

FIAP – CENTRO UNIVERSITÁRIO
CONSELHO DE ENSINO, PESQUISA E EXTENSÃO - CEPE
PROGRAMA DE INICIAÇÃO CIENTÍFICA E DE INOVAÇÃO TECNOLÓGICA

TALK-E

André Alves da Silva

Isabela Ivanoff

Juliana Moreira da Silva

Lucas Basto

Sabrina do Couto Xavier Lima

Joseffe Barroso de Oliveira

SÃO PAULO

2024

André Alves da Silva - RM 552639

Isabela Ivanoff - RM 96824

Juliana Moreira da Silva - RM 554113

Lucas Basto - RM 553771

Sabrina do Couto Xavier Lima - RM 552728

TALK-E

Este documento tem como objetivo apresentar a pesquisa e o desenvolvimento do entregável referente ao Projeto de Iniciação Científica, realizado sob a orientação do Professor Joseffe Barroso de Oliveira, e submetido ao Conselho de Ensino Pesquisa e Extensão – CEPE do FIAP - Centro Universitário.

SÃO PAULO

2024

RESUMO

O Talk-E é um dispositivo projetado para auxiliar no aprendizado de inglês por meio de interações baseadas em inteligência artificial (IA). Composto por um Raspberry Pi, microfone, alto-falante e estrutura impressa em 3D, o dispositivo se conecta à nuvem para processar os modelos de IA, proporcionando respostas precisas e em tempo real. Ele permite que o usuário pratique a conversação, faça perguntas sobre gramática ou vocabulário e receba correções instantâneas. Este projeto de Iniciação Científica visa desenvolver um dispositivo inovador, acessível e eficaz para a prática do inglês, facilitando o aprendizado contínuo e interativo. O Talk-E utiliza recursos tecnológicos modernos para criar um ambiente de aprendizado dinâmico e motivador, além de contribuir para a disseminação de soluções baseadas em IA na educação. Os principais resultados incluem o desenvolvimento de um protótipo funcional e a implementação de modelos eficientes de reconhecimento de voz e processamento de linguagem natural, com impacto direto na melhoria das habilidades linguísticas dos usuários.

Palavras-chave: Dispositivo interativo. Aprendizado de idiomas. Inteligência Artificial.

ABSTRACT

Talk-E is a device designed to help learn English through interactions based on artificial intelligence (AI). Consisting of a Raspberry Pi, microphone, speaker and 3D-printed structure, the device connects to the cloud to process AI models, providing accurate answers in real time. It allows the user to practice conversation, ask questions about grammar or vocabulary and receive instant corrections. This Scientific Initiation project aims to develop an innovative, accessible and effective device for practicing English, facilitating continuous and interactive learning. Talk-E uses modern technological resources to create a dynamic and motivating learning environment, as well as contributing to the dissemination of AI-based solutions in education. The main results include the development of a functional prototype and the implementation of efficient speech recognition and natural language processing models, with a direct impact on improving users' language skills.

Keywords: Interactive device. Language learning. Artificial Intelligence.

1. INTRODUÇÃO	1
2. OBJETIVOS	2
2.1. OBJETIVO GERAL	2
2.2. OBJETIVOS ESPECÍFICOS	2
3. ESTADO DA ARTE	3
3.1. HISTÓRICO DE TRABALHOS RELACIONADOS	3
3.1.1. Uso de Inteligência Artificial no Ensino de Línguas	3
3.2. FUNDAMENTAÇÃO TEÓRICA	3
3.2.1. Aprendizado Baseado em Dispositivos Interativos	3
3.2.2. Computação em Nuvem para Modelos de IA	3
4. JUSTIFICATIVAS	4
5. CRONOGRAMA	5
5.1 Planejamento e Requisitos (Mês 1)	5
5.2 Arquitetura e Design (Mês 2)	6
5.3 Desenvolvimento e Implementação (Mês 3, Mês 4, Mês 5)	7
6. RELATO DO DESENVOLVIMENTO TÉCNICO	9
6.1. HARDWARE	9
6.2. SOFTWARE	11
6.3. IMPRESSÃO 3D	12
7. CONSIDERAÇÕES FINAIS	16
8. REFERÊNCIAS BIBLIOGRÁFICAS	17

1. INTRODUÇÃO

O Talk-E é um dispositivo tecnológico que combina hardware e software de ponta para facilitar o aprendizado de inglês. O projeto envolve a integração de um Raspberry Pi com dispositivos periféricos, como microfone e alto-falante, além de uma estrutura física desenvolvida por impressão 3D. Através da computação em nuvem, o Talk-E processa modelos de inteligência artificial especializados em reconhecimento de voz e processamento de linguagem natural, permitindo que os usuários interajam de forma conversacional. A capacidade de identificar e corrigir erros em tempo real torna o Talk-E uma ferramenta poderosa para quem deseja aprimorar suas habilidades linguísticas, oferecendo um feedback imediato sobre gramática e vocabulário.

As aplicações do Talk-E são amplas, abrangendo desde estudantes que desejam praticar inglês até profissionais que precisam manter suas habilidades afiadas. O dispositivo atua como um tutor acessível, disponível a qualquer momento para responder perguntas, corrigir erros e engajar em diálogos produtivos. Além disso, o uso de IA contribui para a evolução contínua do dispositivo, tornando-o cada vez mais eficiente e personalizado conforme o uso.

2. OBJETIVOS

Os objetivos principais de um projeto de IC, independentemente da sua área de atuação, é a inserção do aluno no mundo da Pesquisa. Seguem alguns exemplos de objetivos para projetos de Iniciação Científica:

- Contribuir para a formação e inserção de estudantes em atividades de pesquisa, desenvolvimento tecnológico e inovação;
- Contribuir para a formação de recursos humanos que se dedicarão ao fortalecimento da capacidade inovadora das empresas no País, e
- Contribuir para a formação do cidadão pleno, com condições de participar de forma criativa e empreendedora na sua comunidade.

2.1. OBJETIVO GERAL

Desenvolver um dispositivo dinâmico baseado em inteligência artificial que auxilie e complemente no aprendizado de inglês, permitindo a prática de conversação e correção de erros gramaticais e de vocabulário em tempo real.

2.2. OBJETIVOS ESPECÍFICOS

- Implementar um sistema de correção de erros gramaticais e de vocabulário em tempo real;
- Desenvolver uma interface de áudio para comunicação fluida entre o usuário e o Talk-E;
- Utilizar computação em nuvem para acessar e processar os modelos de inteligência artificial;
- Criar uma estrutura física utilizando impressão 3D que encapsule os componentes de forma funcional e estética;
- Desenvolver a arquitetura de hardware do dispositivo utilizando Raspberry Pi, microfone, alto-falante e uma estrutura 3D.

3. ESTADO DA ARTE

3.1. HISTÓRICO DE TRABALHOS RELACIONADOS

3.1.1. Uso de Inteligência Artificial no Ensino de Línguas

O uso de IA para aprendizado de idiomas vem crescendo, especialmente com o surgimento de assistentes virtuais como o Duolingo, que combina IA e gamificação para o ensino de várias línguas. Um exemplo importante é o uso de chatbots que simulam conversação em diferentes contextos, proporcionando ao usuário a oportunidade de praticar de forma mais natural.

Duolingo AI-powered Chatbots: Criado para permitir a prática de conversação em diferentes idiomas, os chatbots de Duolingo adaptam-se ao nível do usuário, proporcionando feedback em tempo real sobre o desempenho.

3.2. FUNDAMENTAÇÃO TEÓRICA

3.2.1. APRENDIZADO BASEADO EM DISPOSITIVOS INTERATIVOS

Papert (1993) argumenta que o uso de dispositivos tecnológicos pode promover uma aprendizagem mais ativa e engajadora, especialmente quando os aprendizes interagem diretamente com a tecnologia para solucionar problemas.

3.2.2. COMPUTAÇÃO EM NUVEM PARA MODELOS DE IA

Os avanços da computação em nuvem permitem a execução de modelos de IA com maior eficiência, devido à escalabilidade e ao poder computacional disponíveis. Armbrust et al. (2010) discutem o impacto da computação em nuvem no desenvolvimento de aplicações educacionais.

4. JUSTIFICATIVAS

A importância do Talk-E está na crescente demanda por ferramentas que possam auxiliar no aprendizado de línguas de forma prática, acessível e interativa. O domínio do inglês, sendo um idioma global, é uma habilidade essencial no mercado de trabalho e na comunicação internacional. Não obstante, muitos estudantes e profissionais encontram dificuldades em praticar o idioma regularmente, especialmente em regiões onde o inglês não é amplamente falado.

O Talk-E vem preencher essa lacuna, oferecendo uma solução que combina tecnologia de ponta com uma abordagem funcional e amigável. Além de possibilitar a prática diária de conversação, o dispositivo oferece correções em tempo real, algo que nem sempre é possível em aplicativos ou cursos tradicionais. Ao democratizar o acesso a uma ferramenta eficiente de aprendizado de inglês, o Talk-E pode contribuir para melhorar a proficiência de uma ampla gama de usuários.

O projeto também explora a integração de hardware e software como uma solução educacional acessível, combinando a potência de modelos de IA com um dispositivo físico, de baixo custo, que pode ser utilizado em diversas configurações e ambientes.

5. CRONOGRAMA

Etapa	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
1. Planejamento e Requisitos	X											
2. Arquitetura e Design		X										
3. Desenvolvimento e Implementação			X	X	X							

Abaixo, cada etapa foi dividida por subitens para detalhar o processo, materiais, e tecnologias utilizadas.

5.1 PLANEJAMENTO E REQUISITOS (Mês 1)

Objetivo:

- Definir o escopo do projeto e as funcionalidades principais.
- Selecionar tecnologias adequadas para alcançar os objetivos.

Atividades:

- Brainstorming para definir as principais funcionalidades e metas do projeto.
- Escolha das tecnologias: Cloud para escalabilidade, Python como linguagem principal, IA generativa para melhorar a interação, STT (Speech-to-Text) e TTS (Text-to-Speech) para criar um sistema de comunicação baseado em voz.

Ferramentas e Tecnologias:

- **Python** para desenvolvimento inicial e integração com bibliotecas de API.
- **Serviços de Cloud** para hospedagem e uso de serviços em nuvem.
- Estudo e experimentação com APIs de **STT** e **TTS**.

5.2 ARQUITETURA E DESIGN (Mês 2)

Objetivo:

- Criar a arquitetura do sistema e definir os módulos principais.

Atividades:

- Desenho da arquitetura que integra a IA generativa, STT e TTS, utilizando serviços em cloud.
- Divisão das responsabilidades dos módulos, como a parte de interação por voz (STT e TTS) e a lógica de IA generativa.

Ferramentas e Tecnologias:

- **Python** para criação dos primeiros scripts e componentes modulares.
- **Infraestrutura Cloud Azure** para suportar o processamento e a comunicação de dados.
- Ferramentas de design Draw.IO.

5.3 DESENVOLVIMENTO E IMPLEMENTAÇÃO (Mês 3, Mês 4, Mês 5)

Objetivo:

- Implementar as principais funcionalidades do projeto.

Atividades:

- Desenvolvimento do sistema de STT (Speech-to-Text) para interpretar comandos de voz.
- Implementação da lógica da IA generativa para fornecer respostas inteligentes.
- Integração de TTS (Text-to-Speech) para a resposta em voz.
- Integrar com o mini-computador o software.

Ferramentas e Tecnologias:

- **Python** para implementação e integração de IA generativa e integração com STT e TTS.
- **APIs de STT e TTS** (Microsoft Azure Cognitive Services).

- **API IA** (ChatGPT e Gemini)
- **Cloud Services** para usar serviços e processamento da IA (STT, TTS e IA generativa)
- **Postman** para testar APIs de integração.

6. RELATO DO DESENVOLVIMENTO TÉCNICO

Hardware

O Talk-E foi construído utilizando um Raspberry Pi como base. Esse microcomputador foi escolhido devido ao seu custo benefício e à sua flexibilidade para suportar múltiplos periféricos e processar as tarefas necessárias para o funcionamento do dispositivo. O Raspberry Pi foi integrado com um microfone USB para captar a voz do usuário e um alto-falante que permite a resposta do dispositivo.

Software e Integração com IA

Para o reconhecimento de voz e processamento de linguagem natural, foi utilizado um serviço de IA baseado em computação em nuvem, o que permite que o dispositivo tenha poder de processamento para lidar com interações complexas de forma eficiente. Modelos como o GPT foram adaptados para as necessidades do Talk-E, garantindo uma experiência de interação fluida e precisa.

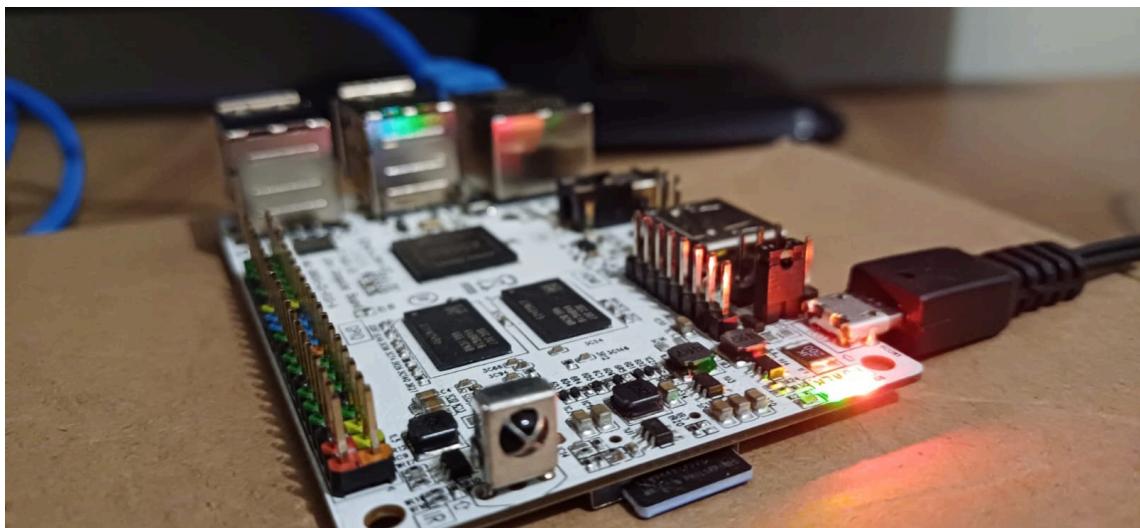
6.1. HARDWARE



Conexão dos componentes eletrônicos



Integração do Hardware

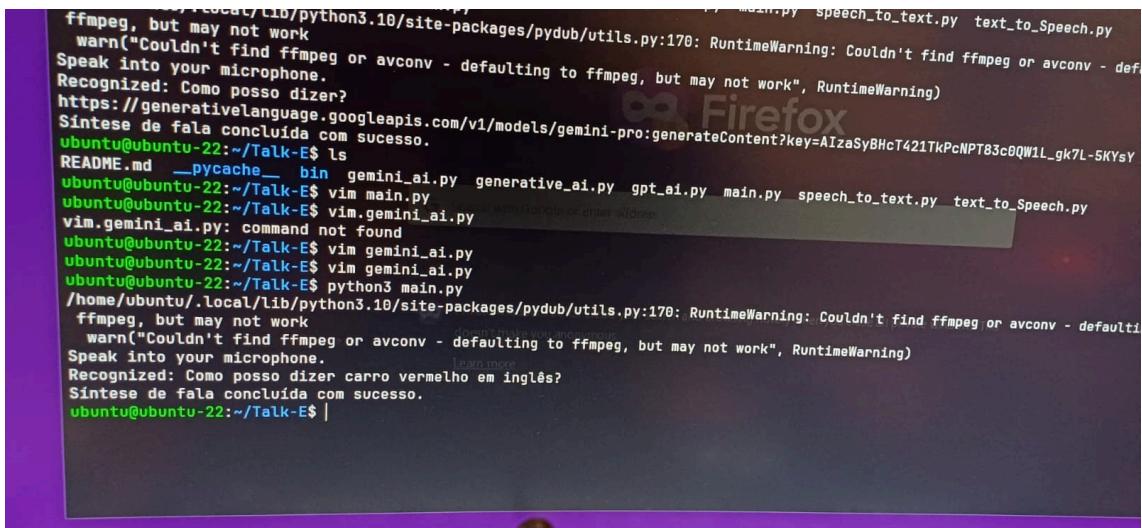


Raspberry Pi



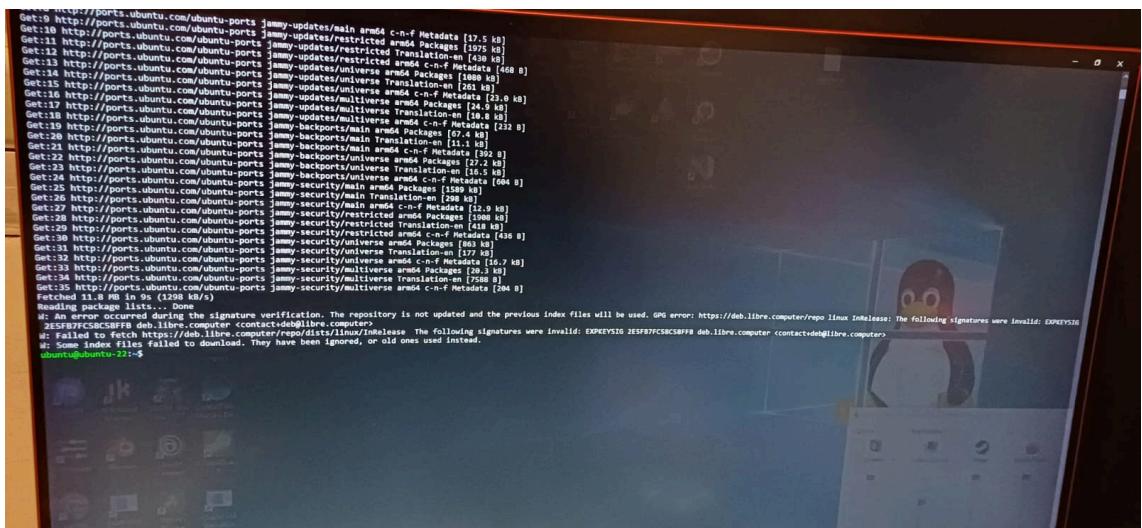
Teste do sistema de áudio e microfone

6.2. SOFTWARE



```
ffmpeg, but may not work
  Warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Speak into your microphone.
Recognized: Como posso dizer?
https://generativelanguage.googleapis.com/v1/models/gemini-pro:generateContent?key=AIzaSyBHCt421TkPcNPT83c8QW1L_gk7L-SKYsY
ubuntu@ubuntu-22:~/Talk-E$ ls
README.md __pycache__ bin gemini_ai.py generative_ai.py gpt_ai.py main.py speech_to_text.py text_to_Speech.py
ubuntu@ubuntu-22:~/Talk-E$ vim main.py
ubuntu@ubuntu-22:~/Talk-E$ vim gemini_ai.py
vim.gemini_ai.py: command not found
ubuntu@ubuntu-22:~/Talk-E$ vim gemini_ai.py
ubuntu@ubuntu-22:~/Talk-E$ vim gemini_ai.py
ubuntu@ubuntu-22:~/Talk-E$ python3 main.py
/home/ubuntu/.local/lib/python3.10/site-packages/pydub/utils.py:170: RuntimeWarning: Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work
  warn("Couldn't find ffmpeg or avconv - defaulting to ffmpeg, but may not work", RuntimeWarning)
Speak into your microphone.
Recognized: Como posso dizer carro vermelho em inglês?
Sintese de fala concluída com sucesso.
ubuntu@ubuntu-22:~/Talk-E$ |
```

Teste do software de reconhecimento de voz



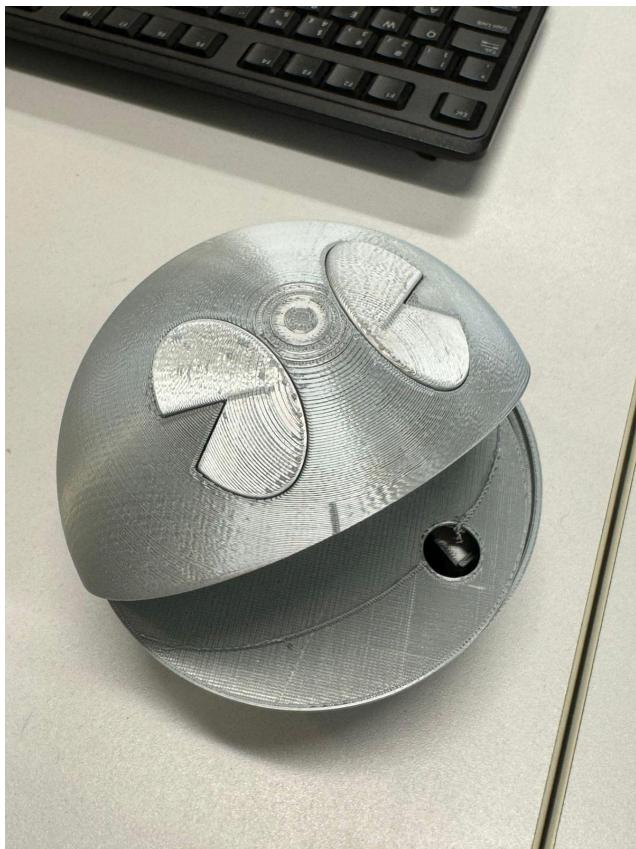
```
Get:9 http://ports.ubuntu.com/ubuntu-ports jammy-updates/main arm64 c-n-f Metadata [17.9 kB]
Get:10 http://ports.ubuntu.com/ubuntu-ports jammy-updates/restricted arm64 Packages [392 kB]
Get:11 http://ports.ubuntu.com/ubuntu-ports jammy-updates/restricted Translation-en [436 kB]
Get:12 http://ports.ubuntu.com/ubuntu-ports jammy-updates/universe arm64 Packages [468 kB]
Get:13 http://ports.ubuntu.com/ubuntu-ports jammy-updates/universe arm64 Translation-en [261 kB]
Get:14 http://ports.ubuntu.com/ubuntu-ports jammy-updates/universe Translation-en [261 kB]
Get:15 http://ports.ubuntu.com/ubuntu-ports jammy-updates/multiverse arm64 Packages [23.0 kB]
Get:16 http://ports.ubuntu.com/ubuntu-ports jammy-updates/multiverse arm64 Translation-en [18.8 kB]
Get:17 http://ports.ubuntu.com/ubuntu-ports jammy-updates/multiverse Translation-en [232 kB]
Get:18 http://ports.ubuntu.com/ubuntu-ports jammy-backports/main arm64 c-n-f Metadata [232 kB]
Get:19 http://ports.ubuntu.com/ubuntu-ports jammy-backports/main arm64 Translation-en [232 kB]
Get:20 http://ports.ubuntu.com/ubuntu-ports jammy-backports/main Translation-en [11.1 kB]
Get:21 http://ports.ubuntu.com/ubuntu-ports jammy-backports/universe arm64 c-n-f Metadata [392 kB]
Get:22 http://ports.ubuntu.com/ubuntu-ports jammy-backports/universe arm64 Translation-en [392 kB]
Get:23 http://ports.ubuntu.com/ubuntu-ports jammy-backports/universe Translation-en [16.5 kB]
Get:24 http://ports.ubuntu.com/ubuntu-ports jammy-security/main arm64 c-n-f Metadata [604 kB]
Get:25 http://ports.ubuntu.com/ubuntu-ports jammy-security/main arm64 Translation-en [298 kB]
Get:26 http://ports.ubuntu.com/ubuntu-ports jammy-security/main Translation-en [298 kB]
Get:27 http://ports.ubuntu.com/ubuntu-ports jammy-security/restricted arm64 c-n-f Metadata [12.9 kB]
Get:28 http://ports.ubuntu.com/ubuntu-ports jammy-security/restricted arm64 Translation-en [12.9 kB]
Get:29 http://ports.ubuntu.com/ubuntu-ports jammy-security/restricted Translation-en [418 kB]
Get:30 http://ports.ubuntu.com/ubuntu-ports jammy-security/universe arm64 c-n-f Metadata [436 kB]
Get:31 http://ports.ubuntu.com/ubuntu-ports jammy-security/universe arm64 Translation-en [177 kB]
Get:32 http://ports.ubuntu.com/ubuntu-ports jammy-security/universe arm64 Translation-en [177 kB]
Get:33 http://ports.ubuntu.com/ubuntu-ports jammy-security/multiverse arm64 Packages [20.1 kB]
Get:34 http://ports.ubuntu.com/ubuntu-ports jammy-security/multiverse arm64 Translation-en [7580 kB]
Get:35 http://ports.ubuntu.com/ubuntu-ports jammy-security/multiverse arm64 c-n-f Metadata [204 kB]
Fetched 11.8 MB in 95 (1290 kB/s)
Reading package lists... Done
W: GPG error: https://deb.libre.computer/repo linux InRelease: The repository is not updated and the previous index files will be used. GPG error: https://deb.libre.computer/repo linux InRelease: The following signatures were invalid: EXPKEYSIG 2E5F97FCS8FB deb.libre.computer ccontact@deplibre.computer
W: Failed to fetch https://deb.libre.computer/repo/dists/linux/InRelease  The following signatures were invalid: EXPKEYSIG 2E5F97FCS8FB deb.libre.computer ccontact@deplibre.computer
W: Some index files failed to download. They have been ignored, or old ones used instead.
ubuntu@ubuntu-22:~$
```

Verificação e processamento do sistema

6.3. IMPRESSÃO 3D



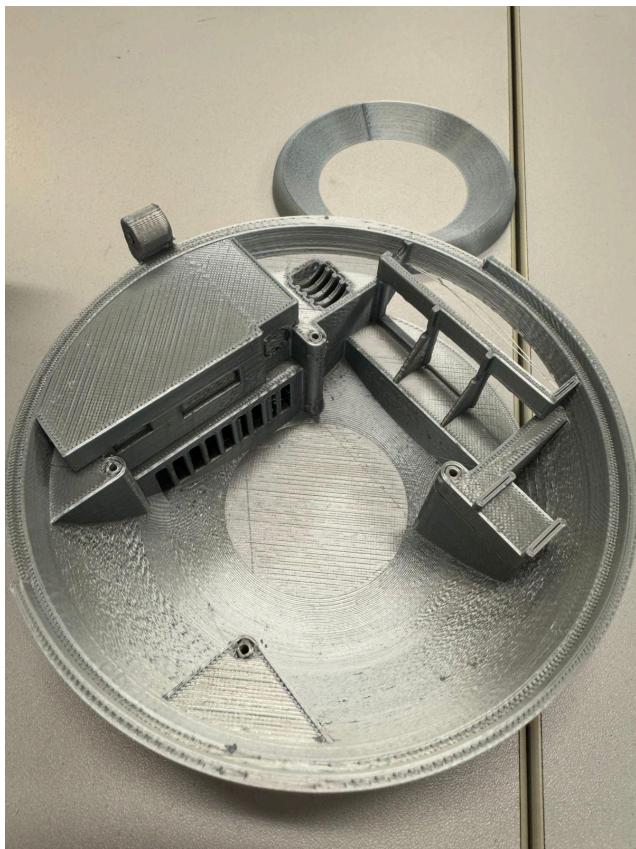
Protótipo em impressão 3D



Dispositivo com design compacto



Estrutura base do Talk-E



Interior do equipamento



Montagem interna do dispositivo



Estrutura 3D

7. CONSIDERAÇÕES FINAIS

O projeto Talk-E representa um passo significativo na convergência entre tecnologia e educação, atendendo à crescente demanda por soluções inovadoras no ensino de idiomas. A motivação inicial do desenvolvimento foi criar um dispositivo que pudesse auxiliar no aprendizado da língua inglesa de maneira viável e interativa utilizando Inteligência Artificial como base para proporcionar um feedback imediato e eficiente aos usuários. Desde o início, buscou-se não apenas um produto funcional, mas também uma nova abordagem para a prática da conversação e correção gramatical.

Durante o processo de desenvolvimento, foi possível superar diversos desafios técnicos relacionados à integração entre hardware e software. A construção do protótipo envolveu a configuração de um Raspberry Pi para processar comandos de voz, que, conectados a um sistema de computação em nuvem, executam algoritmos de processamento de linguagem natural. Além disso, a fase de testes revelou importantes aspectos a serem refinados, como a melhoria na precisão de reconhecimento de voz e a eficácia das correções gramaticais.

Com a evolução do projeto, foi possível não apenas criar uma ferramenta inovadora, mas também abrir novas portas para o desenvolvimento futuro de dispositivos interativos no contexto educacional. O Talk-E demonstra o potencial do uso de IA no aprimoramento de habilidades linguísticas e sugere que, com o avanço contínuo dessa tecnologia, novas oportunidades poderão surgir. Este projeto oferece uma solução concreta e acessível, com um impacto direto no processo de aprendizado, e posiciona-se como uma alternativa eficaz para a prática de inglês, podendo evoluir para incorporar novos idiomas e funcionalidades no futuro. Assim, o Talk-E contribui de maneira significativa para o campo da tecnologia educacional, ao mesmo tempo em que inspira novas pesquisas e desenvolvimentos nessa área.

8. REFERÊNCIAS BIBLIOGRÁFICAS

Aqui você irá listar todas as fontes bibliográficas utilizadas/citadas em outros itens, como artigos, livros, vídeos, sites e tantos outros.

ARMBRUST, Michael, et al. "A View of Cloud Computing." Communications of the ACM, 2010.

Azure AI services documentation. Disponível em:

<https://learn.microsoft.com/en-us/azure/ai-services/>. Acesso em: 01 out 2024.

Começar a usar a API Gemini. 2024. Disponível em:

<https://ai.google.dev/gemini-api/docs?hl=pt-br>. Acesso em: 01 out 2024.

Linguagem de marcação de síntese de voz (SSML). 2024. Disponível em:

<https://cloud.google.com/text-to-speech/docs/ssml?hl=pt-br>. Acesso em: 01 out 2024.

PAPERT, Seymour. "The Children's Machine: Rethinking School in the Age of the Computer." Basic Books, 1993.

Urban, Eric. Mehrotra, Nitin. "Visão geral do SSML (Speech Synthesis Markup Language)". 24/09/2024. Disponível em:

<https://learn.microsoft.com/pt-br/azure/ai-services/speech-service/speech-synthesis-markup>. Acesso em: 01 out 2024.

Urban, Eric. Mehrotra, Nitin. Zoubaolian. "Customize voice and sound with SSML". 24/09/2024. Disponível em:

<https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-synthesis-markup-voice>. Acesso em: 01 out 2024.

Urban, Eric. Mehrotra, Nitin. Zoubaolian. Yulin Li. "What is text to speech?".

24/09/2024. Disponível em:

<https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech>. Acesso em: 01 out 2024.

Von Ahn, Luis, and Hacker, Severin. "Duolingo: Learn a Language for Free While Helping to Translate the Web." Proceedings of the 2013 conference on Intelligent User Interfaces, 2013.