

Verbaliz.AI

Integrantes: Pedro Matos | Davi Mendes | Cauã Henrique

1. Resumo Executivo

O projeto "Verbaliz.AI" propõe o desenvolvimento de um serviço web de alta acessibilidade que gera descrições textuais de imagens para pessoas com deficiência visual. Utilizando um modelo de Inteligência Artificial (*Image Captioning*), a ferramenta permite que um usuário faça o upload de uma imagem e receba, em tempo real, uma frase que descreve o conteúdo visual da cena. A solução visa promover a inclusão digital, fornecendo autonomia para que pessoas cegas ou com baixa visão possam compreender o conteúdo de imagens em redes sociais, sites e documentos, que de outra forma seriam inacessíveis.

2. Introdução

Nossa equipe, formada por estudantes de Análise e Desenvolvimento de Sistemas, acredita no potencial da tecnologia como ferramenta de transformação social. O cenário digital atual é predominantemente visual, o que cria barreiras significativas para milhões de pessoas com deficiência visual que dependem de leitores de tela para navegar. Este projeto é motivado pela necessidade de criar uma solução prática e de fácil acesso que ajude a diminuir essa barreira, aplicando conhecimentos de ponta em IA e desenvolvimento de software para tornar o mundo digital mais inclusivo.

3. Problema/Oportunidade

O problema central é a falta de acessibilidade em conteúdos visuais na internet. Pessoas cegas utilizam softwares leitores de tela que convertem texto em áudio. Quando uma imagem em um site ou rede social não possui uma descrição textual alternativa (o atributo "alt text"), o leitor de tela não consegue interpretá-la, privando o usuário de informação, contexto e participação plena na conversa digital.

A oportunidade é desenvolver um serviço gratuito e centralizado que funcione como um "tradutor" universal de imagens para texto. Essa ferramenta não só resolve um problema prático do dia a dia, como também empodera seus usuários, dando-lhes mais independência para consumir conteúdo online sem depender da ajuda de terceiros para descrever uma imagem.

4. Objetivos

- **Específico:** Implementar uma página web com um único campo de upload de imagem e uma área de texto para exibir a descrição gerada, seguindo as diretrizes de acessibilidade WCAG 2.1.
- **Mensurável:** Atingir um tempo médio de resposta (do upload da imagem à exibição da descrição) inferior a 8 segundos no protótipo.
- **Alcançável:** Utilizar um modelo de *Image Captioning* pré-treinado e de código aberto (como o BLIP, da Salesforce, via Hugging Face Transformers), o que torna o desenvolvimento da IA viável sem o custo e tempo de treinar um modelo do zero.

- **Relevante:** O projeto é uma aplicação direta das disciplinas do curso: IA (Visão Computacional + PLN), Estrutura de Dados (manipulação de tensores de imagem), Segurança da Informação (privacidade do usuário) e DevOps (disponibilização do serviço).
- **Temporal:** Ter o serviço web funcional, testado e publicamente acessível em um ambiente de nuvem dentro de 4 meses.

5. Escopo do Projeto

1. **API RESTful:** Um endpoint que recebe um arquivo de imagem e retorna um objeto JSON contendo a descrição textual.
 2. **Interface Web (Front-end):** Uma página web minimalista, de alto contraste e totalmente operável via teclado, que consome a API.
 3. **Módulo de IA (Back-end):** O serviço que carrega o modelo de *Image Captioning* e realiza a inferência (processamento da imagem).
 4. **Pipeline de CI/CD:** Um fluxo de trabalho automatizado (usando GitHub Actions) para testar e implantar a aplicação.
- O sistema não armazenará as imagens ou as descrições geradas após o processamento, para garantir a privacidade total do usuário.
 - A descrição será limitada a uma ou duas frases em português.
 - O modelo pode apresentar limitações na interpretação de imagens muito abstratas, texto dentro de imagens ou conceitos culturais muito específicos.
 - O protótipo não incluirá um sistema de login ou contas de usuário.

6. Metodologia e Plano de Ação

A abordagem de desenvolvimento será baseada em metodologias ágeis (Scrum), com foco na entrega de valor incremental.

- **Etapas Principais:**
 1. **Pesquisa e Planejamento (Sprint 0):** Definição da arquitetura, escolha das tecnologias (Python, Flask/FastAPI, React/HTML puro, Hugging Face Transformers) e configuração do ambiente de desenvolvimento e do repositório Git.
 2. **Desenvolvimento do Back-end (Sprints 1-2):** Criação da API e do serviço de IA. A imagem recebida será tratada como uma matriz multidimensional (tensor), uma estrutura de dados fundamental, para ser processada pelo modelo.
 3. **Desenvolvimento do Front-end (Sprints 3-4):** Construção da interface web com foco total em acessibilidade (HTML semântico, atributos ARIA para feedback dinâmico ao leitor de tela).
 4. **Implementação de DevOps e Segurança (Sprints 5-6):** Criação do pipeline de CI/CD para automatizar testes (unitários para a API, de acessibilidade para o front-end) e o deploy. Implementação de medidas

de segurança como *rate limiting* (para evitar abuso) e configuração de HTTPS.

- 5. **Testes e Refinamento (Sprint 7):** Testes de ponta a ponta, especialmente com o uso real de leitores de tela (NVDA, VoiceOver) para validar a experiência do usuário e fazer os ajustes finais.

7. Cronograma

Fase / Atividade	Setembro	Outubro	Novembro	Dezembro
1. Planejamento e Arquitetura	✓			
2. Desenvolvimento do Back-end e IA		✓		
3. Desenvolvimento do Front-end Acessível		✓		
4. Implementação de DevOps e Segurança		✓		
5. Testes Integrados e Lançamento			✓	
Marco 1: API Funcional			✓	
Marco 2: Interface Web Integrada				✓
Marco 3: Protótipo Final no Ar				✓

8. Recursos

Recursos Humanos:

Pedro Matos: Foco no desenvolvimento do Back-end, na implementação do modelo de IA e na arquitetura da API.

Davi Mendes: Foco no desenvolvimento do Front-end, garantindo a aplicação das diretrizes de acessibilidade (WCAG).

Cauã Henrique: Foco na configuração do pipeline de DevOps

Recursos Materiais:

Computadores pessoais com ambiente de desenvolvimento configurado (Python, Node.js, Docker).

Contas em plataformas de nuvem com nível gratuito (ex: Heroku, Vercel, AWS Free Tier) para hospedagem do protótipo.

Software leitor de tela (NVDA para Windows, VoiceOver para macOS) para a realização dos testes de acessibilidade.

Conta no GitHub para controle de versão e GitHub Actions.

Recursos Financeiros (estimativa):

O projeto foi planejado para ter um custo de **R\$ 0,00**, utilizando exclusivamente ferramentas de código aberto e serviços de nuvem com camadas gratuitas que são suficientes para um protótipo de escopo acadêmico. Não há previsão de custos com licenças de software ou APIs pagas.

Objetivo Geral: Desenvolver um protótipo funcional de um serviço web que gera descrições textuais precisas e contextuais para imagens enviadas pelos usuários, garantindo uma interface totalmente acessível e compatível com leitores de tela.