UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

# Uso de IA com Multiagentes para apoio no ensino e desenvolvimento da fala em pessoas com Trissomia 21

**Fabiano Chaves Loures**

Monografia - MBA em Inteligência Artificial e Big Data

ICMC USP SÃO CARLOS

**Fabiano Chaves Loures**

# Uso de IA com Multiagentes para apoio no ensino e desenvolvimento da fala em pessoas com Trissomia 21

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientadora: Profa. Mariana Caravanti

**Versão original**

**São Carlos**

**2025**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

L892u

Loures, Fabiano Chaves
uso de inteligencia artificial com multiagentes para apoio no ensino e desenvolvimento da fala em pessoas com trissomia 21 / Fabiano Chaves Loures; orientadora Mariana Caravanti de Souza. -- São Carlos, 2025.
101 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2025.

1. Trissomia 21. 2. Síndrome de Down. 3. Mutiagentes inteligentes. 4. Desenvolvimento da fala. 5. Inclusão social. I. Caravanti de Souza, Mariana, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

LOURES, Fabiano Chaves
S856m Uso de IA com Multiagentes para apoio no ensino e desenvolvimento da fala em pessoas com Trissomia 21 / Fabiano Chaves Loures ; orientadora Mariana Caravanti. – São Carlos, 2025.
101 p.

Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2025.

1. Trissomia 21. 2. Inteligência Artificial. 3. Multiagentes. 4. Desenvolvimento da Fala. 5. Inclusão Educacional. I. Caravanti, Mariana, orient. II. Título.

**Fabiano Chaves Loures**

# Use of Artificial Intelligence with Multi-Agent Systems to Support Speech Teaching and Development in People with Trisomy 21

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Profa. Mariana Caravanti

**Original version**

**São Carlos**

**2025**

*Para Júlia A. Loures, por quem este trabalho foi sonhado e construído. Sua luz, seu amor e cada sorriso são a motivação para buscar sempre o melhor. Que a Inteligência Artificial e a colaboração de multiagentes inteligentes, tema desta pesquisa, se tornem aliados poderosos no auxilio para ensino e desenvolvimento da fala, abrindo novas portas e celebrando cada vitória. Que este estudo seja uma pequena contribuição para um futuro ainda mais promissor, abençoado por Deus, repleto de conquistas para você e para uma sociedade melhor e inclusiva."*

## AGRADECIMENTOS

A realização deste Trabalho de Conclusão de Curso somente foi possível graças ao apoio, incentivo e colaboração de diversas pessoas, que, direta ou indiretamente, contribuíram para cada etapa deste processo. Agradeço, sempre primeiramente, a Deus, pela força, saúde e perseverança ao longo desta jornada. Expresso minha sincera gratidão aos professores, colegas, familiares e amigos que caminharam ao meu lado, oferecendo apoio intelectual, emocional e moral. Em especial, dedico este trabalho às seguintes pessoas, cuja contribuição foi essencial:

- A Dra. Renata da Silvia Santos - Fonoaudióloga
- Ao Sr. Gilberto F. C. Avó
- A Sra M. Sueli F. C. Avó
- Mariana Caravanti, Solange Rezende, pela orientação, paciência e valiosas contribuições acadêmicas

A todos vocês, minha eterna gratidão.

*"A verdadeira inclusão acontece quando cada voz, não importa quão diferente, é ouvida e valorizada."*

*Autor: Desconhecido*

# RESUMO

LOURES, F. **Uso de IA com Multiagentes para apoio no ensino e desenvolvimento da fala em pessoas com Trissomia 21**. 2025. 101 p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

A Trissomia 21 (T21) está frequentemente associada a atrasos no desenvolvimento da fala e da linguagem, decorrentes de comprometimentos cognitivos, motores e auditivos. Tais limitações impactam diretamente a comunicação, a alfabetização e a inclusão social de indivíduos com essa condição. Diante desse cenário, este trabalho propõe uma solução tecnológica baseada em Inteligência Artificial (IA), estruturada por meio de uma arquitetura de agentes inteligentes multimodais, com o objetivo de apoiar o ensino da fala e o desenvolvimento da linguagem em crianças com T21. A proposta integra modelos de linguagem de larga escala (LLMs), reconhecimento automático de fala (como o *Whisper*) e Geração Aumentada por Recuperação (RAG – *Retrieval-Augmented Generation*), articulados por *frameworks* como o *LangChain*. Esses recursos possibilitam a atuação de agentes autônomos capazes de interagir com os usuários por múltiplos canais de comunicação: voz, texto, imagem, vídeo e música. O ambiente de aprendizagem é composto por avatares interativos, jogos educacionais adaptativos e experiências com realidade aumentada, promovendo uma abordagem lúdica, responsiva e centrada no usuário. Além de gerar feedbacks automáticos e dados analíticos para apoio de profissionais da saúde e da educação, a ferramenta visa criar uma rede colaborativa envolvendo famílias, educadores e terapeutas. Com distribuição gratuita e foco em acessibilidade, a iniciativa busca ampliar a autonomia comunicativa de crianças com T21, contribuindo para sua inclusão educacional e digital em contextos diversos.

**Palavras-chave**: Trissomia 21. Inteligência artificial. Multiagentes inteligentes. Desenvolvimento da fala. Tecnologias assistivas. Inclusão social. Processamento de linguagem natural. Educação inclusiva.

## ABSTRACT

LOURES, F. **Use of Artificial Intelligence with Multi-Agent Systems to Support Speech Teaching and Development in People with Trisomy 21**. 2025. 101 p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

Trisomy 21 (T21) is often associated with delays in speech and language development due to cognitive, motor, and auditory impairments. These limitations directly affect communication, literacy, and social inclusion for individuals with this condition. In light of this scenario, this work proposes a technological solution based on Artificial Intelligence (AI), structured through a multimodal intelligent agent architecture aimed at supporting speech teaching and language development in children with T21. The proposal integrates large language models (LLMs), automatic speech recognition (such as *Whisper*), and Retrieval-Augmented Generation (RAG), orchestrated through frameworks such as *LangChain*. These resources enable autonomous agents capable of interacting with users through multiple communication channels: voice, text, images, video, and music. The learning environment comprises interactive avatars, adaptive educational games, and augmented reality experiences, promoting a playful, responsive, and user-centered approach. In addition to generating automatic feedback and analytical data to support health and education professionals, the tool aims to create a collaborative network involving families, educators, and therapists. With free distribution and a focus on accessibility, the initiative seeks to enhance communicative autonomy for children with T21, contributing to their educational and digital inclusion across diverse contexts.



**Keywords**: Trisomy 21. Artificial Intelligence. Intelligent Multiagents. Speech Development. Assistive Technologies. Social Inclusion. Natural Language Processing. Inclusive Education.


**Keywords**: LaTeX. USPSC class. Thesis. Dissertation. Conclusion course paper. Report.

## LISTA DE FIGURAS

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| USPSC: | Universidade de São Paulo Unidade São Carlos |
| ABNT: | Associação Brasileira de Normas Técnicas |
| USP: | Universidade de São Paulo |
| IA: | Inteligência Artificial |
| T21: | Trissomia 21 |
| LLMs: | Modelos de Linguagem de Larga Escala (*Large Language Models*) |
| RAG: | Geração Aumentada por Recuperação (*Retrieval-Augmented Generation*) |
| WER: | Word Error Rate |
| CER: | Character Error Rate |
| SD: | Síndrome de Down |
| ST: | Desenvolvimento Típico |
| LGPD: | Lei Geral de Proteção de Dados |
| PcD: | Pessoa com deficiência |
| ABA: | Applied Behavior Analysis |

## SUMÁRIO

# 1 INTRODUÇÃO

O Dia Mundial da Síndrome de Down, celebrado em 21 de março, tem como principal objetivo valorizar a vida das pessoas com essa condição e promover a conscientização global sobre a importância de sua inclusão plena na sociedade. A data simboliza a Trissomia do cromossomo 21 (T21) e reforça a necessidade de combater estigmas e garantir que as pessoas com Síndrome de Down tenham acesso aos mesmos direitos e oportunidades que todas as demais.

A Síndrome de Down, também conhecida como T21, é uma alteração genética presente na espécie humana desde sua origem. Foi descrita pela primeira vez em 1866 pelo médico britânico John Langdon Down, que identificou o conjunto de características clínicas que hoje definem a síndrome como uma condição com identidade própria. É importante ressaltar que a Síndrome de Down não é uma doença, mas sim uma condição genética causada pela presença total ou parcial de um terceiro cromossomo 21.

Apesar de não ser uma enfermidade, a Trissomia 21 pode estar associada a alterações em diferentes sistemas do organismo, exigindo cuidados e acompanhamento multiprofissional. Entre as possíveis manifestações estão alterações oftalmológicas, auditivas, cardiopatias congênitas, distúrbios do sistema digestório, alterações endócrinas, neurológicas, do aparelho locomotor, hematológicas e ortodônticas.

As características físicas associadas à Síndrome de Down não estão relacionadas ao grau de comprometimento intelectual de uma pessoa. O desenvolvimento de cada indivíduo é influenciado por múltiplos fatores, especialmente pelos estímulos e incentivos recebidos nos primeiros anos de vida, bem como pela carga genética herdada dos pais, assim como ocorre com qualquer outra pessoa.

Segundo o Ministério da Saúde, embora não haja um número exato de pessoas com Trissomia 21 no Brasil, estima-se que ocorra um caso da síndrome a cada 700 nascimentos, o que corresponde a aproximadamente 270 mil pessoas com a condição no país[1]. Além disso, o Ministério da Mulher, da Família e dos Direitos Humanos, por meio de notícia publicada em 2019, informou que, segundo dados do último Censo realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2022b), há aproximadamente 300 mil pessoas com Síndrome de Down no Brasil

Esses dados reforçam a importância de políticas públicas e ações voltadas para a promoção da inclusão social, educacional e comunicacional dessas pessoas, garantindo-lhes acesso a direitos, serviços e oportunidades em igualdade de condições com os demais cidadãos[2].

Existem diferentes formas de comunicação, como a fala, a escrita e os gestos,

todas desempenham um papel fundamental no desenvolvimento social, educacional e emocional dos indivíduos. No entanto, pessoas com Trissomia 21 (Síndrome de Down) podem enfrentar barreiras significativas, especialmente no que diz respeito à aquisição e ao desenvolvimento da fala. Essas dificuldades impactam diretamente sua participação em contextos educacionais e sociais e se refletem em indicadores preocupantes. Conforme demonstrado na Figura 1, um Gráfico de Indicadores Educacionais no Brasil (IBGE, 2022b), as taxas de analfabetismo e os níveis mais baixos de escolarização entre pessoas com deficiência evidenciam a necessidade de apoio contínuo e a adoção de estratégias inclusivas desde a infância.

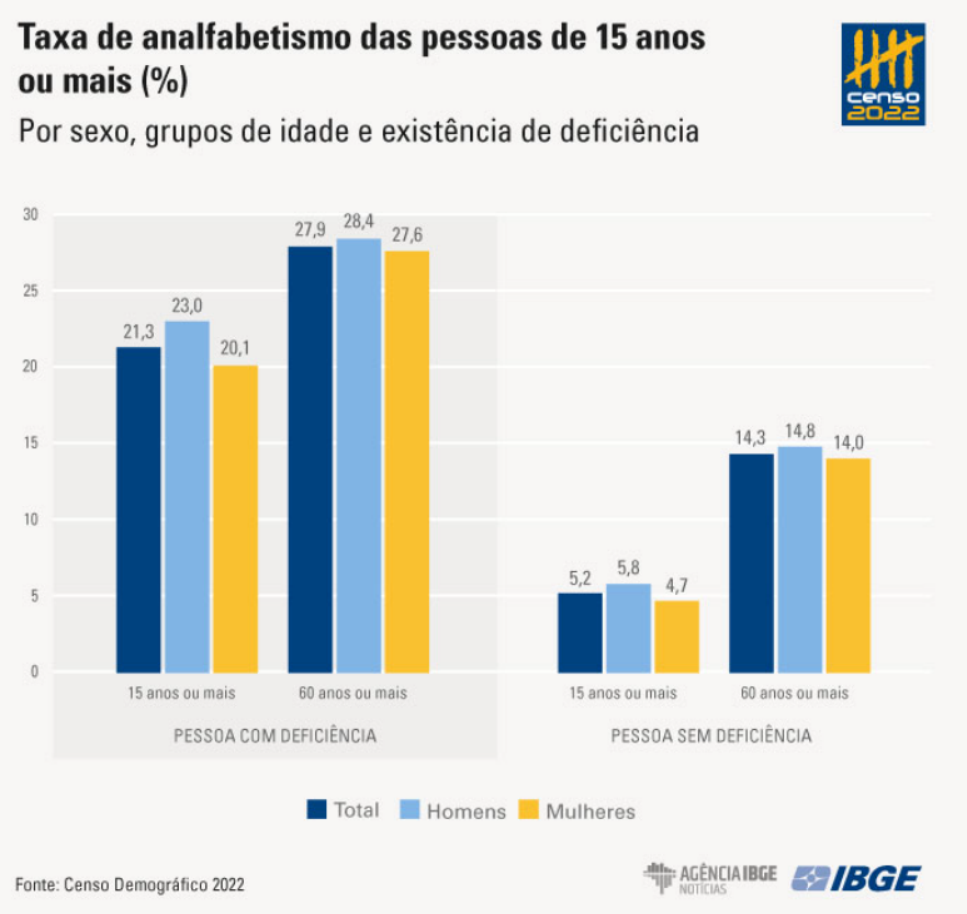


Figura 1 – Gráfico Indicadores Educacionais no Brasil (IBGE, 2022b)

Considerando o cenário apresentado, a proposta deste trabalho consiste no desenvolvimento de um modelo de negócio inovador, fundamentado em Inteligência Artificial e Reconhecimento de Fala, voltado ao apoio no ensino da fala de crianças com necessidades específicas, como aquelas com Trissomia 21.

O sistema, denominado Interface, busca integrar tecnologias de agentes inteligentes a uma plataforma de exercícios estruturados e adaptativos, que equilibram acessibilidade, usabilidade e motivação. Através de uma interação intuitiva, que combina estímulos visuais

e auditivos, a aplicação permite a reprodução de exercícios de fala, a análise automática das tentativas e o acompanhamento contínuo da evolução do usuário. Dessa forma, pretende-se oferecer uma solução escalável e personalizada, que contribua tanto para o desenvolvimento individual das crianças quanto para a criação de novas oportunidades de aplicação da IA no contexto educacional e terapêutico.

# 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, exploramos os conceitos e abordagens essenciais relacionados ao uso da Inteligência Artificial (IA) no suporte à alfabetização e comunicação de indivíduos com Síndrome de Down (T21). A relevância de compreender as características da população-alvo é inegável, pois as soluções tecnológicas propostas devem ser desenvolvidas com foco nas necessidades específicas desses indivíduos. Serão detalhadas as Modelos de Linguagem de Larga Escala (LLMs), arquiteturas multitarefa e sistemas baseados em múltiplos agentes inteligentes. Além disso, destacamos o papel de ferramentas como o Whisper, para reconhecimento de fala, e o uso de arquiteturas de Processamento de Linguagem Natural (PLN). Todas essas tecnologias visam possibilitar interações personalizadas, inclusivas e multimodais, que são cruciais para atender à diversidade da população com Síndrome de Down. O objetivo desta fundamentação teórica é, portanto, sustentar o desenvolvimento e a aplicabilidade de soluções tecnológicas voltadas à educação inclusiva e acessível, sempre considerando as particularidades da população que se busca beneficiar.

## 2.1 População, Gênero e Idade

A análise das características sociodemográficas da população com deficiência é fundamental para compreender a complexidade e a heterogeneidade desse grupo no Brasil. Ao examinar variáveis como sexo, faixa etária e cor ou raça, é possível identificar padrões de prevalência e potenciais iniquidades que informam a elaboração de políticas públicas mais eficazes e inclusivas. Os gráficos a seguir fornecem um panorama detalhado dessas dimensões, utilizando dados recentes que refletem a realidade brasileira. A Figura 2 apresenta a distribuição da população com deficiência por região.

A Figura 3 fornece um panorama detalhado da composição da população com deficiência no Brasil, discriminada pelos diferentes tipos de deficiência. Ao apresentar a distribuição percentual por categoria (visual, auditiva, física, intelectual, etc.), este gráfico é essencial para entender a predominância de certas condições e as necessidades específicas que emergem de cada tipo de deficiência. Essa informação é crucial para o planejamento de serviços de reabilitação, acessibilidade e apoio, assegurando que as políticas públicas sejam sensíveis às particularidades e desafios enfrentados por cada segmento da população com deficiência.

## 2.2 Cenário Nacional da Educação Especial (2022)

No contexto da inclusão social e das políticas públicas voltadas para pessoas com deficiência no Brasil, a análise de dados sociodemográficos, educacionais e laborais nos traz

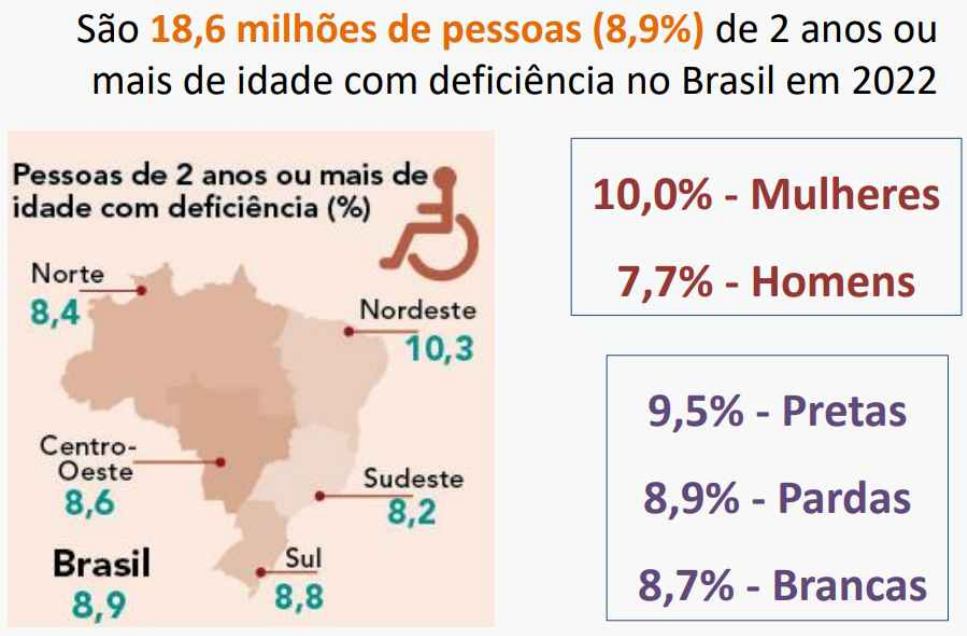


Figura 2 – Distribuição da população com deficiência por região do Brasil (IBGE, 2022a).

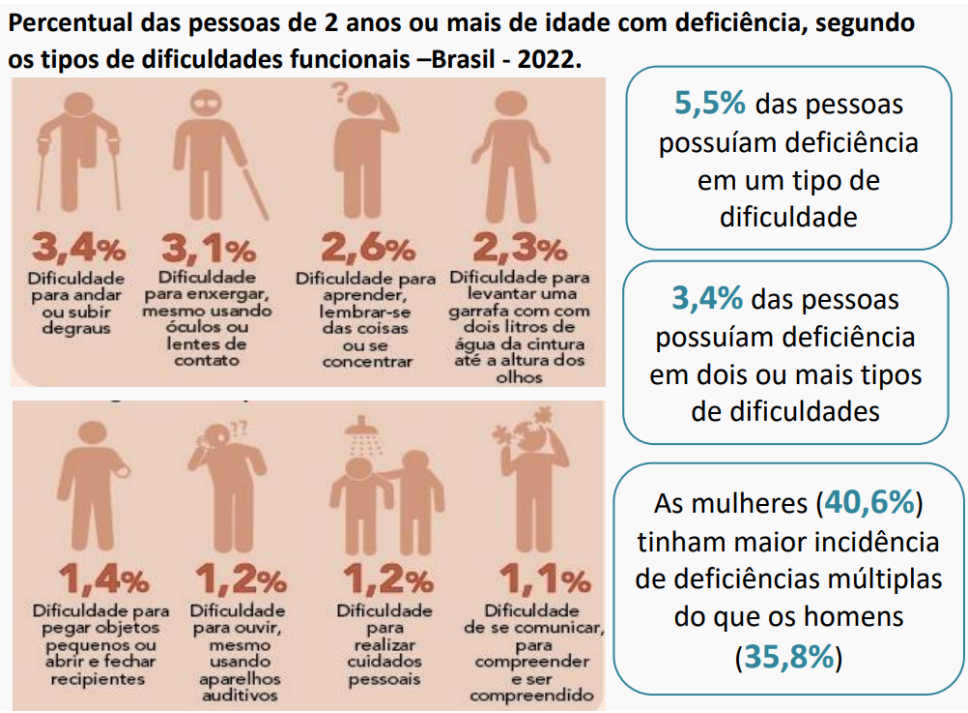


Figura 3 – Gráfico Distribuição da População por Tipo de Deficiência (IBGE, 2022a)

um entendimento melhor para compreender o grupo populacional. Este estudo utiliza dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) de 2022, conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para investigar as disparidades entre pessoas com e sem deficiência em três dimensões principais: características sociodemográficas, acesso à educação e inserção no mercado de trabalho. A comunicação exerce um papel essencial no processo de desenvolvimento humano, especialmente no

contexto educacional e social. Segundo o Ministério da Educação (Ministério da Educação (MEC), 2023), até o ano de 2023 foram registradas 1.771.430 matrículas na educação especial no Brasil, das quais estima-se que pessoas com Síndrome de Down representem aproximadamente 10% a 15% da população com deficiência intelectual. Considerando os 952.904 registros relacionados à deficiência intelectual, isso representa uma estimativa entre 95.290 e 142.936 estudantes com Síndrome de Down (T21) matriculados na rede de ensino em 2023.

Com base nesse cenário educacional, diversos estudos vêm investigando os fatores que influenciam o desenvolvimento educacional de pessoas com deficiência, em especial aquelas com T21. A pesquisa realizada por Eliana Caos, publicada em novembro de 2024, utilizou recursos de Inteligência Artificial (IA) para identificar elementos que impactam a capacitação de pessoas com autismo e Síndrome de Down. Os resultados destacam a relevância das terapias especializadas e do suporte familiar no processo de alfabetização de pessoas com T21, sobretudo em contextos mediados por tecnologias, como o Ensino a Distância (EAD).

Um dos aspectos centrais apontados no estudo refere-se ao fortalecimento da autonomia dos estudantes, o que está diretamente ligado à criação de ambientes acolhedores, à escuta ativa por parte dos educadores e ao respeito ao tempo de aprendizado individual. A comunicação adaptada, nesse sentido, é um fator determinante para a aprendizagem e inclusão. A forma como as instruções são transmitidas, utilizando diferentes linguagens, como verbal, gestual, visual e tecnológica, exerce grande influência na forma como o conteúdo é compreendido e absorvido. Recursos visuais, por exemplo, têm se mostrado altamente eficazes na organização de rotinas e na compreensão de tarefas, promovendo maior independência dos alunos. A escuta ativa e o respeito aos tempos de resposta são estratégias importantes para garantir que os estudantes se sintam acolhidos, compreendidos e motivados. Assim, a comunicação deixa de ser apenas um canal de transmissão de conteúdo e passa a representar uma ponte para o fortalecimento de vínculos e para a efetiva inclusão no processo de aprendizagem.

Para ilustrar essas diferenças, são apresentados gráficos que oferecem uma visualização clara e acessível dos indicadores analisados, permitindo identificar padrões, tendências e lacunas que demandam atenção em políticas públicas e intervenções sociais. Na Figura 4 é apresentado um gráfico comparativo da taxa de analfabetismo entre Pessoas com e sem deficiência (2022), revelando disparidades significativas. Na Figura 5 é apresentada a taxa de escolarização por faixa etária e condição de deficiência (2022), destacando o impacto da deficiência na permanência escolar. Na Figura 6 é apresentada a lacuna na frequência escolar líquida ajustada entre pessoas com e sem deficiência (2022), evidenciando defasagens idade-série.

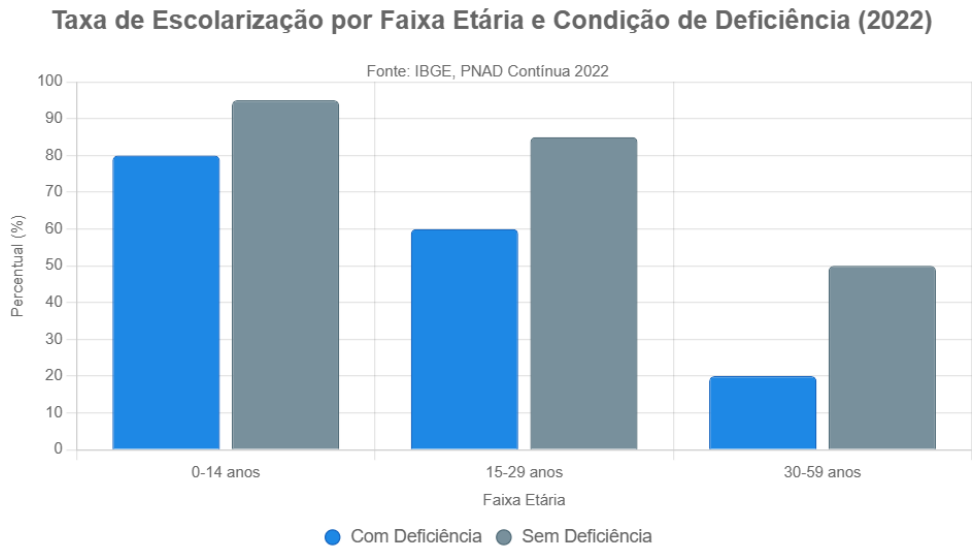


Figura 4 – Taxa de Analfabetismo Pessoas com e sem Deficiência (IBGE, 2022a)

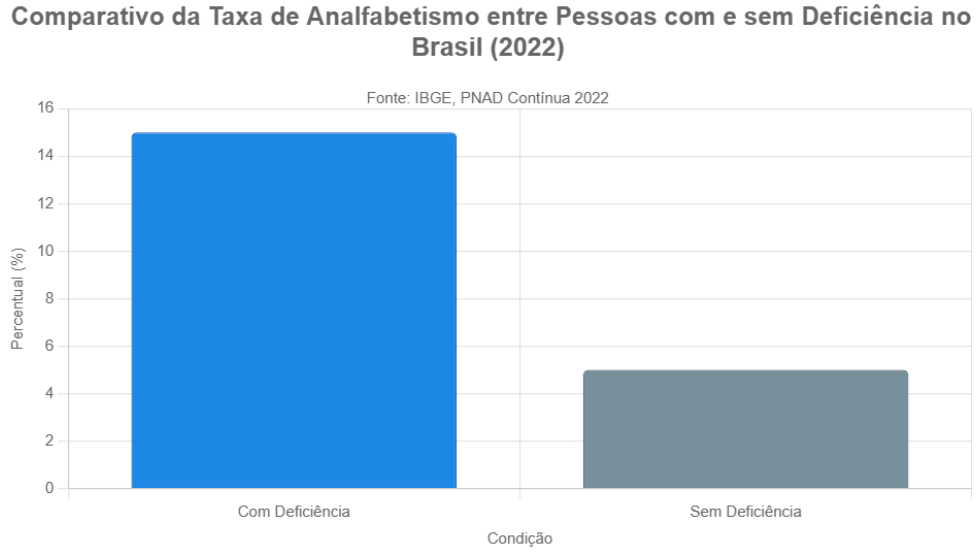


Figura 5 – Gráfico Taxa de Escolarização por Faixa Etária e Condição de Deficiência) (IBGE, 2022a)

## 2.3 Tecnologias de Inteligência Artificial Aplicadas à Inclusão

Nesse contexto, o uso de tecnologias baseadas em Inteligência Artificial tem se mostrado uma alternativa promissora para ampliar as possibilidades de aprendizado e de expressão comunicativa de pessoas com deficiência. Um exemplo relevante é a biblioteca Whisper, desenvolvida pela OpenAI em 2022, que foi treinada com mais de 680 mil horas de áudio multilíngue e multitarefa. A ferramenta destaca-se por sua capacidade de realizar transcrições automáticas de fala, com alta precisão, mesmo diante de variações linguísticas, sotaques e diferentes níveis de clareza na articulação vocal. Essa tecnologia tem grande potencial para ser aplicada em contextos educacionais, especialmente no apoio à comunicação de crianças com dificuldades na fala. Essa base teórica reforça a importância

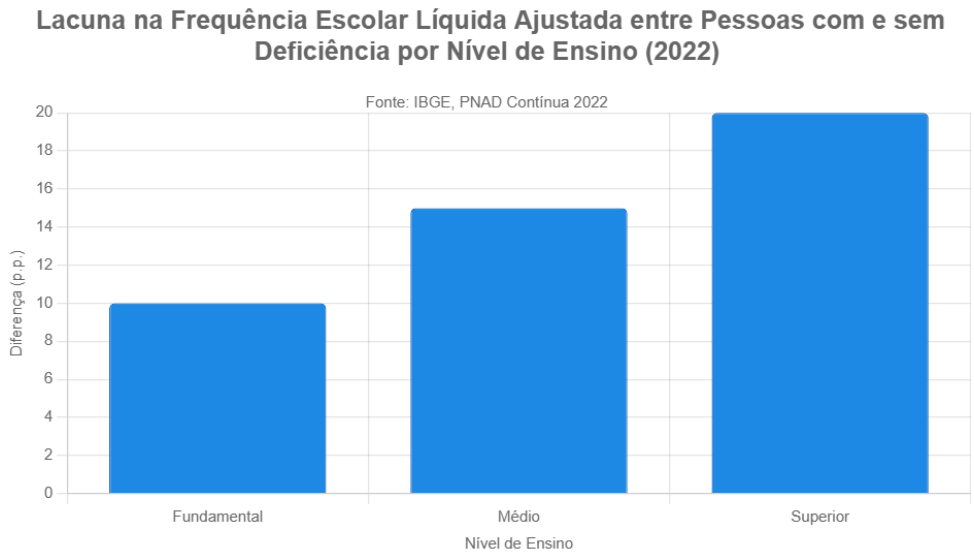


Figura 6 – Diferença na Frequência Escolar Líquida Ajustada por Nível de Ensino (IBGE, 2022a)

do desenvolvimento de soluções tecnológicas inclusivas, personalizadas e sustentadas por modelos de IA, que contribuam não apenas para o aprendizado acadêmico, mas também para o fortalecimento da autonomia, da expressão e da inclusão social de pessoas com Trissomia 21.

## 2.4 Bases neurocognitivas do aprendizado em pessoas com Trissomia 21

Cada criança com Trissomia 21 percebe e aprende o mundo de um jeito próprio. O cérebro dela processa informações de forma diferente, especialmente em áreas como atenção, percepção visual e linguagem. Compreender essas diferenças permite criar estratégias de ensino que realmente façam sentido para ela, e não apenas para o padrão esperado.

Neste trabalho, essa compreensão guia o desenvolvimento de agentes inteligentes que acompanham a criança durante seu aprendizado. Imagine uma criança tentando aprender uma nova palavra: os agentes podem mostrar imagens claras e coloridas, reforçar o som da palavra com voz suave e repetir de maneira adaptativa, respeitando o ritmo de aprendizado de cada criança. Assim, a tecnologia não é apenas uma ferramenta; ela se torna um parceiro, atento às necessidades, ao tempo e ao estilo de aprendizado da criança.

A Figura 7 representa, de forma figurativa, como o aprendizado acontece em pessoas com T21, relacionando conceitos do homúnculo de Penfield (Penfield, 1940) com a personalização de estímulos sensoriais por meio dos agentes inteligentes. Ciência, tecnologia e cuidado humano se unem para tornar o aprendizado mais acessível, motivador e efetivo.

Figura 7 – Homúnculo de Penfield para T21
(Penfield, 1940)

### 2.4.1 Erros gramaticais comuns em pessoas com Trissomia 21

Na Figura 8 é apresentado o gráfico de Distribuição dos Tipos de Erros Gramaticais mais comuns em crianças com T21, conforme (Chapman; Hesketh, 2000). O gráfico evidencia a frequência relativa dos diferentes tipos de erros observados, mostrando que erros de morfologia e sintaxe são os mais prevalentes, seguidos por problemas de concordância e omissão de verbos auxiliares.

Na Figura 9 é apresentada uma comparação no desempenho lexical de crianças com Síndrome de Down (SD) e crianças com desenvolvimento típico (DT) em diferentes categorias de vocabulário e comunicação. Crianças com SD apresentam pontuações inferiores àquelas com DT na maioria das categorias, refletindo maiores dificuldades na expressão verbal e construção de frases. Em algumas categorias, como processos de substituição e campos conceituais, as diferenças não são significativas, indicando habilidades semelhantes

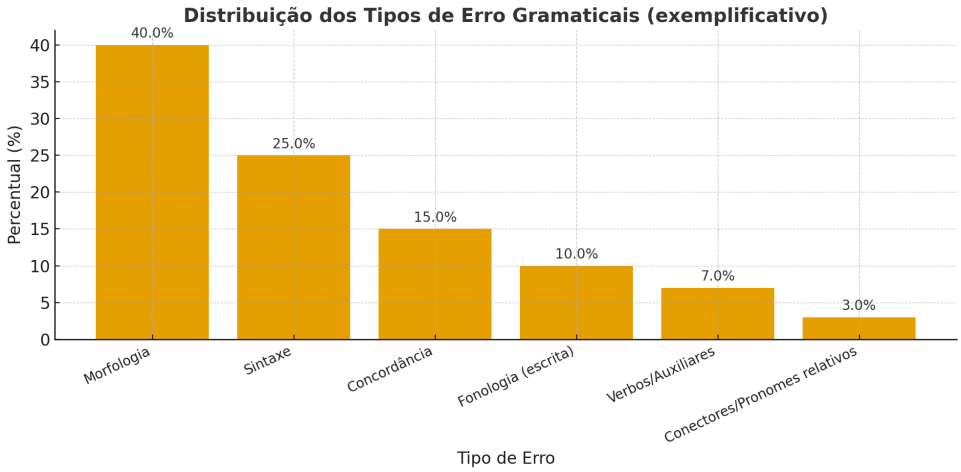


Figura 8 – Distribuição dos tipos de erros gramaticais em crianças com T21. Predominância de erros de morfologia e sintaxe. Fonte: (Chapman; Hesketh, 2000)

entre os grupos.

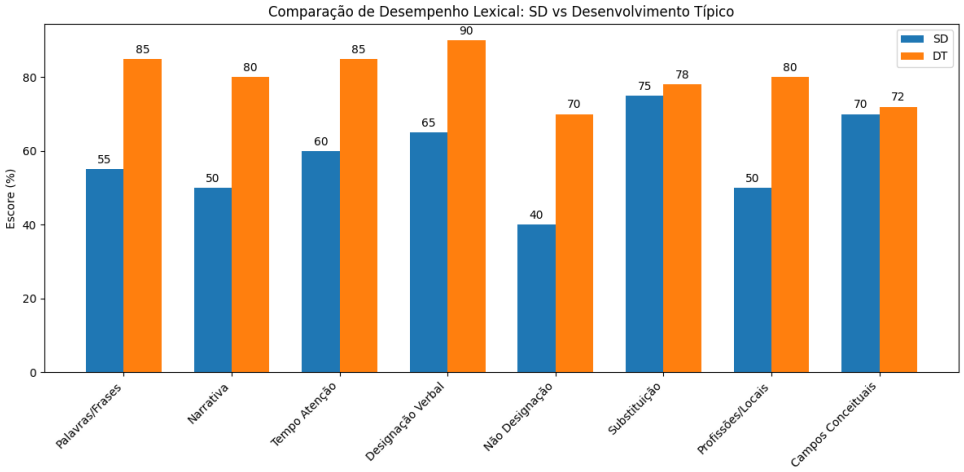


Figura 9 – Comparação do desempenho lexical entre crianças com SD e crianças com desenvolvimento típico. Fonte: elaborado a partir de Ferreira e Lamônica (2012) (Ferreira; Lamônica, 2012)

### 2.4.2 Proporção de Aquisição de Morfemas: DT vs T21

Na Figura 10 é apresentada a proporção de aquisição de morfemas, comparando crianças com T21 e crianças de desenvolvimento típico. O gráfico mostra a aquisição de elementos como plural, artigos e tempos verbais, evidenciando que crianças com T21 tendem a aprender esses morfemas mais tardiamente e de forma menos consistente.

O gráfico da Figura 11 mostra que a omissão de verbos em frases é mais frequente em crianças com T21 do que em crianças de desenvolvimento típico. Isso indica que crianças com T21 tendem a deixar de incluir verbos em algumas frases, comprometendo a construção completa das sentenças.

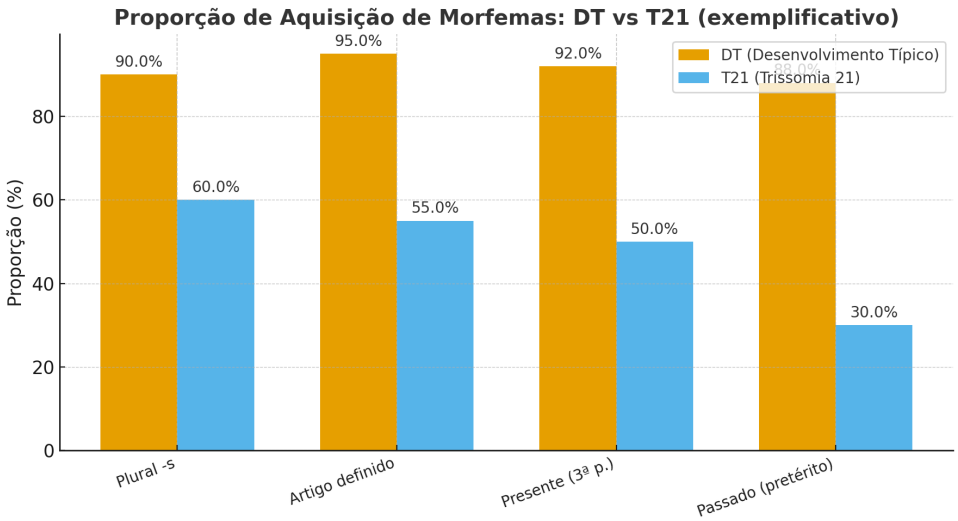


Figura 10 – Comparação da aquisição de morfemas entre crianças com T21 e desenvolvimento típico. Fonte: (Vicari; Caselli; Tonucci, 2000)

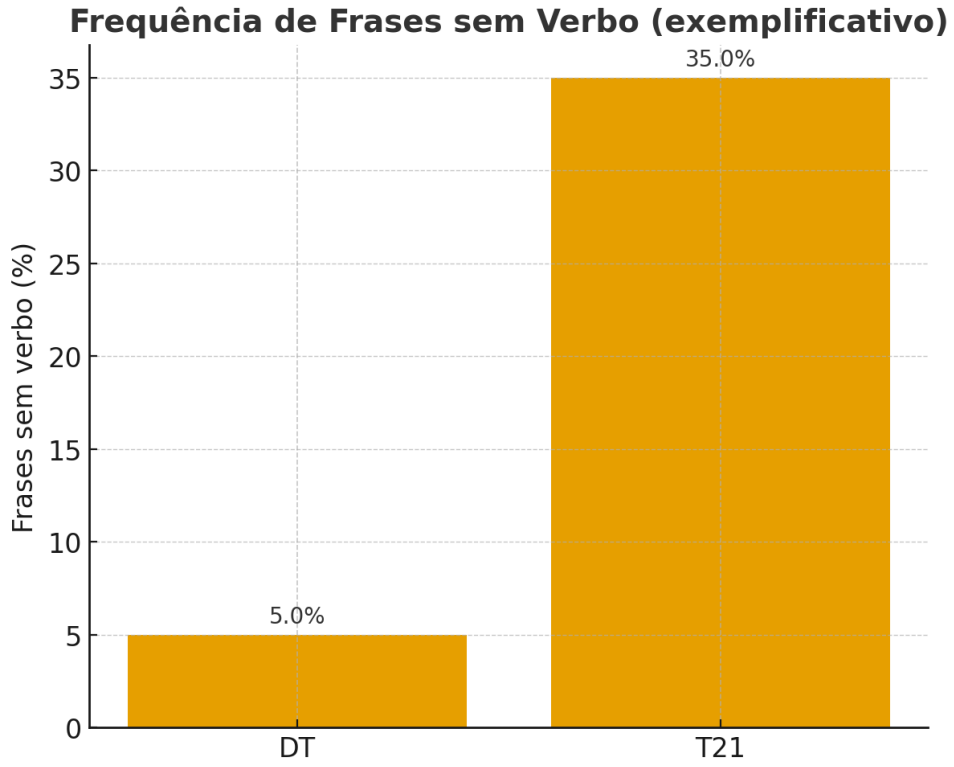


Figura 11 – Omissão de verbos em frases, comparando crianças com T21 e crianças de desenvolvimento típico. Fonte: Eadie *et al.*

O MLU (Mean Length of Utterance), que mede o tamanho médio das sentenças contando o número de palavras, indica que crianças com T21 costumam produzir frases mais curtas e com estruturas menos complexas, como orações subordinadas, refletindo menor complexidade gramatical (Figura 12).

Além disso, a taxa de erro gramatical apresenta relação com a capacidade de

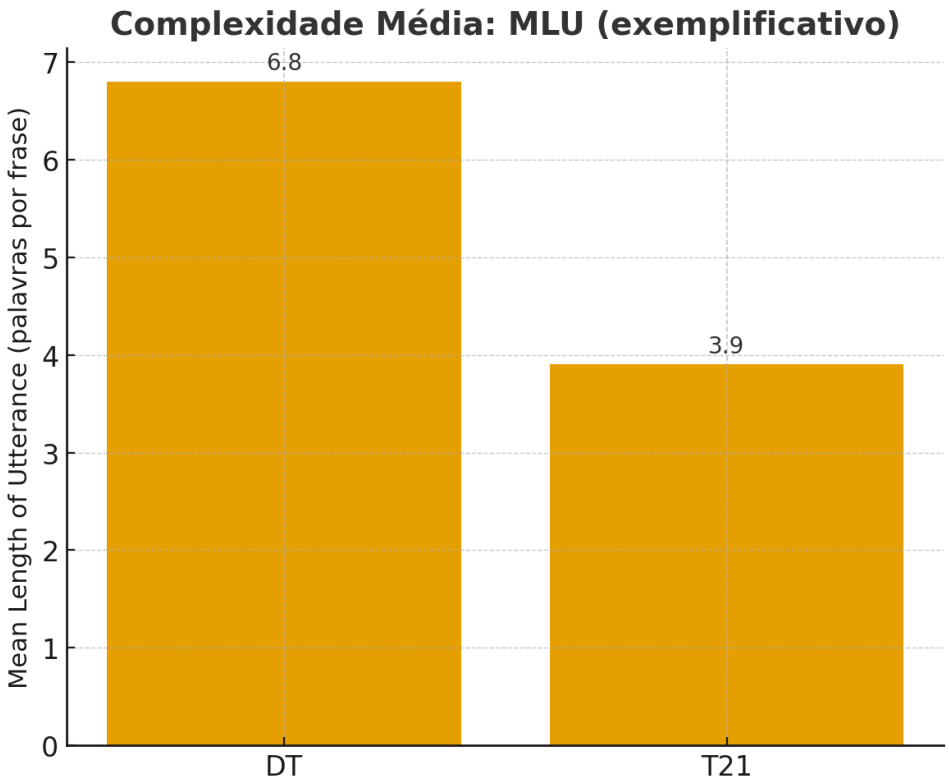


Figura 12 – Média de palavras por sentença (MLU) em crianças com T21 comparadas a crianças de desenvolvimento típico. Fonte: (Thordardottir; Chapman; Wagner, 2002)

memória de trabalho. O gráfico da Figura 13 mostra que crianças com memória de trabalho baixa, que conseguem reter poucas palavras ou informações simultaneamente, cometem mais erros gramaticais. Por outro lado, crianças com memória de trabalho mais alta apresentam menor taxa de erros, evidenciando a influência dessa habilidade cognitiva na produção linguística.

## 2.5 Modelos de Linguagem de Grande Escala (LLMs) e suas Aplicações Educacionais

A representação textual é etapa fundamental em sistemas de Processamento de Linguagem Natural (PLN), responsável por converter informações textuais ou faladas em estruturas que possam ser interpretadas por algoritmos. Conforme destacado por Jurafsky e Martin (2023), a representação semântica da linguagem natural está no cerne das aplicações modernas de inteligência artificial, permitindo o desenvolvimento de sistemas mais robustos e sensíveis ao contexto.

Os **Modelos de Linguagem de Grande Escala** (*Large Language Models* — LLMs) são redes neurais profundas treinadas com enormes volumes de dados textuais, oriundos de fontes como livros, artigos científicos, páginas da internet e repositórios digitais. Seu objetivo é aprender padrões e estruturas da linguagem humana, possibilitando compreender contextos, associar significados e gerar textos de forma coerente e contextualizada,

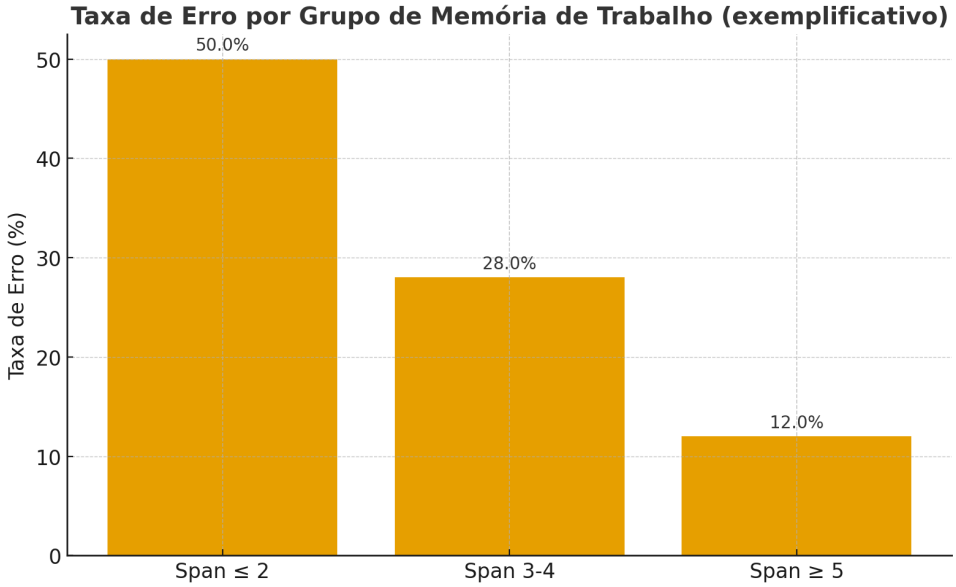


Figura 13 – Correlação entre capacidade de memória de trabalho (quantidade de informações lembradas simultaneamente) e taxa de erros gramaticais em crianças com T21. Fonte: (Laws; Gunn, 2002)

tendo como base a arquitetura Transformer (Vaswani *et al.*, 2017).

Após esse treinamento, os LLMs tornam-se capazes de executar uma ampla gama de tarefas linguísticas, como responder perguntas, redigir textos, traduzir conteúdos, resumir informações e realizar análises semânticas. Seu funcionamento baseia-se em mecanismos de predição, a partir de um contexto, o modelo estima a probabilidade da próxima palavra ou sequência, simulando o comportamento linguístico humano.

Um dos aspectos mais notáveis dos LLMs é a quantidade de parâmetros utilizados no treinamento, muitas vezes na ordem de bilhões. Esses parâmetros representam ajustes internos que refinam a capacidade de generalização e precisão do modelo. Quanto maior o número de parâmetros e a diversidade dos dados de entrada, maior tende a ser sua complexidade e desempenho, viabilizando aplicações avançadas, como a personalização de conteúdos educacionais e assistivos.

No contexto deste trabalho, os LLMs são considerados componentes centrais para a criação de ferramentas adaptativas de apoio à alfabetização e à comunicação de pessoas com Trissomia 21, atuando em sinergia com tecnologias de reconhecimento de fala e interfaces multimodais.

## 2.6 Como os Modelos Multitarefa Aprendem a Ouvir e Transcrever

O **Whisper** (Radford *et al.*, 2023), desenvolvido pela OpenAI, é um modelo de reconhecimento automático de fala (ASR) de código aberto, treinado em grande volume de dados multilíngues e multitarefa. Seu treinamento incluiu transcrições em inglês, traduções

de fala de diversos idiomas, transcrições multilíngues e até sons sem fala, como música de fundo e ruídos. Essa diversidade confere ao modelo ampla capacidade de generalização e a habilidade de executar múltiplas tarefas em um único sistema.

O processo de transcrição inicia com a conversão do áudio em um espectrograma, representação visual do som ao longo do tempo. Esse espectrograma é processado por uma arquitetura baseada em *Transformers* (Vaswani *et al.*, 2017), que opera no formato *sequence-to-sequence*, prevendo a próxima palavra ou símbolo com base no contexto já analisado.

Para diferentes finalidades, como transcrição, tradução ou marcação temporal, o Whisper utiliza um formato padronizado de entrada. Esse padrão inclui especificações sobre o idioma falado, a tarefa desejada e informações temporais, permitindo ao modelo executar funções como geração de legendas sincronizadas, tradução simultânea, identificação de idioma e detecção de fala versus ruído.

A Figura 14 ilustra a arquitetura do modelo e sua abordagem multitarefa. Todas essas funcionalidades são integradas por meio de tokens padronizados que atuam como sinalizadores de tarefa, substituindo diversos componentes isolados de um pipeline tradicional por uma única solução otimizada.

Essa arquitetura mostra-se eficiente em aplicações de acessibilidade e comunicação assistiva, oferecendo flexibilidade para lidar com diferentes idiomas e precisão temporal necessária para interações em tempo real. Em contextos educacionais e terapêuticos, como o apoio a pessoas com dificuldades na fala, incluindo aquelas com Trissomia 21, tais recursos podem transformar a forma como a informação é acessada e expressa.

Tabela 1 – Resumo adaptado do Modelo Whisper

| **Etapa** | **Descrição** | **Função** |
|---|---|---|
| Captação de Áudio | O sistema recebe o som da voz | Entrada de dados |
| Conversão de Sinal | A voz é transformada em espectrograma | Transformação de dados |
| Processamento Primário | O espectrograma é analisado pelo modelo | Interpretação inicial |
| Processamento Secundário | O sistema define a saída com base no contexto | Tomada de decisão |
| Transcrição | O áudio é convertido em texto | Saída de dados |
| Identificação de Idioma | O modelo detecta automaticamente a língua falada | Classificação |
| Detecção de Ruído | O modelo diferencia fala de ruídos | Filtragem |

Fonte: Desenvolvido pelo autor, a partir do diagrama da Arquitetura do Modelo Whisper e Treinamento Multitarefa.

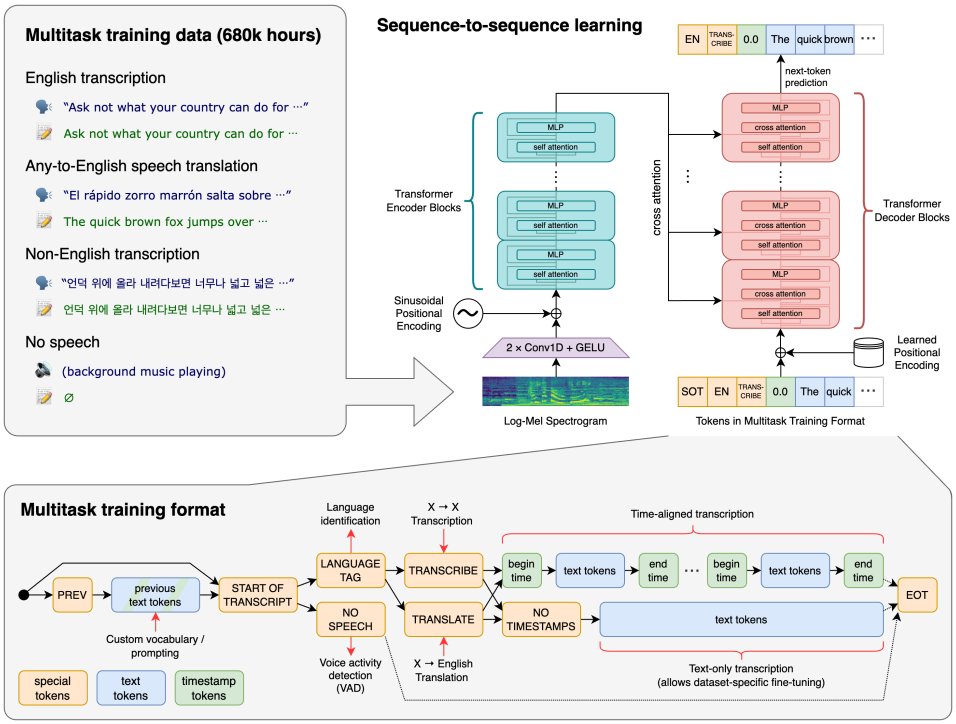


Figura 14 – Diagrama da Arquitetura do Modelo Whisper e Formato de Treinamento Multitarefa

(Radford *et al.*, 2023)

## 2.7 RAG - Geração Aumentada por Recuperação e suas aplicações inclusivas

A Geração Aumentada por Recuperação (Retrieval-Augmented Generation – RAG) constitui uma abordagem híbrida de Processamento de Linguagem Natural (PLN) que combina recuperação de informações contextualmente relevantes com geração de linguagem natural. Essa técnica foi proposta por Lewis *et al.*, (2020) no âmbito da Meta AI, com o propósito de superar limitações inerentes aos Modelos de Linguagem de Larga Escala (Large Language Models – LLMs) em tarefas que exigem conhecimento factual atualizado.

Historicamente, sistemas de retrieval baseados em busca lexical, como BM25 e TF-IDF, foram sucedidos por métodos semânticos, como o Dense Passage Retrieval (DPR) e o Fusion-in-Decoder (FiD), que passaram a utilizar representações vetoriais de alta dimensionalidade (embeddings). A RAG consolida essa evolução ao integrar tais técnicas com LLMs generativos, permitindo que a geração textual ocorra de forma fundamentada em dados externos e não apenas nos parâmetros estáticos do modelo. Assim, o sistema reduz fenômenos de alucinação e aumenta a confiabilidade das respostas (Izacard *et al.*, 2023).

Nos modelos tradicionais, o conhecimento é limitado ao conjunto de dados de treinamento. Em contrapartida, a RAG introduz um componente de busca semântica que acessa bases vetoriais externas, como FAISS, ChromaDB ou Pinecone, onde documentos, artigos e transcrições são indexados. Essa integração cria uma ponte entre memória estática

(modelo) e memória dinâmica (base vetorial), o que amplia a capacidade de generalização e atualização contínua do sistema.

A integração da RAG em contextos educacionais tem se mostrado promissora para a personalização da aprendizagem e para o desenvolvimento de sistemas tutorais inteligentes (Intelligent Tutoring Systems). Através da recuperação dinâmica de conteúdos, a RAG permite que os agentes de IA ofereçam respostas adaptadas ao perfil cognitivo e linguístico do aprendiz, especialmente em públicos com necessidades específicas. O pipeline da RAG é geralmente composto por três etapas:

- **Recuperação (Retrieval)**: o texto de entrada é convertido em embeddings, e o sistema realiza uma busca semântica em uma vector store, retornando os documentos mais relevantes ao contexto da consulta.
- **Leitura e Seleção (Reader)**: Os documentos recuperados são analisados e ranqueados conforme sua relevância, eliminando redundâncias.
- **Geração (Generation)**: O modelo de linguagem concatena o conteúdo recuperado ao prompt, gerando uma resposta coerente, contextualizada e verificável.

Essa sequência, entrada -> busca -> ranqueamento -> geração, caracteriza o fluxo lógico que diferencia a RAG das abordagens de fine-tuning. Enquanto o fine-tuning exige re-treinamento completo do modelo, a RAG injeta conhecimento sob demanda, preservando o custo computacional e a eficiência operacional.

Recentemente, variações como Self-RAG e RAG-Fusion ampliaram esse paradigma, permitindo que o próprio modelo avalie a qualidade das fontes recuperadas e refine suas respostas iterativamente (Shuster *et al.*, 2024).

Os principais benefícios do RAG envolvem redução de alucinações em modelos generativos; atualização contínua de conhecimento sem necessidade de re-treinamento; personalização contextual de acordo com o perfil do usuário; bem como explicabilidade e rastreabilidade, pois as fontes recuperadas podem ser exibidas a profissionais ou responsáveis.

Contudo, desafios persistem. A qualidade das respostas depende da curadoria e representatividade da base vetorial, e há implicações éticas relevantes quanto à privacidade e consentimento na coleta de dados sensíveis, especialmente em crianças, conforme previsto na Lei Geral de Proteção de Dados (LGPD).

Em contextos inclusivos, a RAG atua como um mecanismo de amplificação da empatia comunicacional, pois permite que o sistema compreenda o histórico de interações e recupere exemplos personalizados, fortalecendo o vínculo entre tecnologia, terapeuta e aprendiz.

Entre os desafios e limitações da abordagem, estão: a dependência da qualidade e representatividade da base vetorial; questões éticas e de privacidade, especialmente em dados sensíveis de crianças, conforme a LGPD; e a necessidade de curadoria para evitar vieses e informações desatualizadas.

## 2.8 Arquitetura Multiagentes

A arquitetura multiagente pode ser comparada ao trabalho colaborativo em projetos complexos: diferentes indivíduos, com funções específicas, atuam de forma coordenada para atingir um objetivo comum. De maneira análoga, em sistemas computacionais, agentes são programas de software especializados que podem operar de modo autônomo, tomando decisões a partir de dados e regras, ou sob supervisão, seguindo instruções diretas. A interação entre múltiplos agentes permite dividir problemas complexos e resolvê-los de forma cooperativa.

Essa abordagem mostra-se especialmente útil em contextos de aprendizagem adaptativa e acessibilidade, ao distribuir responsabilidades entre módulos inteligentes, como reconhecimento de voz, geração de linguagem natural, recomendação de atividades pedagógicas e análise de desempenho do usuário. Ao integrar essas funções, a arquitetura multiagente simula processos cognitivos, como aprendizado, raciocínio e tomada de decisão, criando ambientes digitais personalizados e inclusivos.

As Figuras 15 e 16 exemplificam esse conceito. A primeira apresenta uma analogia entre a colaboração humana e o trabalho coordenado de agentes inteligentes. Já a segunda mostra a transposição dessa lógica para a educação especial, destacando como agentes digitais podem apoiar a comunicação e o ensino da fala de crianças com Trissomia 21.

## 2.9 LangChain: Orquestração de Agentes com Modelos de Linguagem de Grande Escala

O LangChain é uma biblioteca open-source desenvolvida para a construção de sistemas complexos baseados em modelos de linguagem de grande escala (LLMs), permitindo a orquestração integrada de múltiplos componentes, tais como agentes autônomos, ferramentas externas, memórias contextuais e repositórios de conhecimento (Chase, 2023). Diferentemente do uso convencional dos LLMs como simples geradores de texto, o LangChain viabiliza fluxos interativos que combinam regras, cadeias de decisão e integração com dados contextuais para respostas mais precisas e contextualizadas. No presente trabalho, o LangChain será utilizado para implementar agentes pedagógicos com as seguintes funcionalidades principais:

- Realizar correção fonética personalizada para usuários com dificuldades de fala;
- Fornecer sugestões de exercícios adaptados ao progresso individual do aprendiz;


Figura 15 – Analogia da colaboração humana aplicada à arquitetura multiagente

- Comunicar familiares e profissionais por meio de relatórios interpretativos em linguagem acessível e clara.

Cada agente operará sobre as representações textuais geradas pelo modelo *Whisper*, consultando bases semânticas especializadas, vocabulários direcionados ao desenvolvimento infantil, além de parâmetros e protocolos definidos por especialistas das áreas de Fonoaudiologia e Educação Especial. Essa arquitetura multiagente, coordenada pelo *LangChain*, potencializa a personalização do atendimento, promovendo uma abordagem mais eficiente e humanizada no apoio ao desenvolvimento da fala em pessoas com Trissomia 21.

## 2.10 Trabalhos Relacionados

Nesta seção são discutidas iniciativas relacionadas ao tema central desse trabalho, buscando compreender e aprofundar aspectos relevantes dessa realidade. Posteriormente, na Tabela 2, é apresentada uma análise comparativa dos trabalhos citados em relação aos objetivos principais, públicos-alvo, tecnologia aplicada, modo de uso, grau de autonomia do usuário, acessibilidade, disponibilidade, abordagem educacional e uso IA e multiagentes inteligentes na implementação.

## 2.11 SofiaFala

Entre uma das iniciativas que buscam contribuir para o desenvolvimento da fala em pessoas com Trissomia 21, destaca-se o SofiaFala(FFCLRP/USP, 2019), um aplicativo desenvolvido por alunos da Universidade de São Paulo (USP) de Riberão Preto. O aplicativo foi criado com o objetivo de apoiar crianças com dificuldades na comunicação

Figura 16 – Aplicação de agentes inteligentes no apoio à educação inclusiva

oral, oferecendo atividades interativas que estimulam a fala por meio de jogos, imagens e desafios adaptados às necessidades de cada usuário.

SofiaFala(FFCLRP/USP, 2019) utiliza uma abordagem que valoriza o engajamento das crianças, apostando no uso de recursos visuais, sonoros e recompensas para incentivar a repetição e o treinamento de fonemas específicos. Embora o aplicativo não utilize tecnologias baseadas em inteligência artificial ou sistemas multiagentes, ele representa um importante avanço no uso da tecnologia como ferramenta de apoio pedagógico e terapêutico na área da fonoaudiologia.

A proposta deste projeto aproxima-se da iniciativa SofiaFala, uma ferramenta voltada para o desenvolvimento da fala em pessoas com Trissomia 21. No entanto, diferencia-se ao incorporar a inteligência artificial com arquitetura baseada em multiagentes como componente central da solução. Esses agentes serão capazes de interagir entre si, promovendo uma abordagem colaborativa e adaptativa. Além disso, o projeto utilizará a técnica de Retrieval-Augmented Generation (RAG), que integra recuperação de conhecimento com geração de linguagem natural, possibilitando respostas mais ricas, contextualizadas e personalizadas às necessidades de cada paciente. Nesse sentido, o SofiaFala surge como uma

importante referência, não apenas por seu foco no desenvolvimento da fala, mas também por demonstrar a viabilidade e o impacto positivo do uso de aplicativos no processo de aprendizado de crianças com necessidades específicas.

Figura 17 – SofiaFala
(FFCLRP/USP, 2019)

## 2.12 ItTakesTwoToTalk

De maneira semelhante, o programa internacional It Takes Two to Talk (Pepper, 2017), desenvolvido pelo Hanen Centre, no Canadá, reforça a importância da participação ativa da família no desenvolvimento da comunicação em crianças com atrasos de linguagem. A iniciativa capacita pais e cuidadores com estratégias baseadas em evidências, utilizando o ambiente familiar e situações cotidianas como base para a aprendizagem natural da fala.

Trata-se de um processo estruturado e interativo, cujo objetivo é empoderar pais e cuidadores como os principais mediadores do desenvolvimento linguístico da criança. Ao invés de focar exclusivamente em intervenções terapêuticas realizadas em consultório, o programa valoriza o papel do ambiente doméstico, transformando atividades rotineiras, como brincadeiras, refeições e leitura de histórias, em oportunidades significativas de aprendizado comunicativo.

A metodologia inclui encontros em grupo, vídeos explicativos, atividades práticas e acompanhamento individualizado, por meio dos quais os responsáveis aprendem a identificar os sinais de comunicação das crianças, responder de forma sensível e intencional, e promover interações que favoreçam a aquisição da linguagem. Essa abordagem contribui não apenas para o desenvolvimento da fala, mas também para o fortalecimento do vínculo familiar, a construção da confiança e a ampliação das habilidades de interação social da criança.

Além disso, seu fundamento são em evidências científicas, com eficácia comprovada na melhora das habilidades comunicativas de crianças com atrasos de linguagem, inclusive aquelas com condições como a Síndrome de Down. Seu foco no cotidiano e na parceria

entre família e profissionais torna-o um modelo de intervenção colaborativa, centrado na criança, e com impacto positivo duradouro no desenvolvimento comunicativo.

Embora utilize recursos visuais, como vídeos e materiais impressos para apoiar o aprendizado dos pais, trata-se de um guia prático (19), não fazendo uso de tecnologias digitais ou plataformas virtuais inteligentes, o que o diferencia de abordagens mais recentes baseadas em IA ou agentes conversacionais.

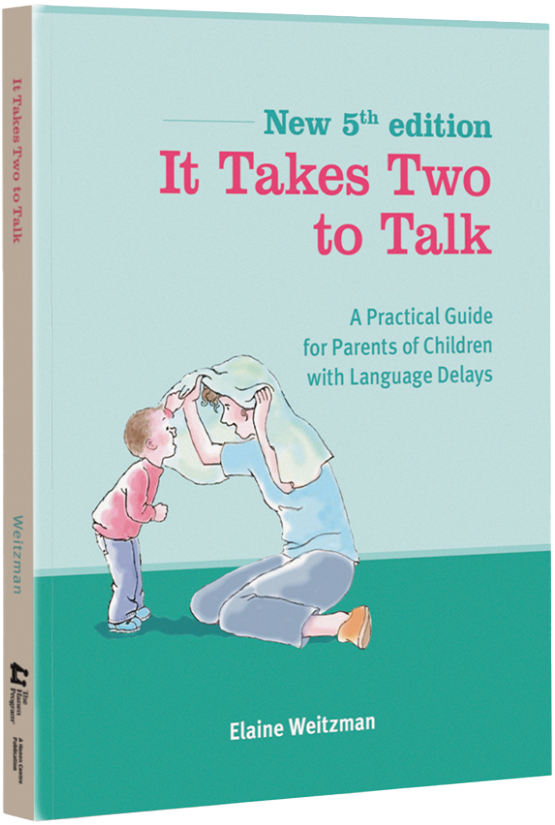


Figura 18 – Guia prático
(Pepper, 2017)

## 2.13 Talkitt

Talkitt (Voiceitt, 2017) é um aplicativo móvel criado pela Voiceitt Company, que usa tecnologias de reconhecimento de voz e inteligência artificial para transformar, em tempo real, sons difíceis de entender em palavras claras e compreensíveis, permitindo que indivíduos com comprometimento da produção da linguagem se comuniquem verbalmente em tempo real.

É uma tecnologia inovadora e assistiva, desenvolvida em Israel, que utiliza inteligência artificial para facilitar a comunicação de pessoas com distúrbios da fala causados

por condições como Trissomia 21, paralisia cerebral, autismo e outras doenças neurológicas. Através de avançados algoritmos de machine learning, o sistema aprende a reconhecer os padrões únicos e pessoais da fala de cada usuário, mesmo quando essa fala não é facilmente compreendida por outras pessoas.

Após um processo de treinamento personalizado, o aplicativo traduz, em tempo real, as vocalizações específicas do usuário em fala clara e inteligível, promovendo uma comunicação mais eficaz com familiares, amigos, educadores e profissionais de saúde. Atualmente, o Talkitt está disponível em vários idiomas, incluindo inglês, português e hebraico, ampliando seu alcance global.

A arquitetura do Talkitt inclui agentes inteligentes que trabalham em conjunto. Para começar a usar o aplicativo, o usuário (ou seu cuidador) precisa "treinar"o Talkitt. Isso envolve gravar palavras e frases que o usuário pronuncia e associá-las às palavras corretas em qualquer idioma desejado. Por exemplo, se uma pessoa diz "choc-lá"quando quer dizer “chocolate”, o aplicativo registra essa associação.

Uma vez que o dicionário personalizado é estabelecido, o Talkitt usa algoritmos de reconhecimento de padrões para identificar as palavras e sílabas que o usuário está tentando pronunciar. Quando o usuário emite um som que pode ser ininteligível para outras pessoas, o aplicativo compara esse som com seu dicionário interno. Ao encontrar uma correspondência, ele traduz essa fala para um discurso claro e compreensível, que pode ser reproduzido em voz alta através do dispositivo.

No aplicativo, são utilizados agentes de aquisição de dados de voz, responsável por capturar a fala do usuário de forma clara e eficiente, agente de análise de padrões, que aprende os padrões de pronúncia únicos do usuário, agente de mapeamento/dicionário, que gerencia e consulta o dicionário personalizado que associa a fala ininteligível à sua forma compreensível, e agente de geração de voz, que Sintetiza a fala traduzida em tempo real, utilizando uma voz que pode ser a do próprio usuário (ou uma voz padrão). Também são considerados agentes de Interface do Usuário, que gerenciam a interação do usuário com o aplicativo, permitindo o treinamento, a configuração e a exibição do texto traduzido.

Esses “agentes” trabalham em conjunto para criar uma experiência de comunicação fluida, permitindo que indivíduos com dificuldades de fala se expressem e sejam compreendidos no dia a dia.

Mais do que uma ferramenta, o Talkitt representa um avanço significativo na inclusão social e na autonomia de pessoas com dificuldades de fala, ao respeitar a forma individual de comunicação de cada usuário sem exigir que este se adapte ao padrão tradicional da linguagem oral. Essa tecnologia vem rompendo barreiras, proporcionando mais qualidade de vida e oportunidades de participação plena na sociedade.

Vale reforçar que o objetivo do aplicativo é se adequar a cada indivíduo e sua

fala, sem forçá-la a aprender novas formas de se comunicar, ao contrário de soluções baseadas em pictogramas ou seleção de palavras pré-programadas. O sistema utiliza machine learning para se adaptar ao vocabulário e entonações específicas de cada usuário, permitindo traduções progressivamente mais precisas com o uso contínuo, entretanto não é uma ferramenta gratuita, tendo custo de assinatura mensal.

Na Figura 19, é apresentada uma linha do tempo que destaca os principais marcos da evolução do Talkitt, incluindo o momento em que a tecnologia passou a incorporar recursos de inteligência artificial para aprimorar o reconhecimento personalizado da fala.

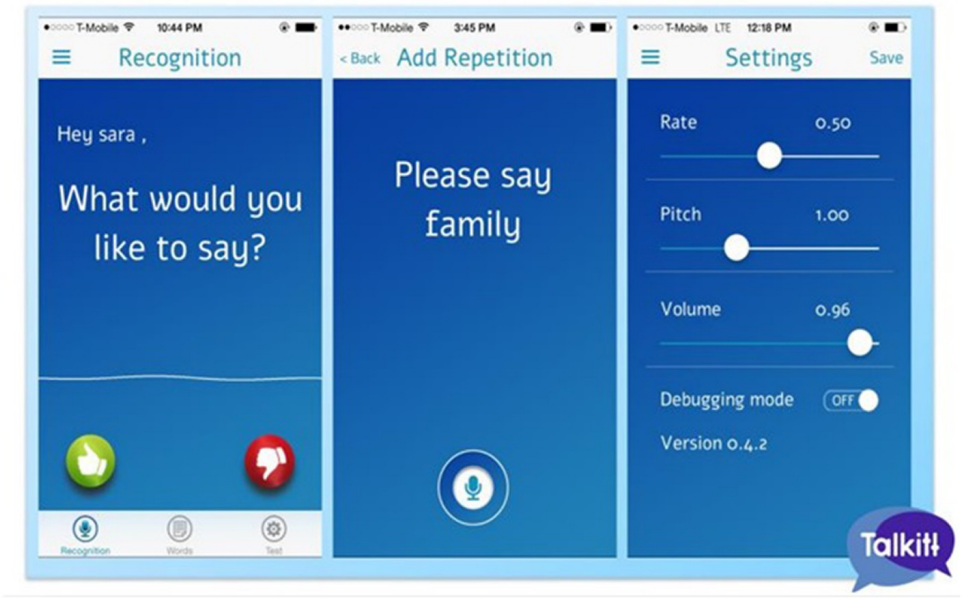


Figura 19 – App Talkitt
(Voiceitt, 2017)

Tabela 2 – Trabalhos Relacionados - Tabela comparativa

| **Critério** | **Talkitt (2015 protótipo / 2020+ piloto)** | **SofiaFala (2022)** | **It Takes Two to Talk** |
|---|---|---|---|
| Objetivo Principal | Traduzir fala atípica em tempo real para comunicação compreensível | Ensinar fala e linguagem com apoio de jogos interativos e lúdicos | Ensinar pais a estimular a fala da criança no dia a dia |
| Público-Alvo | Pessoas com distúrbios na fala (T21, paralisia cerebral, autismo etc.) | Crianças com atraso de linguagem (com ou sem deficiência) | Pais de crianças com distúrbios de comunicação (geralmente até 6 anos) |

| **Critério** | **Talkitt (2015 protótipo / 2020+ piloto)** | **SofiaFala (2022)** | **It Takes Two to Talk** |
|---|---|---|---|
| Tecnologia Aplicada | APP, IA e machine learning com personalização da fala individual | App educacional com recursos multimodais (jogos, TTS, estímulos visuais) | Método de Intervenção comportamental com capacitação presencial de pais |
| Modo de Uso | App aprende a fala do usuário e traduz em tempo real para ouvintes | App gratuito com trilha de atividades educativas | Sessões com fonoaudiólogo + treinamento parental estruturado |
| Grau de Autonomia do Usuário | Alto – comunicação direta e personalizada | Médio – requer mediação de adulto | Baixo – criança depende do estímulo contínuo de pais |
| Acessibilidade | Média – uso restrito por convite ou parcerias institucionais | Alta – gratuito, em português, acessível via Play Store | Média – exige profissional certificado, com custo e presença física |
| Disponibilidade | Assinatura Mensal, sem tradução em português | Gratuito | e-book, livros, sem tradução em português ~2005 |
| Abordagem Educacional | Comunicação funcional e prática | Estímulo à linguagem com apoio visual e sonoro | Interação e vínculo familiar como base para o desenvolvimento da linguagem |
| Uso de IA / Agentes Inteligentes | Não utiliza agentes inteligentes no sentido clássico de "sistemas multiagentes", mas sim modelos baseados em inteligência artificial com foco em machine learning e reconhecimento de padrões individuais de fala. | Não diretamente – utiliza recursos tecnológicos, mas sem uso declarado de IA ou agentes | Não – abordagem tradicional baseada em intervenção humana presencial |

## 2.14 Desafios para Implementação de Soluções com Inteligência Artificial em Contextos Inclusivos

Apesar dos avanços significativos no desenvolvimento de tecnologias baseadas em Inteligência Artificial (IA), ainda persistem desafios importantes que precisam ser enfrentados para garantir sua aplicação ética, eficaz e personalizada em contextos educacionais inclusivos, especialmente no apoio a crianças com **Trissomia 21** e outras condições do neurodesenvolvimento.

Um dos principais desafios é a escassez de estudos científicos de longo prazo que comprovem, de forma robusta, a eficácia das soluções com IA voltadas para esse público. Muitos sistemas ainda não foram validados com amostras suficientemente diversas para assegurar que suas funcionalidades atendam às especificidades de aprendizagem e comunicação dessas crianças, o que levanta dúvidas sobre sua real aplicabilidade pedagógica.

Outro ponto crítico diz respeito à segurança e à privacidade dos dados sensíveis. As tecnologias de IA, por sua própria natureza, dependem da coleta e do processamento de grandes volumes de informações, o que exige a definição de regras claras, mecanismos de proteção confiáveis e comunicação transparente com os responsáveis legais. A ausência de diretrizes bem estabelecidas por parte dos órgãos competentes, como o Ministério da Saúde ou o MEC, faz com que muitas famílias se sintam inseguras ou relutantes em autorizar o compartilhamento de informações, mesmo quando os dados são devidamente anonimizados.

Além disso, observa-se uma carência de soluções personalizadas. A maioria das ferramentas atualmente disponíveis é projetada para o público geral, sem considerar as particularidades cognitivas e comunicativas de pessoas com deficiência intelectual ou múltipla. Dessa forma, torna-se necessário o desenvolvimento de sistemas baseados em IA que sejam treinados com dados específicos desses grupos e que ofereçam adaptações conforme o perfil e o ritmo de aprendizado de cada indivíduo.

Por fim, para que a IA realmente contribua no cotidiano de crianças com deficiência, é fundamental fortalecer a articulação entre tecnologia, profissionais da saúde e da educação, e as famílias. As ferramentas devem ser projetadas para se integrarem às práticas já utilizadas por fonoaudiólogos, psicopedagogos e professores, promovendo sinergia e acompanhamento contínuo. A inclusão de canais de feedback e relatórios personalizados, por exemplo, pode facilitar o trabalho colaborativo entre os agentes envolvidos no processo de ensino-aprendizagem.

## 2.15 Ausência de Bases de Dados Estruturadas

Um dos maiores desafios para o desenvolvimento de soluções tecnológicas de apoio a crianças com Trissomia 21 está na falta de bases de dados organizadas e acessíveis.

Atualmente, não existem repositórios públicos e padronizados que documentem de forma sistemática aspectos como histórico de aprendizagem, evolução de fala, estratégias de intervenção ou exercícios adaptados a esse público.

Essa lacuna impacta diretamente a construção e a validação de sistemas baseados em inteligência artificial, que dependem de dados confiáveis para treinar modelos e oferecer recomendações eficazes. Sem esse suporte, os agentes virtuais encontram limitações para fornecer feedback personalizado, acompanhar o progresso individual ou adaptar-se ao ritmo de aprendizagem de cada criança.

Como alternativa, este projeto considera o uso de dados sintéticos (mockados) para simular cenários de fala, recorrendo a modelos generativos capazes de criar exemplos baseados em padrões linguísticos comuns em pessoas com T21. Além disso, exploramos a possibilidade de utilizar modelos de reconhecimento de fala já pré-treinados, ajustando-os com pequenos conjuntos de dados coletados de forma ética e consentida, aplicando técnicas de aprendizado por reforço com feedback humano.

A criação futura de uma base de dados especializada, construída com rigor científico e respeito às normas éticas e legais, é essencial para garantir a evolução sustentável de tecnologias inclusivas como a aqui proposta. Tal iniciativa permitirá não apenas validar modelos de IA com maior precisão, mas também apoiar famílias e profissionais no processo de alfabetização e comunicação assistida de pessoas com deficiência intelectual.

# 3 PROPOSTA: APLICAÇÃO INTERFACE PARA AUXILIAR O DESENVOLVIMENTO DE FALA EM PESSOAS COM TRISSOMIA 21

O objetivo deste trabalho é a criação de uma arquitetura multiagentes inteligentes, chamada Interface, integrada a um modelo de linguagem de grande escala (LLM). A proposta visa oferecer suporte ao ensino e ao desenvolvimento da fala de pessoas com Trissomia 21, com foco especial em crianças em fase de aquisição da linguagem.

A arquitetura sugerida é composta por dois componentes técnicos principais: o modelo Whisper, responsável pela transcrição automática da fala, e a ferramenta LangChain, que atua como orquestradora dos agentes linguísticos e pedagógicos. Essa combinação permite uma coordenação eficaz entre os módulos de processamento, promovendo interações mais naturais, adaptáveis e centradas nas necessidades do usuário.

Neste capítulo são apresentados protótipos para a avalidação das soluções propostas. Por meio dos protótipos, é possível visualizar e testar o fluxo de interação entre os usuários (crianças, pais, fonoaudiólogos e agentes inteligentes), garantindo que a interface seja intuitiva, amigável e eficaz.

Os protótipos incluem telas de boas-vindas, login, exercícios interativos, análise de desempenho, sugestões personalizadas por IA e feedback contínuo, tudo modelado para refletir a arquitetura proposta com uso de multiagentes inteligentes e tecnologia RAG.

## 3.1 Componentes da aplicação Interface

A construção da plataforma Interface buscou um equilíbrio entre eficiência tecnológica e custo acessível. Para isso, priorizaram-se ferramentas gratuitas ou de código aberto, que oferecem recursos suficientes para o desenvolvimento de um MVP robusto, sem exigir investimentos elevados. A seguir, destacam-se as principais escolhas.

### 3.1.1 Frontend (Interface com o usuário)

Optou-se pelo **React.js**, aliado ao **Next.js** e ao **Tailwind CSS**. Essas ferramentas permitem criar interfaces modernas, rápidas e responsivas, tanto para web quanto para dispositivos móveis. A escolha se deve ao grande suporte da comunidade, facilidade de aprendizado e custo zero.

### 3.1.2 Backend (Lógica e APIs)

Para o backend, foi escolhido o **Python com FastAPI**. Essa tecnologia destaca-se pela rapidez na criação de APIs, ótima integração com modelos de IA e simplicidade de manutenção. Assim, conecta o frontend aos módulos de inteligência artificial de forma prática e segura.

### 3.1.3 Inteligência Artificial

As bibliotecas de IA dão vida aos agentes inteligentes, com foco em baixo custo e acessibilidade. Enquanto o LangChain coordena a comunicação entre os agentes de IA, o Whisper, uma ferramenta open source, realiza a transcrição da fala infantil, inclusive no modo offline. Além disso, a OpenAI (free tier), permite o acesso gratuito inicial a modelos de linguagem avançados, útil para realização de testes e protótipos. Tais ferramentas foram escolhidas por sua maturidade e pela disponibilidade de versões gratuitas.

### 3.1.4 Banco de Dados e Armazenamento

Para armazenamento de dados e arquivos, adotou-se o **Supabase**, que combina banco de dados PostgreSQL, autenticação e espaço em nuvem. Dessa forma, substitui múltiplas ferramentas por uma solução unificada e de ótimo custo-benefício.

### 3.1.5 Hospedagem

A plataforma será hospedada na **Vercel**, que oferece planos gratuitos e integração nativa com React e Next.js. Essa escolha reduz a complexidade da publicação e garante estabilidade nas versões iniciais.

### 3.1.6 Ferramentas de Apoio

Como suporte adicional, serão utilizadas soluções de código aberto, como Grafana, que permite o monitoramento e e construção de dashboards; Chatwoot, que permite a comunicação e feedback com usuários; bem como a ferramenta N8N, que permite a automação de tarefas e integrações.

## 3.2 Classificação Inteligente de Textos e Falas com Apoio de Agentes

A inteligência artificial tem transformado a forma como nos comunicamos. Com ela, sistemas são capazes de “ouvir”, “ler” e até “falar”, interpretando tanto a fala quanto a escrita. Isso se torna especialmente importante quando falamos do apoio ao desenvolvimento da linguagem, como no caso de pessoas com Trissomia 21 ou outras condições que impactam a comunicação.

Na aplicação Iterface, a ideia é que diferentes agentes atuem, de maneira complementar, em funções específicas, como a análise dos sons da fala do usuário, a compreensão do significado das palavras e o acompanhamento da evolução de seu aprendizado. Essa colaboração entre multiplos agentes permitirá que a fala e o texto sejam analisados de forma mais completa, levando em conta as sutilezas do processo de desenvolvimento da linguagem do usuário, algo essencial em contextos inclusivos.

A ideia é que os dados gerados dentro da aplicação sejam úteis para torná-lo mais preciso, por meio da aplicação de critérios de classificação relacionados tanto à fala quanto ao texto. Os critérios considerados na concepção do aplicativo são descritos na Tabela 3, e envolvem aspectos como a classificação do nível do vocabulário, complexidade das frases, tipos de texto e modos de expressão, interesse temático, bem com a adequação fonológica do usuário. Uma vez coletados e rotulados, os dados são usados no treinamento de inteligências artificiais, e juntamente com o apoio de profissionais como fonoaudiólogos e educadores, permitem uma maior personalização e confiabilidade do plataforma, enriquecendo o processo de ensino e aprendizagem da fala do usuário. Entretanto, os testes realizados fazem conceitos apenas diferença entre o texto de referência e o texto produzido.

Tabela 3 – Critérios de classificação linguística aplicáveis à fala e texto

| **Classificação / Rótulo** | **Descrição** |
|---|---|
| **Nível de vocabulário** | Avalia a diversidade lexical, considerando a quantidade e o nível de dificuldade das palavras utilizadas. |
| **Complexidade das frases** | Analisa a estrutura sintática, identificando se há predominância de frases simples ou complexas. |
| **Tipo de texto / Modo de expressão** | Identifica o gênero textual ou a forma de expressão, como: diálogos, instruções, descrições, narrações, músicas ou vídeos. |
| **Interesse temático** | Avalia a aderência do conteúdo ao interesse do usuário (criança), com base em categorias como música, vídeo, texto ou imagem. |
| **Adequação fonológica** | Verifica se os sons e palavras usados estão alinhados com os objetivos fonológicos do processo de ensino da fala. |

## 3.3 Representação de Texto: Do Áudio ao Significado

Como apresentado anteriormente, a arquitetura do Interface será composta por duas camadas principais, que trabalham de forma integrada para viabilizar diferentes níveis de processamento da linguagem. O objetivo é transformar a fala em um conteúdo textual compreensível e contextualizado, quando necessário, bem como realizar o caminho inverso: converter texto em fala.

Esse processo envolve muito mais do que apenas transcrever palavras. Busca-se captar elementos mais profundos da comunicação, como nuances cognitivas, semânticas e pragmáticas, que são fundamentais em contextos educacionais inclusivos. Isso é especialmente relevante quando se trata de apoiar a comunicação de crianças com Síndrome de Down (Trissomia 21), permitindo que os sistemas entendam e se adaptem às particularidades de cada usuário.

Na Figura 20 é apresentado o fluxo de processamento de fala e texto em arquitetura modular que será considerado na construção da aplicação. A imagem demonstra a

transformação de áudio em texto interpretável e, posteriormente, a conversão do texto em fala sintética. O modelo visa aplicações inclusivas, com foco no apoio à comunicação de crianças com Trissomia 21.

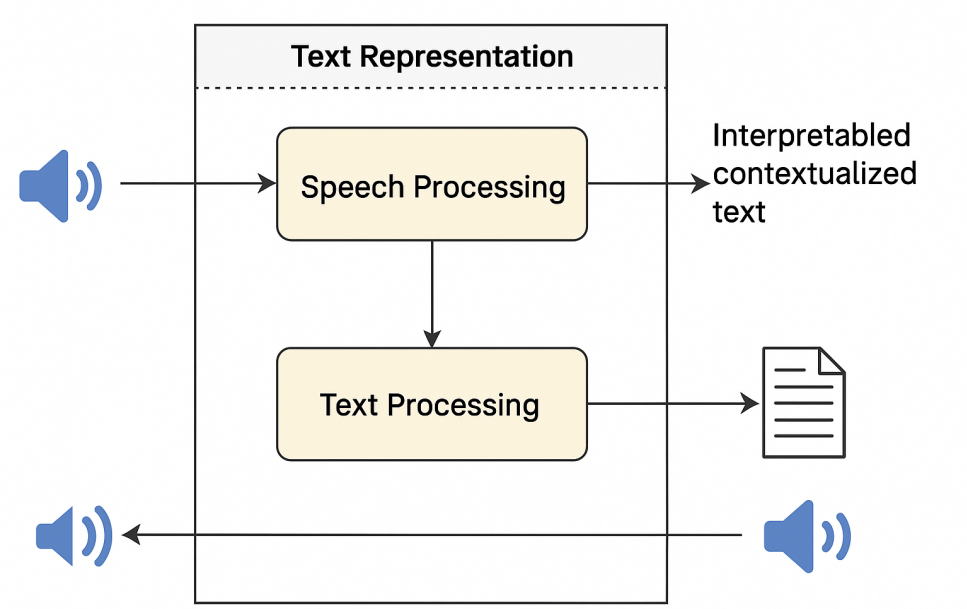


Figura 20 – Arquitetura de Processamento de Fala e Texto para Apoio na Fala
Fonte: Elaborado pelo autor.

## 3.4 Representação no nível de vocabulário e fonética (Compreensão das palavras e de como elas são pronunciadas)

A primeira etapa do processo consiste na transcrição da fala e na representação fonética da entrada de áudio, utilizando o modelo Whisper (Radford *et al.*, 2023). Esse sistema de reconhecimento automático de fala (ASR), conforme apresentado na Seção 2.6, foi treinado com um vasto conjunto de dados multilíngues, o que lhe permite alcançar alta precisão mesmo em ambientes com ruído ou com diferentes sotaques.

Além de gerar a transcrição do texto falado, o Whisper é capaz de extrair informações fonéticas e prosódicas, enriquecendo a análise inicial da fala e aproximando a representação textual da fala espontânea, tal como ela ocorre na prática.

Esse nível de representação é especialmente importante para captar as variações individuais na pronúncia, comuns em pessoas com deficiências intelectuais ou dificuldades na articulação da fala. Assim, garante-se uma transcrição sensível e precisa, minimizando a perda de significado já na etapa inicial do processamento.

## 3.5 Análise do significado real e da intenção por trás das palavras

A segunda camada do sistema aprofunda a interpretação da fala, indo além da simples transcrição literal. Aqui, entra em ação uma cadeia de agentes inteligentes baseada

na plataforma LangChain (Chase, 2022), responsável por coordenar interações com modelos de linguagem de grande porte, como o GPT-4 (OpenAI, 2023).

Nesta etapa, o conteúdo falado começa a ser interpretado de forma contextualizada, levando em consideração não apenas as palavras ditas, mas também o histórico de interações do usuário, seu perfil cognitivo e os objetivos pedagógicos previamente estabelecidos. Por exemplo, se uma criança apresenta dificuldades com determinados fonemas ou conceitos, o sistema pode adaptar sua resposta para reforçar esses pontos, oferecendo interações mais significativas e alinhadas ao seu processo de aprendizagem.

Essa camada atua explorando tanto o nível semântico — ou seja, o significado das palavras e suas conexões — quanto o nível pragmático, que considera a intenção da fala e o contexto em que ela ocorre. Isso permite que o sistema vá além de respostas genéricas e passe a agir de maneira personalizada, sensível às particularidades de cada criança.

Como afirmam Clark (1996) e Tomasello (2003), a linguagem é uma prática social profundamente enraizada no contexto. Por isso, qualquer tecnologia que se proponha a apoiar o desenvolvimento da linguagem precisa considerar essa complexidade. A arquitetura aqui proposta reflete esse princípio ao construir uma ponte entre o som e o significado, entre a palavra falada e a sua plena compreensão — tudo isso ancorado em valores de inclusão, respeito à individualidade e adaptação pedagógica.

Mais do que reconhecer corretamente o que foi dito, essa abordagem busca compreender como e por que algo foi dito, oferecendo respostas que apoiem o desenvolvimento linguístico e cognitivo de pessoas com Trissomia 21, de forma sensível, inteligente e personalizada.

## 3.6 Representação Multimodal

A comunicação humana é inerentemente multimodal, envolvendo diferentes formas de expressão como fala, texto, gestos e imagens. Nesse sentido, o sistema proposto adota uma abordagem de *representação multimodal*, integrando diferentes canais de entrada e saída para enriquecer a interação com a criança com Trissomia 21.

O fluxo de entrada envolve áudio da fala da criança, que é processado por um sistema de reconhecimento automático de fala (ASR), como o *Whisper*. Em seguida, o conteúdo é interpretado em formato textual por modelos de linguagem e reconvertido em saídas multimodais, como respostas em voz (TTS), animações, personagens visuais ou elementos gráficos interativos em *dashboards* e jogos.

A representação multimodal permite que os agentes inteligentes compreendam o contexto de forma mais ampla e ofereçam estímulos diversos, respeitando a pluralidade cognitiva e sensorial das crianças com Trissomia 21, o que contribui significativamente para a inclusão e o aprendizado eficaz, conforme a Figura 21.

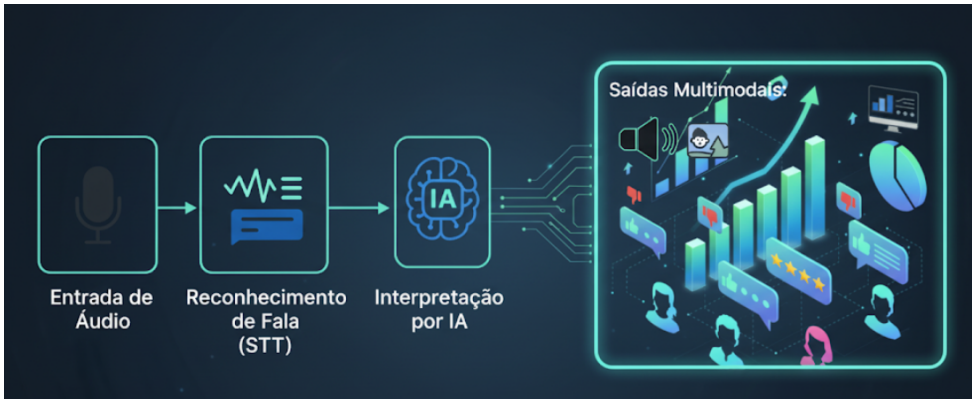


Figura 21 – Representação multimodal do processamento de fala no sistema Interface, entrada por áudio às saídas multimodais (voz, imagem, vídeos e dashboard)
Fonte: Elaborado pelo autor.

A estruturação da interface considera não apenas aspectos tecnológicos, mas também fundamentos neurocognitivos que explicam como pessoas com T21 processam informações. O desenvolvimento de uma ferramenta de apoio a profissionais, familiares e pacientes envolve diferentes meios de comunicação, como sons, imagens, vídeos e gestos. A aproximação entre a tecnologia e as necessidades reais dos pacientes (usuários da ferramenta) busca promover a interação entre os envolvidos. A capacidade de interpretar dados e integrar múltiplos canais de comunicação torna o aprendizado mais eficaz, respeitando as particularidades de cada indivíduo.

Um exemplo de funcionamento do aplicativo seria o sistema emitir um áudio com palavras e solicitar que o paciente as repita. A IA, então, identificaria a fala, realizaria a pronúncia correta e ofereceria um feedback imediato.

O mesmo princípio pode ser aplicado a vídeos e sons. Ao identificar a necessidade específica, a IA poderia criar, com base nas particularidades do paciente, um vídeo com música ou palavras, por exemplo, e solicitar que o paciente repita. Isso enriquece a experiência do usuário, personaliza os exercícios e auxilia no aprimoramento da pronúncia, entre outros aspectos.

Além disso, a ferramenta permitiria o compartilhamento das informações por meio de relatórios individuais ou coletivos entre os agentes envolvidos, garantindo que ações e orientações estejam sempre alinhadas.

Além disso, a ferramenta permitiria o compartilhamento das informações por meio de relatórios individuais ou coletivos entre os agentes envolvidos, garantindo que ações e orientações estejam sempre alinhadas.

Para a criação da base de dados, é recomendável uma arquitetura híbrida, capaz de armazenar arquivos de diferentes extensões — textos, vídeos, imagens e áudios. Nesse contexto, bancos vetoriais como o Weaviate apresentam vantagens, pois permitem armazenar e buscar vetores de alta dimensionalidade, possibilitando consultas híbridas que combinam

proximidade vetorial com filtros estruturados. Também oferecem indexação automática, que acelera buscas complexas, e suporte para GraphQL, o que facilita consultas. Além disso, podem gerar embeddings automaticamente ao armazenar textos usando modelos como os da OpenAI.

Outro diferencial é a interface amigável com LLMs, sem a necessidade de um pipeline adicional, além de ser open source, sem custos de licença. Por outro lado, essa solução pode apresentar uma curva de aprendizado mais elevada e demandar maior capacidade de CPU e memória devido ao processo de indexação, sendo mais indicada para dados imutáveis ou de crescimento apenas por acréscimo (append-only).

A proposta de arquitetura do sistema baseia-se em modelos de IA treinados e ajustados para considerar as particularidades fonológicas e cognitivas desse público, utilizando diversos recursos tecnológicos. O modelo Whisper pode ser empregado para transcrição de fala, enquanto a API da OpenAI, acessada via biblioteca openai, viabiliza a geração de respostas em linguagem natural. Adicionalmente, bibliotecas como sounddevice, wave e soundfile podem ser utilizadas para captura e processamento de áudio.

Para a gestão das interações, a biblioteca LangChain representa um facilitador importante, com componentes como o ChatOpenAI v4.0 (que requer créditos para uso) ou alternativas baseadas em outras LLMs sem custo de licenciamento.

A solução proposta é tecnicamente viável e pode ser aplicada como ferramenta complementar em contextos terapêuticos e educacionais. Ao combinar tecnologias acessíveis com uma abordagem centrada na pessoa, o sistema mostra-se capaz de oferecer suporte real ao desenvolvimento da comunicação em pessoas com T21 ou com dificuldades de fala, contribuindo para sua autonomia, inclusão e qualidade de vida.

Observa-se, entretanto, que em alguns casos de T21 o aprendizado é potencializado pela associação entre fala e estímulos visuais, como vídeos acompanhados de música e dança. Nesse sentido, a integração da Inteligência Artificial com a técnica RAG (Retrieval-Augmented Generation) permite criar interações dinâmicas e conteúdos personalizados, como exercícios e vídeos, adaptados a diferentes níveis de aprendizado. Todo o processo é armazenado em banco de dados, possibilitando que o sistema se autoajuste de forma contínua. Assim, responsáveis e usuários podem escolher como as atividades serão executadas, enquanto o sistema identifica desempenhos abaixo do esperado e sugere ajustes proativos para otimizar o aprendizado.

Outro diferencial está na disponibilidade multicanal: além de dispositivos amplamente utilizados, como celulares e tablets, o sistema também pode ser acessado em televisores, ampliando a imersão. A combinação de áudio, imagem e texto proporciona uma experiência mais rica, que pode ser ainda mais envolvente com a inclusão de personagens animados (como super-heróis ou figuras familiares) e o uso de músicas e melodias com

palavras específicas, tornando o aprendizado lúdico e personalizado.

A escolha da arquitetura e da base de dados constitui um ponto crítico do projeto, especialmente pela necessidade de integração com mecanismos RAG e pelo uso de multiagentes, o que exige um planejamento cuidadoso para garantir escalabilidade, eficiência e usabilidade.

# 4 ANÁLISE DE MERCADO: CANVAS, ANÁLISE SWOT E BENCHMARKING TECH

Neste capítulo é apresentada uma análise de mercado com o intuito de comprovar a viabilidade de construção da aplicação proposta, cujo intuito é auxiliar no desenvolvimento da fala de pessoas com Trissomia 21.

## 4.1 Canvas da Proposta Voltada às Famílias

Na Figura 22 é apresentado o *Canvas* da proposta voltada às famílias e cuidadores de crianças com Trissomia 21. Os **parceiros-chave** do projeto incluem instituições de ensino, organizações sem fins lucrativos e entidades especializadas em apoio ao desenvolvimento infantil, que colaboram na validação pedagógica e na consolidação dos objetivos educacionais da solução.

As **atividades centrais** estão voltadas à promoção de experiências interativas e lúdicas que estimulem o desenvolvimento da fala, aliando tecnologia e metodologias inclusivas de aprendizagem. A solução propõe mecanismos de acompanhamento e adaptação contínua, de modo que as atividades possam ser ajustadas conforme as necessidades e o ritmo de cada criança. Além disso, o sistema prevê formas de engajamento que incentivam a participação da família no processo de evolução e aprendizado.

A **proposta de valor** busca oferecer às famílias segurança, confiança e tranquilidade, permitindo que acompanhem o progresso da criança em casa de forma simples e compreensível. A ferramenta visa tornar o momento de aprendizado mais agradável, motivador e acessível, reforçando o vínculo afetivo entre a criança e seus cuidadores, e promovendo um ambiente de estímulo positivo.

Os **relacionamentos com os usuários** serão construídos com base em suporte empático e comunicação contínua, contemplando atualizações de conteúdo, orientações práticas e um espaço digital voltado à troca de experiências entre famílias e comunidade de apoio.

Os **canais de acesso e comunicação** incluem uma plataforma digital interativa, desenvolvida para oferecer experiências de uso simples e intuitivas, além de suporte complementar por meio de redes sociais e meios institucionais.

O **segmento de usuários** é composto por famílias e cuidadores de crianças com Trissomia 21, que buscam recursos acessíveis e inclusivos para apoiar o desenvolvimento da fala e da comunicação de maneira positiva e eficaz.

A **estrutura de custos** envolve principalmente o desenvolvimento e manutenção tecnológica, bem como o aprimoramento contínuo da experiência do usuário e o suporte

às famílias participantes.

Por fim, os **fluxos de sustentabilidade** poderão incluir parcerias institucionais, colaborações com programas sociais e, futuramente, modelos de apoio financeiro que assegurem a continuidade e acessibilidade da proposta, mantendo seu caráter inclusivo e de impacto social.


Figura 22 – Canvas
Fonte: Elaborado pelo autor.

## 4.2 Análise SWOT

Na Figura 23 são apresentas as forças, fraquezas, oportunidades e ameaças que podem comprometer o desenvolvimento da aplicação. Dentre as forças, encontram-se aspectos como o uso de multiagentes e a integração de multiplas modalides, como voz, texto, imagem e vídeo. Além disso, o aplicativo será construído com o apoio de profissionais, podendo ser expandido posteriormente para o atendimento de um número ilimitado de usuários.

Como fraquezas, estão listados aspectos como a necessidade de acesso estável à internet, a necessidade de recursos para implementação, treinamento e testes da aplicação e uma base de dados inicialmente limitada, o que pode comprometer a qualidade do modelo gerado.

Como oportunidades, são listados aspectos como o crescimento do mercado de *healthcare*, possíveis parcerias com clínicas e universidades, expansão do aplicativo para

outros transtornos e condições, bem como o apoio de políticas públicas de inclusão, como programas governamentais voltados à acessibilidade educacional.

Como ameaças são apontados aspectos como a rápida evolução tecnológica, questões relacionadas a dados e privacidade e possíveis resistências do profissionais ou familiares no uso da aplicação.

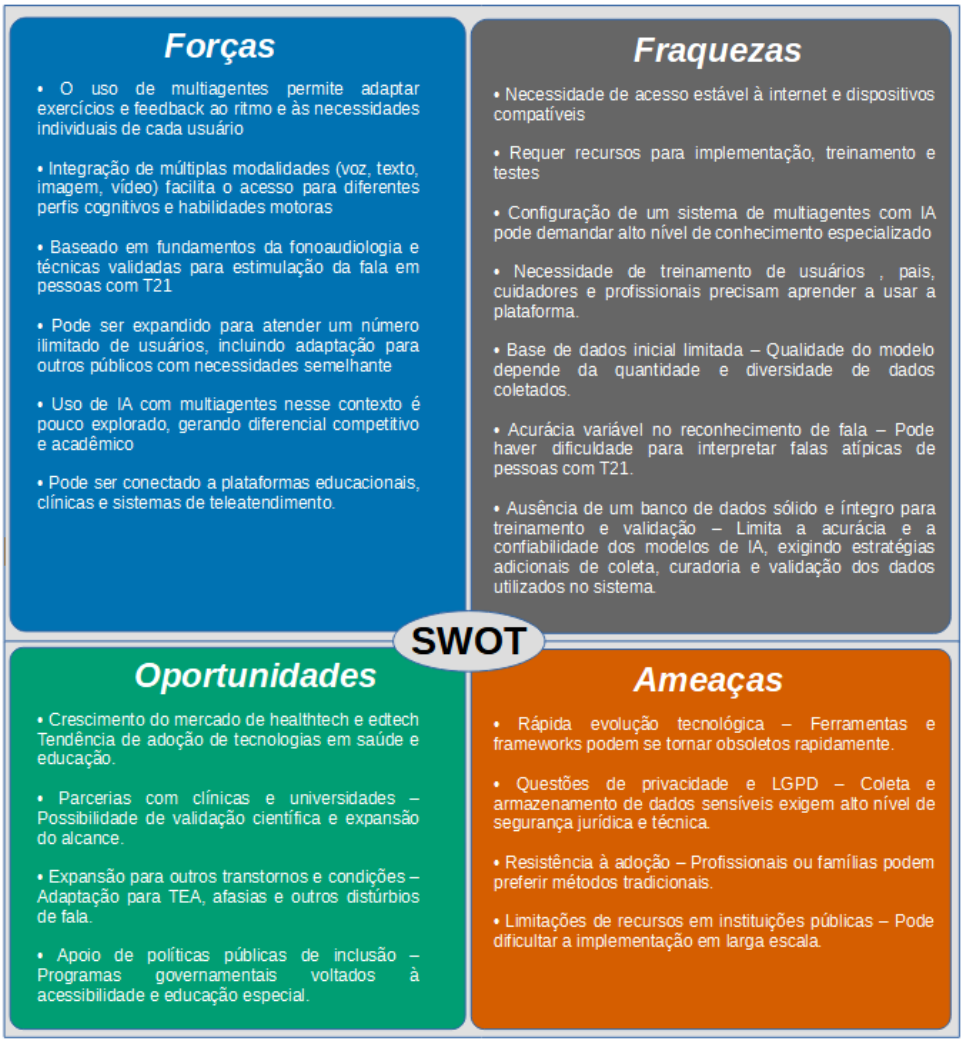


Figura 23 – Swot
Fonte: Elaborado pelo autor.

## 4.3 Benchmarking Tech

Este estudo reúne uma análise comparativa das tecnologias utilizadas no projeto Interface, voltado ao desenvolvimento da fala em crianças com Trissomia 21. A comparação segue os quatro principais tipos de benchmarking – Interno, Competitivo, Funcional e Genérico. O Objetivo está em identificar soluções tecnológicas que utilizam inteligência artificial e agentes inteligentes, avaliando como essas ferramentas podem contribuir de forma eficaz para a evolução da fala em crianças com T21.

### 4.3.1 Benchmarking Internos

O benchmarking interno tem como foco a análise dos próprios componentes, fluxos e decisões tecnológicas adotadas na construção da plataforma, com o objetivo de identificar pontos fortes, gargalos e oportunidades de aprimoramento dentro do ecossistema da solução.

Dentre os módulos do aplicativo, o módulo de reconhecimento de fala, implementado com base no Whisper, apresentou excelente desempenho na transcrição de áudios em língua portuguesa, especialmente quanto à acurácia em ambientes controlados. Entretanto, nos testes iniciais, verificou-se a necessidade de ajustes específicos no pré-processamento de áudio, a fim de adaptar o sistema às características vocais de crianças com Trissomia 21, cujos padrões de fala podem diferir significativamente dos contemplados nos modelos convencionais de treinamento.

Outro módulo analisado é o de síntese de fala (TTS), responsável por converter texto em áudio para possibilitar o retorno interativo com o usuário. Atualmente, são empregadas soluções gratuitas como o Coqui TTS e os modelos de TTS disponíveis na HuggingFace, que apresentam bom desempenho e ampla acessibilidade. Contudo, identificou-se que, embora funcionais, essas soluções ainda carecem de maior naturalidade vocal e expressividade emocional — aspectos fundamentais para estabelecer uma comunicação empática e eficaz com o público infantil-alvo da plataforma.

Assim, o benchmarking interno da Interface revela um cenário promissor em termos de potencial tecnológico, ao mesmo tempo em que aponta caminhos claros para evolução, como a adaptação de modelos de IA para fala atípica e o aprimoramento da qualidade da síntese de voz.

### 4.3.2 Benchmarking Competitivo

Soluções tradicionais, como o programa **It Takes Two to Talk**, têm grande relevância ao capacitarem pais e cuidadores no apoio à comunicação de crianças com distúrbios de linguagem. No entanto, por dependerem de encontros presenciais e seguirem uma estrutura fixa, acabam limitando o alcance e a flexibilidade do processo de aprendizagem, especialmente em contextos com menos acesso a recursos especializados.

Nesse mesmo cenário, destaca-se o **SofiaFala**, um aplicativo gratuito criado na USP de Ribeirão Preto com o objetivo de apoiar o desenvolvimento da fala em crianças, especialmente aquelas com síndrome de Down. Idealizado por uma mãe cientista da computação para atender às necessidades comunicativas de sua filha, o projeto evoluiu e passou a beneficiar diversas famílias e profissionais da área. A ferramenta combina inteligência artificial, visão computacional e aprendizado de máquina para reconhecer sons, fornecer feedbacks interativos e acompanhar a evolução do usuário. Dividido em dois

módulos — um voltado para o paciente e outro para o fonoaudiólogo, o sistema possibilita treinamentos domiciliares mediados por pais ou cuidadores, promovendo uma rotina de prática mais acessível, eficaz e contínua.

O Talkitt possui uma proposta centrada na autonomia comunicativa de pessoas com distúrbios severos de fala. Desenvolvido em Israel, o aplicativo utiliza inteligência artificial com módulos especializados, que funcionam de maneira semelhante a agentes inteligentes, incluindo: Agente de Reconhecimento de Padrões de Fala Personalizada, Agente de Tradução e Interpretação Contextual, Agente de Treinamento Contínuo/Aprendizado Adaptativo e Agente de Interface e Feedback.

Embora não seja um sistema multiagentes no sentido estrito, o Talkitt integra esses módulos de forma coordenada, criando uma experiência personalizada que respeita os padrões únicos de vocalização de cada usuário, traduzindo-os em tempo real em uma fala clara e compreensível. A integração entre os componentes, destacada em (Voiceitt, 2023), permite o reconhecimento de padrões não convencionais de fala, interpretação de contextos linguísticos, adaptação ao uso contínuo e fornecimento de feedback interativo.

Dessa forma, o Talkitt promove inclusão, sem exigir que os usuários se adaptem à linguagem oral padrão. Atualmente, o aplicativo está disponível em inglês e hebraico, ampliando seu potencial de impacto global.

Na Tabela 4 é apresentado o benchamrking competitivo entre as aplicações Talkitt, SofiaFala, It Takes Two to Talk e a nossa proposta (Interface).

A aplicação Interface consiste em uma plataforma modular baseada em inteligência artificial e sistemas multiagentes, desenvolvida para apoiar o ensino e o desenvolvimento da fala em crianças com Trissomia 21.

Diferenciando-se de outras iniciativas, a aplicação utiliza uma arquitetura distribuída composta por agentes especializados — como Agente Fonoaudiólogo Virtual, Agente Tutor de Fala, Agente de Avaliação Contínua, Agente de Feedback Multimodal e Agente de Aprendizado Adaptativo — que interagem entre si para promover uma experiência personalizada, responsiva e orientada ao progresso do usuário.

A plataforma Interface integra tecnologias como Whisper para transcrição automática de fala, LangChain para raciocínio contextual e sistemas de TTS (text-to-speech), criando um ciclo de comunicação fluida e bidirecional entre a criança e o agente. Por meio de dashboards para responsáveis e ferramentas de acompanhamento em tempo real, o Interface oferece um ambiente inclusivo e acessível para o treino diário da fala, mesmo em contextos com poucos recursos especializados. Sua estrutura aberta permite a incorporação de novos agentes e módulos conforme a necessidade pedagógica, tornando-se uma solução escalável, extensível e centrada na criança.

Com esses avanços, a integração entre inteligência artificial, agentes inteligentes

Tabela 4 – Benchmarking Competitivo entre Talkitt, SofiaFala, It Takes Two to Talk e Interface

| Critérios | Talkitt | SofiaFala | It Takes Two to Talk | Interface (Proposta) |
|---|---|---|---|---|
| **1. Uso de IA para reconhecimento de fala** | Sim, com IA proprietária para sons ininteligíveis | Sim, com IA adaptada ao português | Não aplica (abordagem tradicional) | Sim, com Whisper e LLMs adaptados ao perfil infantil |
| **2. Personalização ao perfil do usuário** | Limitada | Parcial | Alta (via intervenção humana) | Alta, com IA e aprendizado contínuo por histórico do usuário |
| **3. Coesão entre os módulos e fluxos** | Foco em tradução de fala | Plataforma integrada, mas com escopos limitados | Muito coeso, porém sem tecnologia embarcada | Altamente coeso com arquitetura modular de microsserviços |
| **4. Acessibilidade e inclusão** | Média (não nativo em português) | Boa | Alta (via interação pais-terapeutas) | Muito alta, com foco em T21 e linguagem simplificada |
| **5. Base em evidências terapêuticas** | Baixa (tecnológica) | Média (apoio clínico) | Muito alta (modelo Hanen) | Alta, com apoio fonoaudiológico e validação contínua |
| **6. Custos e acesso ao público** | Alta mensalidade (em dólar) | Acessível | Alto custo (presencial) | Gratuito e open source |
| **7. Integração com responsáveis e terapeutas** | Não | Parcial | Sim, essencial ao modelo | Sim, com dashboards e relatórios automáticos |
| **8. Integridade dos dados e segurança** | Boa, mas centralizada | Razoável | Alta (manual) | Alta, com armazenamento seguro e LGPD compliance |

autônomos e a técnica RAG (Retrieval-Augmented Generation) inaugura uma abordagem promissora para o apoio educacional. Em vez de trilhas fixas de aprendizagem, essa combinação possibilita experiências dinâmicas, personalizadas e evolutivas, acompanhando o ritmo e as necessidades específicas de cada indivíduo. Por se tratar de uma plataforma aberta, ela se integra facilmente a outras ferramentas e bases de dados, ampliando as possibilidades de uso e facilitando sua aplicação em diferentes contextos, da prática clínica à educação inclusiva.

Assim, a união entre teoria, tecnologia e design centrado no usuário transforma a proposta de ensino da fala em algo mais acessível, engajador e inclusivo, respeitando as singularidades de cada indivíduo e potencializando o papel da IA como aliada na comunicação humana.

### 4.3.3 Análise de proximidade: aplicações mais próximas e mais distantes do propósito do Interface

O SofiaFala é a ferramenta que mais se aproxima do propósito do Interface, auxiliando no desenvolvimento da fala e na terapia fonoaudiológica, especialmente para crianças com síndrome de Down. Ambos têm como objetivo servir de apoio tanto ao paciente quanto ao profissional, utilizando a tecnologia para coletar dados, personalizar exercícios e gerar relatórios. A principal diferença, como observado, está na forma como a IA é

estruturada para atingir esses objetivos.

Por outro lado, o Talkitt se distancia da proposta do Interface, pois seu foco principal não é o ensino ou o desenvolvimento da fala, mas sim a tradução em tempo real. Trata-se de uma ferramenta de comunicação assistiva, que facilita a interação imediata do usuário com outras pessoas, sem necessariamente visar a melhora da fala a longo prazo.

### 4.3.4 Benchmarking Funcional

O benchmarking funcional analisa ferramentas que, mesmo não sendo concorrentes diretas, oferecem funcionalidades semelhantes.

Entre as ferramentas consideradas, destacam-se: o Whispr AI (não relacionado à OpenAI), que realiza transcrição e tradução automática de fala para texto; a Google Speech-to-Text API, que apresenta alta precisão em transcrição, mas envolve custos e não é voltada para fala atípica; e o TTS da Amazon Polly, que oferece síntese vocal de alta qualidade, porém é uma solução paga e voltada para usos comerciais.

O Interface, por sua vez, se beneficia ao adotar soluções open source e gratuitas, ajustadas para cenários de fala não padronizada, combinando funcionalidades similares com maior acessibilidade econômica e foco pedagógico, atendendo de forma mais direta às necessidades de seu público-alvo.

### 4.3.5 Benchmarking Genérico

Por fim, o benchmarking genérico consiste em comparar práticas e tecnologias utilizadas em outros setores, mas que podem ser aplicadas ao contexto do Interface.

Um exemplo é o uso de chatbots educacionais com IA em plataformas de ensino a distância, como o Khan Academy com GPT e o Duolingo Max, que demonstram como agentes conversacionais podem personalizar a experiência de aprendizagem e adaptar conteúdos às necessidades individuais dos usuários.

## 4.4 Discussão

O trabalho evidencia a relevância do uso de tecnologias de inteligência artificial no apoio ao desenvolvimento da fala de crianças com Trissomia 21. A plataforma Interface propõe uma base centralizada para pesquisas e diagnósticos, inovando ao adotar uma arquitetura de multiagentes inteligentes — como os agentes fonoaudiólogo, de personalização e de feedback — que atuam de forma autônoma, colaborativa e especializada. Essa abordagem favorece a escalabilidade, manutenção, evolução contínua e reutilização de componentes, além de permitir fluxos de aprendizagem individualizados, fundamentais para esse público.

O projeto também se apoia em ferramentas gratuitas e de ponta, garantindo viabilidade financeira e alinhamento com os princípios da ciência aberta e da acessibilidade digital. Com isso, estabelece-se um modelo avançado, acessível e socialmente responsável, que sustenta o MVP e possibilita futuras expansões, como novos agentes especializados ou integração com dispositivos móveis, wearables e sistemas escolares.

# 5 METODOLOGIA

Este capítulo descreve os procedimentos adotados para o avaliar a viabilidade da construção do MVP da proposta. Foram empregados modelos de linguagem e agentes inteligentes para a simulação dos dados que alimentariam e seriam gerados por usuários do sistema. Sobre os dados simulados, foram avaliadas possíveis métricas que poderiam ser extraídas, com o intuito de analisar o comportamento de usuários fictícios da plataforma frente aos exercícios realizados. A abordagem metodológica foi planejada para alinhar recursos tecnológicos às necessidades reais do público-alvo, considerando tanto aspectos técnicos quanto pedagógicos.

## 5.1 Plataforma Interface

A proposta consiste na criação de um modelo de negócio inovador, baseado em Inteligência Artificial (IA) e Reconhecimento de Fala (ASR), capaz de facilitar e personalizar a intervenção entre os usuários, especialmente para crianças com necessidades específicas, como as com T21 (Síndrome de Down). A plataforma, chamada Interface, se baseia em um ciclo de interação projetado para ser intuitivo e motivador para a criança, cujo fluxo de funcionamento é apresentado na Figura24.

O aplicativo disponibiliza uma base de dados estruturada especializada contendo uma lista hierarquizada de exercícios de fala, que inclui vogais, sílabas, consoantes, palavras e frases, cuidadosamente selecionados e classificados por nível de dificuldade e idade. Além disso, o Interface conta com interação guiada, na qual o sistema apresenta o exercício de fala (o "Texto Proposto") ao usuário de forma visual e auditiva.

O objetivo da aplicação é que a criança (ou usuário) tente realizar a reprodução da fala solicitada. O aplicativo captura o áudio da tentativa em tempo real e fornece uma análise de desempenho imediatamente após a reprodução, por meio da tecnologia de Avaliação Automática da Fala (ASR), realizando uma análise detalhada da resposta. Por fim, é feita a avaliação de proximidade fonética e lexical da fala reproduzida com o que foi solicitado.

## 5.2 História e Caracterização da Agente Renata

A Agente Renata foi desenvolvida a partir de interações no ChatGPT, utilizando técnicas de Processamento de Linguagem Natural (PLN). Seu papel é atuar como um agente especializado em fonoaudiologia, responsável pela criação de exercícios de fala personalizados para crianças com Trissomia 21 (T21). Suas principais características são sumarizadas na Tabela 5.

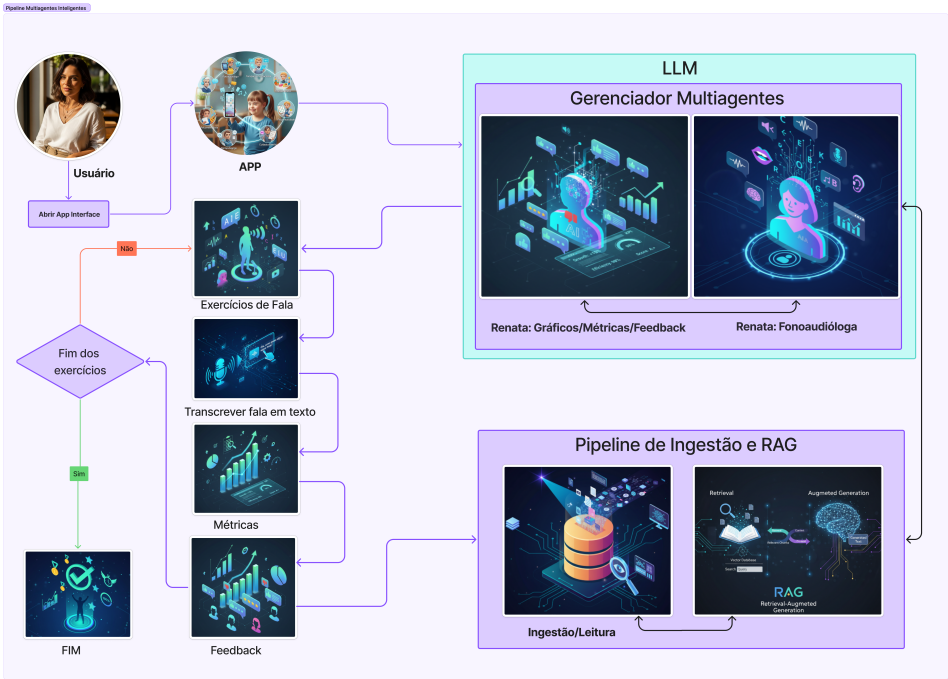


Figura 24 – Pipeline Interação Usuário » Multiagentes » Rag

Embora concebida inicialmente como um único perfil, a Agente Renata evoluiu para desempenhar o papel de um sistema de multiagentes integrados. Mais do que um componente tecnológico, foi projetada como mediadora que combina rigor técnico com proximidade humana, transmitindo empatia, clareza e estímulo positivo durante as interações, ao mesmo tempo em que mantém precisão na análise dos dados coletados. Renata atua em três papéis principais no fluxo do sistema:

- **Facilitadora de Interação**: conduz atividades de fala de maneira lúdica e adaptada à faixa etária, garantindo que a criança compreenda as instruções e sinta-se motivada.

- **Avaliadora de Progresso**: registra e analisa respostas, aplicando métricas de análise de desempenho baseadas em aproximidade para identificar avanços e pontos de atenção.

- **Provedora de Feedback**: retorna orientações imediatas de forma positiva e construtiva, reforçando conquistas e sugerindo pequenos ajustes no aprendizado.

Sua concepção valoriza princípios éticos e de responsabilidade. Embora Renata ofereça apoio para acompanhamento e análise, **ela não substitui diagnósticos clínicos realizados por profissionais da saúde.** Seu papel é complementar, fornecendo informações úteis e confiáveis para pais, professores e fonoaudiólogos, sempre dentro dos limites técnicos estabelecidos.

Tabela 5 – Resumo das Características e Funções da Agente Renata

| Categoria | Descrição |
|---|---|
| **Origem** | Desenvolvida no ICMC-USP, integrando linguística computacional, inteligência artificial e fonoaudiologia. |
| **Missão** | Apoiar o desenvolvimento da fala de crianças com T21, promovendo autonomia, autoestima e inclusão social. |
| **Formação** | Bacharelado em Fonoaudiologia (ênfase em linguagem infantil); Pós-graduação em Educação Inclusiva e Tecnologias Assistivas; Mestrado em Neurociências (aquisição da linguagem e plasticidade cerebral); Doutorado (PhD) em Psicolinguística aplicada ao desenvolvimento da fala em T21. |
| **Idiomas** | Português (principal); inglês (avançado); espanhol (intermediário); LIBRAS (intermediário). |
| **Papéis** | Facilitadora de interação; Avaliadora de progresso; Provedora de feedback. |
| **Competências** | Atuação em fonoaudiologia infantil e intervenção precoce; neurociência da linguagem e psicolinguística; educação inclusiva e aplicação de tecnologias assistivas; processamento de linguagem natural e reconhecimento de fala (Whisper, TTS, LangChain); análise de dados para diagnóstico (métricas de WER, CER e evolução por sessão); além de comunicação empática, acessível e motivadora com crianças e cuidadores. |
| **Funções-Chave** | Avaliação inicial; Acompanhamento de desempenho; Proposição de atividades; Orientação a cuidadores e educadores; Integração multimodal com tecnologias. |
| **Ética e Limitações** | Não substitui diagnósticos clínicos; atua como apoio complementar, respeitando privacidade e individualidade da criança. Utiliza linguagem positiva, segura e inclusiva, bloqueando termos ofensivos ou inadequados. Mantém transparência sobre seu caráter digital e fornece feedback sempre construtivo. |

## 5.3 Júlia, a nossa modelo para essa pesquisa

Júlia é uma menina de 4 anos com Síndrome de Down, dona de uma forma muito particular de enxergar e aprender. Ela tem um sorriso que ilumina o ambiente e uma curiosidade que a leva a explorar o mundo ao seu redor do seu jeito. Seu irmão mais velho, com 13 anos, é seu maior amigo e cúmplice em todas as aventuras, incentivando-a a cada pequena descoberta.

Seus pais, com todo amor e dedicação, buscam incansavelmente os melhores caminhos para o desenvolvimento e independência de Júlia. O desafio da comunicação, em especial, motivou a busca por algo que fosse além das abordagens tradicionais e que trouxesse diversão para o aprendizado.

É aí que entra a Agente Renata, uma assistente digital que não oferece apenas atividades, mas brincadeiras e conversas sob medida para o ritmo de Júlia. Essa parceria, apoiada de perto pelos pais e pela fonoaudióloga, não é sobre a tecnologia em si, mas sobre a vida: promover autonomia, qualidade de vida e, acima de tudo, abrir caminhos para que a Júlia se expresse e seja incluída socialmente.

## 5.4 Geração da massa de dados

Para avaliar a viabilidade de construção do MVP do aplicativo, foi elaborada uma massa de dados sintética que simula o uso da ferramenta por crianças com Trissomia 21. Nessa simulação, o aplicativo apresenta palavras em formato de áudio e texto, que devem ser ouvidas e reproduzidas pelo usuário. O áudio do usuário é coletado, e avaliado. O objetivo é que a reprodução constante de diversas palavras contribua para o desenvolvimento da fala do usuário. A estruturação da massa de dados foi orientada pela experiência profissional atribuída à Agente Renata, integrando conhecimentos de fonoaudiologia infantil e psicolinguística aplicada.

No processo de geração dos dados fictícios, foram considerados fatores como a idade, o nível de dificuldade das atividades e os tipos de erros fonológicos mais prováveis. A partir dessa interação simulada, são calculadas métricas de proximidade entre o enunciado solicitado e a reprodução gerada pela criança, permitindo mensurar a acurácia da fala e produzir diferentes formas de feedback. Com isso, torna-se viável realizar uma avaliação preliminar da evolução do usuário na utilização do sistema, mesmo sem dados reais, criando as condições necessárias para testar a arquitetura, validar funcionalidades e apoiar o desenvolvimento do protótipo.

Após a geração dos dados sintéticos, os áudios foram transcritos, segmentados em palavras e fonemas, e normalizados para garantir a consistência das análises.

## 5.5 Coleta e Simulação de Dados

A base de dados sintética criada, capaz de representar de forma aproximada o cenário de uso real por crianças com Trissomia 21, foi concebida de modo a contemplar diferentes dimensões relevantes ao processo de ensino e aprendizagem da fala, incorporando elementos de personalização, variação de complexidade e ocorrência de erros esperados na produção oral. A base resultante contém:

- Perfis fictícios de crianças (com UUID para anonimização).
- Exercícios de consoantes, vogais, sílabas, palavras e frases.
- Níveis de dificuldade (*fácil*, *médio* e *difícil*).

- Simulação de erros fonológicos, estruturais e semânticos (omissão, substituição e inserção).

Foi implementado um modelo probabilístico para atribuir um fator de influência a cada exercício, calculado como a combinação de um termo base, dependente do grau de dificuldade, com multiplicadores que refletem a complexidade do tipo de exercício, a idade da criança, o efeito da repetição e o desempenho histórico. Formalmente, o fator $F$ foi modelado por:

$$F = \text{clip}\left(U(\alpha_d, \beta_d) \cdot w_t \cdot M_{idade} \cdot M_{rep} \cdot M_{hist},\ 0, 1\right)$$

onde:

- $U(\alpha_d, \beta_d)$ é uma variável aleatória uniforme no intervalo definido pelo grau de dificuldade $d$.
- $w_t$ é o peso do tipo de exercício $t \in \{\text{vogal, sílaba, consoante, palavra, frase}\}$.
- $M_{idade}$ é o multiplicador da idade (crianças mais novas têm maior probabilidade de erro).
- $M_{rep}$ é o multiplicador de repetição (tentativas sucessivas reduzem erro).
- $M_{hist}$ é o multiplicador de histórico de desempenho (quanto maior a acurácia anterior, menor a influência de erro).

Os pesos atribuídos aos diferentes tipos de exercícios foram definidos para refletir sua complexidade fonológica e cognitiva.

- Vogal: 0,70 (unidade fonética mais simples).
- Sílaba: 0,90 (combinação básica de fonemas).
- Consoante: 1,00 (maior exigência articulatória).
- Palavra: 1,10 (demanda lexical - conjunto de palavras que uma pessoa conhece e utiliza - e semântica).
- Frase: 1,25 (integração sintática e memória de trabalho verbal).

Embora as bibliotecas *Whisper* e *LangChain* não tenham sido utilizadas diretamente na geração dos dados sintéticos, a metodologia foi estruturada para simular o processamento que essas ferramentas realizariam em um cenário real, permitindo que a massa sintética

represente de forma realista o comportamento esperado dos módulos de transcrição e de raciocínio/decisão integrados à persona clínico-pedagógica da Agente Renata. A Tabela 6 apresenta a síntese das principais bibliotecas em Python utilizadas para a geração e cálculos de métricas da base de dados sintética.

Tabela 6 – Principais bibliotecas e módulos utilizados na geração da massa de dados

| **Biblioteca / Módulo** | **Descrição** |
|---|---|
| pandas | Usada para a manipulação e análise de dados. Permite criar, ler e editar tabelas (DataFrames) que organizam a massa de dados sintética. |
| numpy | Focada em computação numérica e matemática. Útil para gerar arrays e realizar operações de alto desempenho, como criação de valores aleatórios ou cálculos estatísticos. |
| Faker | Principal biblioteca para geração de dados falsos e realistas, como nomes, endereços, e-mails e números de telefone. Fundamental para criar informações pessoais e de cadastro. |
| random | Módulo integrado do Python que gera números e elementos aleatórios. Usado para simular aleatoriedade de eventos, selecionar itens de uma lista ou introduzir variações nos dados. |
| datetime e timedelta | Módulos do Python para trabalhar com datas e horas. *datetime* cria e manipula objetos de data e tempo, enquanto *timedelta* calcula diferenças entre datas ou adiciona/subtrai intervalos de tempo. |
| Levenshtein (distance) | Calcula a distância de Levenshtein entre duas strings, medindo diferenças literais (edições, inserções ou deleções). Útil para simular erros de digitação ou variações textuais. |
| jiwer | Biblioteca específica para avaliar a qualidade de transcrições de fala. Calcula métricas como WER (Word Error Rate) e CER (Character Error Rate), comparando transcrição gerada com a referência correta. |
| time | Módulo para trabalhar com tempo. Usado para medir tempo de execução ou adicionar atrasos, relevante em simulações. |
| tqdm | Cria barras de progresso para loops longos, visualizando o andamento da geração da massa de dados. |
| uuid | Gera identificadores universais únicos. Ideal para criar IDs exclusivos para cada registro, garantindo ausência de duplicatas. |

## 5.6 Métricas de Avaliação

A avaliação da qualidade de um sistema de transcrição automática de fala (STT) é fundamental para garantir a eficácia da solução. No desenvolvimento deste projeto, a definição das métricas de avaliação foi um aspecto crucial. Embora existam diversas métricas e ferramentas disponíveis, cada uma com objetivos específicos, tornou-se necessário estabelecer um escopo adequado para as avaliações do Trabalho de Conclusão de Curso (TCC) e do Produto Mínimo Viável (MVP).

Para a avaliação da aplicação, adotamos a biblioteca `jiwer`, que utiliza o Word Error Rate (WER) e o Character Error Rate (CER) para avaliação da qualidade da transcrição automática da fala.

A métrica Word Error Rate (WER) indica a integridade da transcrição realizada pelo sistema, comparando o texto gerado automaticamente com uma frase de referência considerada correta. Ela é calculada conforme a Equação 5.1, na qual WER = 0 indica transcrição perfeita e WER = 1 indica erro total:

$$\text{WER} = \frac{S + I + D}{N} \tag{5.1}$$

- **S (Substituições)**: palavras trocadas (ex.: “gato” por “mato”).
- **I (Inserções)**: palavras adicionadas que não existiam na frase original.
- **D (Deleções)**: palavras omitidas.
- **N**: número total de palavras na frase de referência.

Já a métrica CER avalia erros no nível dos caracteres, capturando detalhes finos como erros ortográficos ou omissões parciais. A métrica CER é calculada conforme a Equação 5.3, onde $S_c$, $I_c$ e $D_c$ são substituições, inserções e deleções de caracteres, e $N_c$ é o total de caracteres na referência.

$$\text{CER} = \frac{S_c + I_c + D_c}{N_c} \tag{5.2}$$

Também é calculada a Taxa de Aproximidade, que consiste no grau de correspondência entre a fala proposta e a fala registrada.

$$\text{Proximidade} = Max(0, 1 - \frac{S_c + I_c + D_c}{N_c}) \tag{5.3}$$

Nos casos em que CER ou WER excedem 1 (valores anômalos, mas possíveis em tarefas curtas com muitas inserções), optamos por tratar esses resultados como inadequados para análise clínica, já que indicam uma discrepância significativa entre o texto proposto e o falado. A métrica também pode assumir valores negativos, que são interpretados como falha grave e categorizados separadamente.

# 6 RESULTADOS PRELIMINARES DOS DADOS SIMULADOS

Este capítulo apresenta os resultados obtidos a partir dos dados simulados da aplicação, que integra modelos de linguagem e agentes inteligentes para apoiar o ensino da fala de crianças com Trissomia 21. A proposta é concebida como um modelo de negócio voltado ao desenvolvimento do sistema Interface, que busca disponibilizar agentes inteligentes por meio de interfaces intuitivas, capazes de proporcionar uma experiência de uso satisfatória e de evolução para os usuários envolvidos. Nos tópicos seguintes, serão apresentados maiores detalhes sobre essa solução.

Como porposta de MVP, nosso modelo experimental é a Agente Renata, integrando princípios de fonoaudiologia infantil, neurociência da linguagem, psicolinguística aplicada e educação inclusiva, promovendo autonomia, autoestima e inclusão social. Todo modelo experimental se baseia em ligua portuguesa PT-BR.

A agente Renata foi simulada por meio de um avatar que garante acolhimento e empatia e aproveitando características como coragem, curiosidade e capacidade de motivar e engajar os outros, atributos que reforçam a proximidade e o vínculo com as crianças durante as interações. Para validar essa interação, foi utilizado um roteiro de simulação, no qual a Agente Renata interpreta métricas como WER e CER e as traduz em diferentes tipos de feedback, conforme o público-alvo. Dessa forma, a criança: recebe reforço positivo e orientações lúdicas; a fonoaudióloga recebe relatórios técnicos detalhados, padrões identificados e tendências; os responsáveis recebem feedback simplificado em linguagem cotidiana e gráficos de desempenho ao longo do período.

A simulação possibilita uma interação personalizada e acessível a todos os envolvidos, equilibrando suporte técnico e incentivo emocional. Além disso, permite propor diferentes formas de visualização do desempenho dos usuários, bem como acompanhar e mensurar sua evolução ao longo do tempo.

## 6.1 Critérios de Classificação de Desempenho

A avaliação do desempenho das crianças e do sistema segue faixas pré-estabelecidas, com base em percentuais de erros ou acertos, e é apresentada de forma visual por cores para facilitar a interpretação. A Tabela 8 detalha os níveis de desempenho para WER e CER, indicando a faixa de erros correspondente a cada nível. A Tabela 8 apresenta as faixas de acerto, níveis e cores apresentadas aos usuários e seus responsáveis.

As tabelas permitem uma interpretação rápida e intuitiva do desempenho do usuário, tanto do ponto de vista técnico quanto do acompanhamento pedagógico.

Quanto aos indicadores **WER (Word Error Rate)** e **CER (Character Error**

Tabela 7 – Classificação dos Níveis de Desempenho para WER, CER e Proximidade para Profissionais

| **Nível** | **Faixa de Erros (%)** | **Cor** |
|---|---|---|
| Excelente | 0.0 a 0.2 | Verde Escuro |
| Muito Bom | 0.2 a 0.5 | Verde Claro |
| Atenção | 0.5 a 0.8 | Amarelo |
| Crítico | 0.8 a 1.0 | Vermelho |
| Inadequado | $> 1.0$ | Azul |

Tabela 8 – Classificação dos Níveis de Desempenho para Crianças e Responsáveis

| **Nível** | **Faixa de Acertos (%)** | **Cor** |
|---|---|---|
| Excelente | 80 a 100 | Verde Escuro |
| Muito Bom | 50 a 80 | Verde Claro |
| Atenção | 20 a 50 | Amarelo |
| Precisamos brincar mais | 0 a 20 | Vermelho |
| Vamos brincar novamente | $< 0$ | Azul |

**Rate)**, o nível *Inadequado* ocorre quando a taxa de erro ultrapassa 100%. Isso acontece quando o texto falado contém mais operações (substituições, inserções e deleções) do que o número de palavras/caracteres da referência, caracterizando um desvio total da tarefa proposta (ex.: fala completamente diferente ou com excesso de palavras adicionais).

Já na Taxa de Aproximidade, os valores são normalizados no intervalo $[0, 1]$, de modo que o nível *Inadequado* corresponde apenas a resultados igual a 0 %, porém com wer maior que 100%, indicando ausência de correspondência entre a fala e o estímulo proposto.

## 6.2 Exemplos de Aplicação: Cartões de Desempenho

Foram gerados cartões gráficos para representar visualmente os níveis de desempenho, facilitando a interpretação dos resultados e o acompanhamento pedagógico. Cada cartão indica a faixa de desempenho por cor e exemplos de erros típicos dos usuários:

- **Crítico (Vermelho)** – erros graves, como alterações significativas na frase (ex.: “a casa é grande” → “caça é”).
- **Inadequado (Azul)** – troca total de palavra, indicando que o enunciado original foi substituído por outro sem relação (ex.: “sapato” → “elefante”).
- **Atenção (Amarelo)** – alterações estruturais ou reorganização das palavras (ex.: “mamãe quero brincar” → “mamãe banana quero”).
- **Bom (Verde Claro)** – desempenho satisfatório, com pequenos desvios que não comprometem a compreensão.
- **Excelente (Verde Escuro)** – reprodução correta e completa da frase, sem erros.

Dessa forma, os cartões fornecem feedback imediato e intuitivo tanto para crianças quanto para fonoaudiólogas e responsáveis.

## 6.3 Visualizações Complementares

Para apoiar a interpretação dos resultados e facilitar o acompanhamento do progresso, foram desenvolvidas visualizações complementares que sintetizam diferentes aspectos do desempenho das crianças:

- **Painel de Cartões de Desempenho** – oferece uma visão geral do progresso da criança, destacando níveis de acertos e áreas que precisam de atenção.
- **Nuvem de Palavras** – ilustra a frequência de palavras corretamente reproduzidas (verde) e erros cometidos (laranja), permitindo identificar padrões de acerto e dificuldade.
- **Evolução Mensal (WER)** – gráfico de linha que mostra a evolução do desempenho ao longo do tempo, evidenciando melhorias ou retrocessos.
- **Resumo para Pais** – gráfico simplificado por zonas de cores (verde, amarelo, vermelho), fornecendo uma visão rápida e intuitiva do desempenho da criança.

Essas visualizações complementares tornam o acompanhamento mais intuitivo e permitem que crianças, familiares e profissionais interpretem os resultados de forma rápida e precisa.

## 6.4 Mapa de Calor indicando Proximidade

A Figura 25 apresenta o mapa de calor da proximidade média entre fala proposta pelo aplicativo e a fala registrada (o que a criança reproduziu), agrupada por tipo de exercício e período de coleta, permitindo que o usuário ou responsável seja capaz de identificar padrões de evolução temporal. O mapa de calor, criado a partir de interações da criança Júlia, destaca que:

- Houve estabilidade em **palavras** e **vogais**, com médias próximas de 1.0.
- **Sílabas** tiveram variação entre os períodos, refletindo alternância entre acertos e erros típicos.
- **Frases** e **consoantes** mostraram maior dispersão, reforçando os desafios nessas categorias.
- **Quadrantes na cor cinza** mostraram que não houve atividades para o período ou não houve a prática aplicada pela criança, seja por qualquer motivo.

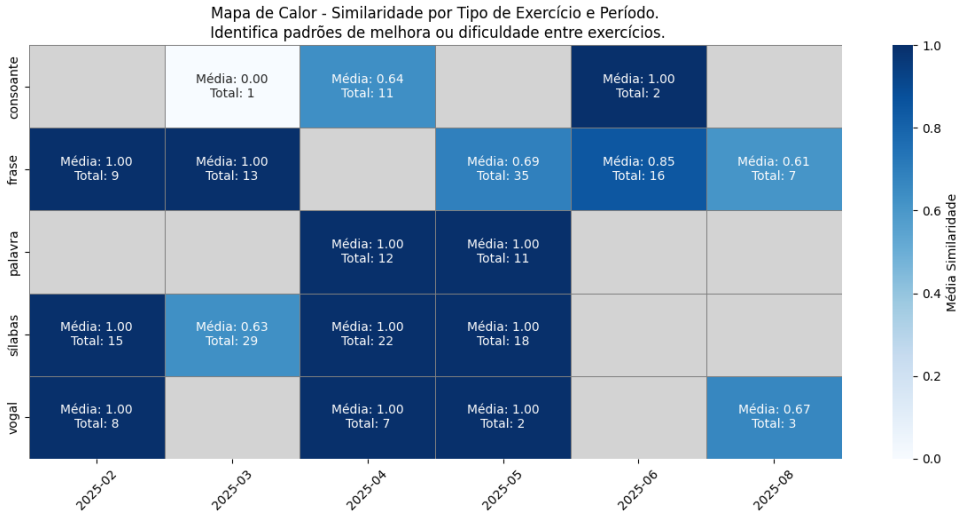


Figura 25 – Mapa de Calor por Aproximidade
Fonte: Elaborado pelo autor.

## 6.5 Mapa de Calor por CER

O gráfico 26 mostra o comportamento do CER (Character Error Rate) médio para o usuário considerando diferentes tipos de exercícios realizados ao longo dos meses de fevereiro a agosto de 2025. Neste gráfico, que representa a evolução da criança Júlia, é possível:

- Identificar padrões de melhoria ou dificuldade nos diferentes tipos de exercícios (consoante, frase, palavra, sílabas e vogal) ao longo do tempo.
- Notar que, em **consoantes**, houve uma dificuldade inicial em março, com CER médio de 6,00 em apenas um exercício, seguido de redução expressiva nos meses seguintes (1,18 em abril, chegando a 0 em maio e junho). Nos exercícios de **frases**, o desempenho foi estável em fevereiro e março (CER médio 0), mas apresentou valores médios em abril (0,31), maio (0,14) e aumento em agosto (0,45). Para **palavras**, o CER médio foi consistentemente 0 em todos os meses analisados, indicando domínio nessa categoria. Nos exercícios de **sílabas**, em março o CER médio foi de 0,36 (29 exercícios), caindo para 0 nos meses seguintes, o que mostra recuperação rápida. Já nos exercícios de **vogais**, o desempenho foi perfeito entre fevereiro e maio (CER médio 0), com uma leve queda em agosto (CER médio 0,33 em 3 exercícios).

Os resultados evidenciam que Júlia apresenta maior estabilidade em palavras e sílabas, enquanto frases e vogais demonstram oscilações que merecem acompanhamento. O caso de consoantes mostra clara evolução, superando a dificuldade inicial. Desta forma, a aplicação pode recomendar reforçar práticas em frases e vogais, auxiliando responsáveis a investigar os fatores que levaram às quedas registradas em agosto.

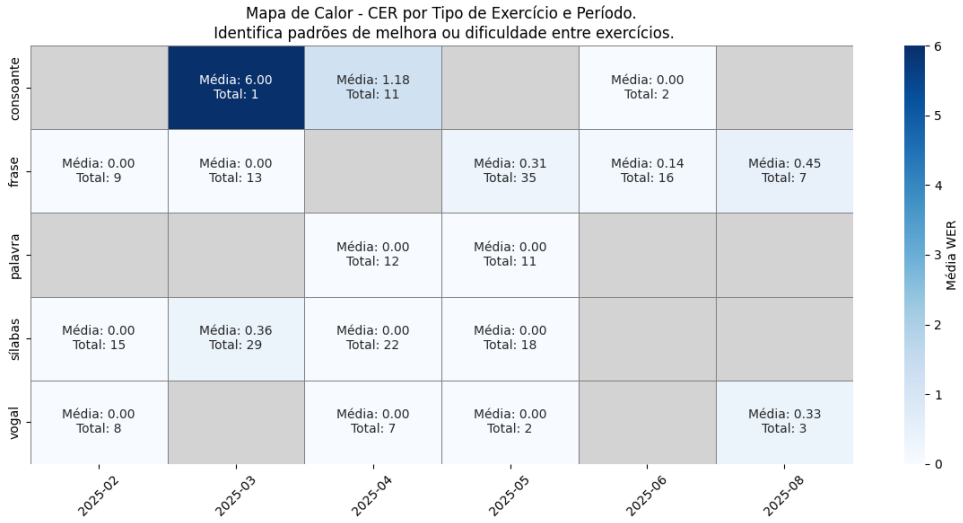


Figura 26 – Mapa de Calor por CER
Fonte: Elaborado pelo autor.

## 6.6 Acurácia da Taxa Geral de Sucesso nos Exercícios

Na Figura 27 é apresentada a proporção de exercícios que Júlia concluiu com sucesso versus aqueles que resultaram em erros. O objetivo deste gráfico é oferecer uma visão geral e imediata do desempenho global de Júlia.

Com uma taxa de sucesso de 78,73%, o sistema obteve 174 acertos em um total de 221 exercícios, o que indica um desempenho geral positivo. Esse resultado oferece um panorama inicial consistente e objetivo, servindo como base sólida para análises mais aprofundadas ao longo do relatório.

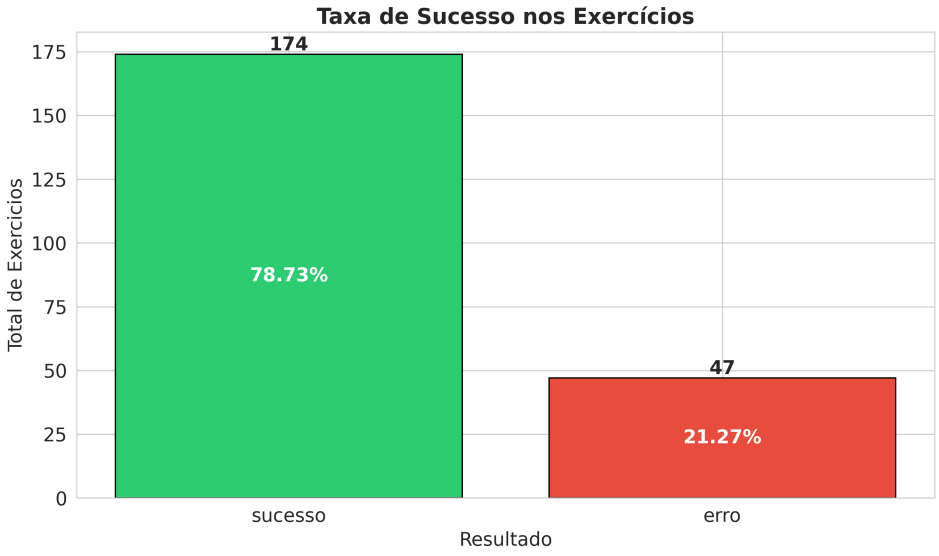


Figura 27 – Acurácia global
Fonte: Elaborado pelo autor.

## 6.7 Acurácia por Tipo de Exercício

Na Figura 28 é apresentada a comparação da acurácia de Júlia em cinco tipos distintos de exercícios: palavras, vogais, sílabas, consoantes e frases. O objetivo deste gráfico é identificar as áreas de maior e menor desempenho. Essas informações podem ser relevantes para direcionar o foco da terapia do usuário, por exemplo.

Neste gráfico, é possível observar que Júlia demonstra um desempenho Excelente (acima de 80%) em exercícios de palavras, vogais e sílabas. No entanto, a acurácia cai para a faixa de "Bom" (50%-80%) em exercícios de consoantes e frases, indicando que estas são as áreas que exigem mais atenção e prática. O gráfico orienta para a tomada de decisões terapêuticas, sugerindo a necessidade de focar em exercícios que combinam consoantes e frases para melhorar a fluência e a precisão.

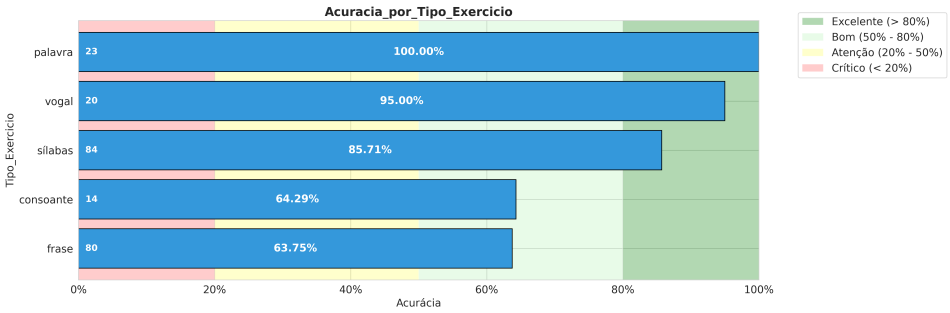


Figura 28 – Acurácia por tipo de exercício
Fonte: Elaborado pelo autor.

## 6.8 Acurácia por Repetição

Na Figura 29 é apresentada a análise da acurácia de Júlia em função do número de repetições realizadas em cada exercício. O objetivo é avaliar a relação entre a quantidade de repetições e a precisão do desempenho.

Por meio do gráfico, é possível observar que a acurácia de Júlia melhora de forma consistente com o aumento das repetições. Inicialmente, em exercícios sem repetição, o desempenho foi de 73,68%. À medida que o número de repetições aumenta, a acurácia evolui progressivamente, atingindo 100% quando os exercícios são repetidos 12 vezes ou mais. Esses resultados evidenciam que a prática contínua é um fator determinante para o processo de aprendizado.

A personalização do plano terapêutico sugere que a terapia de Júlia pode se beneficiar da inclusão de um maior número de repetições, especialmente ao enfrentar novos desafios.

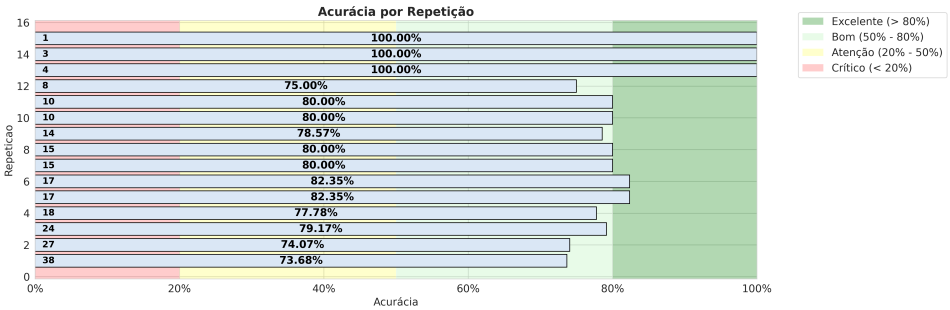


Figura 29 – Acurácia em função do número de repetições
Fonte: Elaborado pelo autor.

## 6.9 Acurácia por Tipo de Exercício

O gráfico de linha apresentado na Figura 30 mostra a evolução da acurácia média de Júlia em todos os exercícios, considerando a variação mês a mês. O objetivo deste gráfico é acompanhar o progresso geral ao longo do tempo e identificar possíveis tendências no desempenho.

Observa-se melhora significativa da acurácia entre fevereiro (100%) e abril (92,31%). Entretanto, ocorreram quedas relevantes em maio (68,18%) e, de forma ainda mais acentuada, em agosto (50%). As setas presentes no gráfico destacam claramente os períodos de alta e baixa no desempenho. As reduções observadas nos meses de maio e agosto representam pontos de atenção, indicando a necessidade de investigar possíveis causas e identificar quais exercícios foram realizados nesses períodos para compreender os fatores associados à diminuição do desempenho.

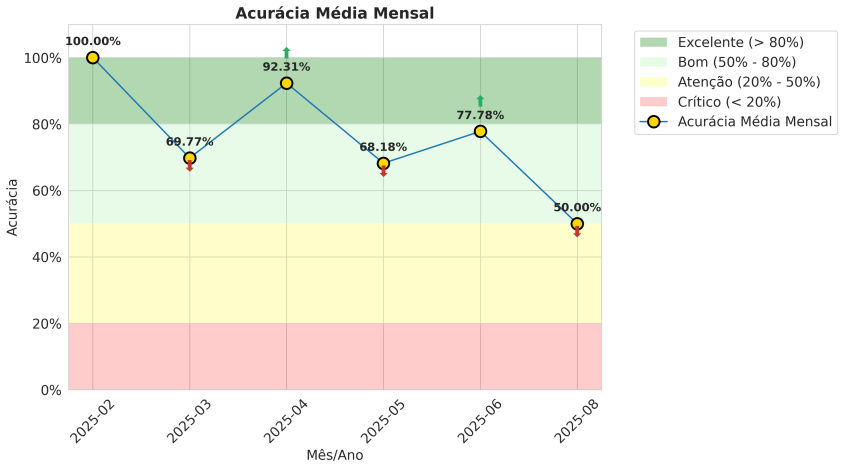


Figura 30 – Evolução mensal da acurácia nos exercícios
Fonte: Elaborado pelo autor.

## 6.10 Acurácia Métrica (WER)

O gráfico apresentado na Figura 31 mostra a taxa de erro de palavras (WER) de Júlia ao longo dos meses, configurando-se como uma métrica complementar à acurácia. O objetivo deste gráfico é apresentar uma perspectiva alternativa sobre o progresso, com foco na taxa de erro. Observa-se uma melhora significativa no desempenho entre março (30,23% de erro) e abril (7,69%). No entanto, a taxa de erro aumentou de forma expressiva em agosto (41,53%), confirmando a tendência de queda já identificada no gráfico de acurácia.

O crescimento da taxa de erro em agosto requer atenção especial, indicando a necessidade de revisar e ajustar o plano terapêutico para reverter a tendência negativa observada.

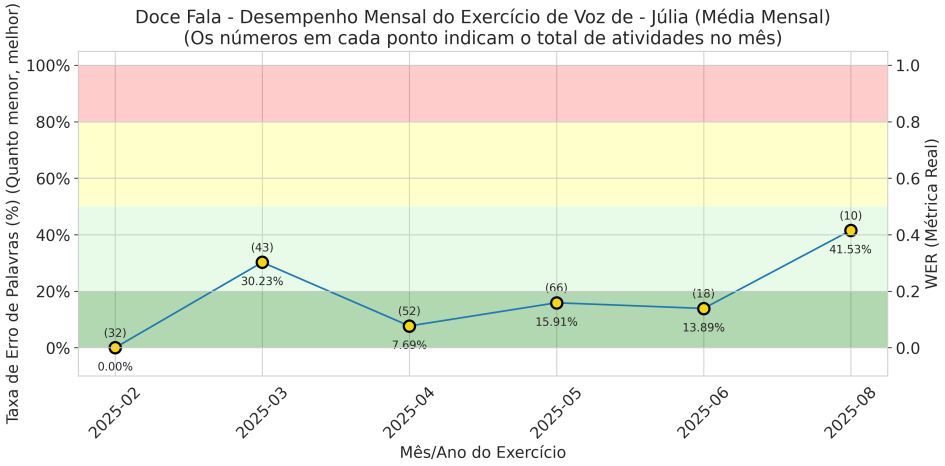


Figura 31 – Desempenho por palavras (WER)
Fonte: Elaborado pelo autor.

## 6.11 Evolução mensal da Acurácia por tipo de exercício - Consoantes

O gráfico apresentado na Figura 32 mostra a evolução da acurácia de Júlia em exercícios que focam em consoantes ao longo do tempo.

O objetivo deste gráfico é rastrear o progresso em um tipo de exercício específico. Nele é possível observar uma progressão significativa no desempenho. Em março de 2025, Júlia iniciou com 0% de acurácia, possivelmente por se tratar de exercícios novos ou desafiadores. Em abril, sua acurácia evoluiu para 63,64%, atingindo a faixa *Bom*, e, em junho, alcançou 100%, entrando na faixa *Excelente*. Essa trajetória evidencia uma melhora notável e consistente, representando um exemplo expressivo de progresso e superação de desafios.

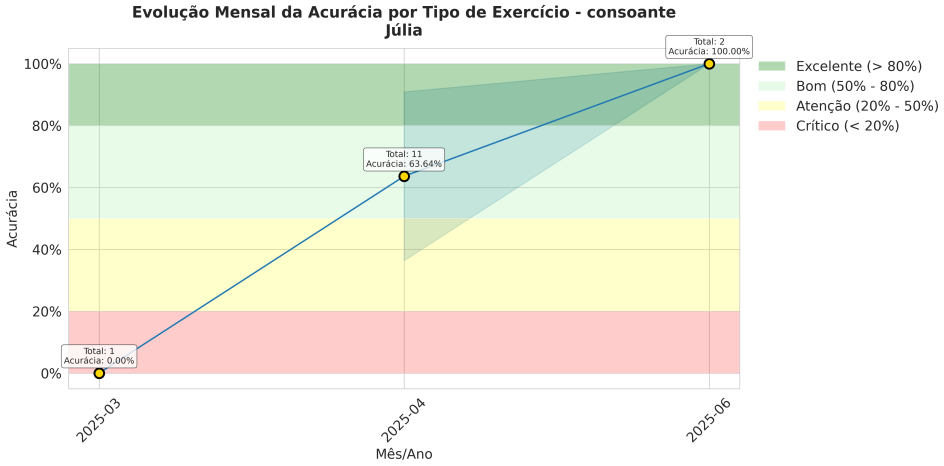


Figura 32 – Desempenho em Consoantes
Fonte: Elaborado pelo autor.

## 6.12 Evolução mensal da Acurácia por tipo de exercício - Frases

O gráfico 33 evidencia a acurácia de Júlia em exercícios de frases, mostrando as flutuações de desempenho ao longo dos meses. O gráfico tem como objetivo o acompanhamento de desempenho em exercícios que exigem a articulação de sentenças completas.

O desempenho iniciou de forma excelente, com 100% de acurácia em fevereiro e março de 2025. Observou-se, entretanto, queda acentuada em maio (40%), posicionando o resultado na faixa de *Atenção*. Em junho, houve recuperação para 75% (faixa *Bom*), mas em agosto a acurácia voltou a cair, atingindo 42,86%. Essas oscilações indicam a necessidade de investigar fatores como complexidade das frases, contexto emocional ou consistência da prática.

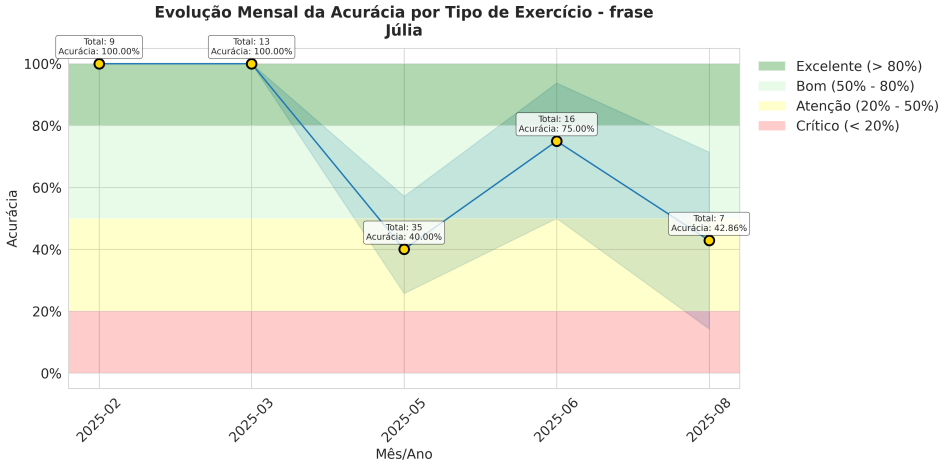


Figura 33 – Desempenho em Frases
Fonte: Elaborado pelo autor.

## 6.13 Evolução mensal da Acurácia por tipo de exercício - Palavras

A Figura apresentada no gráfico 34 mostra o desempenho de Júlia em exercícios de palavras no período de abril a maio. O objetivo é rastrear o desempenho em um tipo específico de exercício. Neste gráfico, o desempenho de Júlia foi perfeito, mantendo 100% de acurácia em ambos os meses analisados, evidenciando domínio total e indicando que está preparada para desafios mais complexos.

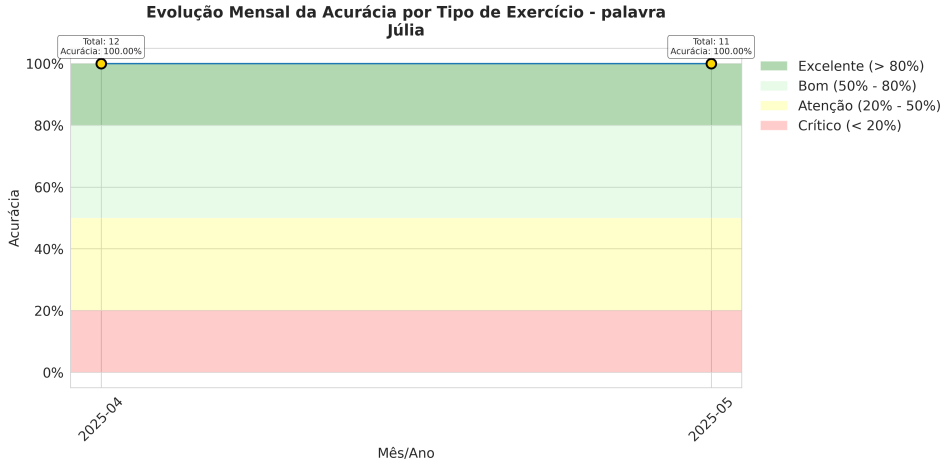


Figura 34 – Desempenho em Palavras
Fonte: Elaborado pelo autor.

## 6.14 Evolução mensal da Acurácia por tipo de exercício - Sílabas

O gráfico apresentado na Figura 35 detalha a acurácia de Júlia em exercícios de sílabas ao longo do tempo. O objetivo é rastrear o progresso em um tipo de exercício específico. Neste gráfico, em fevereiro, a usuária Júlia obteve 100% de acurácia. Em março, observou-se queda para 58,62%. Apesar dessa redução, houve recuperação total nos meses de abril e maio, indicando que a dificuldade foi temporária e superada com a continuidade das práticas.

## 6.15 Evolução mensal da Acurácia por tipo de exercício - Vogais

O gráfico apresentado na Figura 36 mostra a evolução da acurácia de Júlia em exercícios de vogais. O objetivo é acompanhar o progresso em um tipo específico de exercício. Neste gráfico, a acurácia manteve-se em 100% de fevereiro a maio, porém caiu para 66,67% em agosto. Essa redução sugere que a dificuldade pode estar associada ao tipo de exercício ou a fatores externos, sendo importante analisar detalhadamente o contexto desse período.

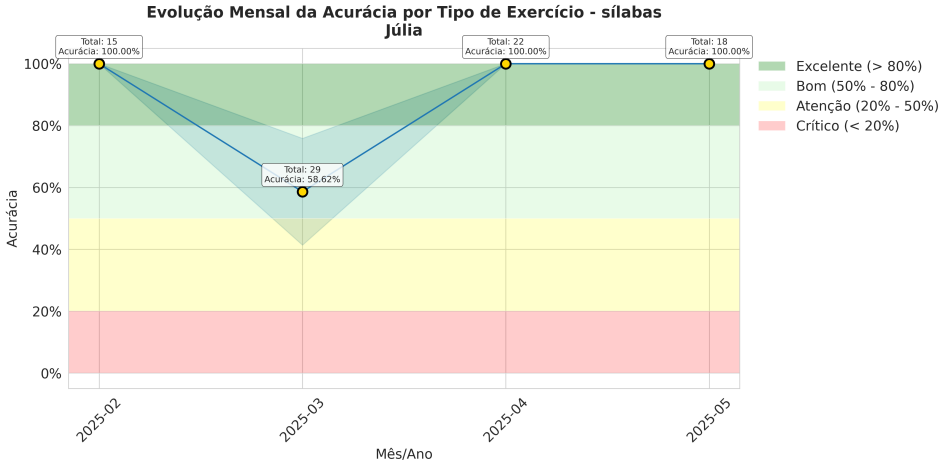


Figura 35 – Desempenho em Sílabas
Fonte: Elaborado pelo autor.

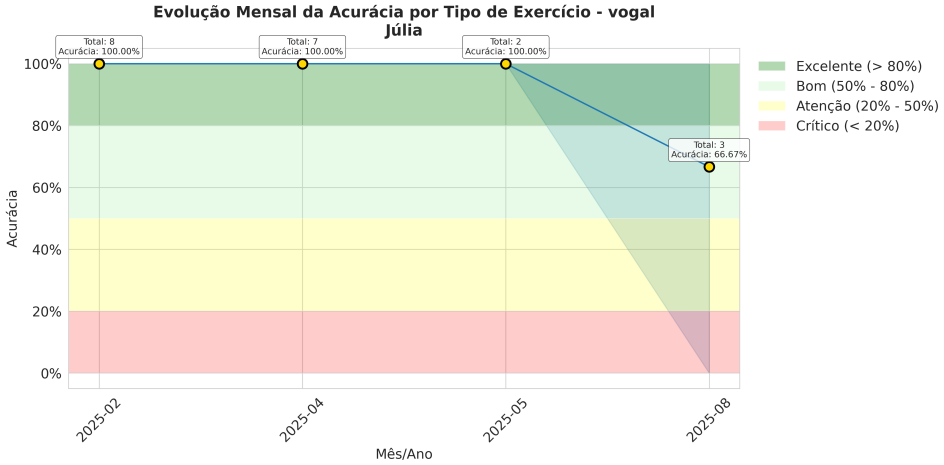


Figura 36 – Desempenho em Vogais
Fonte: Elaborado pelo autor.

## 6.16 Discussão sobre a interpretação dos Resultados

As métricas avaliadas indicam que exercícios de **palavras** e **vogais** atingiram *nível Excelente*, sugerindo consolidação fonológica, ou seja, produção correta dos sons da fala.

Enquanto **sílabas** se mantiveram próximas do nível *Muito Bom*, mas com oscilação temporais, **Consoantes** e **frases** apresentaram maiores desafios, classificadas como *Bom*, o que mostra uma boa evolução, porém podemos observar que teve uma queda considerável no volume de exercícios. O nível *Inadequado* (WER/CER >100%) foi simulado em cenários extremos, como falas totalmente diferentes do estímulo proposto, mas não ocorreu nas médias gerais.

# 7 CONCLUSÕES

A realização deste estudo demonstrou o potencial transformador da **inteligência artificial com multiagentes** no apoio ao desenvolvimento da fala de crianças com Síndrome de Down (T21). Por meio do uso de ferramentas de IA acessíveis, foi possível simular padrões de fala e orquestrar fluxos conversacionais complexos, oferecendo uma base sintética conceitual sólida para futuras aplicações clínicas, educacionais e familiares.

Os agentes foram desenvolvidos como um *protótipo conceitual*, destinado a simplificar a interação com a arquitetura multiagente complexa e permitir a visualização de seu potencial prático. A validação do modelo utilizou métricas objetivas (*Word Error Rate - WER* e *Character Error Rate - CER*) e medidas de aproximação semântica, evidenciando que a arquitetura proposta é tecnicamente viável, adaptável e capaz de refletir as necessidades individuais de cada criança.

Este trabalho reforça que **tecnologias aplicadas de forma ética e responsável** podem se tornar instrumentos poderosos de inclusão social. Além de comprovar a viabilidade técnica, os resultados demonstram como abordagens computacionais podem apoiar cuidadores, educadores e profissionais da saúde, identificando tendências, padrões e recomendações personalizadas de exercícios.

## 7.1 Próximos Passos e Expansão do Produto para Diferentes Usuários

Como próximos passos para o projeto, considerando visões de usuários profissionais e institucionais, serão realizados:

- Integração com interfaces multimodais (voz, vídeo e texto) para permitir interações naturais e intuitivas com crianças, cuidadores e profissionais de apoio.
- Desenvolvimento de **dashboards interativos** voltados a fonoaudiólogos, educadores e terapeutas, possibilitando acompanhamento em tempo real do desempenho e evolução dos usuários atendidos.
- Realização de testes com usuários profissionais e instituições, permitindo análises longitudinais e ajustes nas estratégias pedagógicas e terapêuticas.
- Expansão da base de dados para abranger maior diversidade linguística, fonológica e cognitiva, garantindo robustez, generalização e aplicabilidade em diferentes contextos clínicos e educacionais.

Em síntese, este estudo representa um passo inicial na construção de soluções que combinam ciência, tecnologia e propósito social, abrindo oportunidades para ampliar a comunicação,

expressão e inclusão de crianças com Trissomia 21. A abordagem proposta serve como referência conceitual e técnica para futuras pesquisas e desenvolvimentos, reforçando a importância de sistemas de IA **acessíveis, seguros e centrados em diferentes perfis de usuário**, desde famílias até profissionais da saúde e educação.

# REFERÊNCIAS


CHAPMAN, R. S.; HESKETH, L. J. Behavioral phenotype of individuals with down syndrome. **Mental retardation and developmental disabilities research reviews**, Wiley Online Library, v. 6, n. 2, p. 84–95, 2000.

CHASE, H. **LangChain: Building Applications with Large Language Models through Composable Components**. 2022. Disponível em: https://www.langchain.com/.

EADIE, P. A. *et al.* Profiles of grammatical morphology and sentence imitation in children with specific language impairment and down syndrome. **Journal of Speech, Language, and Hearing Research**, v. 45, n. 4, p. 720–732, 2002.

FERREIRA, A. T.; LAMÔNICA, D. A. C. Comparação do léxico de crianças com síndrome de down e com desenvolvimento típico de mesma idade mental. **Revista CEFAC**, SciELO Brasil, v. 14, p. 786–791, 2012.

FFCLRP/USP. **SofiaFala**. 2019. Disponível em: https://sites.usp.br/sofiafala/#demo.

IBGE. **População com deficiência: Brasil tem 18,6 milhões de pessoas com deficiência, mostra Censo 2022**. 2022. Acesso em: 21 jul. 2025. Disponível em: https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/0a9afaed04d79830f73a16136dba23b9.pdf.

IBGE. **População Educacional PCD Censo 2022**. 2022. Acesso em: 21 jul. 2025. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/43463-censo-2022-brasil-tem-14-4-milhoes-de-pessoas-com-deficiencia.

IZACARD, G. *et al.* **Atlas: Few-Shot Learning with Retrieval-Augmented Language Models**. 2023. Disponível em: https://arxiv.org/abs/2208.03299.

LAWS, G.; GUNN, D. Relationships between reading, phonological skills and language development in individuals with down syndrome: A five year follow-up study. **Reading and Writing**, Springer, v. 15, n. 5, p. 527–548, 2002.

LEWIS, P. *et al.* **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks**. 2020. Disponível em: https://arxiv.org/abs/2005.11401.

Ministério da Educação (MEC). **Matrículas na educação especial chegam a mais de 1,7 milhão**. 2023. Disponível em: https://www.gov.br/mec/pt-br/assuntos/noticias/2024/marco/matriculas-na-educacao-especial-chegam-a-mais-de-1-7-milhao.

OpenAI. **GPT-4 Technical Report**. 2023. Disponível em: https://openai.com/research/gpt-4.

PENFIELD, W. **Homúnculo de Penfield para T21**. 1940. Disponível em: https://exemplo.com/homonculo-t21.

PEPPER, J. **Guide it-takes-two-to-talk**. 2017. Disponível em: https://www.hanen.org/programs/it-takes-two-to-talk.

RADFORD, A. *et al.* Robust speech recognition via large-scale weak supervision. *In*: PMLR. **International conference on machine learning**. [*S.l.: s.n.*], 2023. p. 28492–28518.

SHUSTER, K. *et al.* **Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection**. 2024. Disponível em: https://arxiv.org/abs/2401.00300.

THORDARDOTTIR, E. T.; CHAPMAN, R. S.; WAGNER, L. Complex sentence production by adolescents with down syndrome. **Applied psycholinguistics**, Cambridge University Press, v. 23, n. 2, p. 163–183, 2002.

VASWANI, A. *et al.* Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VICARI, S.; CASELLI, M. C.; TONUCCI, F. Asynchrony of lexical and morphosyntactic development in children with down syndrome. **Neuropsychologia**, Elsevier, v. 38, n. 5, p. 634–644, 2000.

VOICEITT. **Talkitt**. 2017. Disponível em: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1176683/full.

Voiceitt. **Talkitt: integração com Alexa**. 2023. Disponível em: https://nocamels.com/2023/08/new-web-app-gives-a-voice-to-people-with-speech-impairment/?utm_source=chatgpt.com.

# A QUESTIONÁRIO APLICADO AOS PROFISSIONAIS DE FONOAUDIOLOGIA

Este questionário foi elaborado com o objetivo de levantar percepções de profissionais de fonoaudiologia acerca dos desafios, estratégias e possibilidades de uso de Inteligência Artificial no apoio ao desenvolvimento da fala em pessoas com Trissomia 21 (T21).

## Questionário

1. Qual é a sua especialidade profissional?

( ) Fonoaudiologia ( ) Terapia Ocupacional ( ) Outro: _____

2. Qual é o número total de pacientes PCD matriculados em algum grau escolar?

- ( ) 1º Ano do Ensino Fundamental
- ( ) 2º Ano do Ensino Fundamental
- ( ) 3º Ano do Ensino Fundamental
- ( ) 4º Ano do Ensino Fundamental
- ( ) 5º Ano do Ensino Fundamental
- ( ) 6º Ano do Ensino Fundamental
- ( ) 7º Ano do Ensino Fundamental
- ( ) 8º Ano do Ensino Fundamental
- ( ) 9º Ano do Ensino Fundamental
- ( ) 1º Ano do Ensino Médio
- ( ) 2º Ano do Ensino Médio
- ( ) 3º Ano do Ensino Médio
- ( ) Curso superior incompleto
- ( ) Curso superior completo
- ( ) Pós-graduação incompleta
- ( ) Pós-graduação completa

3. Quais são hoje os maiores desafios no desenvolvimento da fala em crianças com T21 que você observa no consultório?

4. Quais estratégias ou exercícios você mais utiliza para estimular a fala e quais considera mais eficazes?

- ( ) Exercícios de sopro e respiração
- ( ) Estimulação auditiva
- ( ) Técnicas de imitação
- ( ) Exercícios de fortalecimento oral
- ( ) Jogo de imitação facial
- ( ) Leitura interativa
- ( ) Uso de brinquedos educativos
- ( ) Conversas em contexto real
- ( ) Uso de gestos e linguagem de sinais
- ( ) Treinamento de sequência de palavras
- ( ) Tecnologia assistiva (aplicativos de fala e comunicação)
- Outro: _____

5. Existe algum tipo de apoio que a família poderia oferecer em casa para reforçar o trabalho clínico?

6. Como você imagina que uma ferramenta baseada em Inteligência Artificial poderia ajudar no seu trabalho ou no apoio à criança?

- ( ) Apoio personalizado ao desenvolvimento da fala
- ( ) Acompanhamento em tempo real
- ( ) Criação de rotinas individualizadas
- ( ) Estimulação lúdica da linguagem
- ( ) Comunicação facilitada com os pais
- ( ) Treinamento em habilidades sociais

- ( ) Análise preditiva do progresso
- ( ) Suporte à equipe multidisciplinar
- ( ) Assistência na personalização de estratégias
- ( ) Todas as alternativas anteriores
- Outro: _____

7. Você vê alguma preocupação ética ou prática no uso de IA para esse público específico?

8. Quais métricas ou indicadores você acompanha no progresso do desenvolvimento da fala?

- ( ) Número de palavras/frases produzidas
- ( ) Compreensão auditiva
- ( ) Qualidade da articulação
- ( ) Variação do vocabulário
- ( ) Fluência na fala
- ( ) Formação de frases completas
- ( ) Capacidade de generalização
- ( ) Tempo de atenção e concentração
- ( ) Todas as alternativas anteriores
- Outro: _____

9. Existe algum tipo de registro (áudio, vídeo, relatórios) que você utiliza ou gostaria de utilizar?

10. De que maneira é passada a evolução e acompanhamento dos pacientes?

- ( ) Reuniões com equipe multidisciplinar
- ( ) Reuniões regulares com outros profissionais
- Outro: _____

11. Você estaria disposto(a) a colaborar como consultor(a) informal para validar protótipos ou ideias da ferramenta?

# B RESULTADOS DO QUESTIONÁRIO

As respostas fornecidas foram analisadas de forma descritiva. A seguir são apresentadas as principais métricas e representações gráficas extraídas da análise do questionário.

### Principais desafios apontados

- Déficit cognitivo associado (impacta aquisição da linguagem).
- Dificuldades respiratórias e controle do sopro.

### Estratégias mais utilizadas

Foram mencionadas nove estratégias principais, classificadas em três categorias:

- **Motoras**: sopro/respiração, fortalecimento oral, imitação facial.
- **Auditivo-linguísticas**: estimulação auditiva, imitação verbal, sequência de palavras, conversas em contexto real.
- **Sociais/lúdicas**: brinquedos educativos, gestos/linguagem de sinais.

### Métricas acompanhadas

Indicadores quantitativos e qualitativos utilizados:

- Número de palavras/frases produzidas
- Compreensão auditiva
- Qualidade da articulação
- Variação do vocabulário
- Fluência na fala
- Formação de frases completas
- Capacidade de generalização
- Tempo de atenção e concentração

### Preocupações éticas levantadas

- Dosagem do uso de tecnologias, devido ao risco de vício em telas.

- Respeito à individualidade da criança.

### Gráficos de apoio


Figura 37 – Estratégias mais utilizadas na intervenção fonoaudiológica (T21).

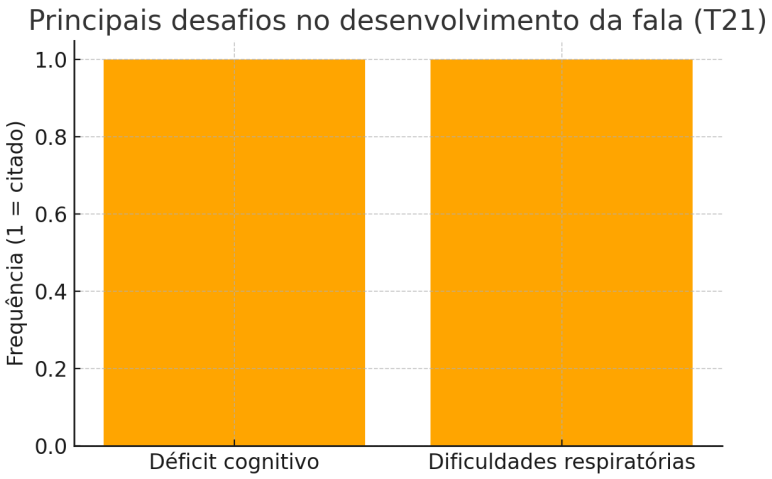


Figura 38 – Principais desafios no desenvolvimento da fala (T21).


Figura 39 – Métricas acompanhadas no desenvolvimento da fala (T21).