nota : 8,3

Observação / Apreciações:

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu **Ricardo Rodrigues Dias de Lima** lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Documento assinado eletronicamente por:

- **Rodrigo Candido Borges** (002.592.511-35), em **03/09/2025 23:49:29** com chave **c6d5b38c893911f0b0d4005056a537a4**.
- **Victor Hugo Lazaro Lopes** (940.053.971-15), em **02/09/2025 17:41:14** com chave **2b483ad6883d11f0a82b005056a537a4**.
- **Ricardo Rodrigues Dias de Lima** (991.163.241-53), em **02/09/2025 16:57:59** com chave **209469b2883711f0a99c005056a537a4**.

Este documento foi emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QrCode ao lado ou acesse https://suap.ifg.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Ata de Projeto Final

Data da Emissão: 03/09/2025

Código de Autenticação: dbaf96

ATA DE DEFESA DE MONOGRAFIA

Na presente data realizou-se a sessão pública de defesa da Monografia intitulada **Ferramenta para a criação automatizada de frases com base na análise de emoções e reconhecimento da expressão facial**, sob orientação de Ricardo Rodrigues Dias de Lima , apresentada pelo aluno **Pedro Gabriel Cunha Mendes (20211030180292)** do Curso **Bacharelado em Engenharia de Software (Câmpus Inhumas)**. Os trabalhos foram iniciados às **20:00** do dia **26/08/2025** pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

- **Ricardo Rodrigues Dias de Lima** (Presidente)
- **Rodrigo Candido Borges** (Examinador Interno)
- **Victor Hugo Lazaro Lopes** (Examinador Interno)

A banca examinadora, tendo terminado a apresentação do conteúdo da Monografia, passou à argüição do candidato. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelo aluno, tendo sido atribuído o seguinte resultado:

[X] Aprovado [ ] Reprovado Nota : 8,3

Observação / Apreciações:

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu **Ricardo Rodrigues Dias de Lima** lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Documento assinado eletronicamente por:

- **Rodrigo Candido Borges** (002.592.511-35), em **03/09/2025 23:48:45** com chave **ad2ad8c2893911f0bfa8005056a537a4**.
- **Victor Hugo Lazaro Lopes** (940.053.971-15), em **02/09/2025 17:41:01** com chave **2339d9da883d11f0a71f005056a537a4**.
- **Ricardo Rodrigues Dias de Lima** (991.163.241-53), em **02/09/2025 16:57:51** com chave **1b54ba60883711f0abb5005056a537a4**.

Este documento foi emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QrCode ao lado ou acesse https://suap.ifg.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Ata de Projeto Final

Data da Emissão: 03/09/2025

Código de Autenticação: 15cd8a

AGRADECIMENTOS

Agradecemos a Deus, aos nossos familiares, aos nossos professores e aos nossos orientadores.

Resumo

Este trabalho apresenta o desenvolvimento de uma ferramenta que automatiza a criação de frases a partir da análise de sentimentos e do reconhecimento de expressões faciais. A solução proposta integra técnicas de Visão Computacional e Processamento de Linguagem Natural (PLN), utilizando algoritmos de redes neurais profundas para identificar emoções em imagens faciais e, em seguida, gerar textos coerentes e contextualmente adequados. Para a implementação, foram utilizadas bases de dados públicas de imagens e frases, que permitiram treinar e validar os modelos empregados. Os resultados demonstraram que a ferramenta é capaz de reconhecer emoções com bom nível de precisão e produzir frases com consistência semântica e relevância contextual. Além disso, a pesquisa evidencia os desafios relacionados à qualidade e diversidade dos dados, bem como às questões éticas no uso de informações biométricas faciais e emocionais. O estudo contribui para o avanço de aplicações que unem inteligência artificial e automação de conteúdo digital, oferecendo uma solução inovadora para contextos pessoais e corporativos.

Palavras-chave: Inteligência Artificial, Reconhecimento Facial, Análise de Sentimentos, Processamento de Linguagem Natural, Conteúdo Digital.

Abstract

This work presents the development of a tool that automates the creation of social media posts based on sentiment analysis and facial expression recognition. The proposed solution integrates Computer Vision and Natural Language Processing (NLP) techniques, employing deep neural network algorithms to identify emotions in facial images and subsequently generate coherent and contextually appropriate texts. Public datasets of images and sentences were used to train and validate the implemented models. The results showed that the tool is capable of recognizing emotions with a good level of accuracy and producing posts with semantic consistency and contextual relevance. Furthermore, the research highlights challenges related to data quality and diversity, as well as ethical concerns regarding the use of biometric and emotional information. This study contributes to the advancement of applications that combine artificial intelligence and digital content automation, offering an innovative solution for both personal and corporate contexts.

Keywords: Artificial Intelligence, Facial Recognition, Sentiment Analysis, Natural Language Processing, Social Media.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

INTRODUÇÃO

A evolução das tecnologias de inteligência artificial tem impulsionado significativamente o desenvolvimento de ferramentas para análise e automação de conteúdo em redes sociais. Esse avanço tem permitido que empresas e organizações obtenham *insights* valiosos sobre a percepção do público em relação a seus produtos, serviços e marca, além de otimizar suas estratégias de comunicação digital (BROWN et al., 2020). No contexto atual, marcado pela crescente digitalização das interações sociais e comerciais, a capacidade de analisar e responder rapidamente às manifestações dos usuários nas plataformas *online* tornou-se um diferencial competitivo. Além disso, o reconhecimento facial tem emergido como uma tecnologia complementar, permitindo a identificação de emoções e sentimentos diretamente a partir de expressões faciais dos usuários, enriquecendo ainda mais a análise de dados em redes sociais.

A análise de sentimentos, em particular, emergiu como uma ferramenta poderosa para compreender as nuances emocionais expressas, permitindo uma abordagem mais precisa e personalizada na gestão do relacionamento com o cliente. Conforme destacado por Fan et al. (2023), publicações com sentimentos positivos tendem a gerar maior engajamento e reações positivas, criando um ciclo de *feedback* benéfico. Por outro lado, postagens que expressam emoções negativas podem atrair comentários de apoio, mas também perpetuar sentimentos adversos se não forem bem gerenciados.

Paralelamente, a automação de postagens e interações nas redes sociais tem se mostrado uma estratégia para manter uma presença constante e engajadora, sem comprometer a eficiência operacional das empresas. Essa abordagem, quando implementada de forma equilibrada, pode resultar em uma comunicação mais consistente e relevante, aumentando o alcance e o impacto das mensagens transmitidas (OLIVEIRA; JUNIOR, 2023).

No entanto, é importante ressaltar que a implementação dessas tecnologias apresenta desafios significativos, tanto do ponto de vista técnico quanto ético. A necessidade de garantir a precisão na interpretação dos sentimentos expressos pelos usuários, bem como a manutenção de um equilíbrio entre automação e autenticidade nas interações, são aspectos críticos que demandam atenção constante (GORWA; BINNS; KATZENBACH, 2020).

Ferramentas de análise emocional, como o *Google Cloud Natural Language API* e o *IBM Watson Tone Analyzer*, utilizam processamento de linguagem natural (PLN) para avaliar sentimentos em tempo real. Essas tecnologias permitem *insights* valiosos, como a detecção de padrões emocionais e a eficácia de campanhas publicitárias, além de identificarem sinais precoces de distúrbios emocionais, como destacado por Lopes (2024). O autor ainda afirma que esses sistemas enfrentam desafios, incluindo nuances linguísticas, gírias e sarcasmo, que podem afetar a precisão das análises. Além disso, questões éticas, como privacidade e consentimento, tornam-se cada vez mais relevantes nesse contexto.

Por outro lado, as ferramentas de análise emocional enfrentam desafios significativos. Um dos principais contras é a precisão das análises, que pode ser afetada por nuances linguísticas, gírias, sarcasmo e contexto cultural, tornando difícil a interpretação correta das emoções (LOPES, 2024). Além disso, há preocupações com a privacidade e a ética do monitoramento emocional, especialmente se os dados forem usados sem o consentimento dos usuários. De acordo com Oliveira e Junior (2023) outro desafio é a necessidade constante de atualização dos algoritmos para acompanhar as mudanças na linguagem e nas tendências de uso das redes sociais. Por fim, a dependência dessas ferramentas pode levar a uma compreensão superficial das emoções, que precisa ser complementada por análises qualitativas mais profundas para obter uma visão completa e precisa.

Diante dessa realidade, este trabalho justifica-se pela crescente necessidade de empresas e indivíduos manterem suas redes sociais atualizadas de forma eficiente, sem demandar grande esforço. O dinamismo das redes sociais exige soluções que combinam praticidade e inovação para acompanhar o ritmo das interações e das demandas por personalização. Além disso, a pesquisa busca contribuir para o desenvolvimento de um software inovador que automatize a criação de conteúdo com base em reconhecimento facial e análise de emoções, destacando sua relevância social e tecnológica.

Neste contexto, o trabalho propõe uma solução de software que utiliza inteligência artificial para identificar emoções a partir de expressões faciais, gerando automaticamente frases de acordo com a valência emocional identificada, não considerando o ambiente em que ele está. Essa abordagem não apenas otimiza estratégias de *marketing* digital, mas também aprimora a personalização das interações e eleva os padrões de segurança *online*. Ao integrar reconhecimento facial e análise de emoções, a proposta inova ao unir ferramentas tecnológicas de ponta em um sistema que atende tanto a necessidades comerciais quanto sociais, ampliando o impacto dessa solução no mercado digital e na experiência dos usuários.

Se um software com técnicas de inteligência artificial for capaz de realizar a análise de sentimentos por reconhecimento facial, então ele poderá gerar frases personalizadas, otimizando estratégias de *marketing* digital por meio da identificação de padrões emocionais e da adaptação das mensagens ao perfil do público-alvo.

O objetivo geral é implementar um *software* de reconhecimento facial com análise de

emoções para gerar frases automáticas com polaridades positiva ou negativa. Os objetivos específicos incluem: (1) Selecionar e implementar técnicas para análise de emoções utilizando imagens faciais, integrando algoritmos de visão computacional e processamento de linguagem natural; (2) Desenvolver um sistema que identifique emoções por meio do reconhecimento facial e, com base nisso, automatize a criação de frases com valência positiva ou negativa, garantindo relevância contextual e adequação emocional; (3) Avaliar a eficácia do sistema proposto em termos de precisão na identificação de emoções e qualidade do conteúdo gerado, comparando com métodos existentes de análise de sentimentos e geração de conteúdo.

O trabalho está dividido em cinco capítulos. No Capítulo 2, será apresentada a fundamentação teórica com base em estudos de autores relevantes para temas sobre automação de processos baseada em software, reconhecimento facial, análise de emoções e técnicas e processos sobre o uso de algoritmos de Aprendizado Profundo, de modo específico, Redes Neurais Artificiais Profundas. O Capítulo 3 detalhará a metodologia, incluindo as etapas de implementação dos algoritmos baseados em técnicas de reconhecimento facial, análise de emoções e as arquiteturas de redes neurais profundas utilizadas nos processos de análise de emoções por meio de reconhecimento facial para gerar frases mediante o sentimento identificado. No Capítulo 4, serão discutidos os resultados experimentais e as limitações do processo. E, no Capítulo 5, serão apresentadas as considerações finais, destacando as contribuições do projeto e sugestões para trabalhos futuros.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

Nesta seção serão apresentados os principais conceitos que serão abordados neste trabalho.

2.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) constituem um paradigma computacional fundamentado em modelos matemáticos inspirados na estrutura e no funcionamento do sistema nervoso biológico, em particular do cérebro humano. Esses sistemas são compostos por uma série de unidades de processamento interconectadas, denominadas neurônios artificiais, que operam de forma paralela e distribuída. Tais características conferem às RNAs a capacidade de aprender padrões complexos a partir de dados, adaptar-se a novas informações e realizar tomadas de decisão com base no conhecimento adquirido durante o processo de aprendizado (FLECK, 2016).

De acordo com Bezerra (2016), a arquitetura de uma RNA é definida pela topologia de conexões entre suas unidades de processamento, as quais são organizadas em camadas. Essa estrutura permite que as redes neurais artificiais sejam capazes de resolver problemas complexos, armazenar conhecimento de forma distribuída e aplicá-lo em tarefas específicas, como classificação, regressão e reconhecimento de padrões.

O desenvolvimento teórico das RNAs teve um marco significativo com o trabalho pioneiro de McCulloch e Pitts (1943), que propuseram um modelo matemático simplificado para representar o funcionamento de um neurônio biológico, conforme ilustrado na Figura 2.1. Esse modelo serviu como base para a concepção do neurônio artificial, unidade fundamental das redes neurais contemporâneas. O neurônio artificial opera por meio da combinação linear de entradas, ponderadas por seus respectivos pesos sinápticos, seguida pela aplicação de uma função de ativação não linear. Essa função é responsável por introduzir não linearidades ao sistema, permitindo que a rede aprenda e modele relações complexas entre variáveis de entrada e saída.

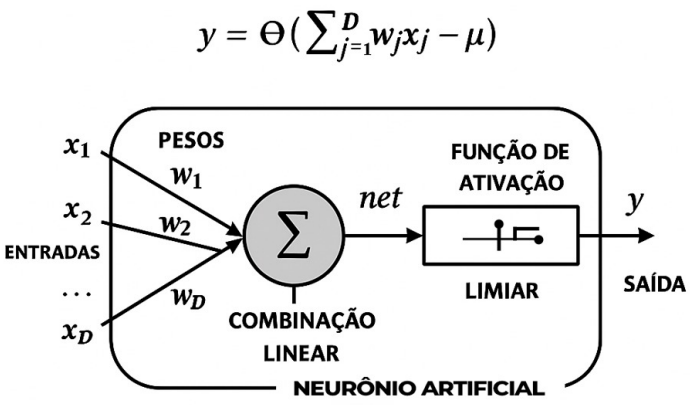

Figura 2.1: Modelo de Neurônio artificial de Kovacs, McCulloch e Pitts
Fonte: McCulloch e Pitts (1943)

O modelo de neurônio proposto por McCulloch e Pitts (1943) representou um marco fundamental no desenvolvimento das Redes Neurais Artificiais, influenciando diretamente a criação e aprimoramento de diversas funções de ativação. A função de ativação tem como objetivo principal regular a amplitude do sinal de saída de um neurônio artificial, restringindo-a a intervalos específicos, como [0, 1] ou, em alguns casos, [-1, 1]. Essa limitação é essencial para garantir a estabilidade e a eficácia do processo de aprendizado da rede, além de permitir a modelagem de comportamentos não lineares ((HAYKIN, 2007).

De acordo com Rauber (2005), existem diversas funções de ativação utilizadas em RNAs, cada uma com características específicas que as tornam adequadas para determinadas aplicações. Dentre as mais comumente empregadas, como apresentadas na figura 2.2, destacam-se:

- Função Limiar ou Escada: Esta função matemática é amplamente utilizada em algoritmos de aprendizado de máquina para transformar valores contínuos em saídas binárias, geralmente no intervalo [0, 1]. Ela é definida por uma transição entre dois estados, sendo particularmente útil em modelos de classificação binária.

- Função Linear: A função linear mantém uma relação proporcional entre a entrada e a saída do neurônio, sem introduzir não linearidades ao sistema. Embora simples, sua aplicação é limitada em redes neurais profundas, uma vez que a ausência de não linearidades impede a modelagem de relações complexas entre variáveis.

- Função *Sigmoid*: A função *sigmoid* é uma das mais utilizadas em Redes Neurais Artificais devido à sua natureza de relação diferenciável. Ela mapeia valores de entrada para o intervalo [0, 1], sendo eficaz em problemas de classificação probabilística. Sua derivada facilita o processo de otimização durante o treinamento da rede, em técnicas baseadas em gradiente.

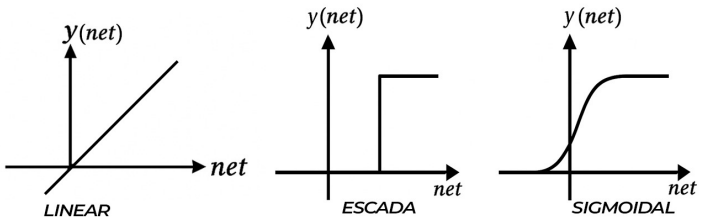

Figura 2.2: Funções de ativação.
Fonte: Adaptado de Silva e Vieira (2022)

O primeiro modelo de Rede Neural Artificial (RNA) foi proposto por Rosenblatt (1958) é denominado *Perceptron*. Esse modelo consistia em um neurônio artificial com a capacidade de identificar superfícies de separação para padrões linearmente separáveis, representando um avanço significativo no campo da inteligência artificial.

O *Perceptron*, como apresenta Silva e Vieira (2022), foi concebido para simular o funcionamento de um neurônio biológico, incorporando suas principais características, tais como a capacidade de processar entradas ponderadas e gerar uma saída com base em uma função de ativação. O modelo é ilustrado na Figura 2.3. Com o desenvolvimento do *Perceptron*, tornou-se possível conectar múltiplas unidades para formar redes neurais artificiais mais complexas, capazes de aprender padrões não lineares e resolver problemas de maior complexidade.

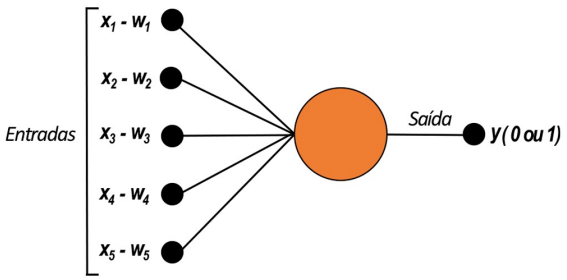

Figura 2.3: Modelo de *Perceptron*
Fonte: Rosenblatt (1958)

O modelo de RNA é composto por quatro componentes principais, conforme descrito por Fleck (2016):

- Conjunto de Sinapses: Representa as conexões entre os neurônios artificiais, onde cada sinapse é associada a um peso sináptico. Esses pesos determinam a intensidade e o tipo de influência (excitatória ou inibitória) que uma entrada exerce sobre o neurônio.

- Integrador: Responsável pelo processo de soma ponderada dos sinais de entrada. O integrador combina as entradas multiplicadas por seus respectivos pesos sinápticos, gerando um valor único que será processado pela função de ativação.

- Função de Ativação: Define a amplitude do sinal de saída do neurônio com base no valor resultante do integrador. A função de ativação introduz não linearidades ao sistema, permitindo que a rede modele relações complexas entre variáveis de entrada e saída.

- Bias (Tendência): Um valor externo aplicado a cada neurônio, que tem o efeito de deslocar a função de ativação. O bias é essencial para ajustar o comportamento do neurônio e melhorar a capacidade de aprendizado da rede.

O processo de aprendizado em uma RNA ocorre por meio de iterações sucessivas, nas quais os pesos sinápticos são ajustados de forma sistemática. Esse ajuste visa minimizar o erro entre a saída gerada pela rede e o valor desejado, permitindo que a RNA generalize soluções para o problema em questão. De acordo com Miranda, Freitas e Faggion (2009), treinar uma rede neural consiste em otimizar sua matriz de pesos de modo que o vetor de saída corresponda, de maneira aproximada, ao valor esperado para cada vetor de entrada. Esse processo é fundamental para garantir que a rede seja capaz de realizar inferências precisas em dados nunca antes vistos.

Os mecanismos de aprendizado em RNA podem ser classificados em duas categorias principais: aprendizado supervisionado e aprendizado não supervisionado. No aprendizado supervisionado, a rede é treinada utilizando um conjunto de dados rotulados, onde cada entrada está associada a uma saída desejada. Durante o treinamento, a rede ajusta seus pesos comparando suas previsões com os valores esperados, utilizando algoritmos como o *backpropagation* para minimizar o erro (BEZERRA, 2016).

Já no aprendizado não supervisionado, a RNA não recebe saídas predefinidas para comparação. Em vez disso, o algoritmo identifica padrões entre os dados de entrada, agrupando-os em classes ou categorias com base em similaridades. Esse método é muito utilizado em tarefas como *clustering* e redução de dimensionalidade (MIRANDA; FREITAS; FAGGION, 2009).

Ambas as abordagens contribuem significativamente para o desenvolvimento e aprimoramento de Redes Neurais Profundas (*Deep Learning*), permitindo a aplicação desses modelos em uma ampla gama de problemas complexos, desde reconhecimento de padrões até o processamento de linguagem natural.

2.2 Redes Neurais com Aprendizado Profundo

O Aprendizado Profundo (*Deep Learning*) constitui um subcampo da Inteligência Artificial (IA), mais especificamente do Aprendizado de Máquina (*Machine Learning*), que visa capacitar sistemas computacionais a aprimorar seu desempenho por meio da experiência e da análise de dados. Esse paradigma de aprendizado se destaca por sua capacidade de processamento computacional avançado, permitindo a representação do mundo em diferentes níveis de complexidade, desde conceitos simples até abstrações sofisticadas (REIS, 2021).

De acordo com Goodfellow, Bengio e Courville (2016), o Aprendizado Profundo é uma técnica de Aprendizado de Máquina fundamentada em Redes Neurais Artificiais (RNA) com múltiplas camadas, também denominadas Redes Neurais Profundas. Essas redes se diferenciam das redes neurais tradicionais pelo número elevado de camadas, as quais realizam um processamento hierárquico dos dados, permitindo o reconhecimento de padrões em múltiplos níveis de abstração. Cada camada em uma Rede Neural Profunda é responsável por extrair e transformar características dos dados de entrada em representações cada vez mais abstratas, o que possibilita a captura de informações complexas e não lineares presentes nos dados (FLECK, 2016). A Figura 2.4 apresenta um exemplo de uma RNA *Perceptrons* de múltiplas camadas (MLP).

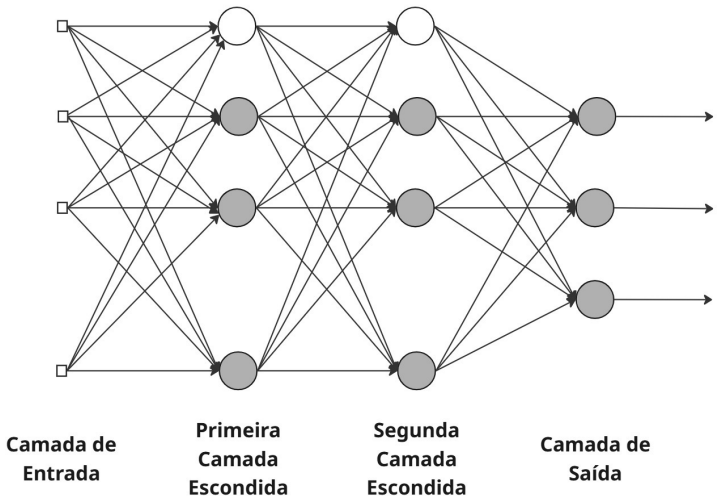

Figura 2.4: Arquitetura MLP múltiplas camadas
Fonte: Fleck (2016)

Ferreira (2017) complementa que o Aprendizado Profundo engloba um conjunto de técnicas que utilizam camadas não lineares para processar informações, viabilizando tanto o aprendizado supervisionado quanto o não supervisionado. Essas técnicas são amplamente aplicadas em tarefas como análise de padrões, classificação de dados e reconheci-

mento de características, demonstrando eficácia em ambientes que demandam processos de generalização e precisão.

Nilsen (2015) destaca que as Redes Neurais Profundas (RNP) têm se mostrado uma ferramenta importante na resolução de uma ampla gama de problemas complexos, tais como reconhecimento de imagens, classificação de dados, reconhecimento de fala e processamento de linguagem natural, entre outros. Essas redes são caracterizadas por sua capacidade de aprender representações hierárquicas e abstratas dos dados, o que as torna particularmente adequadas para tarefas que envolvem grandes volumes de informações.

Conforme Faceli et al. (2021), as RNP se diferenciam das redes neurais tradicionais por possuírem mais de uma camada oculta, podendo incluir diversas camadas intermediárias entre a entrada e a saída. Cada uma dessas camadas é composta por um número variável de neurônios, cuja quantidade e configuração dependem diretamente da aplicação específica da rede. Essa arquitetura multicamadas permite que as RNP realizem um processamento sequencial e incremental dos dados, onde cada camada é responsável por extrair e transformar características em níveis crescentes de abstração.

De acordo com Faceli et al. (2017), as arquiteturas RNP podem ser classificadas em quatro categorias principais, cada uma com características e aplicações específicas. Essas categorias refletem a evolução e a diversificação das técnicas de Aprendizado Profundo, permitindo a resolução de problemas complexos em diferentes domínios. A seguir, descrevem-se as principais arquiteturas:

- Redes Neurais Profundas (RNP): As RNP são a forma mais básica de redes neurais profundas, caracterizadas por múltiplas camadas ocultas entre a entrada e a saída. Um exemplo clássico é o *Perceptron* Multicamadas (*Multilayer Perceptron* - MLP), que utiliza uma estrutura feedforward, onde os dados fluem em uma única direção, da entrada para a saída. Essas redes são amplamente utilizadas em tarefas de classificação e regressão, sendo fundamentais para o desenvolvimento de modelos de aprendizado supervisionado.

- Redes Neurais Convolucionais (RNC): As RNC são especializadas no processamento de dados com estrutura espacial, como imagens. Sua arquitetura inclui camadas de convolução, que aplicam filtros para extrair características locais, camadas não lineares (como ReLU) para introduzir não linearidades, e camadas de agrupamento (*pooling*) para reduzir a dimensionalidade dos dados. Essas redes são amplamente utilizadas em tarefas de reconhecimento de imagens, detecção de objetos e processamento de vídeo.

- Redes Neurais Recorrentes (RNR): As RNR são projetadas para processar dados sequenciais, como séries temporais, texto e áudio. Sua principal característica é a presença de conexões cíclicas, que permitem a manutenção de informações ao longo

do tempo, conferindo-lhes uma "memória"de longo prazo. Variantes como as redes LSTM (*Long Short-Term Memory*) e GRU (*Gated Recurrent Units*) foram desenvolvidas para mitigar problemas como o desaparecimento do gradiente, ampliando sua eficácia em tarefas como tradução automática, geração de texto e reconhecimento de fala.

- Arquiteturas Emergentes: Essa categoria engloba redes neurais profundas que surgiram para resolver problemas específicos ou combinar vantagens de diferentes arquiteturas. Exemplos incluem Redes Neurais Recorrentes Multidimensionais, que processam dados em múltiplas dimensões, e Autoencoders Convolucionais, que combinam técnicas de CNNs com métodos de aprendizado não supervisionado para tarefas como redução de dimensionalidade e geração de dados. Essas arquiteturas refletem a constante evolução do campo, adaptando-se às demandas de aplicações cada vez mais complexas.

As redes de aprendizagem profunda (*Deep Learning*) permitem a modelagem de padrões complexos por meio da organização hierárquica de abstrações, características e conceitos. Nesse contexto, à medida que se avança para níveis mais altos de uma rede neural, as representações dos dados tornam-se progressivamente mais abstratas e não lineares (FLECK, 2016). Cada nível superior é construído com base nas representações aprendidas nos níveis inferiores, onde as camadas iniciais capturam características de baixo nível, como bordas e texturas, enquanto as camadas mais profundas identificam padrões de alto nível, como formas e objetos complexos. Essa hierarquia de representações é fundamental para a capacidade das redes neurais de resolver problemas de alta complexidade.

De acordo com Oliveira e Junior (2023), a organização desses níveis hierárquicos é realizada por meio das camadas que compõem a arquitetura de uma RNA. A profundidade de uma rede neural está diretamente relacionada ao número de camadas que realizam operações não lineares sobre os dados de entrada. Quanto maior o número de camadas, maior é a profundidade da rede e, consequentemente, sua capacidade de representar relações não lineares e complexas. Essa característica é essencial para o sucesso de modelos de aprendizagem profunda em tarefas como reconhecimento de imagens, processamento de linguagem natural e predição de séries temporais.

A flexibilidade arquitetural das Redes Neurais Profundas, aliada ao seu poder de generalização, tem possibilitado avanços significativos em áreas como visão computacional, onde são utilizadas para reconhecer padrões em imagens, e no processamento de linguagem natural, onde contribuem para a tradução automática e a geração de texto. Além disso, sua aplicação em reconhecimento de fala e classificação de dados têm demonstrado resultados robustos, consolidando-as como uma das principais abordagens no campo do Aprendizado de Máquina.

2.3 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN, do inglês *Natural Language Processing*), como afirma Santos (2015), é um campo da inteligência artificial que tem como principal objetivo permitir que máquinas compreendam, interpretem e gerem linguagem humana de forma significativa possibilitando interações mais naturais entre pessoas e sistemas inteligentes.

O Processamento de Linguagem Natural (PLN) está cada vez mais integrado ao cotidiano, presente em diversas interações digitais. Segundo Lavezzo e Conceicao (2023), essa tecnologia é amplamente utilizada em mecanismos de busca, como *Google* e *Bing*, em assistentes virtuais, como Alexa e *Google Assistant*, em sistemas de recomendação personalizados de plataformas como Amazon e Netflix, além de ser a base dos principais tradutores de idiomas contemporâneos. No entanto, apesar de sua ampla aplicabilidade, o PLN enfrenta desafios significativos devido à complexidade inerente à linguagem humana, que envolve ambiguidades, variações contextuais e nuances culturais, tornando seu desenvolvimento e aprimoramento um campo de estudo dinâmico e interdisciplinar.

O PLN concentra seus processos na interação entre máquinas e a linguagem humana, abordando aspectos fundamentais da comunicação, como som, estrutura e significado (SANTOS, 2015). Por meio de técnicas avançadas, o PLN é capaz de processar grandes volumes de dados textuais, analisando, compreendendo, interpretando e até gerando linguagem que se aproxima ou corresponde à forma como os seres humanos se expressam (LAVEZZO; CONCEICAO, 2023). Essa capacidade torna o uma ferramenta essencial para diversas aplicações que envolvem a manipulação e o entendimento de textos e fala.

O principal objetivo do PLN é processar a linguagem humana de modo a viabilizar a execução de uma variedade de tarefas e aplicações. Entre elas, destacam-se a extração de informações de textos, técnicas de classificação de documentos, tradução automática, geração de resumos, interpretação e geração de linguagem natural, correção ortográfica e reconhecimento de voz, entre outras funcionalidades ((SANTOS, 2015). Essas aplicações têm sido fundamentais para o desenvolvimento de sistemas inteligentes que facilitam a interação humano-computador e ampliam as possibilidades de uso da tecnologia em diferentes contextos.

O PLN é considerado um dos pilares iniciais da pesquisa em inteligência artificial, uma vez que representa a interface natural entre humanos e máquinas. Um dos primeiros marcos na história do NLP foi o experimento realizado em 1954 pela Universidade de Georgetown em parceria com a IBM. Esse projeto pioneiro demonstrou a viabilidade do processamento automático de linguagem ao traduzir mais de 60 frases do russo para o inglês de forma automatizada (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011). Esse feito foi um passo significativo para o avanço das técnicas de NLP e para a consolidação dessa área como um campo essencial no desenvolvimento da inteligência artificial.

2.3.1 Técnicas de Processamento de Linguagem Natural

O desenvolvimento de sistemas baseados em PLN envolve uma série de técnicas fundamentais que permitem a manipulação e interpretação de dados textuais. Essas técnicas são essenciais para que modelos computacionais compreendam a estrutura e o significado da linguagem humana. A seguir, são apresentadas as principais abordagens utilizadas no PLN, conforme discutido por Caseli e Nunes (2024):

- **Tokenização:** Processo de segmentação do texto em unidades menores, como palavras ou frases. A tokenização, ou segmentação de palavras, consiste em dividir sequências de caracteres em um texto, identificando os limites onde uma palavra termina e outra começa, sendo os elementos identificados denominados tokens (PALMER, 2010). Essa técnica é amplamente utilizada em linguística computacional para processar linguagens artificiais e naturais (AHO et al., 2007).

- **Remoção de *Stopwords***: Eliminação de palavras muito frequentes, mas com pouco valor semântico, como "de", "a", "o" e "para". *Stopwords* são palavras muito frequentes que geralmente têm pouco ou nenhum peso semântico no contexto da análise, como artigos, preposições e conjunções. Por exemplo, palavras como "o", "de"e "e"são frequentemente removidas para reduzir a dimensionalidade dos dados e focar nos termos mais relevantes, com isso, é uma etapa essencial no pré-processamento de texto, especialmente em tarefas de Processamento de Linguagem Natural (PLN). Essa técnica é amplamente usada para melhorar a eficiência dos algoritmos, economizar recursos computacionais e permitir uma análise mais focada e relevante dos textos (Pang e Lee, 2008; UFABC Divulga Ciência, 2021).

- **Lematização e *Stemming*:** São duas técnicas que reduzem as palavras às suas formas base. Caseli e Nunes (2024) apresentam que a lematização é uma técnica que visa transformar palavras em suas formas básicas ou "lemas", eliminando flexões verbais, nominais e pluralizações. A técnica de *stemming* vai analisar cada palavra individualmente e reduzi-la à sua raiz.

- **Normalização e Correção Ortográfica:** A normalização textual é um processo fundamental no Processamento de Linguagem Natural (NLP), pois busca a padronização dos dados textuais, eliminando variações desnecessárias, como diferenças de capitalização e a presença de pontuações. Segundo Santos e Oliveira (2020, 36), essa etapa tem o objetivo de aumentar a precisão analítica, sendo idealmente aplicada após a tokenização. Um exemplo clássico desse processo envolve as palavras "Manga" e "manga", que podem apresentar significados distintos dependendo do contexto.

- ***Bag-of-Words (BoW):*** É uma técnica caracterizada pela representação de textos com base na frequência de ocorrência das palavras, desconsiderando sua ordem ou estrutura sintática. Conforme descrito por Fonseca (2020, 12), essa técnica transforma textos em informações numéricas, facilitando a análise e a extração de características relevantes para diferentes aplicações computacionais.

- ***TF-IDF (Term Frequency - Inverse Document Frequency):*** Técnica estatística amplamente utilizada no PLN para a extração de características textuais. Conforme descrito por Fonseca (2020, 21), essa abordagem representa um avanço em relação ao método *Bag of Words* (BoW), pois não apenas contabiliza a frequência de um termo em um documento, mas também avalia sua relevância considerando a distribuição desse termo em um conjunto de documentos.

O PLN abrange um conjunto de técnicas fundamentais que permitem a conversão de textos em dados estruturados para análise computacional. Entre essas técnicas, destacam-se a tokenização, normalização, remoção de *stopwords, Bag of Words (BoW) e TF-IDF*, que são essenciais para que sistemas de inteligência artificial processem e compreendam a linguagem humana.

Conforme observado por Premediba (2021), essas abordagens, quando aplicadas em conjunto, possibilitam a transformação de textos em representações numéricas, viabilizando sua análise por algoritmos de aprendizado de máquina. No entanto, a eficácia de cada técnica pode variar de acordo com o contexto e o objetivo da aplicação, exigindo a escolha apropriada de métodos para garantir uma modelagem eficiente dos dados textuais.

2.4 Reconhecimento Facial e de Emoções

O reconhecimento facial é uma tecnologia baseada em métodos avançados de inteligência artificial e visão computacional, que possibilita a identificação de indivíduos por meio da análise de características biométricas únicas, como a geometria dos traços faciais e a distribuição espacial de pixels em imagens digitais (KOUJAN et al., 2020). O autor ainda destaca que essa técnica tem ganhado destaque devido à sua eficiência em operar em tempo real, mesmo diante de desafios como variações de iluminação, diferentes ângulos de captura e obstruções parciais, demonstrando alta robustez em ambientes não controlados. Além disso, o uso de algoritmos de aprendizado profundo tem aprimorado significativamente a precisão e a confiabilidade desses sistemas, ampliando suas aplicações em áreas como segurança, autenticação biométrica e monitoramento inteligente.

Teixeira, Granger e Koerich (2021) destacam que o avanço dos algoritmos de redes neurais profundas, em especial das redes neurais convolucionais (RNC), tem sido um fator determinante para o aprimoramento de sistemas capazes de extrair padrões complexos e sutis contidos em imagens faciais. Esses algoritmos, que utilizam tanto abordagens de

aprendizado supervisionado quanto não supervisionado, realizam a segmentação e o processamento eficiente de dados visuais, viabilizando a identificação precisa de pontos de referência que definem as características únicas da face humana. Os autores ainda afirmam que essa evolução tem permitido maior precisão e confiabilidade na detecção e reconhecimento facial, mesmo em condições desafiadoras, como variações de pose, iluminação e expressões faciais.

No contexto do reconhecimento facial e da análise de emoções, é essencial diferenciar emoções de sentimentos para aprimorar a precisão dos sistemas de interpretação afetiva. Segundo Ekman (1992), as emoções são respostas biológicas e automáticas a estímulos externos, manifestadas por meio de padrões faciais inatos e universais. Ekman (1992) identifica seis emoções básicas, apresentadas na Figura 2.5 – medo, nojo, raiva, surpresa, felicidade e tristeza – que se expressam de forma consistente em diferentes culturas, sugerindo uma base evolutiva compartilhada para essas reações. Em contraste, os sentimentos representam a interpretação consciente e subjetiva dessas respostas emocionais, envolvendo processos cognitivos que influenciam a percepção individual dos estados afetivos. Essa distinção auxilia no desenvolvimento de sistemas de reconhecimento facial capazes de capturar não apenas as expressões faciais, mas também de contextualizá-las em relação aos estados emocionais e sentimentais subjacentes.

Figura 2.5: Seis expressões faciais emocionais básicas.
Fonte: Ekman (1992)

Essa diferenciação é contribui para o desenvolvimento de tecnologias de reconhecimento facial, pois permite que os sistemas não apenas identifiquem as expressões faciais, mas também interpretem a intensidade e as nuances dos estados emocionais subjacentes. Estudos mais recentes têm integrado os fundamentos propostos por Ekman com avanços em aprendizado profundo, demonstrando que tais abordagens podem melhorar a acurácia dos algoritmos em contextos reais – como segurança, saúde mental e análise de comportamento do consumidor – ao reconhecer e classificar emoções em tempo real (DHALL et al., 2016). Dessa forma, ao combinar os conceitos clássicos de Ekman com inovações tecnológicas, os sistemas de reconhecimento facial podem oferecer uma análise mais sofisticada e adaptativa, considerando tanto as manifestações automáticas quanto as interpretações subjetivas dos estados afetivos (EKMAN, 1992).

As aplicações práticas dessas tecnologias são amplas e diversificadas. Em segurança pública, por exemplo, câmeras inteligentes equipadas com sistemas de reconhecimento facial e de emoções podem identificar e monitorar comportamentos suspeitos, contribuindo

para a prevenção de atividades criminosas (TEIXEIRA; GRANGER; KOERICH, 2021). No setor de marketing, a análise de emoções permite a avaliação da resposta dos consumidores a produtos, serviços e campanhas publicitárias, possibilitando o ajuste de estratégias comerciais de acordo com os perfis emocionais identificados (COSTA et al., 2021).

Entretanto, o avanço dessas tecnologias impõe desafios significativos, sobretudo no que tange à privacidade, à segurança dos dados e à ética. O processamento de informações biométricas sensíveis pode representar riscos de invasão de privacidade e de discriminação, exigindo o estabelecimento de normas e regulamentações rigorosas para o uso responsável (COSTA et al., 2021). Além disso, a presença de vieses nos algoritmos — que podem levar a menores índices de acurácia para determinados grupos étnicos ou de gênero — demanda uma constante revisão e aprimoramento dos modelos computacionais, de forma a promover maior inclusão e equidade ((KOUJAN et al., 2020).

Perspectivas futuras apontam para a integração desses sistemas com outras tecnologias emergentes, como a Internet das Coisas (IoT) e plataformas de *big data*, ampliando o potencial de análise em larga escala e a personalização de serviços. A evolução contínua dos algoritmos, associada à disponibilidade crescente de bases de dados diversificadas, deverá contribuir para o aprimoramento da acurácia e da confiabilidade dos sistemas de reconhecimento facial e de emoções, fomentando aplicações inovadoras e socialmente responsáveis ((TEIXEIRA; GRANGER; KOERICH, 2021).

2.4.1 Fundamentos do Reconhecimento Facial

Os fundamentos do reconhecimento facial envolvem um conjunto de técnicas que possibilitam a identificação e verificação de indivíduos por meio da análise de suas características faciais. Esse processo inicia-se com a detecção do rosto em uma imagem, seguido pela extração e representação dos traços faciais, e, por fim, pela comparação desses dados com padrões previamente armazenados em uma base de dados (WANG; DENG, 2018). O autor destaca que os algoritmos modernos utilizam redes neurais profundas – especialmente as convolucionais – para aprender representações discriminantes dos rostos, o que permite alcançar elevados níveis de acurácia mesmo em condições adversas, como variações de iluminação e ângulos de visão.

De acordo com Hassaballah e Aly (2015), os sistemas de reconhecimento facial funcionam por meio da identificação e extração de características únicas presentes nas imagens dos rostos, como contornos, texturas e posições de elementos faciais. Esses algoritmos são treinados com grandes volumes de dados para aprender a reconhecer padrões faciais, tornando-se cada vez mais robustos diante de variações, como diferenças de iluminação, pose e expressões. Esse processo permite que os sistemas se adaptem às sutilezas presentes na identificação de cada indivíduo.

Silva (2019) destaca que a Rede Neural Convolucional (RNC) é a mais utilizada no

processo de reconhecimento facial. Essas redes são compostas por diferentes camadas, incluindo a camada de entrada, as camadas de convolução, as camadas de *pooling*, as camadas totalmente conectadas e a camada de saída. Cada uma dessas camadas desempenha um papel específico na propagação do sinal de entrada. Além disso, cada camada é formada por múltiplos planos celulares, nos quais os neurônios compartilham características idênticas. Isso significa que eles executam a mesma função em diferentes regiões da imagem analisada, permitindo que a rede aprenda padrões visuais de forma eficiente. A estrutura da RNC é apresentada na Figura 2.6.

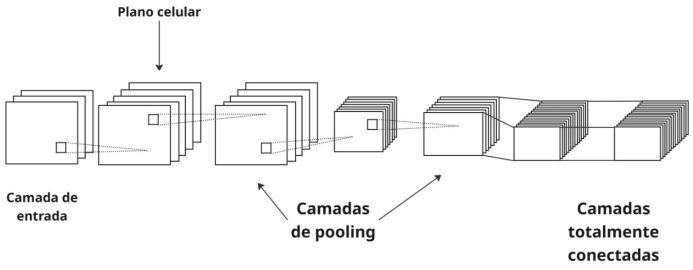

Figura 2.6: Estrutura da RNC e seus componentes: camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas.

Fonte: Silva (2019)

Com o avanço da inteligência artificial, métodos baseados em *deep learning*, especialmente RNC, tornaram-se predominantes no reconhecimento facial. O *FaceNet*, desenvolvido por Schroff, Kalenichenko e Philbin (2015), utiliza uma RNC para mapear imagens faciais em um espaço vetorial de alta dimensão, onde a similaridade entre rostos é determinada pela distância euclidiana entre seus respectivos vetores. A abordagem adota a função de perda *Triplet Loss*, juntamente com um método de mineração *online* de tripletos, para otimizar a precisão do modelo. Como resultado, o FaceNet alcançou 99,63 por cento de acurácia no conjunto de dados *Labeled Faces in the Wild* (LFW), estabelecendo um novo patamar para a área na época.

O *DeepFace*, desenvolvido por Taigman et al. (2014), introduz um processo de alinhamento facial 3D antes da extração de características, o que aumenta a robustez do modelo frente a variações de pose e iluminação. Esse modelo emprega uma rede neural profunda de nove camadas, com mais de 120 milhões de parâmetros, sendo treinado com um conjunto massivo de quatro milhões de imagens de usuários do Facebook. Como resultado, o *DeepFace* atingiu 97,35 por cento de acurácia no LFW, reduzindo significativamente a taxa de erro em relação a abordagens anteriores.

O *OpenFace*, uma implementação de código aberto baseada nos conceitos do *FaceNet*, foca na eficiência computacional, tornando o reconhecimento facial mais acessível para

dispositivos com recursos limitados. Ele utiliza a biblioteca *Torch* e aplica a função Triplet Loss para otimizar o aprendizado da representação facial. Embora sua precisão possa ser ligeiramente inferior em relação ao *FaceNet* e *DeepFace*, o *OpenFace* é amplamente utilizado devido ao seu equilíbrio entre desempenho e custo computacional.

A avaliação de desempenho em sistemas de reconhecimento facial é fundamental para garantir a eficácia da identificação e verificação de indivíduos. Entre as principais métricas utilizadas nesse contexto, destacam-se a precisão, a revocação e o *F1-score*, que permitem mensurar a acurácia dos algoritmos e sua capacidade de generalização.

A precisão refere-se à proporção de verdadeiros positivos em relação ao total de instâncias identificadas como positivas. Essa métrica contribui com sistemas que exigem baixa taxa de falsos positivos, como em aplicações de segurança e autenticação biométrica. Formalmente, como apresentado abaixo, o cálculo de precisão é expresso por TP, do qual representa os verdadeiros positivos e FP que representa os falsos positivos (RIZQYAWAN et al., 2021).

$$\text{Precisão} = \frac{TP}{TP + FP}$$

A revocação, por sua vez, mede a proporção de verdadeiros positivos recuperados em relação ao total de instâncias positivas existentes. Em sistemas de reconhecimento facial, uma alta revocação é desejável quando se busca minimizar a taxa de falsos negativos, garantindo que todas as faces corretas sejam identificadas. Sendo FN os falsos negativos, Rizqyawan et al. (2021) define o cálculo de recall como:

$$\text{Revocação} = \frac{TP}{TP + FN}$$

O *F1-score* representa a média harmônica entre precisão e revocação, fornecendo um equilíbrio entre ambas as métricas. Essa métrica é especialmente útil quando há um desequilíbrio entre os dados positivos e negativos, proporcionando uma visão mais abrangente da performance do sistema. Rizqyawan et al. (2021) apresenta a fórmula do *F1-score* por:

$$F1\text{-}score = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

No estudo de Rizqyawan et al. (2021), foi realizada uma análise comparativa de diferentes classificadores supervisionados aplicados ao reconhecimento facial, incluindo *Support Vector Machines (SVM)*, *Random Forest*, *Decision Tree* e *AdaBoost*. Os resultados indicaram que o SVM obteve o melhor desempenho em todas as métricas avaliadas, alcançando 98 por cento de precisão, revocação e *F1-score*, superando os demais classificadores. Esse alto desempenho está associado à capacidade do SVM de encontrar hiperplanos otimizados para separação das classes, reduzindo erros de classificação e aumentando a robustez do modelo.

Outro aspecto essencial para o sucesso dos sistemas de reconhecimento facial é a normalização das imagens, que pode incluir alinhamento e ajustes geométricos para reduzir variações de iluminação e pose. De acordo com Gomes e Araujo (2021), técnicas de pré-processamento, como filtros de mediana e equalização de histograma, contribuem na melhoria da qualidade das imagens faciais, contribuindo para uma extração de características mais consistente e aumentando a precisão das comparações entre diferentes rostos.

Em resumo, os fundamentos do reconhecimento facial residem na detecção, normalização e classificação de padrões faciais. Conforme Cao et al. (2018), essas etapas são realizadas por meio de técnicas que envolvem desde transformações geométricas para alinhar imagens até métodos avançados de normalização para lidar com variações ambientais. Essa abordagem melhora a precisão da identificação de indivíduos, apesar dos desafios impostos por variações ambientais, tornando os sistemas cada vez mais eficientes para diversas aplicações tecnológicas.

2.4.2 Técnicas para Reconhecimento de Emoções em Imagens

O reconhecimento de emoções em imagens é um campo que combina visão computacional, aprendizado de máquina e psicologia para interpretar estados emocionais a partir de expressões faciais. Esse processo tem aplicações variadas, como interação humano-computador, segurança e análise comportamental. Pesquisas transculturais demonstram que certas expressões, como alegria, tristeza, raiva, medo, surpresa e nojo, são reconhecidas de maneira semelhante em diferentes contextos culturais, sugerindo que há um padrão biológico na percepção emocional (COSTA-VIEIRA; SOUZA, 2014).

A extração de características faciais é essencial para que sistemas computacionais consigam identificar emoções com precisão. O modelo *Facial Emotion Recognition using Convolutional Neural Networks (FERC)*, desenvolvido por Mehendale (2020), utiliza uma abordagem baseada em redes neurais convolucionais para analisar expressões faciais. Esse modelo opera em duas etapas principais, como apresentada na Figura 2.7: na primeira(a), remove-se o fundo da imagem para reduzir ruídos e focar no rosto; na segunda(b), são extraídas características faciais relevantes para a classificação emocional. O autor afirma que esse processo permitiu atingir uma taxa de acerto de 96 por cento na identificação de cinco emoções distintas, demonstrando a eficácia do uso de aprendizado profundo no reconhecimento facial de emoções.

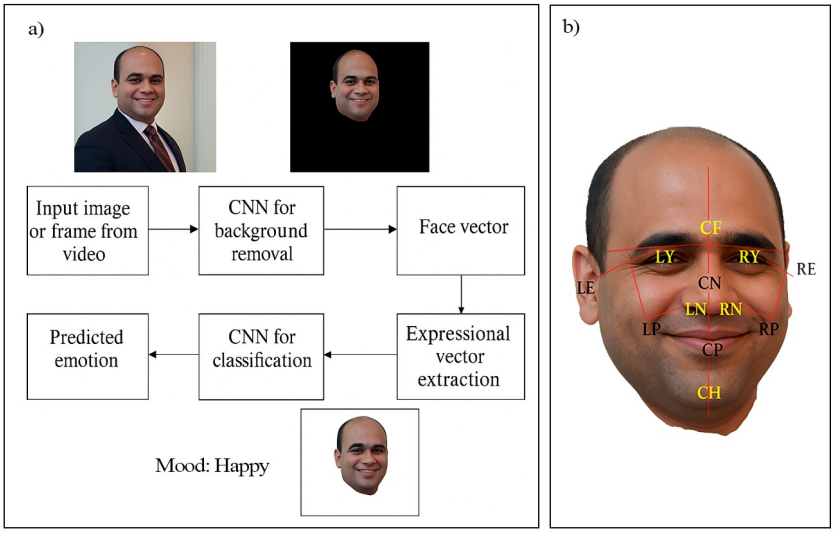

Figura 2.7: Extração de características faciais.
Fonte: Mehendale (2020)

Outra estratégia relevante para melhorar a precisão desses sistemas é o uso de mapas de saliência, que destacam as regiões da face mais expressivas, combinados com redes neurais convolucionais. Essa abordagem foi explorada por Canario (2016) como uma forma de enfatizar pontos críticos na análise das expressões emocionais, tornando a detecção mais robusta. Além disso, o autor apresenta que o uso do *Facial Action Coding System (FACS)*, um sistema de codificação desenvolvido por Ekman para mapear movimentos musculares associados a emoções, tem sido amplamente empregado para otimizar a identificação de padrões expressivos. O FACS permite que algoritmos de aprendizado profundo reconheçam variações sutis nas expressões faciais, aumentando a precisão do reconhecimento automático.

Segundo Canario (2016), a Figura 2.8 ilustra as Unidades de Ação (AU) em uma expressão facial, onde cada AU é associada a um número específico relacionado a um músculo facial. As letras atribuídas a cada AU indicam o grau de intensidade do movimento, variando de 'A' (menor intensidade) até 'E' (maior intensidade). A Figura 2.8 foi adaptada de Whitehill, Bartlett e Movellan (2008).

Figura 2.8: Unidades de Ação (AU) em uma expressão facial
Fonte: Canario (2016)

A avaliação do desempenho dos modelos de reconhecimento emocional, de acordo com Canario (2016) é realizada por meio de métricas como curvas ROC e matrizes de confusão, que ajudam a verificar a taxa de acertos e erros na classificação das expressões. No entanto, ainda há desafios a serem superados. A variabilidade das condições de captura das imagens, incluindo iluminação, ângulos de visão e a presença de elementos obstrutivos, pode afetar a precisão do reconhecimento. Técnicas como remoção de fundo, normalização de imagens e uso de filtros de detecção de bordas são estratégias que ajudam a mitigar esses problemas e a melhorar o desempenho dos modelos (MEHENDALE, 2020).

A avaliação do desempenho dos modelos de reconhecimento emocional é realizada por meio de métricas como curvas ROC e matrizes de confusão, que ajudam a verificar a taxa de acertos e erros na classificação das expressões (CANARIO, 2016). No entanto, ainda há desafios a serem superados. A variabilidade das condições de captura das imagens, incluindo iluminação, ângulos de visão e a presença de elementos obstrutivos, pode afetar a precisão do reconhecimento. Técnicas como remoção de fundo, normalização de imagens e uso de filtros de detecção de bordas são estratégias que ajudam a mitigar esses problemas e a melhorar o desempenho dos modelos (MEHENDALE, 2020).

O reconhecimento facial aplicado à análise de emoções enfrenta desafios significativos que comprometem sua eficácia. Li e Deng (2020) destacam que a transição do reconhecimento de expressões faciais de ambientes controlados para condições mais desafiadoras, como aquelas encontradas no mundo real, introduz variações de iluminação, poses de cabeça e vieses de identidade que afetam negativamente o desempenho dos sistemas de reconhecimento facial. Esses desafios ressaltam a necessidade de abordagens mais robustas e éticas no desenvolvimento de tecnologias de reconhecimento facial para a análise de emoções.

O reconhecimento de emoções em imagens tem evoluído significativamente devido aos avanços no aprendizado profundo e no processamento de imagens. A combinação de CNNs, mapas de saliência e normalização de imagens contribuem para aumentar a precisão dos sistemas, permitindo aplicações práticas em diversas áreas. Contudo, desafios como variações na captura das expressões e diferenças individuais na manifestação emocional ainda precisam ser solucionados para tornar esses sistemas mais confiáveis e aplicáveis em contextos reais. A pesquisa na área continua avançando, explorando novas estratégias para aprimorar a identificação de emoções e fortalecer a interação entre humanos e máquinas por meio da visão computacional (COSTA-VIEIRA; SOUZA, 2014).

2.5 Geração de Conteúdo Automatizado

A evolução das tecnologias de inteligência artificial tem impulsionado a automação de diversas atividades, incluindo a geração de conteúdo automatizada. Com o avanço dos modelos de aprendizado profundo, tornou-se possível produzir textos coesos e contextual-

mente relevantes sem a necessidade de intervenção humana direta (BROWN et al., 2020). Esse avanço tem implicações significativas em diversas áreas, como jornalismo, marketing e atendimento ao cliente, onde a necessidade de escalabilidade e personalização do conteúdo é cada vez mais evidente (RAFFEL et al., 2020).

A automação na geração de conteúdo fundamenta-se em modelos de aprendizado de máquina treinados em extensos volumes de dados textuais. Modelos como o *Gemini* têm demonstrado capacidade de produzir respostas coerentes a partir de instruções em linguagem natural, evidenciando o potencial da inteligência artificial na criação textual.

O uso de arquiteturas baseadas em *Transformer*, como o T5, auxiliam na geração e transformação de texto, permitindo que diversas tarefas linguísticas sejam tratadas de maneira unificada (RAFFEL et al., 2020). Esse modelo adota uma abordagem *text-to-text*, na qual qualquer tarefa de processamento de linguagem pode ser formulada como um problema de conversão de texto, facilitando a personalização do conteúdo gerado e ampliando a aplicabilidade em diferentes domínios.

Além disso, estratégias como a introdução de variações morfológicas nos textos têm sido usadas para tornar os modelos de linguagem mais justos, reduzindo preconceitos linguísticos e melhorando sua capacidade de interpretar diferentes formas do idioma (TAN, 2020). Essa abordagem, conforme demonstrado no estudo pelo autor, permite que os sistemas de automação lidem melhor com variações dialetais e contextuais, reduzindo a discriminação linguística e tornando os modelos mais inclusivos.

Apesar dos avanços, a automação de conteúdo enfrenta desafios como a mitigação de vieses e a necessidade de supervisão humana. Um dos principais desafios éticos envolve a possibilidade de perpetuação de preconceitos e desinformação, uma vez que os modelos são treinados em grandes bases de dados que podem refletir padrões indesejáveis existentes nesses conjuntos de dados (BROWN et al., 2020). Além disso, a transparência na utilização dessas tecnologias é um aspecto crítico, pois a falta de clareza sobre a autoria dos textos gerados por Inteligência Artificial (IA) pode levantar questionamentos sobre credibilidade e confiabilidade do conteúdo produzido (TAN, 2020).

A geração automatizada de conteúdo constitui uma inovação significativa no campo da IA aplicada à comunicação. Modelos como o *Gemini* demonstram a viabilidade da automação na produção de textos coerentes e contextualmente apropriados, desempenhando um papel fundamental em tarefas como raciocínio linguístico, resposta a perguntas baseadas em conhecimento e tradução automática. Esses avanços ampliam a eficiência em diversas áreas, possibilitando aplicações em redação assistida, atendimento automatizado e síntese de informações complexas. No entanto, é fundamental levar em conta os desafios relacionados à ética e à necessidade de supervisão humana no uso dessas ferramentas. Integrar a automação com a revisão humana pode ser a melhor estratégia para assegurar a qualidade e a credibilidade do conteúdo produzido por inteligência artificial (TAN, 2020).

2.5.1 Fundamentos da Automação

A automação tem desempenhado um papel essencial na modernização de diversos setores, tornando processos mais rápidos, eficientes e precisos. Desde a indústria até a agricultura, tecnologias automatizadas visam reduzir a interferência humana, minimizar erros e otimizar a produtividade. Isso se torna possível por meio de sensores avançados, algoritmos inteligentes e sistemas de visão computacional, que permitem a captação e o processamento de dados em tempo real (LECLERC et al., 2020).

Um dos pilares da automação é o uso de sensores e sistemas ópticos para análise do ambiente e detecção de padrões. Na indústria, essa tecnologia é aplicada na inspeção de qualidade, monitoramento de processos e otimização da produção. No setor agrícola, sensores ópticos e técnicas de visão computacional são utilizados para mapear e prever a produtividade das plantações, como demonstrado no estudo de Leclerc et al. (2020), que desenvolveu um sistema de monitoramento baseado em sensores RGB-D e algoritmos de visão computacional para estimar a produtividade de árvores de salgueiro em sistemas agroflorestais. Esse tipo de abordagem permite o monitoramento detalhado do crescimento das plantas e melhora a tomada de decisões no manejo agrícola.

Além disso, técnicas de análise óptica, como a espectroscopia de transmissão, também desempenham um papel importante na automação de processos que exigem medições precisas. Nguyen, Phan e Bui (2020) demonstraram a aplicação desse método na determinação do tamanho de microesferas de poliestireno, destacando como a automação pode aprimorar a precisão na análise de materiais. Essa abordagem pode ser expandida para outras áreas, como controle de qualidade em indústrias químicas e farmacêuticas, onde a medição precisa de partículas e compostos é essencial.

Apesar dos avanços tecnológicos, a automação ainda enfrenta desafios, como a necessidade de adaptação a diferentes condições ambientais. No contexto agrícola, Leclerc et al. (2020) ressaltam que fatores como variações de iluminação e complexidade do ambiente podem afetar a precisão dos sensores. Para superar esses obstáculos, soluções como o uso de sensores múltiplos e algoritmos de filtragem de dados vêm sendo exploradas, tornando os sistemas mais robustos e confiáveis.

2.5.2 Geração de Conteúdo para Redes Sociais

Uma das principais vantagens do uso de IA na criação de conteúdo para redes sociais é sua capacidade de adaptação ao contexto e ao tom desejado. De acordo com Akter et al. (2023), modelos avançados como o Gemini demonstram habilidades linguísticas sofisticadas, permitindo a geração de textos fluidos e contextualmente apropriados. Essa capacidade possibilita que marcas e criadores de conteúdo ajustem suas postagens para refletir sua identidade e engajar o público-alvo de maneira mais autêntica. O autor também ressalta que a aplicação da IA nesse contexto exige um equilíbrio entre automação e au-

tenticidade, uma vez que interações excessivamente mecanizadas podem comprometer a naturalidade da comunicação e reduzir a conexão com os usuários.

Outro aspecto essencial nesse processo é o aprendizado transferível, que possibilita o ajuste de modelos para diferentes nichos e estilos comunicacionais. Raffel et al. (2020) explicam que a conversão de diversas tarefas linguísticas para um formato unificado de entrada e saída textuais permite que modelos de NLP sejam aplicados em uma ampla gama de contextos, incluindo a personalização de mensagens para redes sociais. Esse recurso possibilita que empresas adaptem sua comunicação conforme o público-alvo, criando postagens mais relevantes e alinhadas a diferentes segmentos.

A Figura 2.9 apresenta uma arquitetura tradicional para sistemas de entendimento de linguagem natural (*Natural Language Understanding – NLU*). Nesse modelo, o texto de entrada passa inicialmente por uma camada de processamento sintático, na qual são aplicadas operações como detecção de idioma, segmentação de sentenças, tokenização e análise morfológica e sintática. Em seguida, ocorre a transição para a análise semântica, etapa em que são realizadas tarefas como reconhecimento de entidades nomeadas e identificação de expressões multipalavras. O texto analisado sintaticamente é então processado por um analisador semântico, que gera uma representação lógica da informação contida no texto. Essa representação, em conjunto com bases de conhecimento semântico, serve como entrada para um motor de inferência, permitindo que aplicações finais realizem tarefas como extração de informações, sumarização de textos e resposta a perguntas (SENO; PAIVA; SCHROFF, 2023).

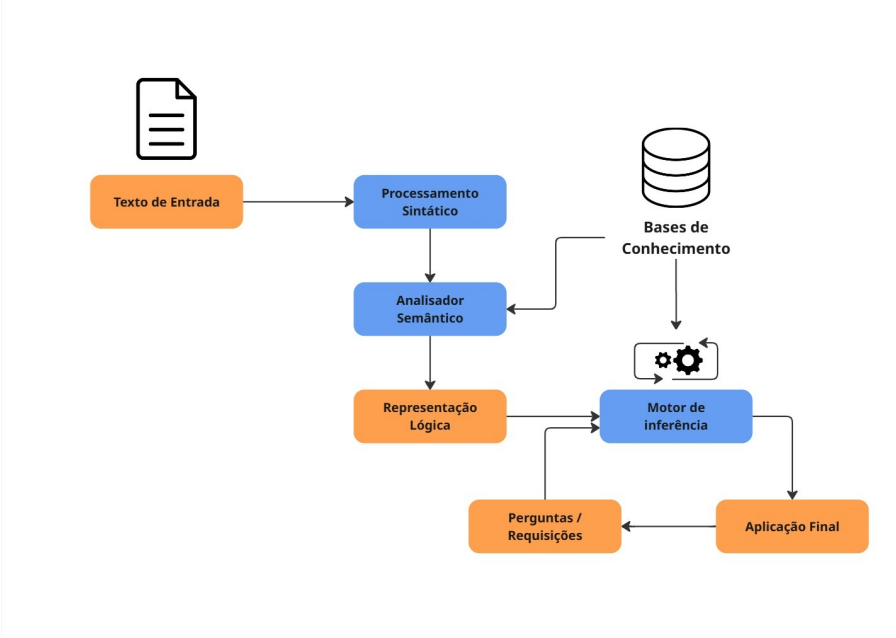

Figura 2.9: Arquitetura tradicional para sistemas de entendimento de linguagem natural.
Fonte: Adapatada de Seno, Paiva e Schroff (2023)

Para avaliar a eficácia do conteúdo gerado por sistemas de NLP, novas métricas têm sido desenvolvidas nos últimos anos, buscando aprimorar a correspondência com avaliações

humanas e capturar nuances mais profundas do texto. Entre essas métricas, destacam-se a *ANLS*, o G-Eval e o NILC-Metrix*, cada uma com abordagens distintas e aplicabilidades específicas.

A ANLS*, proposta por Peer et al. (2024), foi desenvolvida para avaliação de modelos generativos em tarefas como extração de informações e classificação. Diferente de métricas tradicionais baseadas em classificadores discriminativos, a ANLS* utiliza a similaridade normalizada de Levenshtein para comparar a proximidade entre a resposta gerada e a esperada, permitindo lidar melhor com pequenos erros sem penalizar excessivamente o modelo. Além disso, a métrica é compatível com versões anteriores da ANLS, possibilitando comparações diretas com trabalhos anteriores.

Já o G-Eval, introduzido por Liu et al. (2023), adota uma abordagem inovadora ao utilizar modelos de linguagem de grande porte, como o GPT-4, para avaliar a qualidade dos textos gerados. Diferente de métricas tradicionais, ele aplica a técnica de *chain-of-thoughts*, permitindo que o modelo **pense** passo a passo antes de fornecer uma avaliação. Além disso, emprega um paradigma de preenchimento de formulários, estruturando a avaliação com base em critérios específicos, como coerência, relevância e fluência. Essa abordagem resultou em uma correlação de *Spearman* de 0,514 com avaliações humanas em tarefas de sumarização, superando métodos anteriores.

Além da personalização, a capacidade de interagir com diferentes comunidades é um fator determinante para o sucesso de uma estratégia digital. No entanto, como apontado por Tan (2020), modelos de PNL podem apresentar viés linguístico, favorecendo determinadas construções gramaticais e padrões de escrita em detrimento de variações regionais e dialetais. Esse desafio é particularmente relevante nas redes sociais, onde a diversidade de linguagem é um elemento central na construção de comunidades. Para evitar esse problema, é necessário um treinamento contínuo dos modelos, utilizando conjuntos de dados mais representativos e estratégias de ajuste fino que minimizem a exclusão de determinadas formas de expressão.

Para o português brasileiro, uma ferramenta amplamente utilizada é o *NILC-Metrix*, desenvolvido por Leal et al. (2021). Esse sistema compreende 200 métricas extraídas de estudos de discurso, psicolinguística e linguística computacional, permitindo análises detalhadas em diferentes níveis linguísticos. O *NILC-Metrix* se destaca por sua capacidade de medir a coesão, coerência e complexidade textual, sendo aplicado tanto a textos escritos quanto falados. Entre suas métricas, incluem-se índices de diversidade lexical, frequência de palavras, padrões sintáticos e uso de conectivos, proporcionando uma avaliação abrangente da estrutura do texto.

Além da personalização, a capacidade de interagir com diferentes comunidades é um fator determinante para o sucesso de uma estratégia digital. No entanto, como apontado por Tan (2020), modelos de NLP podem apresentar viés linguístico, favorecendo determinadas construções gramaticais e padrões de escrita em detrimento de variações regionais

e dialetais. Esse desafio é particularmente relevante nas redes sociais, onde a diversidade de linguagem é um elemento central na construção de comunidades. Para evitar esse problema, é necessário um treinamento contínuo dos modelos, utilizando conjuntos de dados mais representativos e estratégias de ajuste fino que minimizem a exclusão de determinadas formas de expressão.

Com o avanço da IA, a geração de conteúdo para redes sociais está se tornando mais sofisticada, personalizada e responsiva. No entanto, sua aplicação exige a integração entre automação e curadoria humana, garantindo que os textos gerados mantenham autenticidade, sensibilidade cultural e impacto positivo na interação com os seguidores. O uso inteligente dessas ferramentas pode potencializar a presença digital de marcas e criadores de conteúdo, tornando as redes sociais um espaço mais dinâmico e acessível para diferentes audiências (RAFFEL et al., 2020).

2.6 Ética no Uso de Inteligência Artificial em Redes Sociais

A ética no uso da IA em redes sociais tem se tornado um tema cada vez mais relevante, especialmente diante do crescimento da moderação automatizada de conteúdo. Com a pressão crescente sobre as grandes plataformas para conter discurso de ódio, desinformação e outros tipos de conteúdo prejudicial, muitas empresas têm adotado sistemas baseados em aprendizado de máquina para moderar publicações em larga escala (GORWA; BINNS; KAITZENBACH, 2020). Embora essa tecnologia apresente vantagens como rapidez na detecção de conteúdos problemáticos, seu uso levanta preocupações éticas importantes, como a falta de transparência nos critérios de decisão e o risco de censura indevida.

Conforme Gorwa, Binns e Kaitzenbach (2020), a moderação automatizada funciona principalmente por meio de sistemas de hash-matching e algoritmos preditivos, que identificam conteúdos suspeitos e tomam decisões sobre sua remoção. No entanto, a natureza dessas decisões pode ser questionável, pois os critérios utilizados pelos algoritmos nem sempre são claros para os usuários, e erros podem resultar na exclusão indevida de postagens legítimas. Além disso, a crescente dependência dessas ferramentas pode reforçar a falta de prestação de contas por parte das plataformas, tornando ainda mais difícil contestar decisões equivocadas.

Outro ponto crítico é o impacto que a automação tem sobre a justiça e a equidade na moderação de conteúdo. Como esses algoritmos são treinados em grandes volumes de dados históricos, eles podem reproduzir vieses já existentes na moderação humana. Isso significa que determinados grupos podem ser desproporcionalmente afetados pelas remoções automáticas, enquanto outros conteúdos potencialmente problemáticos passam despercebidos. Segundo Gorwa, Binns e Kaitzenbach (2020), um dos principais desafios

da moderação algorítmica é a dificuldade de garantir que os sistemas sejam justos e transparentes, especialmente quando operam em grande escala e sem supervisão humana.

O viés algorítmico nas redes sociais tem gerado impactos éticos significativos, afetando a equidade e a transparência nas plataformas digitais. Yapo e Weiss (2018) destacam que sistemas de machine learning podem reproduzir e amplificar preconceitos existentes na sociedade, dado que os dados utilizados para treinar esses algoritmos frequentemente refletem desigualdades estruturais.

Um caso emblemático foi identificado em um estudo da *Carnegie Mellon University*, que demonstrou que a empresa Google exibia menos anúncios de empregos bem remunerados para mulheres do que para homens, evidenciando um viés algorítmico na distribuição de oportunidades. Além disso, o uso de reconhecimento facial por grandes empresas de tecnologia tem sido alvo de críticas devido à sua menor precisão na identificação de pessoas negras, o que pode reforçar discriminações sistêmicas (YAPO; WEISS, 2018). Esses exemplos mostram como a falta de supervisão ética na implementação de IA pode perpetuar desigualdades e prejudicar grupos já marginalizados.

Diante desses desafios, diversas iniciativas regulatórias têm sido desenvolvidas para garantir a transparência e a responsabilidade no uso da IA em redes sociais. No Brasil, o marco regulatório da inteligência artificial em discussão no Senado propõe diretrizes para mitigar riscos como o viés algorítmico e a falta de transparência nas decisões automatizadas (AJAYI; UDEH, 2024).

Além disso, a União Europeia implementou regulamentações como o Regulamento Geral sobre a Proteção de Dados (RGPD) e a Lei de Serviços Digitais, que impõem maior responsabilidade às plataformas no tratamento de dados pessoais e na moderação de conteúdos. No setor de tecnologia, práticas inovadoras de recrutamento estão sendo reformuladas para minimizar a discriminação algorítmica e promover maior inclusão, destacando uma crescente preocupação ética na aplicação da IA (AJAYI; UDEH, 2024). Essas iniciativas refletem um esforço global para equilibrar inovação tecnológica e proteção dos direitos fundamentais, garantindo que a inteligência artificial seja desenvolvida e utilizada de maneira ética e responsável.

Além das preocupações com transparência e justiça, a moderação automatizada também obscurece a dimensão política das decisões tomadas pelas plataformas. De acordo com Gorwa, Binns e Kaitzenbach (2020), a remoção de conteúdos é frequentemente tratada como uma questão puramente técnica, ignorando o fato de que diferentes países e comunidades possuem perspectivas variadas sobre o que é aceitável ou não. A padronização global imposta pelos algoritmos pode entrar em conflito com normas culturais locais, dificultando a construção de um ambiente digital verdadeiramente inclusivo.

Diante desses desafios, é essencial que o uso da IA em redes sociais seja conduzido com responsabilidade e supervisão humana. Estratégias como a revisão manual de casos complexos, a transparência nos critérios de moderação e o desenvolvimento de algoritmos

mais sensíveis a diferentes contextos sociais podem ajudar a minimizar os riscos éticos dessa tecnologia. Embora a automação seja uma ferramenta poderosa para lidar com o volume crescente de publicações, seu uso deve ser equilibrado com mecanismos que garantam justiça e responsabilidade na gestão do conteúdo online (GORWA; BINNS; KAITZENBACH, 2020).

2.7 Considerações finais

A fundamentação teórica discutida neste capítulo abordou os principais conceitos que sustentam o desenvolvimento da ferramenta proposta, incluindo redes neurais artificiais, aprendizado profundo, técnicas de processamento de linguagem natural, reconhecimento facial e emocional, além da geração automatizada de conteúdo. Esses elementos forneceram uma base sólida para compreender as tecnologias envolvidas e suas possíveis aplicações no contexto da criação frases de acordo com a valência emocional positiva ou negativa.

Além dos aspectos técnicos, também foram discutidas as implicações éticas relacionadas ao uso de dados sensíveis, como expressões faciais e emoções humanas, destacando a importância da privacidade, do consentimento e da transparência no desenvolvimento de soluções baseadas em IA. Essa dimensão ética é fundamental para garantir que a aplicação da tecnologia ocorra de maneira responsável e em conformidade com os princípios legais e sociais.

A partir desse embasamento teórico, o trabalho segue agora para a etapa metodológica, onde serão descritas as estratégias adotadas para o desenvolvimento prático da ferramenta. O próximo capítulo apresenta a estrutura da pesquisa, os métodos utilizados na implementação do sistema, os detalhes sobre as bases de dados, a arquitetura da solução, os testes aplicados e as limitações observadas, permitindo uma visão clara e sistematizada do processo de construção da proposta.

Capítulo 3

METODOLOGIA

Este capítulo apresenta, de maneira minuciosa, cada procedimento adotado no desenvolvimento da ferramenta de geração automatizada de frases baseada em reconhecimento facial e análise de sentimentos. A abordagem metodológica foi elaborada para garantir ordenação lógica, justificativa de escolhas técnicas e transparência em todo o processo experimental.

3.1 Estrutura Geral da Pesquisa

O processo de pesquisa foi cuidadosamente estruturado em cinco grandes etapas, interligando teoria e prática: (1) Levantamento bibliográfico aprofundado: Inicialmente, buscou-se, por meio de bases digitais, livros, artigos recentes, dissertações e materiais acadêmicos, reunir conceitos fundamentais sobre redes neurais, reconhecimento facial, análise de emoções e processamento de linguagem natural. Essa revisão permitiu identificar o estado da arte na área e fundamentar as decisões técnicas subsequentes; (2) Definição clara dos requisitos e delimitação do escopo: Após a revisão, foram explicitados os requisitos do sistema e o escopo de funcionalidades. Nessa etapa, detalhou-se, por exemplo, quais emoções deveriam ser reconhecidas, o tipo de frases esperadas (formais ou informais), o formato de entrada e saída do sistema, além de se estabelecerem os critérios mínimos de desempenho para considerar o protótipo satisfatório; (3) Projeto detalhado e implementação modularizada: O desenvolvimento foi dividido em submódulos independentes. Cada um (reconhecimento emocional e geração textual) foi planejado, diagramado e codificado com clareza, documentando-se bibliotecas utilizadas, parâmetros técnicos e integrações; (4) Testes sistemáticos e validação experimental: Com o sistema funcional, elaboraram-se testes em condições controladas e reais para avaliar o comportamento da ferramenta. Essa etapa englobou definição de métricas de avaliação, seleção controlada das amostras de imagens e frases, além de processos repetidos para obtenção de resultados estatisticamente confiáveis; (5) Análise crítica dos resultados e sugestões de aprimoramento: Finalizando, os dados reunidos nos testes e no uso experimental foram

submetidos a análise cuidadosa para apontar forças, limitações, causas de eventuais falhas e caminhos de futuras melhorias.

3.2 Levantamento Teórico

Realizou-se uma extensa busca e estudo em fontes nacionais e internacionais abordando:

- Funcionamento detalhado de redes neurais e, especialmente, redes neurais convolucionais para detecção de padrões visuais em faces.
- Técnicas e desafios do processamento de linguagem natural em português, contemplando tokenização, lematização, classificação sentimental e geração de textos automáticos.
- Estratégias de reconhecimento emocional em imagens, análise de datasets, métricas de avaliação e implicações éticas.
- Limitações identificadas em trabalhos semelhantes, como problemas de viés, influência de qualidade de imagem e nuances contextuais em sentimentos.

Esse levantamento guiou toda a arquitetura do sistema, fornecendo embasamento para as escolhas tecnológicas e determinando os limites e possibilidades do projeto.

3.3 Escopo e Arquitetura da Ferramenta

A ferramenta desenvolvida foi idealizada a partir de requisitos práticos e acadêmicos, sendo composta por dois módulos principais, integrados, porém independentes:

- Módulo de reconhecimento emocional em imagens faciais: Este módulo tem como função receber uma imagem, realizar todo o tratamento necessário (adequando o tamanho, brilho, removendo ruídos), identificar e recortar a face presente, e aplicar um classificador de emoções treinado para as seis classes universais (felicidade, tristeza, raiva, medo, surpresa e nojo).
- Módulo de geração automática de frases: Após a identificação da emoção predominante, utiliza-se uma base de frases com polaridade pré-definida (positiva ou negativa) para treinar um algoritmo generativo que cria frases novas correspondentes ao sentimento identificado.

Passo a passo do fluxo sistêmico (*Pipeline*): (1) Recepção da imagem facial pelo sistema, via interface ou leitura de arquivo; (2) Aplicação de técnicas de pré-processamento: redimensionamento, normalização, detecção e alinhamento do rosto; (3) A imagem tratada é então enviada ao classificador, que retorna a emoção detectada; (4) Utiliza-se um

dicionário semântico para converter a emoção (ex: raiva, felicidade) em uma valência (positiva ou negativa); (5) O valor de valência orienta o módulo textual, que acessa o modelo treinado e gera uma frase compatível, utilizando IA generativa; (6) A frase resultante é apresentada ao usuário, podendo ser editada ou usada diretamente em redes sociais.

O desenvolvimento foi feito em *Python*, aproveitando diversas bibliotecas, como*DeepFace, SpaCy, NLTK e a API Gemini.*

3.4 Base de Dados

Para o desenvolvimento e a validação da ferramenta proposta, utilizam-se duas bases de dados principais: uma voltada para o reconhecimento de emoções em expressões faciais e outra composta por frases classificadas em positivas e negativas.

3.4.1 Base de Dados de Imagens Faciais

Foi escolhida a base FER2013, amplamente reconhecida na área científica. Esta base disponibiliza milhares de imagens (em tons de cinza, 48x48 *pixels*) já rotuladas por emoção predominante. A escolha se deu pela padronização dos dados e relevância acadêmica, permitindo treinar e testar o algoritmo em um ambiente controlado.

Processo de preparo das imagens:

- Separação das imagens conforme rótulo.
- Divisão em conjuntos de treino, teste e validação.
- Aplicação de padronização, alinhamento dos rostos e normalização dos *pixels*, assegurando que o classificador tenha entradas consistentes.

3.4.2 Base de Dados de Frases

São utilizadas duas bases principais:

- IMDB PT-BR: Um grande repositório de *reviews* de filmes em português, já classificados como positivos ou negativos. Permitiu ajustar o modelo generativo à linguagem e contextos reais do português brasileiro.
- Sentiment140: Amplia a diversidade textual ao oferecer frases de *tweets* rotulados, aumentando o repertório do modelo e simulando a variabilidade encontrada nas redes sociais.

Ambas passaram por extenso processo de limpeza — retirando termos irrelevantes, uniformizando expressões e aplicando lematização, para maximizar a eficácia do modelo de geração de frases.

3.5 Implementação do Módulo de Reconhecimento Emocional

Para a implementação do módulo de Reconhecimento Emocional segue-se as seguintes etapas:

- As imagens obtidas são processadas em etapas, começando com o reconhecimento e alinhamento facial automático.
- Utiliza-se a *DeepFace*, que integra modelos de redes neurais convolucionais para identificar a emoção predominante.
- A decisão sobre valência da emoção surpresa é feita considerando o segundo maior escore emocional detectado, reduzindo ambiguidades — caso comum nesta categoria.

Todo procedimento foi parametrizado para permitir replicação da abordagem e ajustes conforme novas demandas experimentais.

3.6 Implementação do Módulo de Geração Automática de frases

Para a implementação do módulo de Reconhecimento Emocional segue-se as seguintes etapas:

- Foi realizado o ajuste fino na *API Gemini* utilizando pares de frase/valência extraídos dos datasets — por exemplo: *prompt* frase positiva gera *outputs* positivos, frase negativa *outputs* negativos.
- O modelo foi treinado com três quantidades diferentes de exemplos (2 mil, 4 mil e 8 mil por classe) para investigar o impacto do volume de dados na criatividade e na coerência das frases.
- Após o ajuste, o sistema permite receber comandos como "Gere uma frase com sentimento positivo" e responde de modo relevante.

Todos os processos foram documentados e *scripts* salvos para eventual reuso.

3.7 Testes e Validação

Após a finalização do desenvolvimento, a ferramenta passa por um conjunto de testes controlados com o objetivo de verificar seu desempenho funcional, os passos realizados para os testes são apresentados abaixo:

- O módulo emocional foi testado com amostras do FER2013, medindo acurácia e consistência da detecção das emoções.
- O módulo textual foi avaliado quanto à coerência, diversidade de frases e alinhamento da frase à emoção alvo, utilizando métricas objetivas (quantitativas) e avaliação qualitativa com exemplos.
- Também foram realizados testes integrados: o pipeline completo foi executado com diversos *inputs* reais, avaliando desde a imagem inicial até a frase gerada, em diferentes condições e grupos de teste.

Resultados, tabelas e exemplos concretos estão apresentados no capítulo seguinte do trabalho.

3.8 Limitações e Considerações Éticas

Reconheceu-se que:

- O reconhecimento emocional apresenta subjetividades e nem sempre capta nuances emocionais sutis em faces.
- Há limitações nos *datasets* utilizados, que podem não expressar toda a diversidade cultural e emocional de personas reais.
- No módulo textual, eventuais respostas podem ser genéricas ou não atingir a sensibilidade esperada, reflexo do treinamento limitado pelo escopo dos dados.

No aspecto ético:

- Todos os dados utilizados são públicos e não envolvem coleta de imagens reais sem consentimento.
- Não são armazenadas informações pessoais de usuários.
- O projeto atende às exigências da LGPD e princípios internacionais de proteção à privacidade.

3.9 Considerações Finais

A metodologia apresentada neste capítulo descreve de forma estruturada o processo de desenvolvimento da ferramenta proposta, desde o levantamento teórico até a preparação para os testes e validações.

A partir desta estrutura metodológica, procede-se à apresentação dos resultados obtidos, os quais envolvem tanto a análise do desempenho do sistema de reconhecimento emocional quanto a avaliação da relevância e adequação das frases geradas. No capítulo seguinte, examinam-se os dados coletados durante os testes e discute-se sua efetividade e limitações frente aos objetivos propostos neste trabalho.

Capítulo 4

RESULTADOS

Este capítulo apresenta os principais resultados obtidos a partir da implementação da ferramenta proposta, com base nos conceitos e métodos descritos nos capítulos anteriores. A análise está dividida em seções que evidenciam o funcionamento da solução desenvolvida, os dados utilizados, o desempenho dos módulos de reconhecimento emocional, e o processo de geração automática de conteúdo.

Inicialmente, é apresentado o diagrama de funcionamento geral do sistema, seguido da descrição das bases de dados utilizadas — tanto para o reconhecimento facial quanto para a análise de frases com valência emocional. Em seguida, são detalhados os resultados relacionados à identificação de emoções faciais, categorização das emoções em termos de valência (positiva, negativa ou neutra), e a lógica de geração automática de frases personalizadas para redes sociais.

Por fim, são discutidas as observações gerais sobre o desempenho da aplicação, incluindo as limitações encontradas, os pontos fortes da abordagem adotada e uma análise crítica sobre o potencial de uso em cenários reais. Esses resultados permitem avaliar a viabilidade técnica da solução e sua aplicabilidade prática no contexto de automação de conteúdo com base em expressões emocionais humanas.

4.1 Desenvolvimento do Algoritmo

Nesta seção apresentamos o processo de desenvolvimento do algoritmo de reconhecimento facial com análise de emoções e geração de frases e análise das frases geradas.

4.1.1 Diagrama de funcionamento

No processo de desenvolvimento do trabalho, como apresentado na Figura 4.1, desenvolvemos o fluxo de interação entre algoritmos de visão computacional e geração de linguagem natural, com o objetivo de gerar frases contextualizadas com base na valência emocional extraída de expressões faciais.

Inicialmente, aplica-se um algoritmo de reconhecimento facial e análise de emoções ao conjunto de imagens, a fim de identificar a emoção predominante expressa em cada rosto. A emoção identificada é então classificada por meio de um dicionário de categorização de valência emocional, o qual determina se a emoção é positiva ou negativa. Essa informação é repassada ao módulo responsável pela construção de *prompts* textuais.

O *prompt* gerado segue o formato **Crie uma frase com a valência emocional (positiva ou negativa)**, onde é preenchido conforme a valência classificada anteriormente. Esse comando é então processado por um algoritmo de geração de linguagem natural, o qual produz automaticamente uma frase compatível com a emoção detectada. Dessa forma, temos a combinação de técnicas de inteligência artificial em visão computacional e processamento de linguagem natural, permitindo uma abordagem integrada para reconhecimento emocional e resposta textual automatizada.

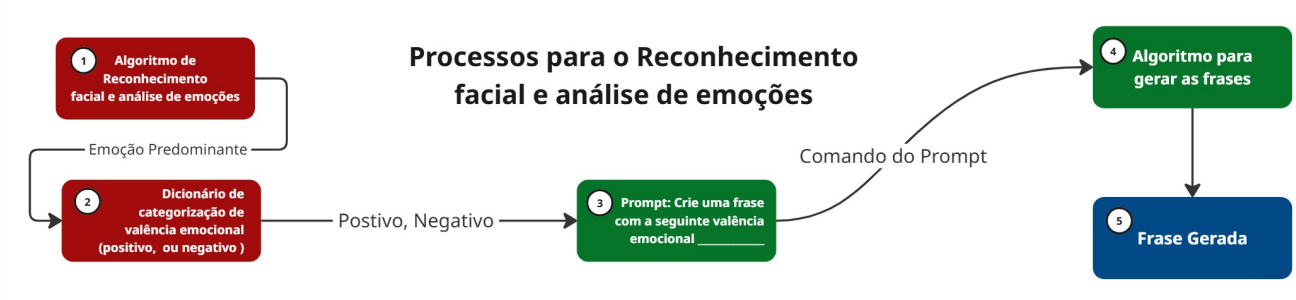

Figura 4.1: Diagrama de fluxo
Fonte: Elaborado pelos autores

4.1.2 Base de dados FER2013

Os algoritmos baseados em inteligência artificial demandam conjuntos de dados específicos para a realização das etapas de treinamento e validação. Neste estudo, foram selecionados dois conjuntos de dados amplamente utilizados na literatura científica. primeiro é o *FER2013 (Facial Expression Recognition 2013)*, disponível publicamente na plataforma *Kaggle* (https://www.kaggle.com/datasets/msambare/fer2013), o qual contém imagens rotuladas de expressões faciais humanas, sendo amplamente empregado em tarefas de reconhecimento de emoções.

O conjunto de dados FER2013 é amplamente utilizado em pesquisas científicas e está licenciado como de uso público, sendo composto por imagens em tons de cinza com resolução de 48x48 *pixel*s. Ele contém um total de 28.709 imagens para treinamento e 3.589 imagens para teste, todas anotadas com rótulos correspondentes a sete categorias emocionais distintas: surpresa, alegria, nojo, medo, tristeza e raiva. A Figura 4.2 apresenta exemplos de expressões faciais de acordo com as emoções reconhecidas extraídas do conjunto de dados FER2013.

Figura 4.2: Expressões de emoções faciais
Fonte: Adaptado da Plataforma *Kaggle*

Este *dataset* é frequentemente empregado em estudos na área de visão computacional, especialmente no desenvolvimento de modelos baseados em algoritmos de inteligência artificial voltados para o reconhecimento de expressões faciais e análise emocional. Neste trabalho, foram utilizadas 30 mil imagens do *dataset* FER2013 para o processo de classificação emocional. As expressões foram organizadas conforme a valência emocional, resultando em um subconjunto composto por 15 mil imagens classificadas como emoções positivas e 15 mil como emoções negativas.

4.1.3 Base de dados IMDB PT-BR

O segundo conjunto de dados utilizado foi retirado do *dataset Sentiment140*(https:// www.kaggle.com/datasets/kazanova /sentiment140/data), composto por 1,6 milhão de *tweets* coletados por meio da API do *Twitter*. Os dados são públicos e foram previamente anotados com rótulos binários de sentimento (0 = negativo, 4 = positivo), possibilitando sua aplicação em modelos de detecção e análise de sentimentos em linguagem natural.

O *dataset* IMDB PT-BR, *dataset* utilizado para classificação de texto. Este *Dataset* é disponibilizado pela plataforma *Kaggle*, a qual contém 50.000 documentos de resenhas (*reviews*) de filmes da plataforma *Internet Movie Database (IMDB)* e apresenta a sua polaridade positiva e negativa, onde possui 25.000 resenhas positivas e 25.000 resenhas negativas (FRED, 2011). Com isso, podemos afirmar que a base de dados está equalizada, ou seja, contém um número igual de dados classificados em positivos e negativos, contribuindo para que não haja enviesamento dos resultados por causa do desbalanceamento das classificações. Este dataset foi utilizado por diversos autores para treinamento e teste de algoritmos de análise sentimental baseados em Inteligência Artificial, como por exemplo, Barbosa et al. (2020) e Goulart et al. (2024).

No pré-processamento dos dados do *dataset* IMDB PT-BR foram aplicadas diversas

técnicas para tratar os dados utilizados, incluindo a remoção de ruídos e de elementos textuais irrelevantes. Essas etapas tiveram como objetivo aprimorar a qualidade dos dados, visando obter melhores resultados na formulação de frases de acordo com a polaridade de sentimento (positiva ou negativa) apresentada.

Os autores Han, Kamber e Pei (2011), abordam as técnicas de pré-processamento, mediante quatro categorias, das quais foram aplicados no dataset IMDB PT-BR, como apresentado no Apêndice A.3 Código para Treinamento do *Dataset*:

- Limpeza e padronização textual: usada para remover ruídos nos dados ou corrigir inconsistências. No *dataset* foram removidos caracteres irrelevantes, tais como, caracteres especiais, caracteres alfanuméricos e sinais de pontuação. Também, foi realizada a padronização do texto para ter somente caracteres minúsculos.

- Remoção de *Stopwords*: *Stopwords* são palavras muito frequentes que geralmente têm pouco ou nenhum peso semântico no contexto da análise, como artigos, preposições e conjunções. Por exemplo, palavras como "o", "de"e "e"são frequentemente removidas para reduzir a dimensionalidade dos dados e focar nos termos mais relevantes, com isso, é uma etapa essencial no pré-processamento de texto, especialmente em tarefas de Processamento de Linguagem Natural (PLN) Pang e Lee (2008). Essas palavras não expressam sentimentos e com isso não agregam ao significado da frase. Para realizar este processo foi utilizada a lista de *stopwords* disponibilizada pela biblioteca de Processamento de Linguagem Natural (PLN) *Spacy*, da qual possui uma lista própria de palavras consideradas irrelevantes.

- Lematização: A lematização é uma técnica em NLP que visa transformar palavras em suas formas básicas ou "lemas", eliminando flexões verbais, nominais e pluralizações. No *dataset* IMDB, por exemplo, termos como "gostando", "gostavam"e "gostam"são reduzidos ao lema "gostar". Isso simplifica os textos e ajuda a unificar variações de palavras, reduzindo a complexidade dos dados e facilitando a análise semântica. No processo de aplicação da técnica de lematização foi utilizada a biblioteca *SpaCy*, da qual oferece recursos robustos para o processo de lematização, onde usa modelos linguísticos treinados para identificar os lemas corretos, mesmo em idiomas complexos como o português brasileiro (UFABC Divulga Ciência, 2021; Hossain et al., 2019). Essa abordagem foi utilizada pois contribui para melhorar a compreensão de textos e a precisão de acertos pelos algoritmos de inteligência artificial utilizados.

- *Label encoding*: os modelos de Aprendizado Profundo e de Aprendizado de Máquina exigem que os dados de entrada e de saída sejam numéricos. Com isso, foi utilizada a biblioteca *Sklearn* , como apresentado no código abaixo, para aplicar a técnica

de *Label encoding* nos dados, transforma as polaridades (positiva e negativa) em valores numéricos (0, 1).

4.1.4 Algoritmo de Reconhecimento facial e de emoções

Para realizar o processo de reconhecimento facial e análise de emoções foi utilizada a biblioteca *DeepFace* (https://pypi.org/project/deepface/0.0.24/), como apresentado no Apêndice A.2 Código para Reconhecimento Facial e Geração de Frase, um *framework open source* desenvolvido em Python, que oferece suporte a diversos modelos de reconhecimento facial e análise emocional. É um modelo originalmente proposto pelo *Facebook AI Research*, que foi um dos primeiros a alcançar uma precisão de 92 por cento quando foi utilizada em testes de desempenho de reconhecimento facial feito por pesquisadores da Universidade de *Massachusetts*, superando o desempenho humano pela primeira vez (TAIGMAN et al., 2014). A biblioteca *DeepFace* utiliza redes neurais profundas para fazer as análises de faces e realizar processos de reconhecimento de características emocionais.

No processo também foi utilizado a base de dados FER2013, composta por imagens faciais em escala de cinza, cada uma rotulada com uma das seis emoções universais. Este *dataset* é composto por milhares de imagens de rostos humanos, em escala de cinza e resolução de 48x48 *pixels*, rotuladas com diferentes emoções. Inicialmente, foi realizado o pré-processamento das imagens, que incluiu a detecção automática das faces utilizando os recursos integrados da *DeepFace*. Após a detecção, as imagens faciais foram alinhadas e normalizadas, garantindo um padrão adequado para a análise posterior.

O processo de reconhecimento facial e a análise de emoções, como apresentado no bloco 1 da Figura 4.1 e apresentado no Apêndice A.2 Código para Reconhecimento Facial e Geração de Frase, iniciou-se com a leitura e pré-processamento das imagens do *dataset*. A partir desse processamento, a biblioteca *DeepFace* foi utilizada com um modelo de RNC previamente treinado para realizar a classificação das emoções presentes nas imagens. A *DeepFace* realizou automaticamente a detecção das faces por meio de técnicas de alinhamento facial com base em pontos de referência.

Para o processo de classificação do sentimento surpresa foi utilizada a seguinte lógica, como apresentado no Apêndice A.2 Código para Reconhecimento Facial e Geração de Frase: será analisado o segundo sentimento predominante, caso seja positvo, a emoção surpresa receberá a valência emocional positiva, caso o segundo sentimento seja negativo, a emoção surpresa receberá a valência emocional negativa.

Em seguida, cada face foi processada pelo módulo de análise emocional, representado pelos blocos 3 e 4 da figura 4.1 e apresentado no Apêndice A.2 Código para Reconhecimento Facial e Geração de Frase, que emprega uma rede neural convolucional previamente treinada para estimar a emoção predominante. A classificação foi feita diretamente sobre os dados do FER2013, sem a utilização de imagens externas ou entradas em tempo real,

permitindo a avaliação da precisão e da consistência do modelo em um ambiente de dados padronizados. Esse processo garantiu a reprodutibilidade dos testes e facilitou a análise do desempenho da ferramenta em um contexto controlado.

Todo o processo de reconhecimento consistiu em detectar o rosto na imagem, processá-lo por meio do classificador emocional e identificar a emoção predominante. Desta forma, a análise emocional foi realizada exclusivamente com as imagens fornecidas pelo *dataset*, sem a inclusão de imagens externas ou em tempo real.

4.1.5 Categorização de valência emocional

A categorização das emoções básicas em termos de valência emocional (positiva e negativa) é uma prática amplamente adotada em sistemas de reconhecimento emocional, especialmente os que operam sobre textos ou expressões faciais. Conforme descrito por Bruna, Avetisyan e Holub (2016), os modelos categóricos são baseados na Teoria das Emoções Discretas, que considera que o ser humano vivencia um número limitado de emoções universais, entre elas as seis propostas por Ekman (1992) — felicidade, tristeza, raiva, medo, nojo e surpresa.

Dentro deste modelo, a felicidade é comumente classificada como uma emoção de valência positiva, por refletir estados de prazer e bem-estar, enquanto tristeza, raiva, medo e nojo são associadas a valência negativa, por estarem ligadas a experiências de sofrimento, aversão ou ameaça. A surpresa, por sua vez, ocupa uma posição ambígua: apesar de seu caráter reativo e inesperado, sua valência depende do contexto em que ocorre, podendo ser positiva ou negativa (BRUNA; AVETISYAN; HOLUB, 2016).

Tabela 4.1: Dicionário de categorização de valência emocional

Polaridade	**Emoções**
positivo	Felicidade, Surpresa
negativo	Tristeza, Raiva, Medo, Nojo, Surpresa

O dicionário de polaridade emocional foi elaborado com o objetivo de categorizar emoções humanas com base em sua valência afetiva predominante. A construção desse dicionário seguiu uma abordagem qualitativa, considerando a associação entre emoções básicas e suas respectivas polaridades — positiva ou negativa — conforme observadas em contextos gerais.

Inicialmente, foram selecionadas emoções básicas. Em seguida, procedeu-se à classificação dessas emoções com base em seu valor afetivo predominante. A emoção felicidade foi classificada como de polaridade positiva, por estar associada a estados de bem-estar e satisfação. As emoções tristeza, raiva, medo e nojo foram categorizadas como negativas, por representarem estados emocionais comumente relacionados ao desconforto, sofrimento ou ameaça. A emoção surpresa foi classificada como positiva ou negativa, podendo variar de acordo com o contexto.

A organização das informações foi realizada em formato tabular, de modo a permitir uma visualização clara e objetiva da correspondência entre as polaridades emocionais e as emoções associadas. O resultado foi um dicionário simplificado, que pode ser utilizado como ferramenta de apoio em análises de sentimentos e classificação emocional de textos no campo do PLN.

4.1.6 Algoritmo de geração de frases

No processo de geração de frases utilizou-se a API da *Google* para realizar o ajuste fino de um modelo da família *Gemini ("flash")* com base em dados de sentimento em português extraídos do *dataset* IMDB PT-BR, como apresentado no Apêndice A.1 Código do Treinamento da *API do Gemini*.

O processo começa com a preparação dos dados em um formato compatível com a *API Generative AI*, onde cada exemplo é estruturado como um par de entrada (*text.input*) e saída esperada (*output*). A entrada guia o modelo com um *prompt* condicional, como **Gere uma frase com o sentimento positivo**, enquanto a saída corresponde à frase real de sentimento positivo ou negativo, conforme anotado no *dataset.* Essa estrutura de *prompting* supervisionado é alinhada com abordagens recentes descritas por Ding et al. (2023), que defendem o uso de instruções explícitas no treinamento supervisionado de LLMs para melhorar o desempenho em tarefas específicas, como classificação ou geração condicional de texto.

A API *creat.tuned.model()* foi utilizada para treinar o modelo em três épocas com um *batch.size* de 4 e taxa de aprendizado de 0.001. Essa configuração leve se enquadra nas chamadas abordagens *parameter-efficient fine-tuning* (PEFT), discutidas por Ding et al. (2023), que mostram que é possível adaptar grandes modelos com poucas atualizações de parâmetros, mantendo a eficiência e a capacidade do modelo de generalizar para que possa reduzir o custo computacional com pequenos conjuntos de dados e treinamento supervisionado simplificado.

Adicionalmente, o balanceamento das classes foi garantido ao limitar o número de exemplos positivos e negativos, mantendo a paridade de cada sentimento, com modelos treinados em configurações de 2000, 4000, e 8000 frases por classe. Essa prática visa mitigar o viés do modelo para classes majoritárias, um desafio crítico em tarefas de classificação subjetiva. Ding et al. (2023), em uma análise abrangente de métodos de delta-tuning para modelos de linguagem de grande escala, destacam que a uniformidade na distribuição de dados é essencial para garantir a eficiência de adaptação paramétrica.

Por fim, o modelo ajustado passa a responder aos prompts, como apresentado nos blocos 3 e 4 da Figura 4.1, personalizados permitindo sua aplicação prática e, com isso, possibilitando a geração de frases de acordo com a categorização de valência emocional (positiva ou negativa) identificada, como é apresentada na Tabela 2.

Tabela 4.2: Exemplos de frases geradas

Polaridade	Emoções
positivo	muito feliz
positivo	melhor filme ja vi
positivo	incrivel voce poderia
positivo	filme melhor comedia
positivo	vida perfeita
postivo	melhor personagem tv
negativo	acho melhor pior filme nao acreditar
negativo	acho filme ruim
negativo	decepcionante realidade
negativo	ok melhor nao
negativo	nao sao tao engracados
negativo	pior filme ja vi

As frases geradas pelos modelos são armazenadas em arquivos .csv estruturados, organizados de forma particionada conforme o tamanho do conjunto amostral (2.000, 4.000 e 8.000 instâncias por arquivo). Esta arquitetura de armazenamento possui três objetivos analíticos principais:

- Avaliação da Coerência Semântica: Verificação sistemática de se as frases mantêm significado lógico e contextual após o processo de geração
- Análise de Diversidade Léxica: Identificação de frases repetitivas ou redundâncias indesejáveis através da quantificação, garantindo variedade expressiva nas saídas.
- Validação da Acurácia Classificatória: Verificação do alinhamento entre o sentimento atribuído pelo modelo e a polaridade real da frase gerada.

4.2 Análise das frases geradas

Para avaliar o desempenho do modelo *Gemini* ajustado com diferentes quantidades de dados por sentimento, foram conduzidos experimentos utilizando conjuntos com 2 mil, 4 mil e 8 mil frases por classe. A análise foi dividida em três métricas principais: coerência semântica da frase gerada (Tem Sentido), acurácia da classificação (Classificação Correta) e diversidade textual (Frase Repetida).

O Gráfico de Coerência Semântica (Tem Sentido), Figura 4.3, refere-se à capacidade do modelo de gerar frases que fazem sentido. Observa-se que, com 8 mil frases por sentimento, o número de frases consideradas coerentes (sim) é significativamente maior. Já com 4 mil frases, o modelo apresentou pior desempenho, com a maioria das frases classificadas como incoerentes (não). O cenário com 2 mil frases apresentou desempenho intermediário. Com isso, conclui-se que o aumento na quantidade de dados utilizados para ajuste fino resultou em maior coerência semântica das frases geradas, com destaque para o cenário de 8 mil frases.

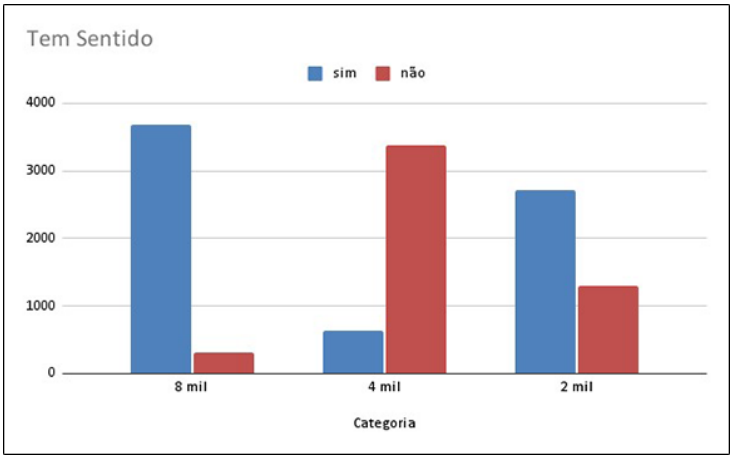

Figura 4.3: Gráfico de Coerência Semântica (Tem Sentido).
Fonte: Elaborado pelos autores.

O Gráfico de Acurácia da Classificação (Classificação Correta), Figura 4.4 apresenta se o modelo conseguiu classificar corretamente os sentimentos. Curiosamente, o melhor desempenho ocorreu com apenas 2 mil frases, onde o número de classificações corretas foi superior ao número de erros. Nos cenários com 4 mil e 8 mil frases, houve um predomínio de classificações incorretas. Diante disso, conclui-se que a melhor acurácia na classificação dos sentimentos foi observada no cenário com menos dados (2 mil frases).

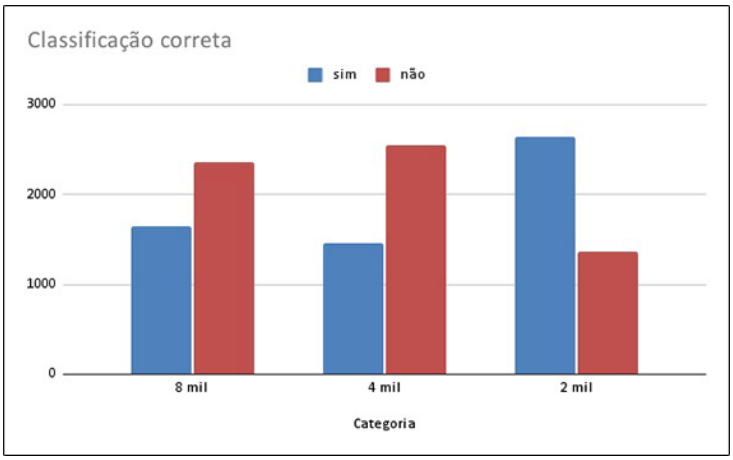

Figura 4.4: Gráfico de Acurácia da Classificação (Classificação Correta).
Fonte: Elaborado pelos autores.

No Gráfico de Diversidade Textual (Frase Repetida), Figura 4.5, avaliou-se a repetição de frases geradas. Neste caso, a palavra sim indica que a frase foi repetida, enquanto não representa uma frase única (não repetida). Assim, quanto maior o número de não, maior a diversidade textual.

O modelo treinado com 2 mil frases obteve o maior número de frases únicas, evidenciando uma maior diversidade. Já o cenário com 8 mil frases apresentou o menor número de frases únicas, indicando um comportamento mais repetitivo. Com isso, conclui-se que

o cenário com 2 mil frases promoveu maior variedade nas frases geradas, enquanto o uso de bases maiores resultou em maior repetição, das frases.

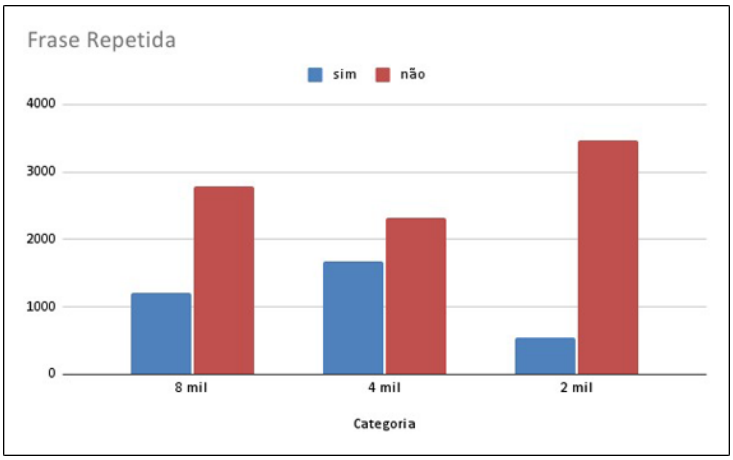

Figura 4.5: Gráfico de Diversidade Textual (Frase Repetida).
Fonte: Elaborado pelos autores.

Os resultados indicam que há um *trade-off* entre as métricas analisadas, com isso, o uso de mais dados (8 mil frases) favorece a coerência semântica, mas reduz a diversidade e não melhora a acurácia da classificação. Por outro lado, menos dados (2 mil frases) resultam em melhor acurácia de classificação e maior diversidade textual, mesmo com uma perda de coerência.

4.3 Discussão

Os resultados obtidos demonstram que a coerência semântica tende a aumentar conforme cresce o volume de dados de treinamento. Esse comportamento está relacionado ao funcionamento dos modelos *Transformer*, que utilizam o mecanismo de autoatenção para capturar dependências de longo alcance em sequências textuais (VASWANI, 2017). Dessa forma, o modelo consegue avaliar a relevância de cada *token* no contexto, gerando textos mais consistentes mesmo em entradas extensas. Em arquiteturas mais avançadas essa capacidade é ainda maior, pois o modelo ativa sub-redes específicas conforme a necessidade do contexto, como ocorre no *Gemini 1.5 Pro*.

Por outro lado, foi identificado o problema de *overfitting* em grandes volumes de dados. Quando a base de treinamento ultrapassa certos limites, como o caso de *datasets* com cerca de 8 mil frases por classe, ocorre uma queda de desempenho. Esse efeito, também observado por Raffel et al. (2020) no modelo T5, acontece porque a repetição de padrões leva o modelo a memorizar informações em vez de generalizar, o que compromete sua aplicabilidade em novos cenários. Além disso, nota-se um paradoxo entre tamanho e diversidade dos dados: embora bases maiores cubram uma gama mais ampla de exemplos, elas também podem gerar homogeneização nas representações internas do modelo.

Esse efeito já foi apontado por Devlin et al. (2019) no modelo *BERT*, e pode reduzir a criatividade e a variedade das respostas geradas.

No que se refere às limitações do estudo, duas questões principais merecem destaque. A primeira é o uso de métricas unidimensionais, como a acurácia. Apesar de úteis para avaliar o desempenho, elas não consideram aspectos mais complexos, como a profundidade semântica ou a relevância do conteúdo. Isso pode gerar o fenômeno conhecido como miragem de fluência (RAFFEL et al., 2020), em que o texto parece coerente, mas carece de conteúdo significativo. A segunda limitação está relacionada ao controle de ruído nos dados, especialmente em bases de sentimentos, onde expressões ambíguas geram alta taxa de discordância entre anotadores. Estudos como o de Pawlowski (2024) apontam que esse índice pode chegar a 40 por cento, afetando a consistência dos resultados.

Dessa forma, percebe-se que o equilíbrio entre volume, diversidade e qualidade dos dados é essencial para alcançar melhores resultados. Além disso, destaca-se a importância do uso de métricas mais abrangentes e da curadoria rigorosa dos *datasets*, garantindo maior confiabilidade na avaliação e maior potencial de aplicação dos modelos baseados em *Transformer*.

4.4 Considerações finais

Os experimentos e análises apresentados ao longo deste capítulo evidenciam que a integração entre reconhecimento facial e análise de sentimentos, aliada a técnicas de geração automática de texto, pode resultar em uma solução funcional para a criação personalizada de conteúdos para redes sociais. Apesar dos resultados promissores, algumas limitações foram observadas, como a sensibilidade a expressões ambíguas e a necessidade de maior diversidade nas frases geradas. Tais aspectos indicam que ainda há espaço para aprimoramentos, tanto na ampliação das bases de dados quanto na sofisticação dos modelos de geração textual. Diante disso, o capítulo seguinte apresenta as conclusões gerais do trabalho, destacando suas principais contribuições e sugerindo caminhos para pesquisas e melhorias futuras.

Capítulo 5

CONCLUSÃO

Este trabalho apresentou o desenvolvimento de uma ferramenta automatizada para geração de frases de acordo com a valência emocional positiva ou negativa, com base no reconhecimento facial e na análise de emoções. A proposta foi implementada com o uso de redes neurais profundas, técnicas de visão computacional e processamento de linguagem natural, permitindo a identificação de emoções a partir de expressões faciais e, a partir disso, a produção de frases com valência emocional correspondente. A abordagem adotada demonstrou ser viável e coerente com os objetivos definidos.

Os resultados obtidos indicaram que a ferramenta é capaz de realizar a detecção emocional com precisão satisfatória dentro das bases de dados utilizadas, bem como gerar conteúdos relevantes e emocionalmente alinhados com o contexto identificado. Além disso, a integração dos módulos e o fluxo automatizado entre detecção facial, categorização emocional e geração textual evidenciam o potencial da solução em cenários reais, especialmente no marketing digital e em estratégias de comunicação personalizada. Ainda assim, limitações foram observadas, como a dependência de bases de dados previamente organizadas e a sensibilidade a imagens com variações fora do padrão.

De modo geral, a proposta alcançou seu propósito central ao oferecer uma solução inovadora, que alia inteligência artificial e usabilidade prática em um contexto de crescente demanda por automação e personalização nas redes sociais. Além da contribuição técnica, o trabalho também levanta reflexões sobre ética, privacidade e responsabilidade no uso de dados emocionais — aspectos essenciais no desenvolvimento de tecnologias centradas no ser humano.

5.1 Contribuição do trabalho

A principal contribuição desta pesquisa é a criação de uma ferramenta que aplica conceitos avançados de inteligência artificial para gerar conteúdo emocionalmente orientado de forma automática. Ao combinar reconhecimento facial com análise de sentimentos e

geração textual, o sistema representa um avanço no uso integrado dessas tecnologias, ampliando as possibilidades de automação com sensibilidade ao estado emocional do usuário.

A solução desenvolvida pode ser útil tanto em contextos comerciais quanto sociais, oferecendo um ponto de partida para sistemas mais empáticos e adaptativos. No campo acadêmico, o trabalho também se destaca pela aplicação prática de diversos conceitos de Engenharia de *Software* e Inteligência Artificial, favorecendo abordagens multidisciplinares com foco na experiência do usuário.

5.2 Trabalhos futuros

Apesar dos resultados satisfatórios alcançados, a ferramenta desenvolvida pode ser aprimorada em diversos aspectos técnicos. O uso de modelos de Visão Computacional mais avançados e redes neurais profundas especializadas em reconhecimento de emoções permitiria maior precisão na identificação de sentimentos, incluindo emoções sutis. Do mesmo modo, o módulo de Processamento de Linguagem Natural (PLN) poderia integrar modelos de linguagem mais recentes, ajustados ao contexto de redes sociais, garantindo produções textuais mais coerentes e adequadas ao perfil de cada usuário. Essas melhorias visam aumentar a robustez do sistema, possibilitando sua aplicação em tempo real e em diferentes cenários.

Além disso, futuras pesquisas podem explorar funcionalidades adicionais, como a identificação do grau de intensidade das emoções (por exemplo, distinguir entre alegria moderada e euforia) e a personalização do estilo textual, permitindo ao usuário definir preferências de linguagem, uso de *emojis ou hashtags*. Outra perspectiva envolve a integração da ferramenta com plataformas externas, como *Instagram* ou *Facebook*, por meio de *APIs*, automatizando o ciclo de geração e publicação de postagens. Para validar a utilidade da solução em cenários reais, recomenda-se a realização de estudos de caso e testes de usabilidade, coletando *feedback* de usuários para guiar aprimoramentos futuros e ampliar o potencial de aplicação em contextos pessoais e corporativos.

Em suma, abordar esses pontos nos trabalhos futuros contribuirá para que a solução evolua de um protótipo promissor para uma ferramenta madura, segura e efetiva na geração automatizada de posts personalizados com base em emoções.

Referências

AHO, A. V. et al. *Compilers: Principles, Techniques, and Tools*. 2. ed. India: Pearson Education India, 2007.

AJAYI, F.; UDEH, C. A. Innovative recruitment strategies in the it sector: A review of successes and failures. *Magna Scientia Advanced Research and Reviews*, v. 10, n. 2, p. 150–164, 2024.

BARBOSA, F. A. et al. Análise de sentimentos em redes sociais: Uma revisão da literatura. *Revista Brasileira de Computação Aplicada*, v. 10, n. 2, p. 45–60, 2020.

BEZERRA, E. Introdução à aprendizagem profunda. *Simpósio Brasileiro de Banco de Dados*, v. 1, 2016.

BROWN, T. et al. Language models are few-shot learners. *Advances in neural information processing systems*, p. 1877–1901, 2020.

BRUNA, O.; AVETISYAN, H.; HOLUB, J. Emotion models for textual emotion classification. *Journal of Physics: Conference Series*, v. 772, n. 2, p. 12–63, 2016.

CANARIO, J. P. P. de S. Uma abordagem deep learning para reconhecimento de expressões faciais. *Dissertação (Mestrado em Ciência da Computação) – Universidade Federal da Bahia, Salvador*, 2016.

CAO, Q. et al. Vggface2: A dataset for recognising faces across pose and age. *IEEE international conference on automatic face gesture recognition.*, p. 67–74, 2018.

CASELI, H. de M.; NUNES, M. das G. V. Processamento de linguagem natural: conceitos, técnicas e aplicações em português. *Universidade de São Paulo - USP*, 2024.

COSTA-VIEIRA, H. A.; SOUZA, W. C. de. O reconhecimento de expressões faciais e prosódia emocional: Investigação preliminar em uma amostra brasileira jovem. *Revista Estudos de Psicologia*, v. 19, n. 2, p. 89–156, 2014.

COSTA, W. et al. Multi-cue adaptive emotion recognition network. *Computer Vision and Pattern Recognition*, 2021.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *Cnference of the North American chapter of the association for computational linguistics: human language technologies.*, v. 1, p. 4171–4186, 2019.

DHALL, A. et al. Emotion recognition in the wild challenge 2016. *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI)*, p. 585–588, 2016.

DING, N. et al. Enhancing chat language models by scaling high-quality instructional conversations. *arXiv preprint arXiv:2305.14233.*, 2023.

EKMAN, P. Are there basic emotions? *Psychological Review*, v. 99, n. 3, p. 550–553, 1992.

FACELI, K. et al. Deep learning in bioinformatics. *Briefings in bioinformatics, Oxford University Press*, v. 18, n. 5, p. 851—-869, 2017.

FACELI, K. et al. Inteligência artificial: uma abordagem de aprendizado de máquina. *Grupo Gen - LTC*, 2021.

FAN, T. et al. Multimodal sentiment analysis for social media contents during public emergencies. *Journal of Data and Information Science*, 2023.

FERREIRA, A. dos S. Redes neurais convolucionais profundas na detecçao de plantas daninhas em lavoura de soja. *Universidade Federal do Mato Grosso do Sul*, 2017.

FLECK, L. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, v. 1, n. 13, p. 47–57, 2016.

FONSECA, C. Introdução a bag of words e tf-idf. *Medium*, 2020.

FRED, L. Imdb pt-br dataset de 50k movie reviews. *https://www.kaggle.com/datasets/luisfredgs/imdb-ptbr*, 2011.

GOMES, J. de S.; ARAUJO, F. H. Análise de técnicas de pré-processamento de imagem para reconhecimento facial baseada em vgg faces e ball tree. *Encontro Unificado de Computação do Piauí (ENUCOMPI).*, p. 136–143, 2021.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. *MIT press*, 2016.

GORWA, R.; BINNS, R.; KAITZENBACH, C. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data e Society*, v. 7, n. 1, p. 1–15, 2020.

GOULART, F. et al. Sentpt: Uma solução personalizada para análise de sentimentos multigênero de textos em português. *Applied Soft Computing, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal*, v. 245, n. 123075, 2024.

HAN, J.; KAMBER, M.; PEI, J. Data mining concepts and techniques. *Journal of Physics: Conference Series*, v. 3, 2011.

HASSABALLAH, M.; ALY, S. Face recognition under unconstrained conditions: challenges and future directions. *IET Computer Vision*, v. 9, n. 4, p. 614–626, 2015.

HAYKIN, S. *Redes Neurais- Princípios e Práticas*. 2. ed. Porto Alegre, RS: Bookman, 2007.

KOUJAN, M. R. et al. Real-time facial expression recognition "in the wild" by disentangling 3d expression from identity. *International Conference on automatic face and gesture recognition (FG 2020) - IEE*, 2020.

LAVEZZO, G. H.; CONCEICAO, G. C. da. AnÁlise de tÓpicos em redes sociais utilizando processamento de linguagem natural (nlp). *Revista Interface Tecnológica*, v. 20, n. 2, p. 65–76, 2023.

LEAL, S. E. et al. Ethical implications of bias in machine learning. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2021.

LECLERC, M. et al. Development of willow tree yield-mapping technology. *Sensors*, v. 20, n. 9, p. 2650, 2020.

LI, S.; DENG, W. Deep facial expression recognition: A survey. *IEEE transactions on affective computing 13.3.*, p. 1195–1215, 2020.

LIU, Y. et al. G-eval: Nlg evaluation using gpt-4 with better human alignment. *arXiv preprint arXiv:2303.16634.*, 2023.

LOPES, A. C. Otimização de aplicações de processamento de linguagem natural para análise de sentimentos. *Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual Paulista (Unesp)*, 2024.

MCCULLOCH, W. S.; PITTS, W. *A logical calculus of the ideas immanent in nervous activity.* 4. ed. São Paulo, SP: The bulletin of mathematical biophysics, 1943.

MEHENDALE, N. Facial emotion recognition using convolutional neural networks (ferc). *Applied Sciences*, v. 2, n. 446, p. 1–9, 2020.

MIRANDA, F. A.; FREITAS, S. R. C.; FAGGION, P. L. Integração e interpolação de dados de anomalias ar livre utilizando-se a técnica de rna e krigagem. *Boletim de Ciências Geodésicas*, v. 15, n. 3, p. 428–443, 2009.

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. *Jamia: Journal of the American Medical Informatics Association*, p. 544–551, 2011.

NGUYEN, T. V.; PHAN, L. T.; BUI, K. X. Size determination of polystyrene sub-microspheres using transmission spectroscopy. *Applied Sciences*, v. 7, n. 1, p. 1–15, 2020.

NILSEN, M. A. Neural networks and deep learning. *Determination Press*, 2015.

OLIVEIRA, G. S.; JUNIOR, O. F. Implantação de um sistema para automação de pedidos atrelado a técnicas de marketing. *Revista Mundi Engenharia, Tecnologia e Gestão*, v. 8, n. 2, 2023.

PALMER, D. D. *Text preprocessing. In Handbook of natural language processing.* 2. ed. Florida: Chapman and Hall/CRC, 2010.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, v. 2, n. 1-2, p. 1–135, 2008.

PAWLOWSKI, M. Annotation noise in sentiment datasets. *Proceedings of the ACL.*, 2024.

PEER, D. et al. Anls*–a universal document processing metric for generative large language models. *arXiv preprint arXiv:2402.03848.*, 2024.

PREMEDIBA, S. M. Guia de nlp: conceitos e técnicas. *Alura*, 2021.

RAFFEL, C. et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, p. 1–67, 2020.

RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, v. 29, n. 1, p. 39, 2005.

REIS, C. H. Otimização de hiperparâmetros em redes neurais profundas. *Universidade Federal de Itajubá - UNIFEI*, 2021.

RIZQYAWAN, M. I. et al. Comparing performance of supervised learning classifiers by tuning the hyperparameter on face recognition. *International Journal of Intelligent Systems and Applications*, v. 10, n. 5, 2021.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Peer Reviewed Journal*, v. 65, n. 6, p. 386, 1958.

SANTOS, C. M. dos. Classificação de documentos com processamento de linguagem natural. *Instituto Superior de Engenharia de Coimbra - Tese de Doutorado*, 2015.

SANTOS, M. C.; OLIVEIRA, A. F. O impacto da análise de sentimentos na reputação online de marcas. *Revista de Administração Contemporânea*, v. 24, n. 6, p. 1021–1034, 2020.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 815–823, 2015.

SENO, E. R. M.; PAIVA, V.; SCHROFF, V. P. Processamento de linguagem natural: Conceitos, técnicas e aplicações em português. *Capítulo 9: Semântica com Técnicas Simbólicas*, 2023.

SILVA, E. P. da. Reconhecimento facial com redes convolucionais em imagens de regiões oculares. *Centro Universitário Campo Limpo Paulista, Unifaccamp, Campo Limpo Paulista.*, 2019.

SILVA, J. C.; VIEIRA, R. O. Introdução às redes neurais profundas com python. *Sociedade Brasileira de Computação*, v. 1, 2022.

TAIGMAN, Y. et al. Deepface: closing the gap to human-level performance in face verification. *IEEE Conference on Computer Vision and Pattern Recognition*, p. 1701–1708, 2014.

TAN, S. It's morphin' time! combating linguistic discrimination with inflectional perturbations. *Association for Computational Linguistics*, p. 2920–2935, 2020.

TEIXEIRA, T.; GRANGER, E.; KOERICH, A. L. Continuous emotion recognition with spatiotemporal convolutional neural networks. *Applied Sciences*, v. 11, n. 24, p. 11738, 2021.

VASWANI, A. Attention is all you need. *Advances in Neural Information Processing Systems.*, v. 30, 2017.

WANG, M.; DENG, W. Deep face recognition: A survey. *Neurocomputing*, 2018.

YAPO, A.; WEISS, J. Nilc-metrix: assessing the complexity of written and spoken language in brazilian portuguese. *International Conference on System Sciences.*, 2018.

Apêndice A

Apêndice

A.1 Código do Treinamento da API do Gemini

Listing A.1: Código para treinamento de modelo com Google Generative AI

```
import pandas as pd
import re
import numpy as np
import os
import google.generativeai as genai
from google.colab import userdata
import time
import random
from google.colab import drive
import seaborn as sns

GOOGLE_API_KEY=userdata.get('GOOGLE_API_KEY')
genai.configure(api_key=GOOGLE_API_KEY)

# Monta o Google Drive
drive.mount('/content/drive')

# Define o caminho da pasta
folder_path = '/content/drive/MyDrive'
print(f'A pasta est  montada em: {folder_path}')

# Define o caminho completo dos arquivos
files_path = []
files_path.append(folder_path + "/Models/C pia de novo_arquivo_limpo.
    csv")

# Verifica se o arquivo existe
if os.path.exists(files_path[0]):
    print('Arquivo encontrado.')
```

```
else:
    print('Arquivo n o encontrado.')

# Carrega o arquivo CSV original
df = pd.read_csv(files_path[0])

# Filtrar os dados para cada classe
pos_data = [
  {"text_input": 'Gere uma frase como o sentimento positivo', "output":
      row['text_pt']}
  for _, row in df.iterrows() if row['sentiment'] == 'pos'
]

neg_data = [
  {"text_input": 'Gere uma frase como o sentimento negativo', "output":
      row['text_pt']}
  for _, row in df.iterrows() if row['sentiment'] == 'neg'
]

# Garantir que cada lista tenha no m ximo 8000 itens
pos_data = pos_data[:8000]
neg_data = neg_data[:8000]

training_data = pos_data + neg_data
random.shuffle(training_data)

# Resultado final
print(len(training_data))  # Deve ser 16000
print(training_data[:10])

base_model = [
    m for m in genai.list_models()
    if "createTunedModel" in m.supported_generation_methods and
    "flash" in m.name][0]
base_model

name = f'generate-num-{random.randint(0,100000)}'
operation = genai.create_tuned_model(
    source_model=base_model.name,
    training_data=training_data,
    id = name,
    epoch_count = 3,
    batch_size=4,
    learning_rate=0.001,
)

for status in operation.wait_bar():
```

```
    print(status)
    time.sleep(2)

result = operation.result()
print(result)

while not operation.done():
    time.sleep(30)  # verifica a cada 30 segundos
result = operation.result()
print(result)

model = operation.result()
snapshots = pd.DataFrame(model.tuning_task.snapshots)

sns.lineplot(data=snapshots, x = 'epoch', y='mean_loss')

model = genai.GenerativeModel(model_name=result.name)
model_id = result.name
save_path = f"{folder_path}/Models/gemini/8000/model_id.txt"

with open(save_path, "w") as f:
    f.write(model_id)
print(f"Modelo salvo em: {save_path}")

result = model.generate_content("Gere uma frase como o sentimento
   positivo de no maximo 5 palavras")
print(result.text)
```

A.2 Código para Reconhecimento Facial e Geração de Frase

Listing A.2: Código para reconhecimento de emoções e geração de frases

```
import cv2
from deepface import DeepFace
import os
from google.colab import drive
import google.generativeai as genai
from google.colab import userdata
import time

drive.mount('/content/drive')

# Define o caminho da pasta
folder_path = '/content/drive/MyDrive'
```

```
13
14 GOOGLE_API_KEY=userdata.get('GOOGLE_API_KEY')
15 genai.configure(api_key=GOOGLE_API_KEY)
16
17 model = genai.GenerativeModel(model_name="tunedModels/generate-num-9389"
       )
18 # model = genai.GenerativeModel(model_name="tunedModels/generate-num
       -6712")
19
20 input_path = '/content/drive/MyDrive/FER2013/test/'
21 subdirectories = [os.path.join(input_path, d) for d in os.listdir(
       input_path)
22                   if os.path.isdir(os.path.join(input_path, d))]
23
24 dictionary = dict(
25     felicidade='positiva',
26     happy='positiva',
27     tristeza='negativa',
28     sad='negativa',
29     raiva='negativa',
30     fear='negativa',
31     medo='negativa',
32     angry='negativa',
33     disgust='negativa',
34     nojo='negativa',
35     surpresa='neutra',
36     surprise='neutra',
37     neutral='neutra',
38     neutror='neutra'
39 )
40
41 # Passo 2: para cada diret rio, pegar todos os arquivos
42 all_files = []
43 for subdir in subdirectories:
44     files = [os.path.join(subdir, f) for f in os.listdir(subdir)
45              if os.path.isfile(os.path.join(subdir, f))]
46     for file in files:
47         img = cv2.imread(file)
48         img = cv2.resize(img, (640,480))
49         result = DeepFace.analyze(img, actions=("emotion",),
               enforce_detection=False)
50         emotion_result = (result[0]['dominant_emotion'])
51         sentment = dictionary[emotion_result]
52         if sentment == 'neutra':
53             emotion_result = (result[1]['dominant_emotion'])
54             sentment = dictionary[emotion_result]
```

```
        result = model.generate_content(f"Gere uma frase como o
            sentimento {sentment} com um maximo de 5 palavras")
        print(result.text)
        time.sleep(5)
```

A.3 Código para Treinamento do Dataset

Listing A.3: Pré-processamento de dados textuais com spaCy e Label Encoding

```
import pandas as pd
import re
from unidecode import unidecode
import spacy
from sklearn.preprocessing import LabelEncoder
import numpy as np

# Carrega o modelo em portugu s do spaCy
nlp = spacy.load('pt_core_news_sm')

def clean_text(text):
    print('Entrada: '+text[:50])
    # Remove caracteres especiais e n meros, substituindo por
        caracteres ASCII
    text = unidecode(text)
    # Remove caracteres n o alfab ticos, deixando apenas letras e
        espa os
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    # Converte para min sculas
    text = text.lower()
    print('Pradoniza  o: '+text[:50])

    # Processa o texto com spaCy para remover as stopwords
    doc = nlp(text)
    text_no_stopwords = ' '.join([token.lemma_ for token in doc if not
        token.is_stop])
    print('Lematizada e sem stopwords: '+text_no_stopwords[:50])
    return text_no_stopwords

# Fun  o para obter a m dia dos vetores das palavras em cada frase
def sentence_to_embedding(sentence, embeddings_index):
    word_vectors = [embeddings_index.get(word) for word in sentence.
        split() if word in embeddings_index]
    if not word_vectors:
        return np.zeros(300)
    return np.mean(word_vectors, axis=0)
```

```
33
34 from google.colab import drive
35
36 # Monta o Google Drive
37 drive.mount('/content/drive')
38
39 # Define o caminho da pasta
40 folder_path = '/content/drive/MyDrive'
41 print(f'A pasta est  montada em: {folder_path}')
42
43 import os
44
45 # Define o caminho completo dos arquivos
46 files_path = []
47 files_path.append(folder_path + "/Models/C pia de imdb-reviews-pt-br.
       csv")
48
49 # Verifica se o arquivo existe
50 if os.path.exists(files_path[0]):
51     print('Arquivo encontrado.')
52 else:
53     print('Arquivo n o encontrado.')
54
55 # Carrega o arquivo CSV original
56 df = pd.read_csv(files_path[0])
57
58 def clean_text(text):
59     print('Entrada: '+text[:50])
60     text = unidecode(text)
61     text = re.sub(r'[^a-zA-Z\s]', '', text)
62     text = text.lower()
63     print('Pradoniza  o: '+text[:50])
64     doc = nlp(text)
65     text_no_stopwords = ' '.join([token.lemma_ for token in doc if not
           token.is_stop])
66     print('Lematizada e sem stopwords: '+text_no_stopwords[:50])
67     return text_no_stopwords
68
69 # Cria um novo DataFrame apenas com a coluna de texto limpa
70 new_df = pd.DataFrame()
71 new_df['text_pt'] = df['text_pt'].apply(clean_text)
72
73 # Aplica o label encoding na coluna de sentimento
74 new_df['sentiment'] = df['sentiment']
75
76 label_encoder = LabelEncoder()
```

```
77 new_df['sentiment_encoded'] = label_encoder.fit_transform(df['sentiment'
       ])
78
79 # Salva o novo DataFrame em um novo CSV
80 new_df.to_csv(folder_path+ '/Models/novo_arquivo_limpo2.csv', index=
       False)
81 new_df
```