INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS - *CAMPUS* BETIM
BACHARELADO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Warley Soares Ribeiro

DESENVOLVIMENTO DE APLICATIVO ANDROID PARA TREINAMENTO DE PRONÚNCIA INTEGRADA AO ANKI

Betim
2024

WARLEY SOARES RIBEIRO

DESENVOLVIMENTO DE APLICATIVO ANDROID PARA TREINAMENTO DE PRONÚNCIA INTEGRADA AO ANKI

Trabalho de Conclusão de Curso apresentado à banca examinadora do curso de Engenharia de Controle e Automação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais *Campus* Betim, como parte dos requisitos para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientador: Prof. Maurício Monteiro da Silva

Coorientador: Prof. Dr. Leandro Freitas de Abreu

Betim
2024

FICHA CATALOGRÁFICA

R484d Ribeiro, Warley Soares
Desenvolvimento de aplicativo android para treinamento de pronúncia integrada ao Anki / Warley Soares Ribeiro. – 2024.

41 f.: il.

Trabalho de conclusão de curso (Bacharelado em Engenharia de Controle e Automação) - Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Campus Betim, 2024.

Orientação: Prof. Esp. Maurício Monteiro da Silva
Coorientação: Prof. Dr. Leandro Freitas de Abreu

1. Inteligência artificial. 2. Dispositivos móveis. 3. Avaliação de pronúncia. 4. Android (Recurso eletrônico). 5. Engenharia de Controle e Automação. I. Ribeiro, Warley Soares. II. Título.

CDU: 681.5

Ficha catalográfica elaborada pelo Bibliotecário Denísio Pereira Marcos CRB-6/3142

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
Campus Betim
Diretoria de Ensino
Docentes Automação Industrial e Tecnologia da Informação
Rua Itamarati - CEP 32677-564 - Betim - MG
3135976360 - www.ifmg.edu.br

ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

BACHARELADO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Aos 19 dias do mês de Dezembro do ano de 2024, às 18:00 horas, no endereço R. Itaguaçu, 595 - São Caetano, Betim - MG, 32677-562, Brasil, iniciou-se a apresentação pública do Bacharelado em Engenharia de Controle e Automação, pelo(a) discente **Warley Soares Ribeiro**, intitulado Desenvolvimento de aplicativo Android para treinamento de pronúncia integrada ao Anki, tendo como orientador(a) o(a) Prof. Mauricio Monteiro da Silva. O início dos trabalhos se deu com a apresentação da Banca Examinadora que foi composta pelos seguintes membros: Prof. Mauricio Monteiro da Silva - Orientador, Prof. Me Virgil del Duca Almeida, Prof. Me Felipe Monteiro Lima e Prof. Dr. Leandro Freitas de Abreu, membros titulares. O discente iniciou sua apresentação, expondo seu trabalho durante 30 minutos. Os membros da banca apresentaram seus questionamentos e sugestões que foram respondidos pelo(a) discente. A seguir, a Banca Examinadora reuniu-se, sem a presença do(a) discente e do público, para fazer a avaliação final do trabalho apresentado. Em conclusão, a Banca Examinadora deliberou que o Trabalho de Conclusão de Curso foi:

(X) Aprovado.

( ) Aprovado com ressalvas*.

( ) Reprovado.

O conceito atribuído ao TCC foi A e a nota 89,0. Eu, Mauricio Monteiro da Silva, Presidente da Banca Examinadora, lavrei a presente ata que será assinada por mim e pelos demais membros da Banca.

*O não cumprimento das exigências pelo(a) discente no prazo estabelecido implicará na sua reprovação.

Prof. Mauricio Monteiro da Silva - IFMG *Campus* Betim - Orientador

Prof. Me. Virgil del Duca Almeida - IFMG *Campus* Betim

Prof. Me. Felipe Monteiro Lima - IFMG *Campus* Santa Luzia

Prof. Dr. Leandro Freitas de Abreu - IFMG *Campus* Betim

Exigências de revisão indicadas pela Banca Examinadora

Prof. Mauricio Monteiro da Silva - IFMG *Campus* Betim - Orientador

Prof. Me. Virgil del Duca Almeida - IFMG *Campus* Betim

Prof. Me. Felipe Monteiro Lima - IFMG *Campus* Santa Luzia

Prof. Dr. Leandro Freitas de Abreu - IFMG *Campus* Betim

Betim, 20 de Dezembro de 2024.

Documento assinado eletronicamente por **Mauricio Monteiro da Silva**, **Professor**, em 20/01/2025, às 17:57, conforme Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Felipe Monteiro Lima**, **Professor**, em 21/01/2025, às 09:12, conforme Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leandro Freitas de Abreu**, **Professor**, em 21/01/2025, às 16:17, conforme Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Virgil Del Duca Almeida**, **Professor**, em 21/01/2025, às 17:54, conforme Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade do documento pode ser conferida no site https://sei.ifmg.edu.br/consultadocs informando o código verificador **2118190** e o código CRC **4BA12463**.

RESUMO

O aprendizado de idiomas em dispositivos móveis tem crescido cada vez mais e ferramentas para aumento de vocabulário e treinamento de pronúncia auxiliam no processo. Este documento apresenta o contexto geral sobre a avaliação de pronúncia através de softwares, sobre o uso de inteligência artificial para a geração de flashcards e o reconhecimento de imagens utilizando OCR e visão computacional. O objetivo do trabalho é o desenvolvimento de um aplicativo Android que integre as funcionalidades de importação de vocabulário por reconhecimento de palavras destacadas utilizando OpenCV, criação de flashcards integrada ao AnkiDroid utilizando a API do Google Gemini e avaliação de pronúncia com os cards disponíveis utilizando o Azure Speech Services. O resultado obtido foi um aplicativo funcional de código aberto publicado no GitHub disponível para versões do Android 10 ou superior. O trabalho destaca as tecnologias utilizadas e as técnicas de implementação e de validação da qualidade dos serviços utilizados nas funcionalidades criadas.

Palavras-chave: Avaliação de pronúncia; Desenvolvimento android; Anki; Api azure; Api gemini.

ABSTRACT

Language learning on mobile devices has been growing more and more, and tools to increase vocabulary and practice pronunciation help in the process. This document presents the general context on pronunciation assessment through software, on the use of artificial intelligence for the generation of flashcards and image recognition using OCR and computer vision. The objective of the work is the development of an Android application that integrates the functionalities of vocabulary import by recognition of highlighted words using OpenCV, creation of flashcards integrated with AnkiDroid using the Google Gemini API and pronunciation assessment with the available cards using Azure Speech Services. The result obtained was a functional open source application published on GitHub available for versions of Android 10 or higher. The work highlights the technologies used and the techniques for implementation and validation of the quality of the services used in the functionalities created.

Keywords: Pronunciation assessment; Android development; Anki; Azure api; Gemini api.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ASR	Automatic Speech Recognition
STT	Speech To Text
EFL	English as a Foreign Language
LLM	Large Language Model
SoC	System on a chip
SR	Spaced Repetition
GOP	Goodness of Pronunciation
MALL	Mobile Assisted Language Learning
UI	User Interface

SUMÁRIO

1 INTRODUÇÃO

O aprendizado de um novo idioma por pessoas adultas pode ser muito desafiador e desgastante. Diferentemente do aprendizado da língua materna na infância, onde a aquisição ocorre de forma natural através da exposição da criança diariamente aos estímulos dos cuidadores, aprender um idioma estrangeiro requer o esforço de buscar ativamente essa exposição.

Segundo (KRASHEN, 1981), a maioria dos programas de aprendizado de idiomas, se sub-divididos em componentes, divide o conteúdo em "quatro habilidades", fala, compreensão auditiva, leitura e escrita, porém essa divisão é artificial pois é impossível focar em uma habilidade e ignorar as outras. Apesar disso a pronúncia aparenta ser o aspecto mais difícil da aquisição de um segundo idioma, pois se aprofunda mais no centro da personalidade dos alunos.

Com a democratização da computação e do acesso a internet, a quantidade de conteúdo disponível para o estudo se tornou massiva, porém mesmo com a disponibilidade de material à distância de uma simples busca, uma pequena parcela da população brasileira tem proficiência em um idioma estrangeiro (BRITISH COUNCIL, 2014). Essa democratização ocorreu em grande parte através dos smartphones, onde de acordo com (GODWIN-JONES, 2011), houve um crescimento do interesse no uso de aplicativos por parte dos educadores de idiomas.

> Parece que todo mundo, de agências do governo federal até sua padaria local, tem um aplicativo disponível. Esse fenômeno, não surpreendentemente, levou a um tremendo interesse entre educadores. O aprendizado móvel (frequentemente “m-learning”) não é novo em si, mas novos dispositivos com capacidades aprimoradas aumentaram drasticamente o nível de interesse, inclusive entre educadores de idiomas. (GODWIN-JONES, 2011) [1]

Um método bastante popular entre os estudantes de idiomas é o uso de softwares de repetição espaçada como o *Anki*, especialmente no início do aprendizado, onde é essencial a aquisição de vocabulário. O software utiliza um algoritmo que mostra os *flashcards* introduzidos recentemente e aqueles que o usuário apresenta maior dificuldade com maior frequência, ao passo que os mais fáceis são revisados com maior espaço temporal, ajudando a consolidar a memória de longo prazo.

A leitura desempenha um papel fundamental no aprendizado de um idioma, pois expõe os estudantes a estruturas gramaticais, vocabulário e contextos culturais de maneira natural e abrangente. Além de desenvolver a compreensão textual, a leitura amplia o repertório linguístico ao apresentar palavras e expressões em contextos reais. Identificar palavras desconhecidas durante a leitura e transformá-las em oportunidades de aprendizado posterior é uma prática essencial

[1] Tradução livre: It seems that everyone from federal government agencies to your local bakery has an app available. This phenomenon, not surprisingly has led to tremendous interest among educators. Mobile learning (often “m-learning”) is in itself not new, but new devices with enhanced capabilities have dramatically increased the interest level, including among language educators.

para a aquisição de vocabulário. De acordo com (NATION, 2001), a retenção de palavras é significativamente maior quando o aprendizado está contextualizado e atrelado a estratégias ativas, como a criação de listas personalizadas ou *flashcards*. Essa abordagem permite que os estudantes consolidem o vocabulário recém-descoberto, promovendo um aprendizado mais eficaz e duradouro.

Conforme o progresso no aprendizado, os estudantes se deparam com a dificuldade de praticar o idioma falado, devido muitas vezes não terem com quem conversar. Os estudantes podem recorrer a prática autônoma, porém a ausência de correção imediata pode levar ao reforço de hábitos de pronúncia incorretos, dificultando a comunicação efetiva.

Segundo (MUNRO; DERWING, 2015), a pronúncia é essencial para a inteligibilidade em um novo idioma, e métodos que oferecem feedback detalhado, como tecnologias de reconhecimento de voz, são cruciais para melhorar a confiança e a precisão na fala. O avanço da tecnologia de conversão de voz em texto possibilitou o desenvolvimento de soluções de avaliação de pronúncia capazes de fornecer feedback a nível de fonemas, permitindo que os usuários melhorem suas habilidades de fala de forma autônoma.

1.1 Objetivos

1.1.1 *Objetivo geral*

A proposta deste trabalho é desenvolver um aplicativo para smartphones Android onde o usuário seja capaz de treinar sua pronúncia usando os flashcards que possui no *Anki*, além da funcionalidade de criação automática de novos cards utilizando inteligência artificial e importação de vocabulário através de imagens.

1.1.2 *Objetivos específicos*

- Permitir que o usuário seja capaz de treinar sua pronúncia com os seus próprios flashcards do *Anki*, gravando a sua voz pronunciando o texto do card e recebendo o feedback da avaliação a nível de fonemas na tela
- Possibilitar a criação automática de novos flashcards utilizando inteligência artificial para gerar frases utilizando uma palavra ou expressão de uma lista de vocabulário
- Criar a funcionalidade de importação de vocabulário por imagem, reconhecendo de forma automática palavras destacadas com um marca-texto em um livro físico a partir de uma foto da página

1.2 Justificativa

Segundo o (BRITISH COUNCIL, 2014) somente 5% da população brasileira declara ter conhecimento na língua inglesa, atualmente a mais falada e estudada no mundo e essencial para oportunidades melhores de trabalho. Diversos fatores como a baixa qualidade do ensino nas escolas públicas, poucos imigrantes em cidades não turísticas e a falta de recursos financeiros para tutores particulares contribuem para isso.

As tecnologias de reconhecimento automático de fala (ASR) evoluíram bastante nos últimos anos e a integração de softwares de voz para texto (STT) no aprendizado do inglês como língua estrangeira (EFL) demostraram resultados consistentes quando comparados com avaliadores humanos (HIRAI; KOVALYOVA, 2023).

Com o advento da inteligência artificial foi possível treinar modelos com grande quantidade de dados de áudio de falantes nativos e não nativos, melhorando ainda mais a confiabilidade dos resultados e proporcionando uma análise contextual da fala, considerando pausas inesperadas, entonação, velocidade e ritmo (MICROSOFT, 2023).

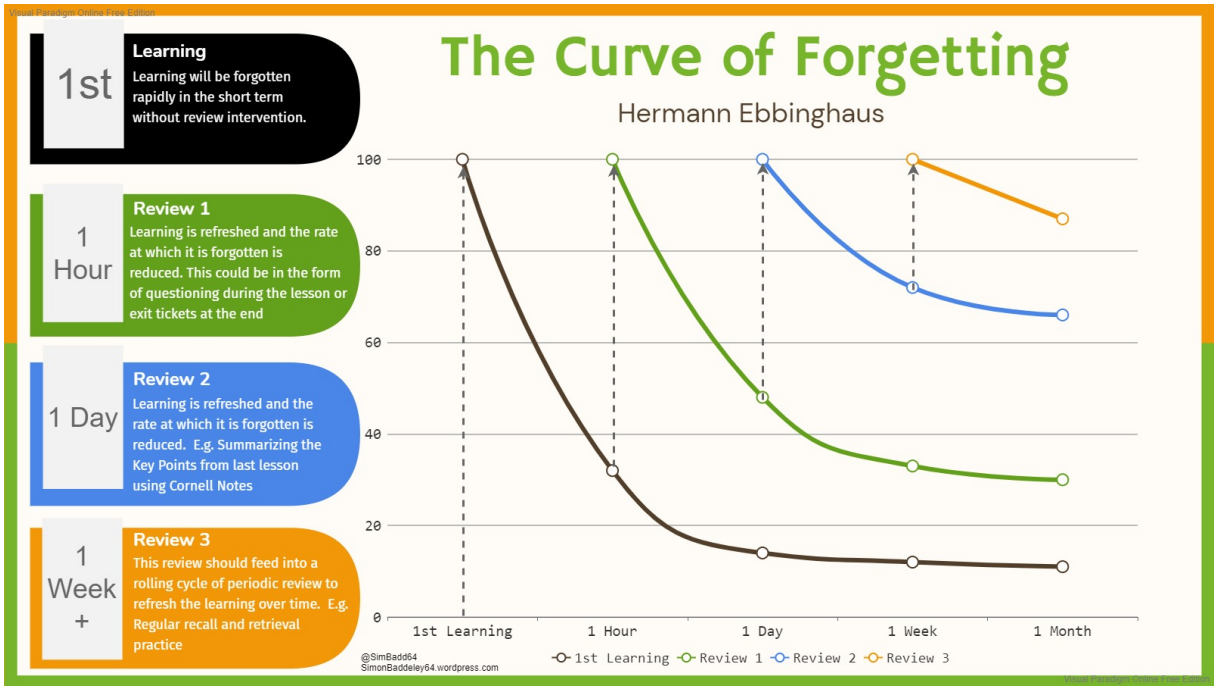
O uso dessas ferramentas em combinação com um software já utilizado por estudantes de idiomas apresenta ser uma ótima proposta para facilitar a prática da pronúncia em um idioma estrangeiro.

2 REVISÃO BIBLIOGRÁFICA

2.1 Uso do Anki para estudo de idiomas

O Anki é um software de repetição espaçada (SR) que se utiliza do conceito de flashcards, que geralmente são baseado em dois campos, um vísivel ao usuário no momento da revisão e outro que é mostrado através de um comando. O primeiro pode ser uma pergunta ou algum outro conteúdo como texto ou até imagens e vídeos para contextualizar, e o segundo serve como a resposta ou lembrete do que está sendo revisado. A repetição espaçada se baseia na curva de esquecimento de (EBBINGHAUS, 1885), onde a retenção do conteúdo estudado pode ser aumentada através de revisões espaçadas. A Figura 1 demonstra essa curva e como revisões espaçadas podem aumentar o porcentagem de retenção.

Figura 1 – Curva de esquecimento.



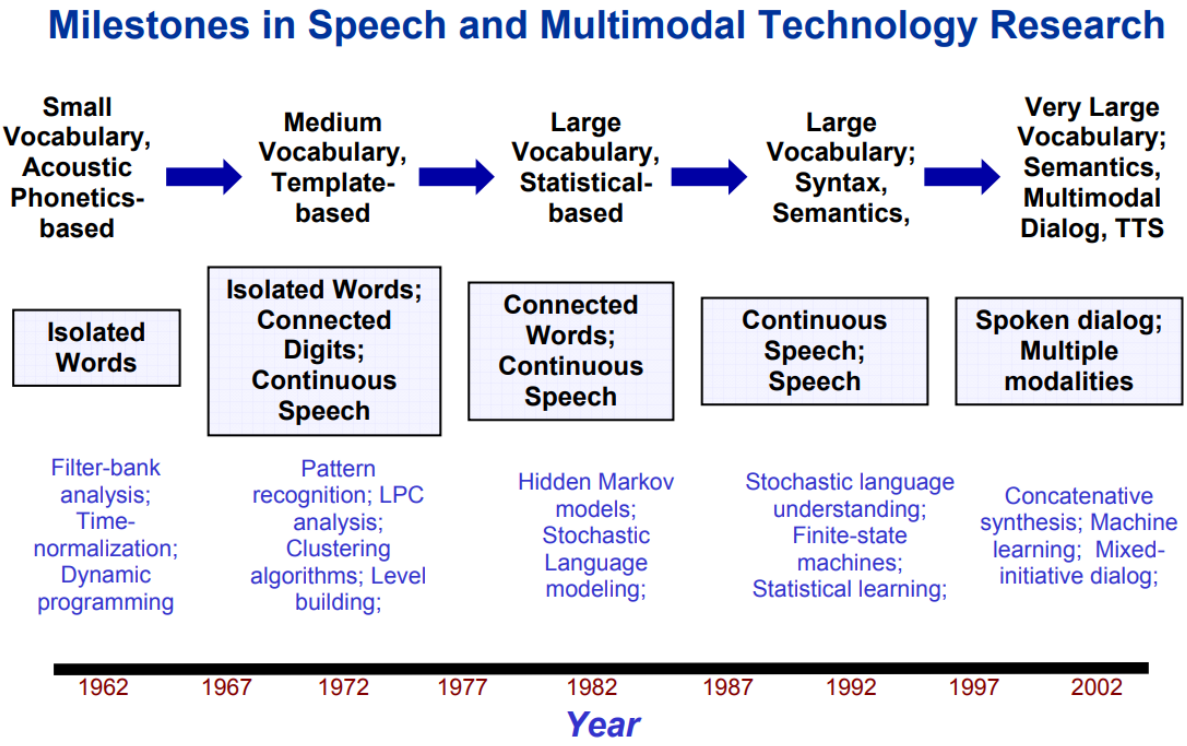
Fonte: (BADDELEY, 2021)

O uso de repetição espaçada para o aprendizado de idiomas se popularizou devido a grande quantidade de vocabulário que este método possibilita reter. O Anki para o estudo de EFL se mostrou mais eficaz no nível de retenção quando comparado com outros métodos MALL e métodos tradicionais (KHOSHSIMA; KHOSRAVI, 2021-2022) e a repetição espaçada vem sendo implementada em aplicativos de idiomas reconhecidos mundialmente como o Duolingo (SETTLES; MEEDER, 2016).

2.2 Avaliação de pronúncia

O princípio fundamental para realizar uma avaliação de pronúncia é o ASR. Essa tecnologia que surgiu na década de 1960 foi sendo aprimorada ao longo do tempo, conforme demonstra a Figura 2, passando por sistemas baseados em fonemas, análise espectral, métodos estatísticos até chegar no aprendizado de máquina que é o mais utilizado atualmente (JUANG; RABINER, 2004). O ASR permitiu o desenvolvimento de aplicações com diversas finalidades como transcrição de áudio, comando por voz, automação de call centers e dentre elas a avaliação de pronúncia.

Figura 2 – Evolução do ASR.

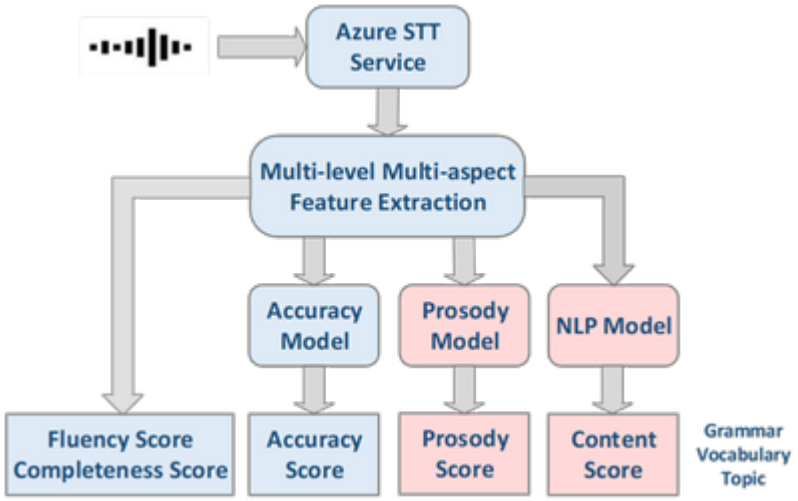


Fonte: JUANG; RABINER, 2004.

Para realizar a avaliação da pronúncia é necessário obter o áudio da pessoa pronunciando um texto de referência, então é reconhecido o texto do áudio e utilizando um dicionário de fonemas é cálculado o Goodness of Pronunciation (GOP), que representa a qualidade da pronúncia reconhecida pelo sistema em relação a pronúncia correta. Os resultados do GOP são utilizados para identificação de erros de pronúncia e podem ser mais precisos utilizando redes neurais (HU *et al.*, 2015).

Há diversos serviços disponíveis atualmente que implementam sistemas de ASR e os disponibilizam através de APIs, o disponibilizado pela *Microsoft Azure*, cuja arquitetura está ilustrada na Figura 3, apresentou a maior acurácia de 98,11% para o português brasileiro (BORGHI; TEIXEIRA; FREITAS, 2023), além de possuir uma API dedicada a avaliação de pronúncia, com múltiplas camadas de análise.

Figura 3 – Arquitetura da API de avaliação de pronúncia.



Fonte: MICROSOFT, 2023.

Sendo uma dessas análises a de prosódia que é a área da fonética que estuda o emprego correto da entonação e ritmo nas palavras, possuindo um papel fundamental na pronúncia correta. Muitos imigrantes apesar de serem capazes de se comunicarem possuem limitações nas interações com falantes nativos devido não conseguirem reproduzir claramente os sons da língua, podendo melhorar significativamente nesse aspecto utilizando de técnicas de repetição (KJELLIN, 1999) como a proposta do aplicativo desenvolvido.

2.3 Criação automática de flashcards

A criação automática de flashcards consiste no uso de inteligência artificial generativa para criação de conteúdo contextualizado para posterior revisão, onde através de um prompt é requisitado a geração de dados em um formato específico que depois pode ser importado em um software de SR. Ao criar cards para estudo de idiomas é essencial que o modelo tenha conhecimentos de múltiplos tópicos para que o conteúdo gerado seja relevante para o estudante.

Uma tecnologia que poderia ser aplicada é o uso de LLMs no próprio dispositivo utilizando o framework MediaPipe, porém os modelos disponíveis atualmente possuem limitações consideráveis no entendimento devido o número limitado de parâmetros. Outra limitação relevante é a necessidade de hardware potente para a execução local das inferências, quando usado em um smartphone com processador *Snapdragon 662* o aplicativo de exemplo não foi capaz de produzir resultados e quando testado em um *Snapdragon 695* apresentou muitos travamentos e demora. A Tabela 1 mostra a performance em tarefas de IA para diferentes processadores de smartphones, deixando claro a limitação dessa tecnologia em dispositivos de entrada.

Devido a essas limitações optou-se por utilizar um serviço em nuvem, sendo escolhido o modelo *Gemini 1.5 Flash* do Google que apresentou 67,3% de acurácia no teste MMLU

Tabela 1 – Performance de processadores em tarefas de IA.

Posição	SoC	Pontuação Total
1	Snapdragon 8 Gen 3	1402798
2	Snapdragon 8 Gen 2	1059060
3	Snapdragon 8+ Gen 1	879683
4	Snapdragon 8 Gen 1	787826
5	Snapdragon 7+ Gen 2	586458
...	...	...
28	Snapdragon 695	83003
...	...	...
34	Snapdragon 662	56466

Fonte: Adaptado de (ANTUTU, 2024)

(HENDRYCKS *et al.*, 2021) que testa a capacidade em 57 diferentes assuntos como história, matemática, direito, biologia etc. Outro fator relevante foi a disponibilidade da API gratuita para o mesmo, em contraste com somente versões de teste limitadas para os principais concorrentes. O modelo pago *Gemini Ultra* apresentou acurácia de 90,04% sendo o primeiro a ultrapasar a marca medida no teste para um especialista humano de 89,8% (GOOGLE, 2024). O detalhamento dos benchmarks dos modelos da família Gemini estão disponíveis na Figura 4.

2.4 Reconhecimento de palavras marcadas

Para reconhecer palavras marcadas é necessário o uso do reconhecimento óptico de caracteres (OCR). Este processo consiste em transformar texto presente em um documento como uma imagem e convertê-lo para texto digital que pode ser posteriormente processado. Também é necessário o uso de visão computacional (CV), que de maneira ampla são um conjunto de métodos para analisar programaticamente, processar e entender uma imagem, muito utilizados em etapas de pré e pós-processamento para melhorar o resultado do caso de uso do OCR (COOK, 2020). A biblioteca OpenCV será usada devido sua licença de código aberto, amplo uso em diversos softwares e suporte ao sistema operacional Android.

Figura 4 – Benchmark dos modelos da família Gemini.

	Gemini Ultra	Gemini Pro	GPT-4	GPT-3.5	PaLM 2-L	Claude 2	Inflect-ion-2	Grok 1	LLAMA-2
MMLU Multiple-choice questions in 57 subjects (professional & academic) (Hendrycks et al., 2021a)	**90.04%** CoT@32* 83.7% 5-shot	79.13% CoT@8* 71.8% 5-shot	87.29% CoT@32 (via API**) 86.4% 5-shot (reported)	70% 5-shot	78.4% 5-shot	78.5% 5-shot CoT	79.6% 5-shot	73.0% 5-shot	68.0%***
GSM8K Grade-school math (Cobbe et al., 2021)	**94.4%** Maj1@32	86.5% Maj1@32	92.0% SFT & 5-shot CoT	57.1% 5-shot	80.0% 5-shot	88.0% 0-shot	81.4% 8-shot	62.9% 8-shot	56.8% 5-shot
MATH Math problems across 5 difficulty levels & 7 subdisciplines (Hendrycks et al., 2021b)	**53.2%** 4-shot	32.6% 4-shot	52.9% 4-shot (via API**) 50.3% (Zheng et al., 2023)	34.1% 4-shot (via API**)	34.4% 4-shot	—	34.8%	23.9% 4-shot	13.5% 4-shot
BIG-Bench-Hard Subset of hard BIG-bench tasks written as CoT problems (Srivastava et al., 2022)	**83.6%** 3-shot	75.0% 3-shot	83.1% 3-shot (via API**)	66.6% 3-shot (via API**)	77.7% 3-shot	—	—	—	51.2% 3-shot
HumanEval Python coding tasks (Chen et al., 2021)	**74.4%** 0-shot (PT****)	67.7% 0-shot (PT****)	67.0% 0-shot (reported)	48.1% 0-shot	—	70.0% 0-shot	44.5% 0-shot	63.2% 0-shot	29.9% 0-shot
Natural2Code Python code generation. (New held-out set with no leakage on web)	**74.9%** 0-shot	69.6% 0-shot	73.9% 0-shot (via API**)	62.3% 0-shot (via API**)	—	—	—	—	—
DROP Reading comprehension & arithmetic. (metric: F1-score) (Dua et al., 2019)	**82.4** Variable shots	74.1 Variable shots	80.9 3-shot (reported)	64.1 3-shot	82.0 Variable shots	—	—	—	—
HellaSwag (validation set) Common-sense multiple choice questions (Zellers et al., 2019)	87.8% 10-shot	84.7% 10-shot	**95.3%** 10-shot (reported)	85.5% 10-shot	86.8% 10-shot	—	89.0% 10-shot	—	80.0%***
WMT23 Machine translation (metric: BLEURT) (Tom et al., 2023)	**74.4** 1-shot (PT****)	71.7 1-shot	73.8 1-shot (via API**)	—	72.7 1-shot	—	—	—	—

Fonte: GOOGLE, 2024.

3 METODOLOGIA

3.1 Tecnologias e ferramentas utilizadas

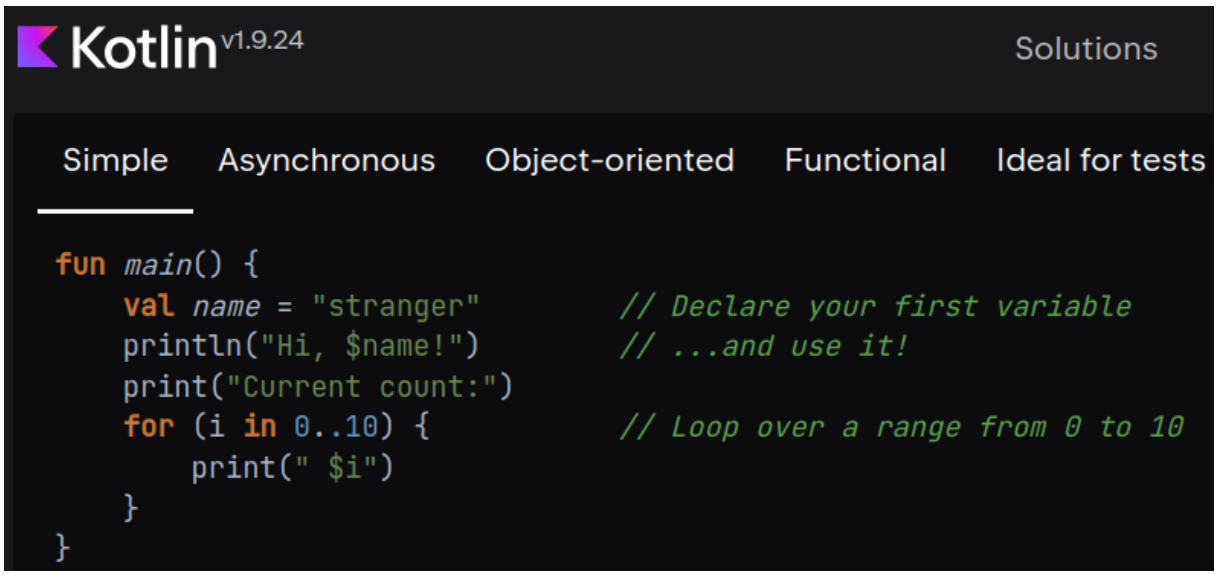
Nesta seção, serão abordadas as principais tecnologias e ferramentas empregadas no desenvolvimento do aplicativo, incluindo a linguagem de programação, a arquitetura adotada, as APIs integradas e os recursos utilizados para a construção da interface do usuário. Cada um desses elementos será descrito detalhadamente, destacando suas contribuições para o funcionamento e o sucesso do projeto.

3.1.1 Linguagem de programação Kotlin

A linguagem de programação Kotlin é uma linguagem de código aberto, estaticamente tipada com suporte a orientação a objetos e programação funcional desenvolvida pela JetBrains. Foi projetada para ser interoperável com o Java, então é possível chamar código escrito em Java no Kotlin e vice-versa, o que a torna fundamental para o desenvolvimento nativo Android, pois não é necessário migrar toda a base de código para Kotlin, podendo desenvolver em ambas linguages.

O Kotlin tem uma sintaxe mais simples e menos verbosa que o Java, conforme exemplo da Figura 5 onde um *for loop* de 0 a 10 é feito com apenas 16 caracteres, tornando o código mais conciso e fácil de realizar manutenção além de possuir diversas funcionalidades que aumentam a produtividade no desenvolvimento. É a linguagem recomendada pelo Google para desenvolvimento de aplicativos Android desde 2019. O aplicativo proposto pelo trabalho foi desenvolvido majoritariamente em Kotlin com o uso do Java para compatibilidade com a API do AnkiDroid.

Figura 5 – Exemplo de sintaxe do Kotlin.



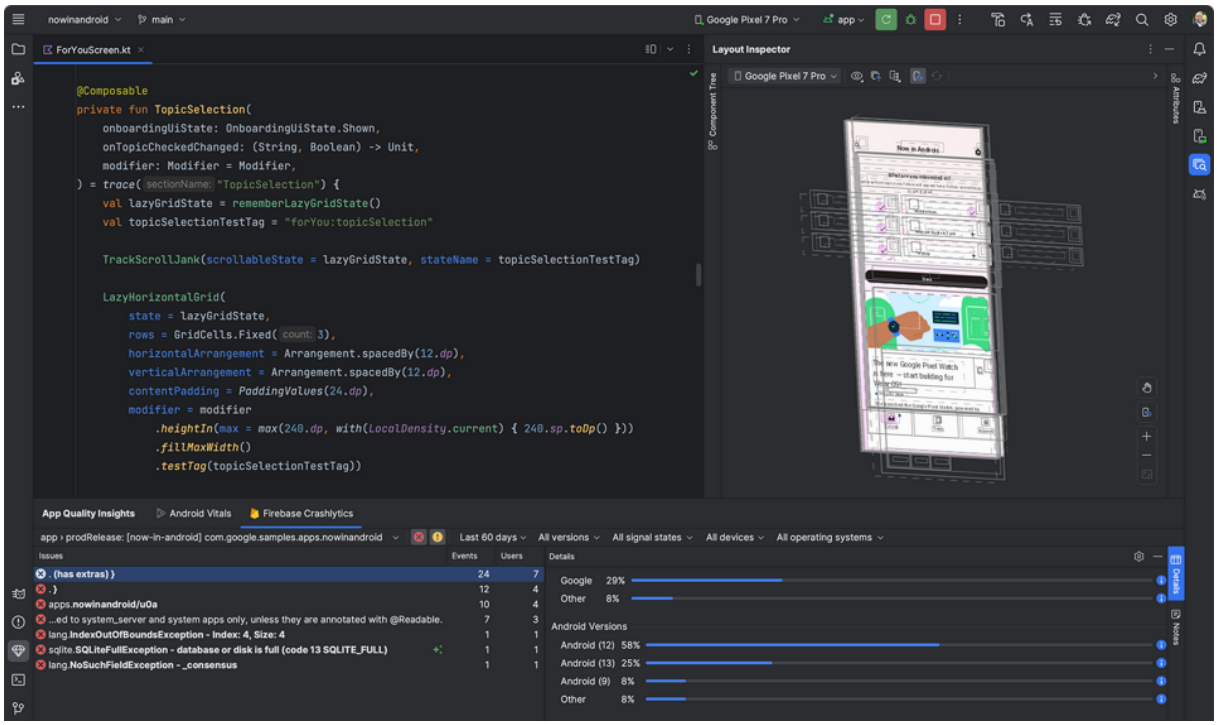
Fonte: (KOTLIN FOUNDATION, 2024)

3.1.2 AndroidStudio IDE

O Android Studio é a IDE oficial do Google para desenvolvimento Android. Baseada no IntelliJ da JetBrains ela conta com ainda mais funcionalidades que facilitam o desenvolvimento como o emulador de dispositivos, sistema de build flexível baseado no Gradle e depuração sem fio em dispositivos físicos.

Todo o desenvolvimento do aplicativo foi realizado utilizando esta IDE, fazendo o uso da vasta quantidade de recursos disponíveis focadas no ecossistema Android, conforme mostra a Figura 6, principalmente para a criação da UI com Compose onde é possível visualizar previamente como a tela será mostrada sem necessidade de realizar o build e inicialização no dispositivo físico ou emulado, o que economiza bastante tempo.

Figura 6 – Interface do AndroidStudio.



Fonte: (GOOGLE, 2024)

3.1.3 Arquitetura MVVM

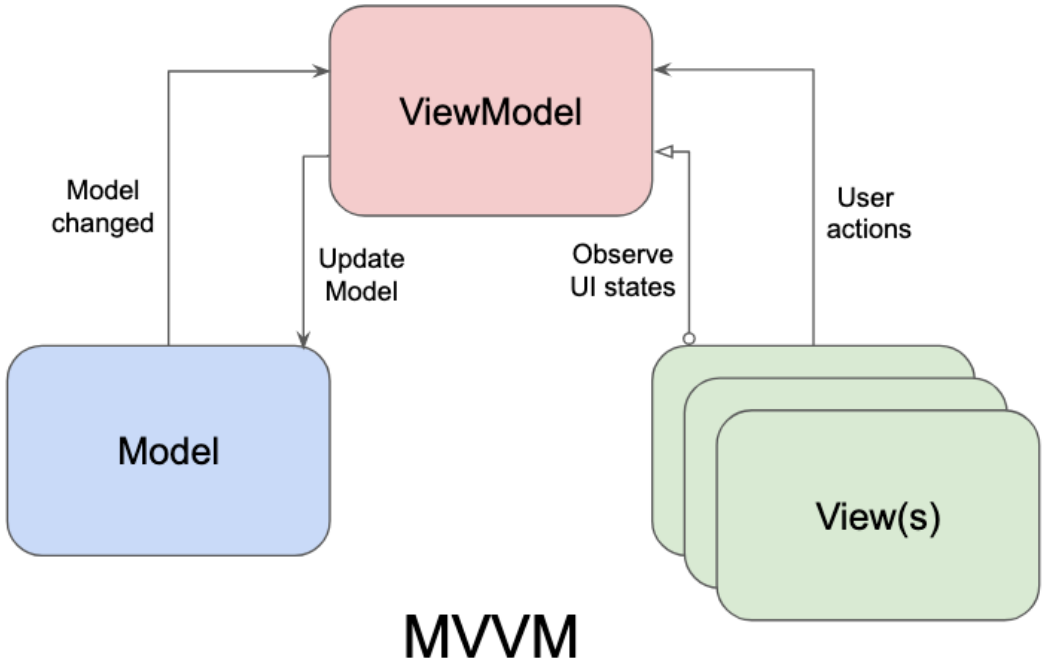
Uma das arquiteturas de software mais utilizadas no desenvolvimento de aplicativos Android é a MVVM, cujo acrônimo significa Model-View-ViewModel, que visa separar as responsabilidades e tornar o código modular e mais testável.

O Model é o que define a abstração da camada de dados. Ele pode utilizar-se de um repositório local ou fazer uma consulta a uma API remota, abstraindo os dados para objetos que fazem parte da estrutura do software e que serão manipulados pela ViewModel.

A View compreende tudo aquilo que é visível através da tela do smartphone. Nela é criada a interface do usuário, que não deve conter regra de negócio embutida, e é também responsável por receber inputs através dos campos, cliques ou gestos e requisitar que a ViewModel processe essas informações.

A ViewModel atua como uma ponte entre a View e o Model, ela é responsável por receber as ações requisitadas pela View e realizar o processamento das mesmas, requisitando a busca ou atualização de dados ao Model, processar esses dados conforme a regra de negócio e adaptar o estado dos atributos que irão ser retornados a View para a atualização da interface conforme ilustra a Figura 7.

Figura 7 – Exemplo arquitetura MVVM.



Fonte: (BHADORIA, 2023)

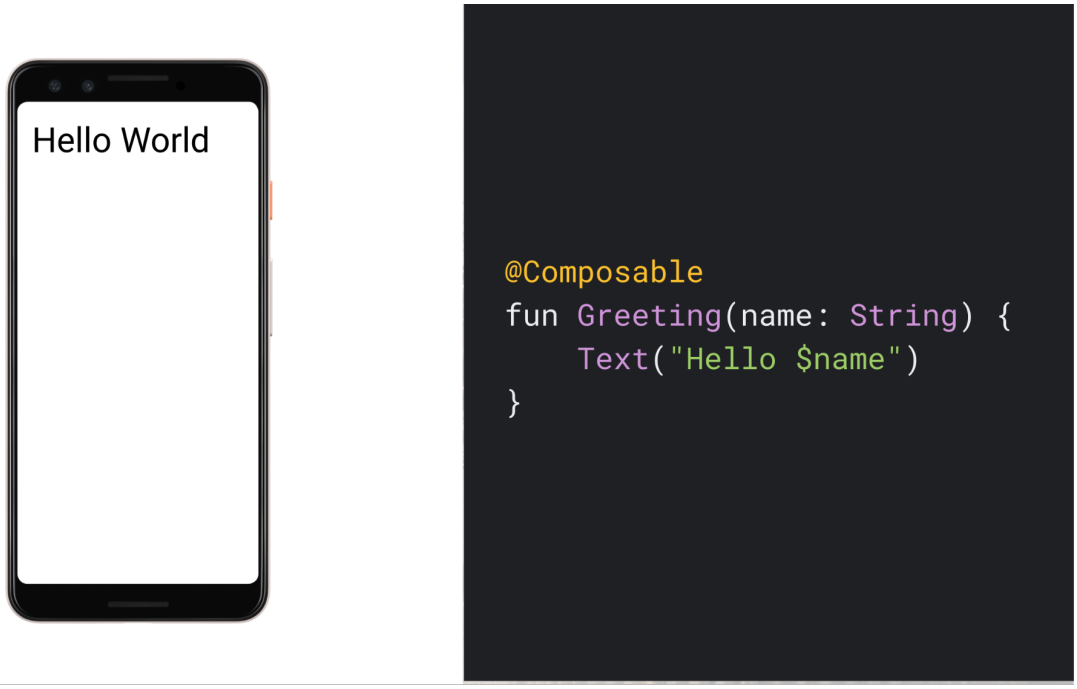
O desenvolvimento do aplicativo foi realizado utilizando esta arquitetura devido a facilidade de implementação e clara separação de responsabilidades, o que torna o código mais robusto e fácil de dar manutenção, permitindo que hajam contribuições para o projeto por parte de outros desenvolvedores por se tratar de um projeto de código aberto.

3.1.4 Construção de UI com Jetpack Compose

O Jetpack Compose é um framework moderno para desenvolvimento de interface de usuário no Android recomendado pelo Google. Ele utiliza o paradigma de programação declara-

tiva o que torna a criação de telas muito mais simples e menos propensa a erros. Na Figura 8 é possível ver a simplicidade de construção de uma tela utilizando o Jetpack Compose, onde uma tela simples com um texto é feita em 3 linhas de código.

Figura 8 – Exemplo básico de tela usando o Jetpack Compose.



Fonte: (GOOGLE, 2024)

O Compose introduziu uma mudança significativa no desenvolvimento de UI no Android em relação ao método historicamente utilizado baseado em hierarquia de visualização. Neste método, o layout da tela era montado através de arquivos XML que definem a posição e atributos de cada componente que deve ser mostrado, enquanto para alterar o contéudo do mesmo é feito acessando diretamente os métodos *setters* disponibilizados pela API.

Já o Compose, utilizando-se do paradigma declarativo é capaz de construir a UI de forma dinâmica usando as funcionalidades da linguagem Kotlin. Então é possível utilizar, por exemplo, laços de repetição para adicionar mais elementos à tela e condicionais para definir se um componente está vísivel ou não.

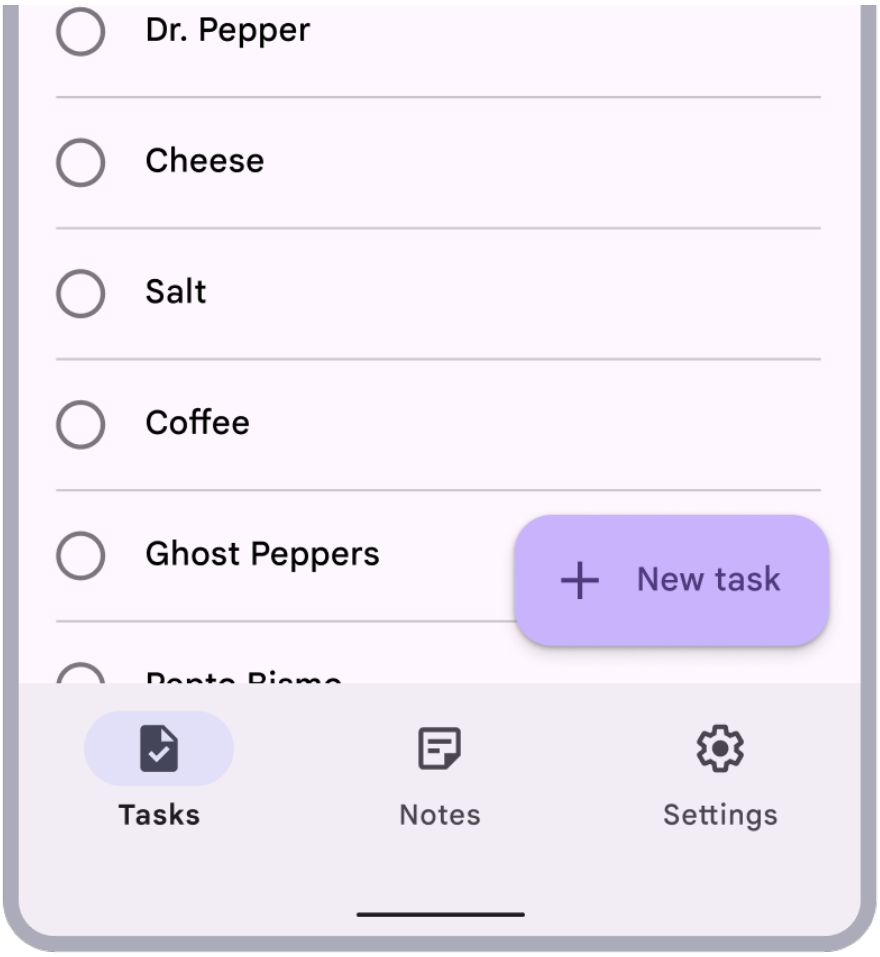
Outra característica do framework é que os componentes não são expostos como objetos para serem manipulados, para atualizá-los na UI é necessário modificar os estados dos mesmos, isto é, alterar os dados observáveis que os representam. O Compose é capaz de identificar quais dados foram modificados e redesenhar somente o que é necessário, evitando o custo computacional de construir toda a tela novamente.

3.1.5 Uso do Material Design 3

O Material Design é um *design system*, isto é, uma coleção de componentes e príncipios para construção de interface de usuário que visa manter a consistência no desenvolvimento de produtos digitais. Introduzido pelo Google em 2014 tem evoluído ao longo dos anos e foi adotado em grande partes dos produtos da companhia.

Apresenta um visual chamativo porém simples e na sua versão mais recente, Material 3, visa trazer experiênciais mais personalizadas para cada usuário através do Material You, que adapta as cores dos aplicativos conforme o plano de fundo definido. O trabalho foi desenvolvido utilizando o M3 com um tema próprio e opção de utilizar as cores dinâmicas e escolher entre os temas claro e escuro. A Figura 9 mostra um exemplo de tela construída usando o M3.

Figura 9 – Exemplo de tela construída com Material Design 3.



Fonte: (GOOGLE, 2024)

3.1.6 Reconhecimento de palavras marcadas em um texto

Para reconhecer uma palavra marcada em um texto através de uma imagem do mesmo é necessário que haja o processamento na imagem para delimitar a região de interesse como

somente a área que contenha a cor do marcador utilizado. O próximo passo é aplicar filtros que facilitem a visualização do texto para posterior identificação por algum mecanismo de OCR.

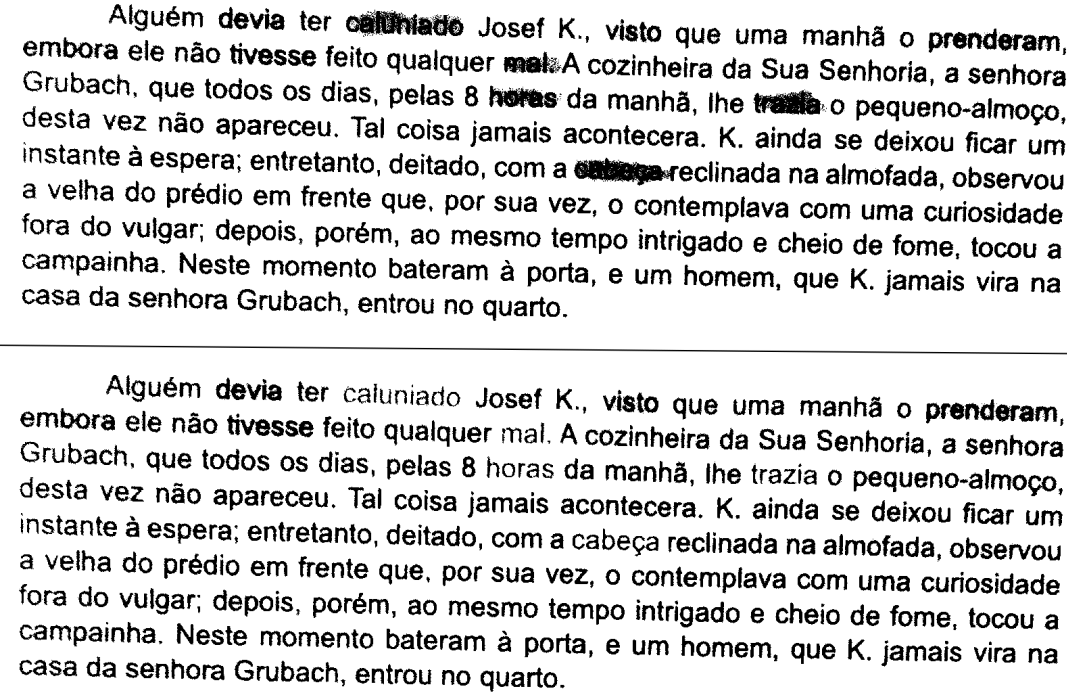
A etapa de processamento foi feita utilizando a biblioteca OpenCV que conta com inúmeras funcionalidades e suporte para o ecosistema Android além de ser amplamente conhecida e de código aberto. Ela utiliza o suporte do Android a código C++ para execução das funções de processamento com maior desempenho, o que torna o uso possível em dispositivos móveis.

O primeiro passo é obter a imagem original e transforma-lá para o espaço de cor HSV (Hue, Saturation, Value) utilizado pelo OpenCV, este espaço é utilizado pois cada cor tem sua tonalidade única o que torna mais fácil a segmentação por cor.

Após a transformação é aplicada uma máscara para delimitar as cores de interesse, então é necessário um intervalo no espaço HSV para ser segmentado, como podem haver variações dependendo do marcador, luminosidade da foto e cor da página foi criado uma tela para que o usuário seja capaz de ajustar esses limites e verificar se a configuração abrange a cor do marca-texto usado pelo mesmo.

Com a máscara aplicada, tanto essa imagem quanto a original são binarizadas, ou seja, transformadas em pixeis pretos ou brancos somente , então é feito uma junção da imagem com a máscara e da imagem sem a máscara a fim de melhorar a imagem de entrada para o OCR. Essa melhoria pode ser vista na Figura 10, onde a parte superior é a imagem original binarizada e o inferior o resultado final.

Figura 10 – Imagem original e final binarizadas.



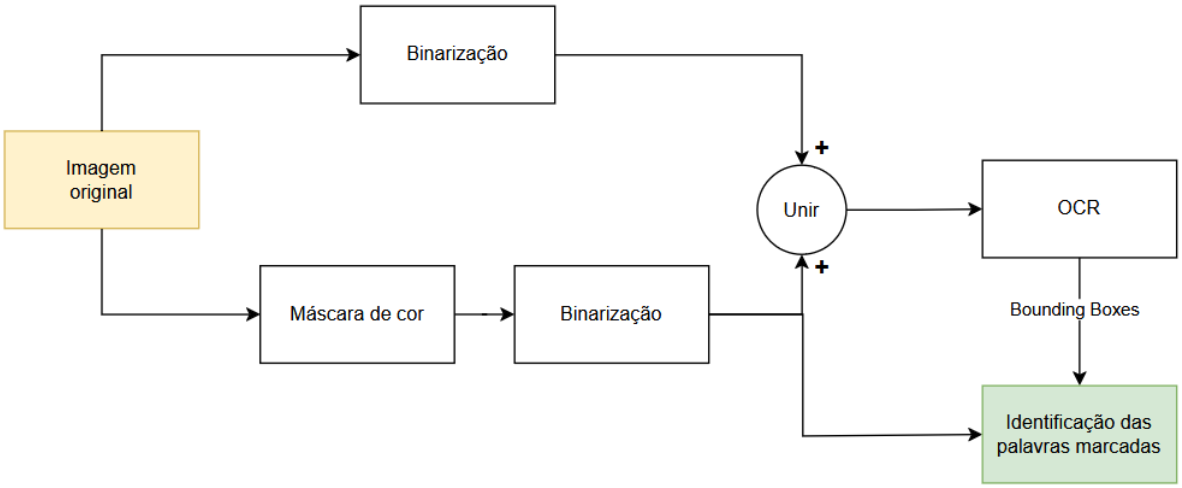
Fonte: Elaborado pelo autor, 2024.

Essas manchas não são observadas em todas as cores de marca-texto, uma cor que apresentou essa característica foi o laranja, pois possui componentes de vermelho e amarelo, que tendem a ter uma intensidade luminosa intermediária no espaço de cores RGB. Durante a conversão para tons de cinza (necessária para a binarização), ele pode gerar valores intermediários que não são suficientemente claros para serem tratados como branco ou suficientemente escuros para serem tratados como preto. Isso causa ambiguidades no algoritmo de limiarização (como o Otsu), gerando as manchas.

O último passo consiste em utilizar um mecanismo de OCR para identificar as palavras e os retângulos limitantes das mesmas, neste trabalho foi utilizado o ML Kit do Google por sua facilidade de integração no Android. Com essas informações é calculado o percentual de pixeis brancos dentro de cada retângulo na imagem com a máscara aplicada e binarizada, se este percentual for maior que um determinado valor a imagem é considerada marcada.

O fluxo completo do reconhecimento de palavras marcadas é ilustrado no diagrama da Figura 11. O processo de união consiste em realizar uma operação *"bitwise or"* entre as duas imagens, ou seja, os pixeis pretos que causam as manchas na imagem original binarizada serão substituídos pelos pixeis brancos da imagem com a máscara de cor e a binarização aplicada, tornando a imagem mais clara para a entrada do OCR. Este por sua vez retorna as palavras reconhecidas e as *bounding boxes* correspondentes e para cada uma delas é feito a contagem dos pixeis brancos dentro da área e se for maior do que 30% a palavra é considerada marcada.

Figura 11 – Diagrama do fluxo do reconhecimento de palavras marcadas.



Fonte: Elaborado pelo autor, 2024.

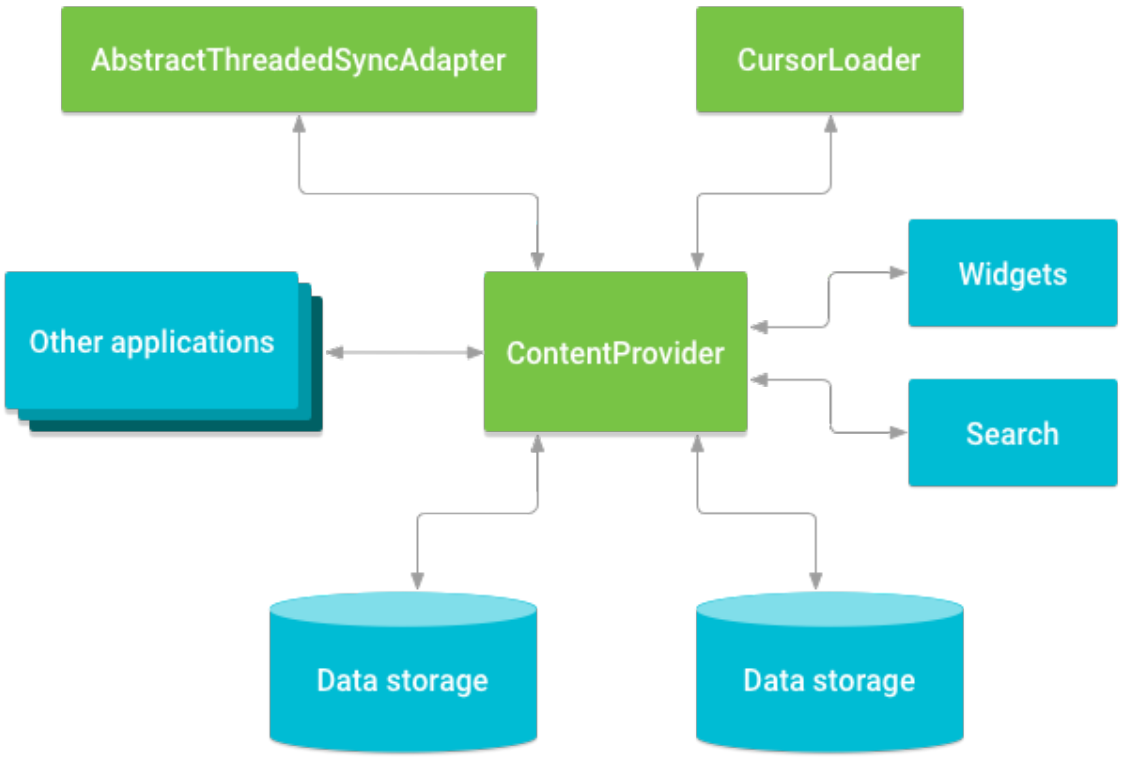
3.1.7 Acessando dados com a API do AnkiDroid

O Anki não possui versão oficial para Android, porém existe o projeto AnkiDroid desenvolvido pela comunidade open-source que possibilita a compatibilidade com o ecossistema do Anki, oferecendo no smartphone a maioria das funcionalidades presentes na versão oficial para computadores. Ele é capaz de sincronizar com o banco de dados proprietário do AnkiWeb, portanto as revisões e alterações feitas em um dispositivo são refletidas nos demais.

Tanto para a funcionalidade de treinamento de pronúncia quanto para a de criação automática de flashcards será necessário utilizar a API disponibilizada pelo AnkiDroid para obter e inserir novos dados. Essa API implementa um ContentProvider conforme demonstrado na Figura 12, que gerencia o acesso a um repositório central de uma aplicação: ele recebe as requisições de dados dos clientes, processa a ação solicitada e retorna os resultados.

Para ter acesso a API é necessário declarar no AndroidManifest que a aplicação desenvolvida requisita dados do pacote do AnkiDroid e também é necessário requisitar em tempo de execução a permissão do usuário para acessar os dados do mesmo.

Figura 12 – Relação entre um ContentProvider e outros componentes.



Fonte: (GOOGLE, 2024)

3.1.8 Uso do Azure Speech SDK

Os serviços de fala da Microsoft Azure, no qual a avaliação de pronúncia utilizada no aplicativo está inclusa, permite o uso através da API REST mas também fornece SDKs em diversas linguagens facilitando a utilização devido o gerenciamento automático das requisições e processamento dos dados.

Uma das linguagens suportadas é o Java que devido a interoperabilidade com o Kotlin é possível de ser utilizada diretamente no código do aplicativo. A Microsoft também disponibiliza o SDK como uma dependência que pode ser integrada via Gradle, sendo assim de fácil integração na aplicação Android.

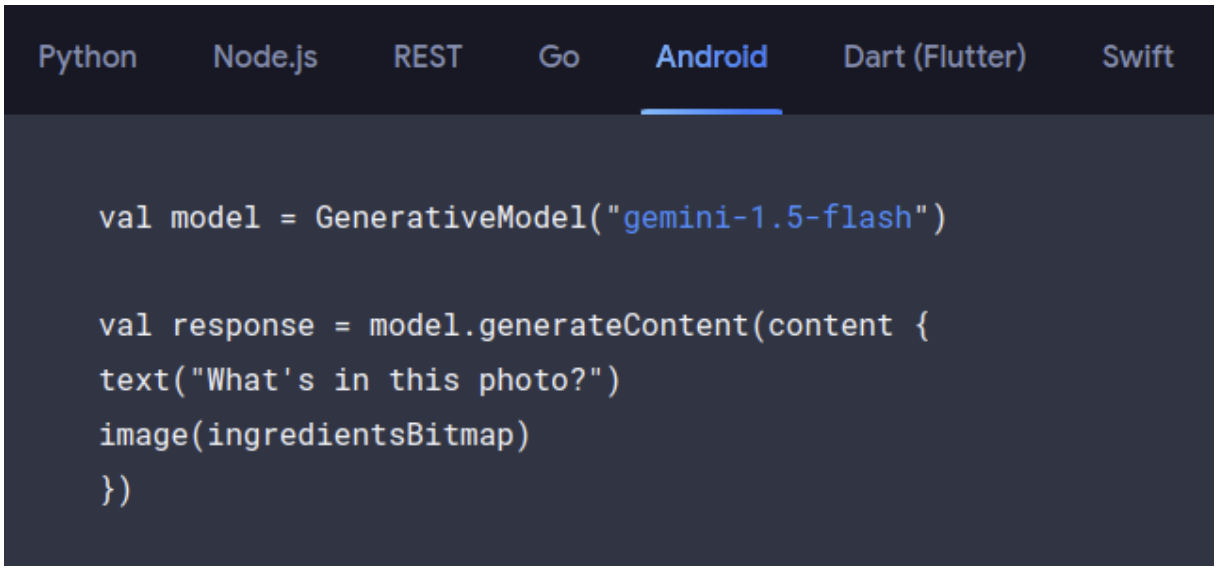
3.1.9 Utilizando LLMs da família Gemini

O Google disponibiliza uma API para realizar inferências nos modelos de inteligência artificial da família Gemini de fácil integração com o Android. Através dela é possível realizar inferências em diferentes modelos de forma multi-modal, configurando os parâmetros e instruções.

Para a criação automática de flashcards é necessário disponibilizar um prompt padrão que o usuário seja capaz de modificar para que o contéudo gerado seja relevante para o mesmo.

Também é essencial que o contéudo gerado pelo modelo siga um padrão de formato para que a aplicação seja capaz de processar os dados e criar os flashcards, para garantir isso é interessante o uso das instruções do sistema, fornecendo entradas e saídas para que o modelo responda da forma correta. A Figura 13 demonstra a facilidade de realizar inferências utilizando o SDK do Gemini no Android, onde é utilizado o modelo *gemini-1.5-flash* para perguntar o que há em uma imagem.

Figura 13 – Exemplo de uso da API do Gemini no Android.



Fonte: (GOOGLE, 2024)

3.2 Validação do funcionamento

Nesta seção será abordado a metodologia utilizada para validação das funcionalidades implementadas no aplicativo: avaliação de pronúncia, importação de vocabulário por imagem e criação automática de flashcards.

3.2.1 Validação da avaliação de pronúncia

Para realizar a validação da funcionalidade de avaliação de pronúncia será utilizado o dataset *speechocean762* (ZHANG *et al.*, 2021) que consiste em áudios de 250 falantes não-nativos pronunciando frases cotidianas em inglês que englobam 2600 palavras. Os aúdios foram classificados por cinco especialistas independentes e possuem pontuação a nível da sentença, palavra e fonema. As métricas de avaliação manual a nível de sentença estão descritas na Tabela 2.

Dos 5 mil áudios disponíveis foram escolhidos aleatoriamente 10 para cada nível de acurácia, fluência e prosódia, exceto para o intervalo de 0 a 2 de acurácia que havia somente 1 áudio disponível nessa faixa, totalizando 121 áudios. Para cada áudio extraído na amostra foi criado um flashcard com a frase de referência da pronúncia.

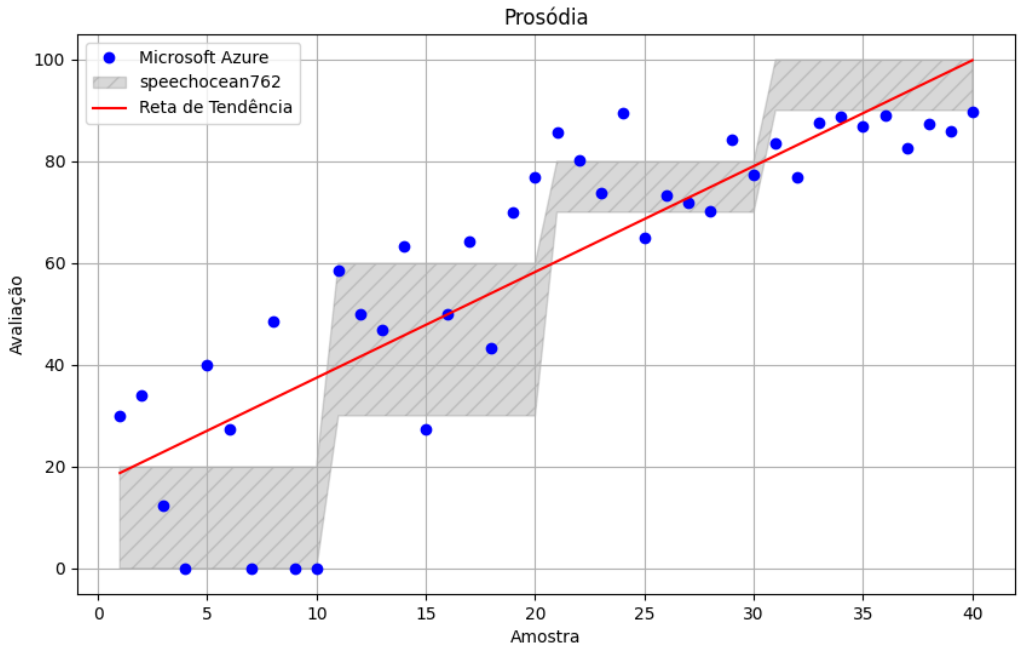
Tabela 2 – Métricas de avaliação manual do dataset para sentenças.

Pontuação	Descrição
	Acurácia
9-10	A pronúncia geral da frase é excelente, sem erros de pronúncia óbvios
7-8	A pronúncia geral da frase é boa, com alguns erros de pronúncia
5-6	A pronúncia da frase tem muitos erros, mas ainda é compreensível
3-4	Pronúncia estranha com muitos erros graves
0-2	A pronúncia da frase é impossível de entender ou não há voz
	Fluência
8-10	Fala coerente, sem pausas perceptíveis, repetições ou gagueiras
6-7	Fala coerente, com algumas pausas, repetições ou gagueiras
4-5	A fala é incoerente, com muitas pausas, repetições ou gagueiras
0-3	O falante não consegue ler a frase como um todo ou não há voz
	Prosódia
9-10	Entonação correta, velocidade e ritmo de fala estáveis
7-8	Entonação quase correta em uma velocidade de fala estável
3-6	Velocidade de fala instável ou a entonação é inadequada
0-2	A leitura da frase é muito gaguejante para avaliar ou não há voz

Fonte: Adaptado de (ZHANG *et al.*, 2021)

O aplicativo foi configurado para realizar avaliação de pronúncia na variante americana do inglês e utilizando os áudios da amostra foram realizadas avaliações nos respectivos flashcards e obtidos os resultados para comparação com a categorização do dataset. A Figura 14 mostra o gráfico de comparação de prosódia do speechocean762 com o resultado da Microsoft Azure a nível de sentença. A região cinza representa as métricas de avaliação do dataset e os pontos azuis o resultado da avaliação da mesma amostra pela API da Microsoft Azure, é possível perceber que os resultados ficaram coerentes conforme indicado pela reta de tendência em vermelho.

Figura 14 – Resultado da comparação da prosódia.



Fonte: Elaborado pelo autor, 2024.

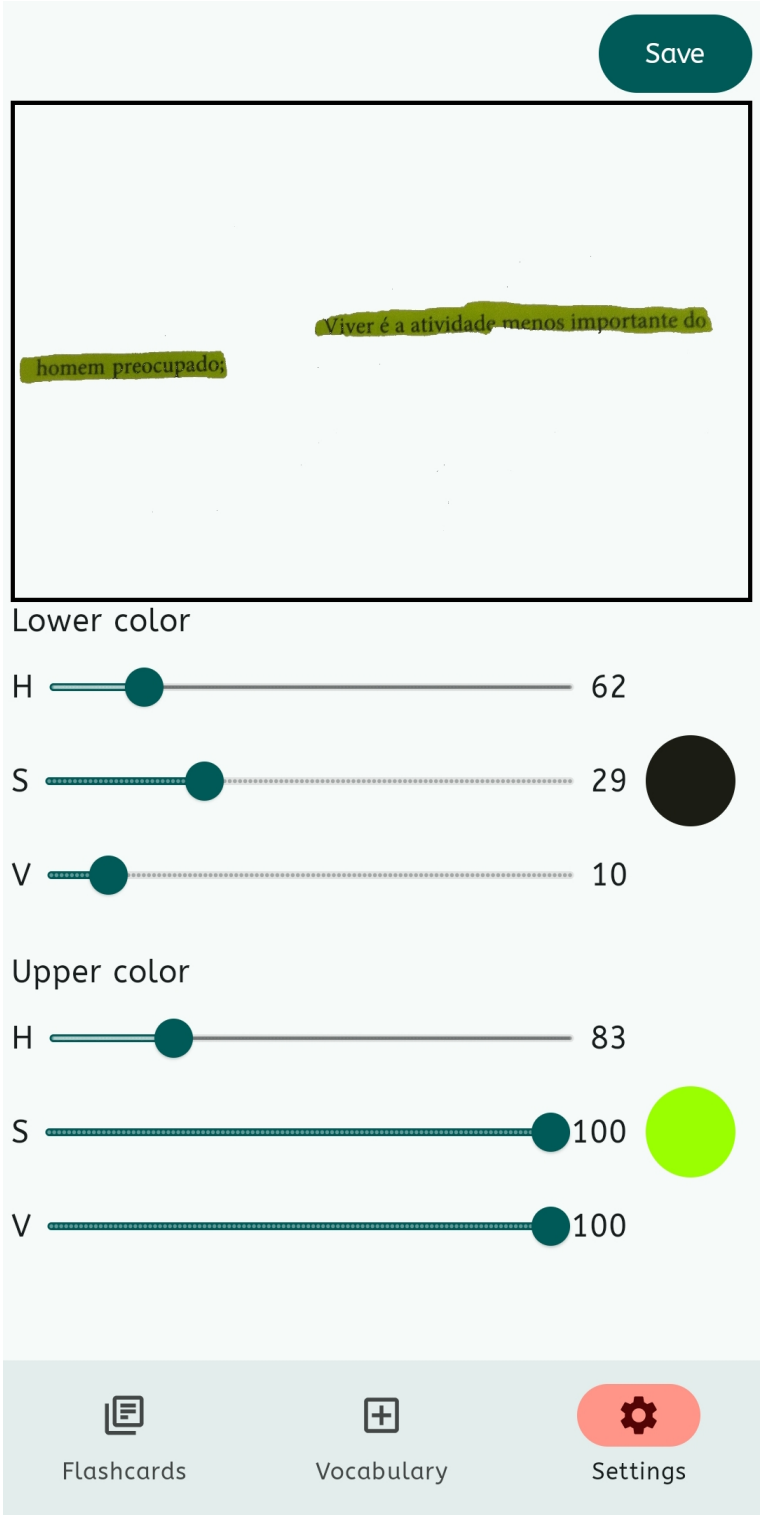
3.2.2 Importação de vocabulário por imagem

Para realizar a validação da funcionalidade de importação de vocabulário por imagem foi criado um dataset próprio contendo fotografias de textos impressos com fontes Arial e Times New Roman, por serem fontes muito comuns. As marcações das palavras foram realizadas utilizando o marca-texto em gel *Tris Twist* em 6 cores: amarelo, laranja, rosa, violeta, azul e verde. As imagens foram adquiridas usando um smartphone com câmera de 48 megapixels em um ambiente com luz natural indireta a uma distância de aproximadamente 20 cm da página.

Foram marcadas 5 palavras em português e 5 inglês para cada cor e para as fontes escolhidas nos tamanhos 8 e 12, totalizando 60 palavras únicas e 240 marcações. Em seguida, utilizando a interface do app desenvolvido, as imagens foram cortadas na região de interesse para então serem processadas e obtido o resultado do reconhecimento.

Para essa etapa ajustou-se cada cor de marca-texto na tela de configuração conforme a Figura 15. Essa tela permite que o usuário escolha uma imagem de referência e ajuste o limite inferior e superior da cor do marca-texto utilizado no espaço HSV e veja o resultado em tempo real da máscara de cor aplicada à imagem, quando somente as palavras marcadas estiverem vísiveis o usuário realiza o salvamento da configuração.

Figura 15 – Tela para ajuste do intervalo de cores para segmentação.



Fonte: Elaborado pelo autor, 2024.

3.2.3 Validação da criação automática de flashcards

Com as palavras importadas, selecionando 5 por vez, foi solicitado através da interface do aplicativo a criação de um flashcard para cada utilizando o modelo Gemini 1.5 Flash com o

seguinte prompt:

"Your task is to create for each word in the following list: {WORD_LIST} an example sentence and an explanation of the meaning, both in {LANGUAGE}. Highlight the word in the example sentence. Give the output in the CSV format separated by '//' without column names."

Onde "{WORD_LIST}" foi substituído pela lista de palavras e "{LANGUAGE}" pelo idioma das palavras selecionadas. Após a criação foi analisado a quantidade criada e se o conteúdo dos cards continham na parte da frente um exemplo usando a palavra e no verso uma explicação razoável do significado da mesma.

4 RESULTADOS

Para avaliar os resultados obtidos na avaliação de pronúncia foi utilizado o método de correlação de postos de Spearman devido os dados do dataset estarem categorizados em intervalos. Esse método é particularmente eficaz para avaliar a consistência dos rankings, uma vez que mede como a ordem dos elementos em um conjunto se correlaciona com outro, independentemente das escalas de medição, além de não exigir a normalidade dos dados nem pressupor uma relação linear entre as variáveis. Os resultados para cada métrica estão descritos na Tabela 3.

Tabela 3 – Resultado da correlação de Spearman para avaliação de pronúncia.

Métrica	ρ	Valor-p
Acurácia	$0,7937$	$6,01 \times 10^{-10}$
Fluência	$0,8182$	$1,14 \times 10^{-10}$
Prosódia	$0,9109$	$3,45 \times 10^{-16}$

Fonte: Elaborado pelo autor, 2024.

Os coeficientes ρ positivos obtidos, por se aproximarem de 1 indicam uma forte correlação entre o resultado da avaliação de pronúncia e o resultado dos avaliadores do dataset e o valor-p próximo a zero indica que esta correlação não foi por acaso. Portanto, o uso do serviço da Microsoft Azure para avialiação de pronuncia mostrou-se eficiente, produzindo resultados coerentes com a de avaliadores humanos.

Para a criação automática de flashcards, o modelo Gemini 1.5 Flash conseguiu criar cards com frente e versos coerentes para todas as palavras em inglês, porém para o português falhou em 9 casos onde não explicou o significado da palavra, ou criando frases incoerentes ou simplesmentes substituindo a mesma por outra semelhante. Também não destacou a palavra na frase de exemplo conforme a instrução do prompt em 7 casos. Os dados por idioma estão resumidos na Tabela 4 e a relação das palavras e do contéudo criado pelo modelo estão detalhadas no Anexo B.

Tabela 4 – Resultado da criação automática de flashcards.

Idioma	Frases coerentes	Palavras destacadas nas frases
Inglês	30	30
Português	21	23

Fonte: Elaborado pelo autor, 2024.

Por depender da inferência em um formato específico pode haver problemas quanto a geração de novos flashcards ao utilizar o mesmo prompt em diferentes modelos. No modelo Gemini 1.5 Pro, por exemplo, o resultado tende a ser mais detalhado evitando que os cards sejam criados corretamente, portanto é essencial que o usuário valide o prompt que deseja utilizar para o modelo escolhido. Esse problema pode ser contornado incluindo intruções de sistema, onde o

modelo recebe input e outputs esperados ajudando a resposta a ter determinado formato, porém usualmente essas instruções são contabilizadas e aumentam a quantidade de tokens utilizados.

A importação de vocabulário por palavra marcada também obteve bons resultados conforme mostra a Tabela 5. As fontes utilizadas obtiveram resultados iguais para o tamanho 12, porém a fonte Arial se saiu melhor para o tamanho 8. Deve-se considerar que a fonte Arial é maior que a Times New Roman para o mesmo tamanho, o que pode ter contribuído para um pior resultado em fontes pequenas.

Tabela 5 – Resultado do reconhecimento de palavras marcadas por cor e fonte.

Cor	Arial 12	Times New Roman 12	Arial 8	Times New Roman 8
Azul	10	10	7	5
Verde	9	8	8	7
Amarelo	10	9	9	7
Rosa	9	9	9	8
Violeta	9	9	7	4
Laranja	8	10	7	8
Total	55 (91,67%)	55 (91,67%)	47 (78,33%)	39 (65,00%)

Fonte: Elaborado pelo autor, 2024.

Enquanto para o tamanho 12 a quantidade de palavras reconhecidas para cada cor variou entre 8 e 10 para o tamanho 8 obteve-se entre 4 e 9, sendo a principal causa a qualidade baixa da imagem de entrada para o OCR. A fonte menor mostrou-se mais suscetível a variações de luminosidade, sendo necessário mais tentativas para capturar uma imagem com boa qualidade. Outra dificuldade encontrada foi na própria marcação manual das palavras, sendo elas muito pequenas a marcação acabava ultrapassando para outras palavras o que causou o reconhecimento de palavras que não estavam marcadas originalmente.

5 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho foi desenvolvido um aplicativo Android para treinamento de pronúncia integrado ao Anki. O usuário é capaz de utilizar o texto dos próprios flashcards que criou durante sua jornada de aprendizado para gravar sua voz e receber o feedback de sua pronúncia. A integração com a API da Microsoft Azure mostrou-se coerente quando comparado com dados de avaliadores humanos, possibilitando a prática da pronúncia de forma autônoma.

Além da funcionalidade principal, o aplicativo conta com o reconhecimento de palavras marcadas a partir de uma imagem, possibilitando o usuário importar novo vocabulário de forma prática e conta também com a criação automática de cards, sendo capaz de utilizar modelos de IA generativa para criar flashcards com frases de exemplo e explicação do significado das palavras aprendidas. O conjunto dessas ferramentas proporciona um excelente apoio para pessoas que já utilizavam o Anki para estudar idiomas.

Portanto, os objetivos propostos no trabalho foram atingidos, resultando em uma solução que combina ferramentas tecnológicas avançadas para abordar desafios comuns no aprendizado de idiomas. O aplicativo desenvolvido não apenas facilita a prática da pronúncia e a aquisição de vocabulário, mas também integra funcionalidades que promovem a autonomia e a eficiência no estudo. Dessa forma, o projeto contribui para enriquecer o ecossistema de aprendizado de idiomas e demonstra o potencial de aplicações móveis em contextos educacionais.

5.1 Trabalhos Futuros

Como trabalhos futuros pretende-se implementar e validar o reconhecimento de palavras marcadas em outros alfabetos além do latino, como o Japonês, Chinês, Coreano e Devanagari, que são suportados pelo ML Kit. Também pretende-se utilizar outros modelos de IA generativa além dos da familia Gemini, integrando também com as APIs da OpenAI e Anthropic. Além disso, pretende-se validar a substituição do OpenCV pelo uso de modelos multimodais, capazes de analisar imagens, para o reconhecimento de palavras marcadas.

REFERÊNCIAS

ANTUTU. **Performance Ranking of V3**. [S.l.: s.n.], 2024. Disponível em: <https://www.antutu.com/en/ranking/ai3.htm>. Acesso em: 29/03/2024.

BADDELEY, S. **The Curve of Forgetting**. [S.l.: s.n.], 2021. Disponível em: <https://simonbaddeley64.wordpress.com/2021/06/09/the-curve-of-forgetting/>. Acesso em: 02/04/2024.

BHADORIA, D. S. **Android MVVM — how to use MVVM in android example?** 2023. Disponível em: <https://medium.com/@dheerubhadoria/android-mvvm-how-to-use-mvvm-in-android-example-7dec84a1fb73>. Acesso em: 07/05/2024.

BORGHI, P. H.; TEIXEIRA, J. P.; FREITAS, D. R. da S. Automatic speech recognition for portuguese: A comparative study. In: PEREIRA, A. I. *et al.* (Ed.). **Optimization, Learning Algorithms and Applications - Third International Conference, OL2A 2023, Ponta Delgada, Portugal, September 27-29, 2023, Revised Selected Papers, Part I**. [S.l.]: Springer, 2023. (Communications in Computer and Information Science, v. 1981), p. 217–232.

BRITISH COUNCIL. **Learning English in Brazil**: Understanding the aims and expectations of the brazilian emerging middle classes. [S.l.], 2014.

COOK, C. **Extract Highlighted Text from a Book using Python**. 2020. Disponível em: <https://dev.to/zirkelc/extract-highlighted-text-from-a-book-using-python-e15>. Acesso em: 30/03/2024.

EBBINGHAUS, H. **Memory: A Contribution to Experimental Psychology**. [S.l.: s.n.], 1885.

GODWIN-JONES, R. Mobile apps for language learning. University of Hawaii National Foreign Language Resource Center, 2011.

GOOGLE. **Components – Material Design 3**. 2024. Disponível em: <https://m3.material.io/components>. Acesso em: 08/05/2024.

GOOGLE. **Content provider basics**. 2024. Disponível em: <https://developer.android.com/guide/topics/providers/content-provider-basics>. Acesso em: 17/08/2024.

GOOGLE. Gemini: A family of highly capable multimodal models. 2024. Disponível em: <https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf>. Acesso em: 07/05/2024.

GOOGLE. **Google AI Gemini API**. 2024. Disponível em: <https://ai.google.dev/#develop-with-gemini>. Acesso em: 17/08/2024.

GOOGLE. **Thinking in Compose | Jetpack Compose**. 2024. Disponível em: <https://developer.android.com/develop/ui/compose/mental-model>. Acesso em: 07/05/2024.

HENDRYCKS, D. *et al.* **Measuring Massive Multitask Language Understanding**. 2021.

HIRAI, A.; KOVALYOVA, A. Using speech-to-text applications for assessing english language learners' pronunciation: A comparison with human raters. In: ____. **Optimizing Online English Language Learning and Teaching**. Cham: Springer International Publishing, 2023. p. 337–355. ISBN 978-3-031-27825-9.

HU, W.; QIAN, Y.; SOONG, F. K.; WANG, Y. Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers. **Speech Communication**, v. 67, p. 154–166, 2015. ISSN 0167-6393.

JUANG, B.; RABINER, L. R. Automatic speech recognition – a brief history of the technology development. 2004.

KHOSHSIMA, H.; KHOSRAVI, M. Vocabulary retention of efl learners through the application of anki, whatsapp and traditional method. **Journal of Foreign Language Teaching and Translation Studies**, v. 6, n. 4, p. 77–98, 2021–2022. ISSN 2645-3592.

KJELLIN, O. Accent addition: Prosody and perception facilitates second language learning. **Proceedings of LP**, v. 98, n. 2, p. 373–98, 1999.

KOTLIN FOUNDATION. **Kotlin Programming Language**. 2024. Disponível em: <https://kotlinlang.org/>. Acesso em: 07/05/2024.

KRASHEN, S. Second language acquisition. **Second Language Learning**, v. 3, n. 7, p. 19–39, 1981.

MICROSOFT. Speech service ignite update – new enhancement features for pronunciation assessment. 2023. Disponível em: <https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/speech-service-ignite-update-new-enhancement-features-for/ba-p/3978093>. Acesso em: 07/05/2024.

MUNRO, M. J.; DERWING, T. M. A prospectus for pronunciation research in the 21st century: A point of view. **Journal of Second Language Pronunciation**, John Benjamins, v. 1, n. 1, p. 11–42, 2015.

NATION, I. **Learning vocabulary in another language**. [S.l.]: Cambridge university press Cambridge, 2001. v. 10.

SETTLES, B.; MEEDER, B. A trainable spaced repetition model for language learning. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics**. [S.l.: s.n.], 2016. p. 1848–1858.

ZHANG, J. *et al.* speechocean762: An open-source non-native english speech corpus for pronunciation assessment. In: **Proc. Interspeech 2021**. [S.l.: s.n.], 2021.

ANEXO A – CÓDIGO FONTE.

O código fonte do aplicativo desenvolvido e o material usado nas validações das funcionalidades estão disponíveis no repositório público do GitHub, acessível em: <https://github.com/warleysr/pronki>.

ANEXO B – CONTEÚDO DOS FLASHCARDS CRIADOS.

Palavra	Frente (Frase de exemplo)	Verso (Explicação do significado)	Coerente
himself	He hurt himself while playing basketball.	Himself refers to the person who is the subject of the sentence, in this case, he. It indicates that the action of hurting was done by the same person who is being talked about.	Sim
back	I put my book back on the shelf.	Back in this sentence means to return something to its original position.	Sim
head	He nodded his head in agreement.	Head here refers to the part of the body that is used to signal agreement or understanding.	Sim
see	Can you see the bird in the tree?	See in this sentence means to perceive something with your eyes.	Sim
slightly	She was slightly taller than her brother.	Slightly means to a small degree or extent.	Sim
little	The little girl giggled at the funny clown.	Little means small in size or amount.	Sim
brown	The brown dog barked at the mailman.	Brown is a color that is a mix of red and yellow.	Sim
domed	The old cathedral had a beautiful domed ceiling.	Domed describes something that has a rounded roof, like a dome.	Sim
arches	The old bridge was made of stone arches.	Arches are curved structures that support weight.	Sim
bedding	I bought new bedding for my guest room.	Bedding refers to sheets, blankets, and other items used on a bed.	Sim
morning	The sun was shining brightly as I woke up this morning.	The word morning refers to the early part of the day, from sunrise to noon.	Sim
woke	I woke up feeling refreshed after a good night's sleep.	The verb woke is the past tense of wake, which means to stop sleeping.	Sim
dreams	Last night, I had the most vivid dreams about flying.	The word dreams refers to the images, thoughts, and feelings that occur in our minds while we are sleeping.	Sim
transformed	The caterpillar transformed into a beautiful butterfly.	The verb transformed means to change completely in form or appearance.	Sim

Continua na próxima página...

Palavra	**Frente (Frase de exemplo)**	**Verso (Explicação do significado)**	**Coerente**
horrible	The food was horrible and I couldn't eat it.	The adjective horrible means very unpleasant or frightening.	Sim
when	When will you be home?	This sentence asks about the time of someone's return home. When refers to a specific point in time.	Sim
troubled	Her troubled past made it difficult for her to trust anyone.	This sentence describes someone who has experienced difficulties or hardships in the past. Troubled implies a state of distress or anxiety.	Sim
found	I found my keys under the couch.	This sentence describes the act of discovering something that was lost. Found means to locate something after a search.	Sim
bed	I went to bed early last night.	This sentence describes the act of going to sleep in a bed. Bed refers to a piece of furniture designed for sleeping.	Sim
vermin	The house was infested with vermin.	This sentence describes the presence of unwanted pests in a dwelling. Vermin refers to small, often harmful animals, such as rats or cockroaches.	Sim
from	I got a letter from my grandmother yesterday.	The word from indicates the origin or source of something.	Sim
lay	I lay down on the couch and took a nap.	The word lay means to put something down or to lie down.	Sim
lifted	He lifted the heavy box with great effort.	The word lifted means to raise something up.	Sim
could	I could not believe my eyes when I saw the amazing view.	The word could expresses ability or possibility.	Sim
stiff	My muscles were stiff after the long hike.	The word stiff means rigid or inflexible.	Sim
pitifully	The old dog looked pitifully at his empty food bowl, his tail drooping low.	Pitifully describes something that evokes pity or sadness.	Sim

Continua na próxima página...

Palavra	**Frente (Frase de exemplo)**	**Verso (Explicação do significado)**	**Coerente**
size	I was surprised by the size of the house, it was much larger than I expected.	Size refers to the physical dimensions or magnitude of something.	Sim
helplessly	The kitten was helplessly trapped in the tree, unable to climb down.	Helplessly means without the ability to act or defend oneself.	Sim
cover	He used a blanket to cover himself from the cold.	Cover means to place something over or on top of something else to hide or protect it.	Sim
slid	The children laughed as they slid down the water slide.	Slide means to move smoothly over a surface.	Sim
trazia	Ela trazia um sorriso no rosto.	Trazia significa carregava ou tinha consigo.	Sim
horas	Ele passou várias horas trabalhando no projeto.	Horas se refere a unidades de tempo.	Sim
cabeça	Ele bateu a cabeça na porta.	Cabeça se refere à parte superior do corpo humano.	Sim
caluniado	Ele foi caluniado pelos seus inimigos.	Caluniado significa acusado falsamente ou difamado.	Sim
mal	Ele se sentiu mal depois de comer muito.	Mal pode significar doente, fraco ou incorretamente.	Sim
momento	Este é o momento perfeito para tirar uma foto.	Momento significa um ponto específico no tempo, um instante.	Sim
homem	O homem alto de chapéu estava na esquina.	Homem significa um ser humano do sexo masculino.	Sim
casa	Minha casa é um lugar acolhedor e confortável.	Casa significa a residência de uma pessoa, o lugar onde mora.	Sim
quarto	O quarto de hóspedes está pronto para receber os visitantes.	Quarto significa um cômodo dentro de uma casa, geralmente usado para dormir.	Sim
curiosidade	A curiosidade matou o gato, mas a satisfação o ressuscitou.	Curiosidade significa o desejo de saber ou aprender sobre algo, interesse intenso por algo desconhecido.	Sim
todos	Todos os alunos foram para a aula hoje.	Significa todos ou cada um.	Não
não	Eu não gosto de brócolis.	Significa não ou não é.	Não

Continua na próxima página...

Palavra	**Frente (Frase de exemplo)**	**Verso (Explicação do significado)**	**Coerente**
coisa	Eu preciso comprar algumas coisas no supermercado.	Significa algo ou uma coisa.	Não
deixou	Ele deixou o livro em cima da mesa.	Significa permitiu, colocou ou abandonou.	Sim
instante	Espere um instante, eu já volto.	Significa momento ou breve período de tempo.	Sim
devia	Ela devia ter estudado mais para a prova.	Ela deveria ter estudado mais para a prova.	Não
visto	O documento foi visto pelo gerente.	O documento foi examinado pelo gerente.	Não
prenderam	Os ladrões foram prenderam pela polícia.	Os ladrões foram presos pela polícia.	Não
embora	Embora estivesse cansado, ele terminou o trabalho.	Apesar de estar cansado, ele terminou o trabalho.	Não
tivesse	Se ele tivesse mais tempo, viajaria para o Brasil.	Se ele tivesse mais tempo, viajaria para o Brasil.	Não
cozinheira	A cozinheira preparou um delicioso jantar para a família.	Cozinheira significa uma mulher que cozinha profissionalmente.	Sim
senhora	A senhora comprou um lindo vestido na loja.	Senhora é um termo respeitoso para se referir a uma mulher.	Sim
alguém	Preciso de alguém para me ajudar com a mudança.	Alguém significa uma pessoa não especificada.	Sim
manhã	Na manhã, o sol brilhava intensamente.	Manhã significa o período do dia entre o amanhecer e o meio-dia.	Sim
qualquer	Você pode pegar qualquer livro da estante.	Qualquer significa que não importa qual seja a escolha.	Sim
jamais	Nunca mais a vi depois daquele dia.	Jamais significa nunca mais, expressando uma ideia de algo que não irá acontecer novamente.	Não
entretanto	Estudava muito, entretanto não conseguia boas notas.	Entretanto indica uma ideia de contraste ou oposição entre duas coisas, geralmente expressando uma situação inesperada ou um obstáculo.	Sim

Continua na próxima página...

Palavra	**Frente (Frase de exemplo)**	**Verso (Explicação do significado)**	**Coerente**
almofada	Encostei minha cabeça na almofada e adormeci.	Almofada é um objeto macio que se utiliza para apoiar a cabeça durante o sono ou para dar conforto em outras situações.	Sim
velha	A velha casa de madeira ainda estava de pé.	Velha indica que algo é antigo, seja em relação à idade de uma pessoa ou a um objeto.	Sim
contemplava	Ele contemplava a paisagem da montanha, admirando sua beleza	Contemplar significa olhar para algo com atenção e admiração, geralmente por um período de tempo prolongado.	Sim

Fonte: Elaborado pelo autor, 2024