

# TRANSCRIÇÃO DE VÍDEOS PARA GERAÇÃO AUTOMATIZADA DE DESCRIÇÕES TEXTUAIS E ATAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

Igor Francisco Lima de Paula<sup>1</sup>

Douglas Donizeti de Castilho Braz<sup>2</sup>

## RESUMO

A pandemia de COVID-19 desencadeou uma mudança global sem precedentes em direção ao trabalho remoto e ao uso intensivo de videochamadas como meio de comunicação. Este cenário trouxe à tona novos desafios, particularmente para servidores e docentes do IFSULDEMINAS Câmpus Poços de Caldas, que enfrentam a necessidade de documentar eficientemente o conteúdo discutido durante reuniões online. Neste contexto, o presente artigo propõe uma solução inovadora utilizando Inteligência Artificial (IA) para a transcrição automatizada de vídeos, visando facilitar a criação de atas e descrições textuais de reuniões.

O trabalho se concentra no desenvolvimento de uma ferramenta de IA que não apenas atende às crescentes demandas de transcrição de vídeo, mas também supera as limitações das soluções existentes em termos de eficiência, precisão e usabilidade. A abordagem metodológica adotada foi uma análise comparativa detalhada de diversas ferramentas de IA disponíveis no mercado, enfatizando critérios como acurácia, usabilidade e custo.

Os resultados obtidos indicam que nossa ferramenta supera as soluções atuais, oferecendo uma relação custo-benefício superior e uma experiência de usuário mais intuitiva. As transcrições geradas pela nossa ferramenta são notavelmente mais precisas, facilitando a documentação e análise do conteúdo das reuniões. Este estudo contribui significativamente para o campo da IA aplicada à transcrição de vídeo, abrindo novas perspectivas para a automatização eficiente de atas e descrições textuais em ambientes educacionais e corporativos.

**Palavras-chaves:** Pandemia COVID-19, Trabalho Remoto, Transcrição de Vídeos, Inteligência Artificial

<sup>1</sup> Graduando em Engenharia de Computação, IFSULDEMINAS - Poços de Caldas. Minas Gerais. E-mail: igor.depaula@alunos.ifsuldeminas.edu.br

<sup>2</sup> Doutor em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo - SP (USP), Mestre em Ciência da Computação pela Universidade Federal de Minas Gerais - MG (UFMG) e Cientista da Computação pela Universidade Federal de Alfenas - MG (UNIFAL). E-mail: douglas.braz@ifsuldeminas.edu.br

## **ABSTRACT**

The COVID-19 pandemic triggered an unprecedented global shift towards remote work and the intensive use of video calls as a means of communication. This scenario has brought new challenges to light, particularly for the staff and faculty of IFSULDEMINAS Campus Poços de Caldas, who face the need to efficiently document the content discussed during online meetings. In this context, this article proposes an innovative solution using Artificial Intelligence (AI) for the automated transcription of videos, aiming to facilitate the creation of minutes and textual descriptions of meetings.

The work focuses on developing an AI tool that not only meets the increasing demands for video transcription but also surpasses the limitations of existing solutions in terms of efficiency, accuracy, and usability. The methodological approach adopted was a detailed comparative analysis of various AI tools available in the market, emphasizing criteria such as accuracy, usability, and cost.

The results obtained indicate that our tool surpasses current solutions, offering a superior cost-benefit ratio and a more intuitive user experience. The transcriptions generated by our tool are notably more accurate, facilitating the documentation and analysis of meeting content. This study significantly contributes to the field of AI applied to video transcription, opening new perspectives for the efficient automation of minutes and textual descriptions in educational and corporate environments.

**Keywords:** COVID-19 Pandemic, Remote Work, Video Transcription, Artificial Intelligence

## **1. INTRODUÇÃO**

A crise desencadeada pela pandemia de COVID-19 transformou radicalmente a maneira como interagimos e comunicamos em ambientes profissionais e educacionais. A transição para o trabalho remoto e o ensino à distância impulsionaram o uso intensivo de videochamadas, trazendo à tona a necessidade crítica de documentação eficiente dessas interações.

Esse novo cenário apresentou desafios específicos para o IFSULDEMINAS Câmpus Poços de Caldas, onde a captura e o registro precisos do conteúdo das reuniões online tornaram-se uma tarefa essencial.

Este estudo responde a esses desafios com uma solução inovadora: o desenvolvimento de uma ferramenta avançada de transcrição de vídeos baseada em Inteligência Artificial (IA). Esta ferramenta foi concebida para superar as limitações das abordagens de transcrição

tradicionais, oferecendo uma maneira eficiente e precisa de converter o conteúdo falado em texto.

Nossa pesquisa se concentrou em explorar o potencial da IA para otimizar este processo, visando melhorar significativamente a precisão das transcrições, a facilidade de uso e a eficiência operacional.

Além de atender às necessidades imediatas do IFSULDEMINAS, a pesquisa buscou contribuir para o campo mais amplo da IA e da tecnologia de processamento de linguagem natural, explorando novas formas de aplicar essas tecnologias em ambientes educacionais e corporativos. Através de uma combinação de análise teórica e desenvolvimento prático, o projeto visou estabelecer um novo paradigma na transcrição automatizada, abrindo caminho para avanços futuros na área.

## **2. FUNDAMENTAÇÃO TEÓRICA**

A ascensão da Inteligência Artificial (IA) representa um marco na evolução tecnológica, impactando significativamente diversos campos, incluindo a transcrição de vídeos. O trabalho presente busca explorar e expandir a aplicação da IA neste domínio, com foco na eficiência e precisão. Esta seção delinea a jornada histórica da IA, desde os primeiros conceitos teóricos até as aplicações práticas mais recentes, enfatizando sua relevância e potencial.

No contexto da pandemia de COVID-19, a IA emergiu como uma ferramenta vital para facilitar a comunicação remota. Com o aumento do uso de videochamadas, surgiram desafios únicos, especialmente para servidores e docentes do IFSULDEMINAS Câmpus Poços de Caldas, que necessitam documentar eficientemente o conteúdo de reuniões online. Este cenário sublinha a importância de desenvolver uma ferramenta de IA capaz de transcrever vídeos de maneira precisa e acessível.

Os avanços na IA, particularmente em áreas como processamento de linguagem natural e análise de áudio, são cruciais para este projeto. Exploramos como estes desenvolvimentos podem ser aplicados à transcrição de vídeos, visando a criação de uma ferramenta que não só atenda, mas também supere as soluções existentes em termos de eficiência e usabilidade.

A integração de diferentes tecnologias de IA abre novas possibilidades para a automatização da transcrição, transformando a maneira como as informações são extraídas e utilizadas.

Além disso, a seção aborda como a demanda crescente por tais ferramentas estimulou a inovação e a concorrência no mercado. Ferramentas populares como Google Meet, Microsoft Teams e Zoom, que tiveram um aumento significativo de uso durante a pandemia, implementaram funcionalidades como legendas automáticas e tradução em tempo real.

No entanto, essas plataformas ainda não oferecem uma solução para a documentação textual pós-reunião, um problema que este trabalho visa resolver.

A fundação teórica do projeto está ancorada na combinação de conhecimentos práticos e teóricos. Exploramos como o aprendizado de máquina e as redes neurais, que requerem grande poder computacional e manipulação de dados, podem ser efetivamente utilizados para a transcrição de vídeos. A discussão inclui também a análise de chatbots e ferramentas de IA como o ChatGPT e o Whisper da OpenAI, ressaltando seu papel na facilitação de tarefas diárias e na transcrição de áudios.

Este estudo representa um esforço para preencher uma lacuna existente no uso de IA para a geração automatizada de atas e descrições textuais de reuniões. O objetivo é desenvolver um sistema que não apenas transcreva, mas também formate esses textos de forma eficiente e intuitiva, demonstrando o poder transformador da IA na resolução de problemas práticos em ambientes educacionais e corporativos.

## **2.1 INTELIGÊNCIA ARTIFICIAL**

No livro *Alan Turing: The Enigma* (1983), Andrew Hodges relata a biografia de Turing e a busca para a solução da descryptografia dos nazistas, que utilizavam a comunicação via rádio encriptografada para que os inimigos não pudessem interceptar e obter informações sobre ataques. Eles utilizavam uma máquina chamada Enigma, que criptógrafos da época achavam ser inquebrável, pois era humanamente impossível decifrar o código criptografado antes que o mesmo trocasse no início do próximo dia.

Para resolver isso, Turing junto de sua equipe criaram uma máquina capaz de resolver problemas mais rápido que humanos, desenvolvendo o conceito da Teoria Computacional, algoritmo e da análise de padrões e formas. Desenvolvimentos esses que tiveram influência em teorias sobre o fundamento da mente e foram precursoras dos estudos sobre a IA (BIRCHENALL, 2009).

A IA ou a elaboração de redes neurais necessitam de uma grande quantidade de poder computacional e de dados, uma vez que análises precisam ocorrer diversas vezes, simulando e treinando infinitas possibilidades.

Em meados dos anos 2000, o pesquisador Geoffrey Hinton descobriu um modo de treinar novas camadas de redes neurais de forma mais eficiente, ele e outros pesquisadores, futuristas e CEOs de tecnologia começaram a falar sobre o enorme potencial do campo para decifrar a fala humana, traduzir documentos, reconhecer imagens, prever comportamentos e tomar decisões através de treinamentos (LEE, Kai-fu. 2019).

Pesquisas no campo juntamente do avanço tecnológico e computacional permitiu que o pensamento desses pesquisadores e futuristas no início dos anos 2000 se tornasse realidade nos dias atuais.

Com o avanço dessa tecnologia, surgiram os *chatbots* que é um modelo de linguagem ajustado com técnicas de aprendizado supervisionado e por reforço de ferramentas onde através da conversa com uma IA é possível obter respostas sobre diversos assuntos aos quais ela tenha sido treinada (GATTUSSO, Christophe. 2023).

Uma das ferramentas mais conhecidas e que se popularizou muito, desde o início de 2023, foi o *Chat GPT*, chamou a atenção por suas respostas detalhadas e articuladas, onde o mesmo possui um treinamento com uma grande base de dados disponíveis online até 2021 (PASSOS, Mauro; JÚNIOR, José. 2023).

O *Chat GPT* é uma ferramenta da empresa OpenAI, empresa especializada na pesquisa e uso de IA. Esta possui outra ferramenta também de IA, que possui treinamento e base de dados diferentes do *GPT*, o chamado *Whisper*, que será de grande foco durante o desenvolvimento deste trabalho e que será detalhado no devido momento.

De forma sucinta, o *Whisper* é uma ferramenta que transcreve e traduz áudios em texto. Com diferentes *prompts*, é possível obter toda a informação transcrita de um áudio, e até mesmo, usando o *prompt* correto, gerar legenda nos formatos srt, extensão essa utilizada nos serviços de hospedagem de vídeos, como o YouTube e o Vimeo.

## **2.1 TRANSCRIÇÃO DE VÍDEO OU ÁUDIO PARA TEXTO**

Com o avanço da inteligência artificial, funções antes realizadas por humanos estão sendo desenvolvidas para serem realizadas com o uso da IA, para poupar tempo e custo operacional. Possivelmente você já viu filmes e notícias que se passam em um tribunal e observaram em cena um profissional escrevendo tudo que está sendo discutido durante a reunião ou assembleia. Pois bem, o nome desse profissional é Taquígrafo e é o profissional capaz de escrever, por meio de sinais convencionais, a uma velocidade bem mais elevada que a escrita comum, reproduzindo, em texto escrito e com fidelidade, as palavras de um orador.

Existe uma maneira de realizar essa escrita de forma mais rápida, mais assertiva através do uso da IA. Um dos campos da IA é o Processamento de Linguagem Natural (PLN), um campo que estuda a análise e reconhecimento de textos em linguagens humanas, ou naturais. O estudo e a pesquisa sobre o PLN não é uma tarefa fácil devido à rica ambiguidade da linguagem natural. Essa ambiguidade torna o PLN diferente do processamento das linguagens de programação de computador, as quais são formalmente definidas evitando, justamente, a ambiguidade. (VIEIRA, Renata; LOPES, Lucelene. 2010).

Ao realizar uma reunião, uma conferência e armazenar a gravação, é possível através do uso de ferramentas de transcrição que utilizam o processamento de linguagem natural, converter todo o áudio desse vídeo para texto, e trabalhar em cima dele, seja identificando os indivíduos e suas frases ou utilizando a transcrição para formatar documentos de diversas frentes.

Isto é a solução que ferramentas do mercado vêm fazendo nos últimos anos, graças ao aumento significativo das ferramentas à disposição dos pesquisadores de PLN permitiu também um aumento significativo nas ambições da área. (VIEIRA, Renata; LOPES, Lucelene. 2010). Uma que se destaca por sua simples integração com outros sistemas é o *Whisper*, serviço de transcrição e tradução de vídeos para texto e que possui em média 99% de precisão nas transcrições, possibilitando uma escrita da oratória de uma reunião de forma simples e rápida.

### **3. METODOLOGIA**

#### **3.1 CONTEXTO E JUSTIFICATIVA**

A emergência da pandemia de COVID-19 remodelou fundamentalmente as práticas de trabalho e educação, impulsionando uma transição abrupta para o trabalho remoto e o ensino a distância. Neste cenário dinâmico, as videochamadas emergiram como uma ferramenta essencial, não apenas para a comunicação, mas também para a documentação e revisão de informações compartilhadas. A necessidade de uma transcrição eficaz e precisa do conteúdo das videochamadas tornou-se uma prioridade, especialmente para o corpo docente e administrativo do IFSULDEMINAS Câmpus Poços de Caldas, que frequentemente lidam com uma grande quantidade de informações durante reuniões e aulas online.

### 3.2 OBJETIVO METODOLÓGICO

O objetivo principal da metodologia adotada neste estudo foi desenvolver uma ferramenta de transcrição automatizada baseada em Inteligência Artificial (IA) que fosse não apenas eficiente e precisa, mas também acessível e fácil de usar. A ideia era criar uma solução que pudesse superar os desafios enfrentados pelas ferramentas de transcrição existentes, principalmente em termos de acurácia e usabilidade.

### 3.3 DESENVOLVIMENTO E TESTE DO PROTÓTIPO

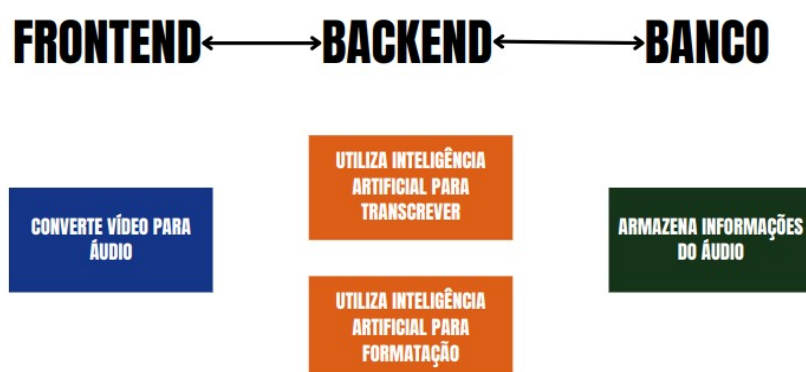
Este protótipo foi construído utilizando as mais recentes tecnologias em IA, processamento de linguagem natural e reconhecimento de voz. Os testes incluíram a verificação da acurácia da transcrição, a avaliação da interface do usuário e a eficiência do sistema em diferentes condições, como em vídeos com ruído de fundo ou em situações com múltiplos falantes.

### 3.4 ANÁLISE DE DADOS

Para analisar os dados coletados durante os testes, utilizamos métodos estatísticos para avaliar a precisão da transcrição e técnicas de análise de usabilidade para entender a experiência do usuário.

A metodologia adotada neste estudo foi integral para o desenvolvimento de uma ferramenta de transcrição de vídeos baseada em IA que não só atende, mas também excede as expectativas em termos de precisão, usabilidade e acessibilidade. Este trabalho representa um avanço significativo no uso da IA para facilitar a transcrição de vídeos, contribuindo para melhorar a eficiência e eficácia na documentação de reuniões e aulas em ambientes educacionais e corporativos.

Figura 1 - Diagrama do sistema implementado



Fonte: O autor (2023)

### **3.4 DESENVOLVIMENTO DO SISTEMA**

O desenvolvimento de uma ferramenta de transcrição de vídeos baseada em Inteligência Artificial (IA) representa um desafio único que cruza os domínios da engenharia de software, linguística computacional e processamento de sinais. O objetivo deste desenvolvimento era criar uma solução robusta e adaptável que pudesse transcrever com precisão o conteúdo das videochamadas e reuniões online, uma necessidade crítica identificada no ambiente acadêmico e corporativo durante a pandemia de COVID-19.

#### **3.4.1 SELEÇÃO DE TECNOLOGIA E DESIGN DE SOFTWARE**

A escolha da tecnologia foi um passo crucial. Optamos por algoritmos já existentes de IA avançados, especialmente aqueles centrados em aprendizado de máquina e redes neurais para o processamento de linguagem natural e reconhecimento de voz. A arquitetura de software foi projetada para ser modular, permitindo integrações futuras e melhorias contínuas.

#### **3.4.2 TESTES E ITERAÇÕES**

Realizamos testes em diferentes cenários através do uso de diferentes amostras, incluindo simulação de conversa, noticiários e assembleias. Cada rodada de testes proporcionou insights valiosos, que foram usados para refinar e melhorar o sistema.

#### **3.4.3 INTEGRAÇÃO E IMPLEMENTAÇÃO FINAL**

Após várias iterações, chegamos a uma versão estável do sistema, capaz de transcrever com alta precisão e eficiência.

Este processo de desenvolvimento resultou em uma ferramenta de IA inovadora que não só atende às necessidades identificadas, mas também estabelece um novo padrão em transcrição de vídeo. A ferramenta representa um avanço significativo na aplicação da IA para melhorar a comunicação e a documentação em ambientes acadêmicos e corporativos.

A ferramenta faz uso dos tipos mais utilizados de arquivo de vídeo para realizar as transcrições, como .mp4, .avi e .mkv.

Vídeo convertido para áudio, é então realizado o armazenamento do arquivo em uma pasta temporária no servidor, e armazenado no banco de dados informações referentes ao arquivo, como o nome e o caminho. Com isso, é possível realizar a transcrição textual utilizando a IA denominada *Whisper*, treinada para realizar transcrições de áudio para texto, legendas e traduções para diferentes idiomas.



É então trabalhado em cima desse retorno, uma formatação de texto, pois o mesmo retorna de uma maneira que o usuário precisará realizar correções. Ainda com o auxílio da IA, é possível realizar duas diferentes funções na ferramenta desenvolvida, obter todo o texto transcrito, para que o usuário possa trabalhar em cima dele da maneira ao qual achar melhor, ou, realizar uma formatação do texto para ATA de uma Assembleia Geral através do uso de IA. Esse retorno aparece diretamente na tela do sistema, não possuindo uma opção de download do arquivo, fica a cargo do usuário copiar o texto e utilizar a ferramenta de documentos que mais lhe agrada, revisar e formatar do seu gosto.

### **3.5 BASE DE DADOS**

As análises foram realizadas através da seleção de algumas ferramentas online para transcrição de vídeo e realização de videoconferências. É importante distinguir estas duas etapas.

A análise dos serviços oferecidos pelas ferramentas de videoconferência focou na geração automática de legendas. Estas ferramentas são gratuitas, no entanto, não oferecem a opção de baixar o texto gerado pelas legendas de forma gratuita.

Para comparar a precisão das legendas, foi gerado um texto através de Inteligência Artificial (IA) que foi lido por indivíduos nas ferramentas Google Meet e Microsoft Teams, simulando uma conversa real entre duas pessoas. Além disso, foi criado um vídeo transcrito nas ferramentas de transcrição para análise posterior.

As ferramentas de transcrição, por outro lado, foram submetidas a uma análise mais detalhada, uma vez que oferecem uma gama maior de serviços, não se limitando apenas à geração de legendas durante uma chamada de vídeo.

Foram utilizadas três amostras de vídeo diferentes nas ferramentas de transcrição. A primeira amostra é a gravação caseira de um diálogo entre dois indivíduos em um ambiente tranquilo com pouco ruído de fundo.

A segunda amostra é um noticiário de um jornal, com um som limpo e uma única fonte de fala. A terceira amostra é uma gravação de uma assembleia geral sobre problemas agrícolas, em um ambiente ruidoso com múltiplas fontes de fala. Todos esses vídeos podem ser facilmente encontrados no YouTube e estão devidamente referenciados como amostras.

### **3.6 MÉTRICAS DE AVALIAÇÃO**

Para analisar a assertividade e qualidade da solução proposta, foram utilizadas as seguintes métricas de avaliação e suas escolhas tiveram auxílio a base do trabalho realizado

por Giovana Carneduto (2020), em seu trabalho de conclusão, ela discute sobre métricas no ambiente de desenvolvimento.

### **3.6.1 DESEMPENHO**

Para medir o desempenho das ferramentas, foi calculado o tempo gasto para realizar as transcrições. Apesar das amostras utilizadas possuírem tempo baixos, essa métrica é necessária visto que reuniões e assembleias podem possuir longas durações.

### **3.6.2 ACURÁCIA**

Será utilizado o texto original para medir a quantidade de palavras transcritas de maneira correta pelas ferramentas, por exemplo, uma ferramenta que transcreve 76 palavras de 100 ditas em um vídeo, de forma correta, possui 76% de acurácia.

### **3.6.3 USABILIDADE**

Contabilização da quantidade de cliques feitos pelo indivíduo ao navegar pelas ferramentas até que o mesmo consiga obter a transcrição dos vídeos.

### **3.6.4 CUSTO**

Cada ferramenta testada possui um valor diferente para suas funcionalidades, precisou-se então definir o tempo que será gasto mensalmente, sendo utilizado como parâmetro o valor de 8 (oito) horas de vídeos transcritos mensalmente, valor estipulado na realização de uma reunião semanal de até 2 (duas) horas e essa métrica irá calcular esse custo.

Assim como o custo de hospedagem futuro que a implementação da ferramenta desenvolvida pode ter ao se tornar um produto, a fim de se ter resultados mais justos, visto que as ferramentas de comparação possuem gastos e que são repassados para seus clientes.

## **4. DESENVOLVIMENTO DO TRABALHO**

Antes do desenvolvimento, veio a criação do *layout*, utilizando o *Figma* e os conceitos de UI e UX para pensar sempre no usuário. O *layout* é importante visto que, com ele, várias etapas do desenvolvimento podem ser agilizadas já que não precisa ser consumido tempo de desenvolvimento pensando como cada componente irá ficar ou se comportar na tela da aplicação.

A maneira simples que o *layout* foi estruturado, foi pensando na maneira mais simples de que, os docentes e servidores, consigam num primeiro momento, já entender todas opções

disponíveis, e como funciona todo o fluxo. Foi proposto uma ferramenta web com *layout* simplificado para garantir melhor usabilidade por parte dos usuários.

#### 4.1 INTERFACE DE USUÁRIO

Com a ideia inicial de como seria a interface do projeto, começou-se o desenvolvimento da interface de usuário. Quando se pensa no desenvolvimento do *frontend* ou a interface de usuário (a parte que o usuário de fato visualiza de um site), logo se pensa nas tecnologias habituais utilizadas como o *HTML* e o *CSS*, e a utilização do *JavaScript* para aprimorar todo o desenvolvimento.

Todas essas tecnologias foram utilizadas, mas, através do uso da biblioteca *React*, foi-se substituindo o uso do *Javascript* pelo *Typescript*, que adiciona tipagem estática opcional ao *Javascript*.

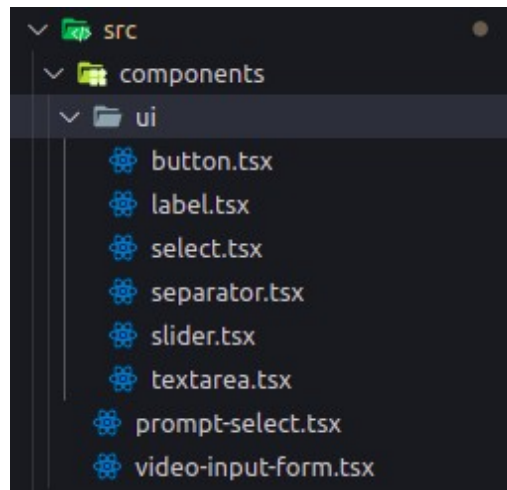
Na documentação do *React*, o mesmo informa que para se criar um novo projeto ou adicionar o *React* a um já existente é necessário o *NodeJS*, que segundo a descrição da documentação do mesmo, é um ambiente de execução *Javascript*, *cross-plataform* e de código aberto, ou seja, ele permite o uso de *Javascript* fora do ambiente web, ao qual o mesmo foi designado.

Com o projeto criado, a estrutura inicial do projeto segue a padrão do *React*, posteriormente foram feitos ajustes, mas, para a maioria dos casos de um projeto *React*, a estrutura inicial padrão já atende os requisitos. Para continuar o desenvolvimento da interface de usuário, foram adicionadas outras bibliotecas e dependências, dessas é importante ressaltar três que são a base da estrutura imaginada para o projeto final, abaixo quais são elas e o passo a passo de como foi a instalação.

- *Tailwind CSS (Framework de CSS)*
- *Radix UI* (Biblioteca de componentes, sem personalização)
- *shadcn/ui* (Biblioteca com componentes estilizados do *Radix UI*)

Seguindo a documentação de instalação do *shadcn/ui* é possível verificar que ao instalar o mesmo, as etapas prévias já realizam a instalação das outras dependências, visto que todas serão utilizadas em conjunto.

Figura 2 - Componentes criados



Fonte: O autor (2023)

Nesta etapa, já se tem um projeto *React*, com suas dependências instaladas e as bibliotecas que serão utilizadas também devidamente instaladas, basta então começar o desenvolvimento da tela. O *React* permite a criação de componentes, como detalhado no tópico respectivo, isso faz com que seja possível separar a interface em blocos, e reutilizá-los se necessário. Foram criados um grande número de componentes, onde cada componente tem suas particularidades, abaixo é possível visualizar a estrutura final de componentes criados e um exemplo do componente de *textarea*.

Figura - Componente *TEXTAREA*

```
import { cn } from "@lib/Utils"

igorflpaula, 2 months ago | 1 author (igorflpaula)
export interface TextareaProps
  extends React.
    TextareaHTMLAttributes<HTMLTextAreaElement> {}

const Textarea = React.forwardRef<HTMLTextAreaElement,
  TextareaProps>({
  ({ className, ...props }, ref) => {
    return (
      <textarea
        className={cn(
          "flex min-h-[60px] w-full rounded-md border
          border-input bg-transparent px-3 py-2 text-sm
          shadow-sm placeholder:text-muted-foreground
          focus-visible:outline-none focus-visible:ring-1
          focus-visible:ring-ring
          disabled:cursor-not-allowed
          disabled:opacity-50",
          className
        )}
        ref={ref}
        {...props}
      />
    )
  }
)
Textarea.displayName = "Textarea"
export { Textarea }
```

Fonte: O autor (2023)

Foram criados um grande número de componentes, onde cada componente tem suas particularidades. Todos esses componentes criados, precisam ser utilizados em algum momento, para dar sentido à interface.

O *React* possui um arquivo chamado de *App.tsx* onde ele é o arquivo principal do projeto, é ele que é iniciado e escutado toda vez que algo acontece, e é justamente nele que iremos fazer toda a lógica do *frontend* e a estilização final da interface.

Os detalhes da utilização de outras tecnologias integradas, a criação de rotas e demais informações serão detalhadas no tópico 4.3, já que é necessário explicar sobre a criação e o funcionamento do *backend* antes, para que tudo faça sentido.

Figura - *Layout* final da Interface do usuário

INSTITUTO FEDERAL  
de São Carlos

GitHub © Desenvolvido por Igor de Paula

Inclua o prompt para a IA.

Selecione um vídeo

Prompt de transcrição

Adicione palavras-chave mencionadas no vídeo separadas por vírgula (,)

Carregar vídeo

Resultado gerado pela IA.

Prompt

Selecione um prompt.

Modelo

GPT 3.5

Temperatura

Valores mais altos tendem a deixar o resultado mais criativo e com possíveis erros.

Lembre-se: você pode utilizar a variável `[transcription]` no seu prompt para adicionar o conteúdo da transcrição do vídeo selecionado

Executar

Fonte: O autor (2023)

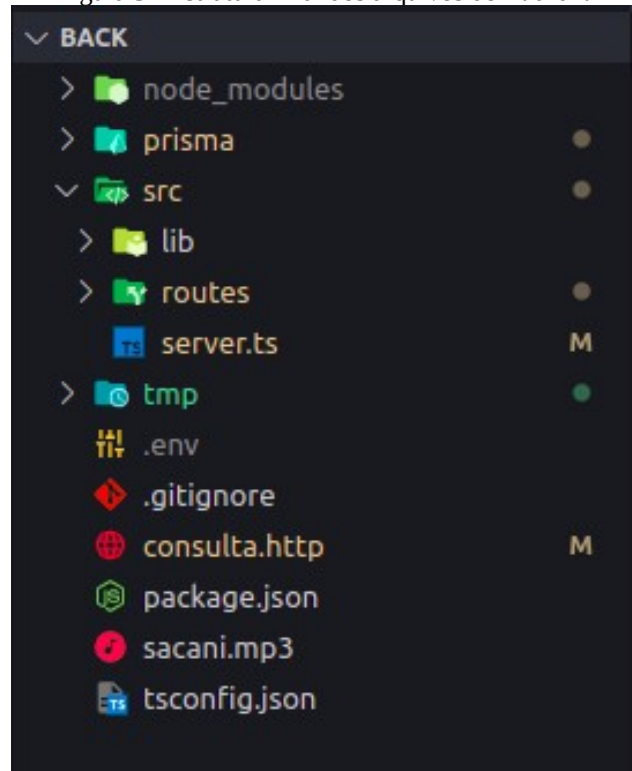
## 4.2 BACKEND

Com o *frontend* já pronto, chegamos a etapa do desenvolvimento do *backend*, o sistema que irá executar as funções por trás da interface de usuário e integrar com as ferramentas mencionadas na fundamentação teórica.

Assim como a estrutura do *frontend*, também precisamos criar um novo diretório para o *backend*, e iniciar o projeto usando *NodeJS* e as informações disponíveis na documentação do mesmo.

A estrutura inicial é simples e com poucos arquivos e pastas, alguns dos arquivos e pastas precisaram ser criados manualmente, ou foram criados em um momento futuro do desenvolvimento, mais detalhes à frente neste tópico.

Figura 3 - Estrutura final dos arquivos do *Backend*



Fonte: O autor (2023)

O arquivo responsável por toda a organização do *backend* é o *server.ts* dentro da pasta *src*, a ideia é a mesma do arquivo *App.tsx* do *React*, ele irá conter todas as informações macro do sistema, parte das importações, em qual porta o sistema irá rodar, e a chamada das rotas que serão criadas.

Figura 4 - Estrutura final dos arquivos do *Backend*

```

const app = fastify();

app.register(fastifyCors, {
  origin: '*',
})

app.get('/', async () => {
  return 'Hello World! 🖐️'
})

app.register(getAllPromptsRoute)
app.register(uploadVideoRoute)
app.register(createTranscriptionRoute)
app.register(generateAICompletionRoute)
app.register(clearTMPFiles)
app.register(createSubtitleRoute)

app.listen({
  port: 3333,
}).then(() => {
  console.log('Servidor HTTP rodando..')
})

```

Fonte: O autor (2023)

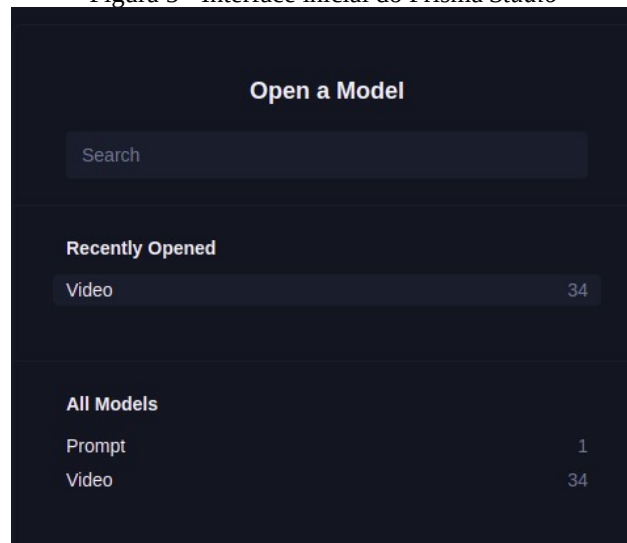
É possível visualizar algumas funções importantes, o uso da constante *app* chamando a função *fastify()*, devidamente importada, onde está constante passa a ser uma instância do *Fastify*, possuindo todas as suas funcionalidades.

A criação do método *GET* na rota inicial do projeto (*/*), retornando o clássico *Hello World*, mostrando que a aplicação está funcionando. Dentro temos as chamadas das rotas, onde cada chamada nos leva para um arquivo de rota individual, deixando o arquivo principal muito mais limpo e simples de entender. Por fim, um *listen*, configurando em qual porta iremos trabalhar, e informando que o Servidor *HTTP* está rodando.

Para a aplicação funcionar corretamente como prevista, é necessário o uso de banco de dados. Existem uma grande variedade de bancos de dados disponíveis, assim como *ORM* e *Workbenches* distintos no mercado, para esta aplicação foi escolhido o uso do Prisma como *ORM*, e do *SQLite* como banco de dados, durante o tópico 2.5 já foram dadas mais informações sobre essas tecnologias.

Abaixo a sequência de comandos para a implementação do Prisma, as configurações necessárias para nossa *ORM* e a criação da estrutura do banco, através das *model* seguem a estrutura convencional de um banco de dados relacional. É possível visualizar a estrutura do banco de dados no apêndice A e B.

Figura 5 - Interface inicial do Prisma Studio



Fonte: O autor (2023)

Para fazer a conexão com o banco de dados, cria-se um arquivo *prisma.ts* dentro da pasta *lib*, que está localizada dentro da pasta *src* na raiz do projeto. Através da criação do arquivo *seed.ts*, dentro da hierarquia de pasta do prisma, é possível deixar pronto *scripts* de execução de banco de dados, é neste momento que foi configurado um *prompt* que será utilizado para solicitar futuramente ao *Chat GPT* a formatação correta das ATAS.

Figura 6 - Prompt utilizado para formatação das ATAS

```
await prisma.prompt.create({
  data: {
    title: 'Criação de ATA',
    template: `Seu papel é formatar uma ATA de AGE.

Abaixo você receberá uma transcrição do que foi discutido durante a AGE, use essa transcrição para formatar a ATA.

Transcrição:
...
{transcription}
...`.trim()
  }
})
```

Fonte: O autor (2023)

Figura 7 - Prompt utilizado para retornar apenas a transcrição



```

    await prisma.prompt.create({
      data: {
        title: 'Transcrição',
        template: `Seu papel é formatar o texto.

Transcrição:
...
{transcription}
...`.trim()
      }
    })
  }
}

```

Fonte: O autor (2023)

Figura 8 - Configuração de conexão com o banco de dados

```

import { PrismaClient } from "@prisma/client";

export const prisma = new PrismaClient()

```

Fonte: O autor (2023)

Para visualização do conteúdo da tabela *Prompts* do nosso banco de dados, e para que o front consiga utilizar essas informações, foi-se criado uma rota para encontrar os *prompts* existentes, como a chamada da *app.register(getAllPromptsRoute)*, função esta que está no arquivo *server.ts* e que chama o conteúdo da rota criada em um arquivo separado.

Figura 10 - Conteúdo da rota getAllPrompts

```

import { FastifyInstance } from "fastify";
import { prisma } from "../lib/prisma";

export async function getAllPromptsRoute(app: FastifyInstance) {
  app.get('/prompts', async () => {
    const prompts = await prisma.prompt.findMany()

    return prompts
  })
}

```

Fonte: O autor (2023)

Para ser realizado a transcrição dos vídeos para texto, e posteriormente formatado para ATAS, precisamos que o nosso *backend* tenha acesso ao vídeo em si, para que seja possível a transcrição. Para isso, foi criado uma rota para o *upload* de arquivos, utilizando o pacote *multipart* do *Fastify*, a configuração do limite máximo do arquivo é configurável, mas recomendada seguindo a documentação de que se limite a 25 MB, e caso seja necessário a transferência de arquivos maiores, que haja a repartição desses.

Figura 11 - Configuração do limite do tamanho do arquivo

```
app.register(fastifyMultipart, {  
  limits: {  
    fileSize: 1_048_576 * 25, //25mb  
  }  
})
```

Fonte: O autor (2023)

### 4.3 INTEGRAÇÕES

Com o vídeo previamente já convertido em áudio pelo *frontend* através do *FFMpeg*, fazemos o *upload* do arquivo via método *POST* na rota detalhada no apêndice C, temos acesso então as informações do vídeo em questão no nosso banco de dados, como o *uid* do áudio e o caminho para o arquivo, precisamos agora gerar a transcrição deste áudio para texto. Para isso, foi criada uma nova rota, ao qual através da integração com o *Whisper* e seguindo com a documentação do mesmo, é possível através de um *create* nos métodos de transcrição de áudio, adicionar parâmetros e obter a transcrição, ou até mesmo caso necessário, a tradução do áudio em texto para outro idioma.

Neste momento é possível utilizar além do áudio em si, alguns *prompts* e parâmetros para que a transcrição seja mais assertiva ou mais criativa, que retorne em um formato específico e demais funções.

Com a transcrição pronta e armazenada em seu respectivo índice do banco de dados, precisamos fazer alguns ajustes e formatação no texto, utilizando o *Chat GPT*. No arquivo *seed.ts* citado anteriormente nesse artigo, foi criado um *prompt* para que com a inserção da transcrição, o *Chat GPT* possa formatar esse texto para a estrutura de uma ATA de Assembleia Geral Extraordinária.

Figura 12 - Transcrição de Áudio para texto através do *Whisper*

```

export async function createTranscriptionRoute(app: FastifyInstance) {
  app.post('/videos/:videoId/transcription', async (request) => {
    const paramsSchema = z.object({
      videoId: z.string().uuid()
    })

    const { videoId } = paramsSchema.parse(request.params)

    const bodySchema = z.object({
      prompt: z.string(),
    })

    const { prompt } = bodySchema.parse(request.body)

    const video = await prisma.video.findUniqueOrThrow({
      where: {
        id: videoId,
      }
    })

    const videoPath = video.path
    const audioReadStream = createReadStream(videoPath)

    const response = await openai.audio.transcriptions.create({
      file: audioReadStream,
      model: 'whisper-1',
      language: 'pt',
      response_format: 'json',
      temperature: 0,
      prompt,
      igorflpaula, 2 months ago • first commit
    })

    const transcription = response.text

    await prisma.video.update({
      where: {
        id: videoId,
      },
      data: {
        transcription
      },
    })

    return { transcription }
  })
}

```

Fonte: O autor (2023)

Para isso, será utilizado a rota *generateAICompletion*, onde enviaremos através da integração da OpenAI, e usando o modelo do GPT 3.5, este disponibilizado sem custos, tanto o *prompt* quanto a transcrição, o mesmo retornará a ATA formatada. O mesmo retornará uma ATA com algumas informações faltantes, visto que nem todas as informações necessárias de uma ATA de AGE são ditas durante as reuniões. É possível de maneira simples, editar essas informações, já que o retorno mostra onde existem informações faltantes, como o nome completo dos membros, o local e o horário da reunião.

## 5. RESULTADOS E DISCUSSÃO

Assim como dito por João Pedro em sua publicação no blog da Filmora (2023), existe um amplo espectro de diferentes produtos de software de transcrição automática que são anunciados gratuitamente, você ainda precisará adquirir uma licença ou uma assinatura se

quiser usar todos os seus recursos, muitas vezes, valores altos em dólar, vindo de contramão a uma das métricas que este projeto visa atender, o baixo custo.

Abaixo é possível visualizar um gráfico de comparação de custo em dólar, sempre usando o mesmo valor de tempo como parâmetro, nele é possível visualizar as três ferramentas mais bem qualificadas na publicação do João Pedro, a solução do Google e da Microsoft e a solução deste artigo.

## **5.1 FERRAMENTAS DE COMPARAÇÃO**

Com o advento da pandemia de COVID-19, a demanda por ferramentas de transcrição automatizada de vídeos aumentou significativamente. Conforme destacado por João Pedro em sua publicação no blog da Filmora (2023), há uma variedade de softwares de transcrição automática disponíveis no mercado.

No entanto, muitos desses produtos, apesar de anunciados como gratuitos, exigem uma licença ou assinatura para acessar todos os seus recursos, frequentemente a custos elevados.

### **5.1.1 SELEÇÃO DE FERRAMENTAS**

Para este estudo, escolhemos uma variedade de ferramentas para análise e comparação. Duas delas são ferramentas de videoconferência, Google Meet e Microsoft Teams, conhecidas por gerar legendas automaticamente durante reuniões.

Outras três são ferramentas de transcrição de vídeo para áudio, que oferecem funcionalidades mais próximas da solução desenvolvida neste artigo. Essas incluem a Trint, conhecida por suas excelentes ferramentas de edição e identificação de falantes, a Happy Scribe e a Sonix, ambas populares por suas soluções baseadas em tempo de transcrição.

### **5.1.2 CRITÉRIOS DE ESCOLHA**

A escolha do Google Meet e do Microsoft Teams foi fundamentada em dados do Comitê Gestor da Internet no Brasil, que indicou um alto índice de familiaridade e uso dessas plataformas. Entretanto, ambas possuem limitações, como a falta de opção para download de transcrições textuais completas, o que dificulta comparações diretas com outras ferramentas.

### **5.1.3 CUSTOS E FUNCIONALIDADES**

Cada ferramenta foi analisada considerando suas funcionalidades, custos e o valor agregado que oferecem. As ferramentas de videoconferência, apesar de oferecerem legendas

automáticas, limitam-se em aspectos como gravação e download de transcrições. Já as ferramentas dedicadas à transcrição, como a Trint, Happy Scribe e Sonix, variam em termos de custo e funcionalidades, com a Trint oferecendo uma assinatura mensal e as outras cobrando com base no tempo de transcrição.

## **5.2 ANÁLISE DE CUSTO**

### **5.2.1 COMPARAÇÃO DE CUSTOS**

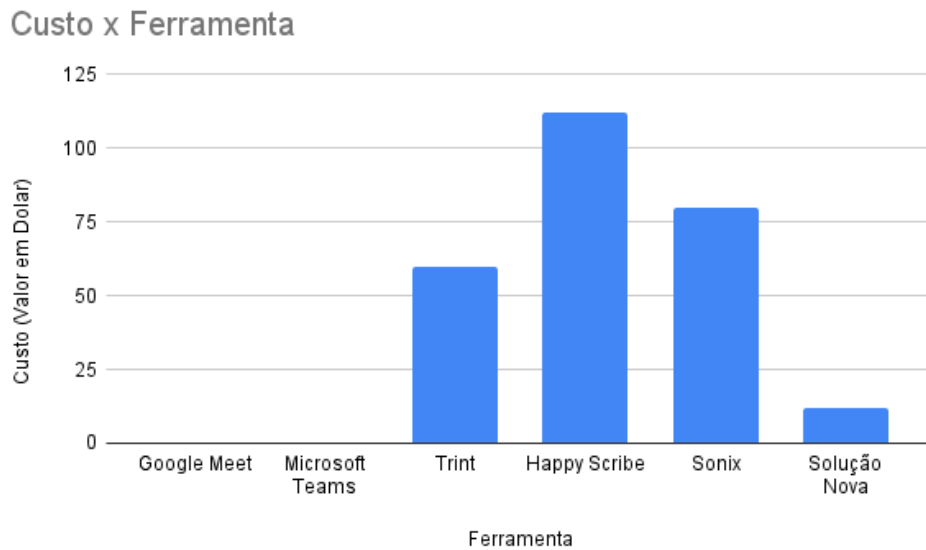
Realizamos uma análise comparativa de custos entre as ferramentas selecionadas. Um gráfico foi desenvolvido para ilustrar esta comparação, usando o tempo como parâmetro constante para todas as ferramentas. As ferramentas do Google e Microsoft foram excluídas desta análise de custo devido à impossibilidade de download das transcrições.

Em um cenário onde a eficiência de custos é crucial, a hospedagem de ferramentas web na AWS proporciona uma solução acessível e escalável. Utilizando a calculadora de preços online da AWS, é possível estimar um valor médio mensal de apenas 6 dólares para uso básico com menos de 2000 requisições mensais. O custo está distribuído entre dois serviços fundamentais, sendo eles o Amazon CloudFront e o Amazon EC2.

### **5.2.2 ANÁLISE DETALHADA**

A análise revelou discrepâncias significativas em termos de custo entre as ferramentas de mercado e a solução desenvolvida neste artigo. A Trint, apesar de ser a mais cara entre as ferramentas de transcrição, oferece transcrições ilimitadas sob sua assinatura mensal. Em contraste, a Happy Scribe e a Sonix cobram por hora de transcrição, com a Sonix apresentando um custo ligeiramente inferior. A solução desenvolvida neste estudo se destaca por seu baixo custo, oferecendo transcrições a menos de US\$ 3 para oito horas de vídeo por mês.

Gráfico 1 - Custo em Dólar x Ferramenta



Fonte: O autor (2023)

É possível ver no gráfico 1 que as soluções da Microsoft e Google não possuem dados, isso se dá pois apesar de ser possível obter em tempo real as legendas do que é dito durante uma videochamada, elas não disponibilizam a opção de *download* desse texto, tornando inviável essa comparação.

Olhando para as três soluções do mercado de transcrição e a nossa nova solução, é possível observar uma discrepância enorme quanto ao custo. A solução da Trint, se torna a mais barata dentre as três, pois nela, é possível adquirir uma assinatura mensal no valor de US\$ 60 dólares, onde é possível obter transcrições ilimitadas, com certas condições que não são relevantes nesse contexto.

Como solução, a da Happy Scribe e da Sonix, foram usados os valores de 2 (duas) horas de vídeos transcritos por semana, visto que elas não possuem uma assinatura e é necessário pagar por tempo de transcrição.

Nesse quesito, o programa aqui desenvolvido custa menos do que US\$ 10 dólares para transcrever 8 (oito) horas de vídeo por mês, valores esses baseados no custo do uso das ferramentas da OpenAI mais o custo de hospedagem com os serviços da Amazon.

### 5.3 ANÁLISE DE DESEMPENHO

Quando se analisar outras medidas como o desempenho e eficiência, é necessário definir alguns critérios de avaliação, visando o melhor desempenho e eficiência, tem-se os seguintes critérios: velocidade da transcrição e acurácia da transcrição. Serão utilizados três

vídeos diferentes, um deles, citado anteriormente neste artigo, utilizado para verificar a acurácia das ferramentas do Google e do Meet, uma discussão real de uma Assembleia Geral Extraordinária (AGE), encontrada no Youtube, e um vídeo aleatório de um noticiário.

Observa-se na tabela 1 que o tempo gasto de cada ferramenta para gerar a transcrição, sendo a Trint é que gasta mais tempo para gerar a transcrição, por sua vez, a mesma permite a edição posterior do vídeo na sua plataforma, podendo recortar e instantaneamente refazer a transcrição.

A solução aqui desenvolvida é a ferramenta mais rápida dentre as avaliadas, gasta menos da metade do tempo da segunda que possui o menor tempo, uma parcela dessa velocidade vai da utilização local de recursos, fazendo com que a conversa entre os serviços seja mais rápida, visto que as outras soluções precisam de um tempo de conversa com o servidor que irá realizar as tarefas.

Vale ressaltar que o vídeo utilizado para a geração do gráfico 2, possuía um minuto e onze segundos de duração, já o tempo do vídeo utilizado para gerar o gráfico 3 possuía um minuto e dezessete segundos e do gráfico 4, cinco minutos e cinco segundos.

Nas colunas dos vídeos 2 e 3, pode-se perceber que a ferramenta da Happy Scribe não possui valores, isso se dá em razão da ferramenta só disponibilizar a transcrição gratuita de um vídeo.

Tabela 1 - Tempo gasto para gerar transcrições

	<b>Vídeo 01 - 01:11</b>	<b>Vídeo 2 - 01:17</b>	<b>Vídeo 3 - 05:05</b>
<b>Ferramenta</b>	<b>Tempo (s)</b>	<b>Tempo (s)</b>	<b>Tempo (s)</b>
Google Meet	-	-	-
Microsoft Teams	-	-	-
Trint	300	315	557
Happy Scribe	65	0	0
Sonix	72	79	225
Solução Nova	19	20	52

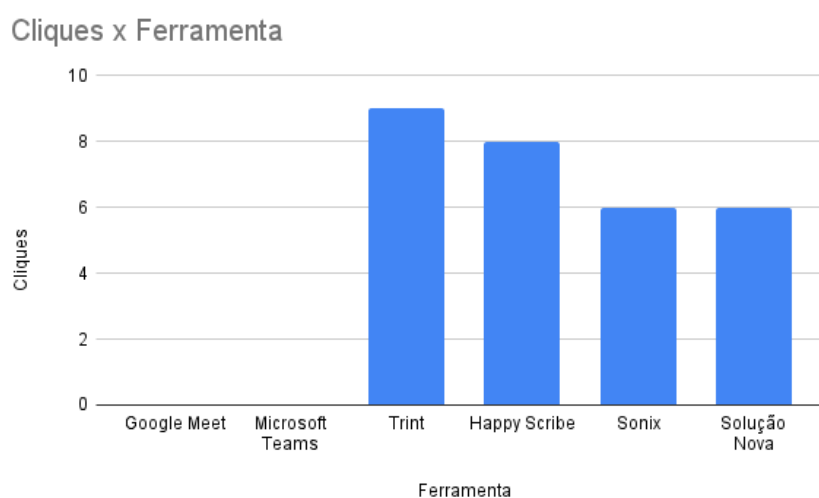
Fonte: O autor (2023)

A ferramenta da Trint possui a configuração de identificar os indivíduos da conversa, onde quando existe mais de um indivíduo no áudio, é possível apresentar os resultados escolhendo as trilhas individuais deles assim como obter o resultado completo com a identificação de cada frase. Esse talvez seja o motivo da quantidade significativa de tempo gasto para suas transcrições, função essa que torna a Trint uma ótima escolha caso este seja o resultado esperado.

## 5.4 ANÁLISE DE USABILIDADE

Medir o quão simples é a sua utilização é necessário, mas esta pode ser uma métrica difícil de se obter boas comparações, visto que a usabilidade distingue muito de indivíduo e de opinião. Esta será abordada da seguinte maneira, quantos cliques foram necessários desde o momento em que o indivíduo abriu a ferramenta até a transcrição estar pronta.

Gráfico 2 - Números de Cliques feitos até que a Transcrição esteja pronta



Fonte: O autor (2023)

Em geral, a quantidade de cliques em todas as ferramentas é bem baixa, todas as soluções possuem uma usabilidade simples e fácil, entregando o resultado em poucos cliques mesmo levando em consideração que, algumas necessitam de criação de contas para utilizar os serviços.

Pode-se observar que mesmo que a solução da Sonix precise da criação de uma conta, utilizando a integração dela com as contas Google, foi possível obter o resultado na mesma quantidade de cliques que nossa solução.

## 5.5 ANÁLISE DE ACURÁCIA

### 5.5.1 CONTEXTUALIZAÇÃO E IMPORTÂNCIA DA ACURÁCIA

Em um mundo cada vez mais conectado e dependente da comunicação digital, a acurácia na transcrição de vídeos tornou-se uma medida essencial de desempenho para as ferramentas de IA.



Esta análise crítica busca quantificar a precisão das ferramentas de transcrição automatizada, um componente crucial para garantir a fidelidade e utilidade das informações capturadas em reuniões online e conteúdo multimídia.

### **5.5.2 METODOLOGIA DE AVALIAÇÃO DA ACURÁCIA**

Para mensurar a acurácia das transcrições, adotamos um método rigoroso, coletando dados sobre o número total de palavras faladas em diferentes vídeos e a quantidade de palavras transcritas corretamente por várias ferramentas. A acurácia é expressa em percentuais, refletindo a proporção de palavras corretamente transcritas em relação ao total de palavras pronunciadas. Esta métrica fornece um indicador direto da qualidade da transcrição.

### **5.5.3 ANÁLISE COMPARATIVA DETALHADA**

Conforme ilustrado na Tabela 2, as ferramentas de videoconferência do Google Meet e Microsoft Teams exibiram níveis mais baixos de acurácia em comparação com as demais ferramentas especializadas. Isso pode ser atribuído à complexidade inerente à transcrição em tempo real durante as videochamadas e a ausência de uma funcionalidade que permita a revisão e edição posterior da transcrição gerada.

A análise das transcrições de um vídeo de notícias revelou uma acurácia surpreendentemente alta para todas as ferramentas, provavelmente devido à clareza e formalidade do discurso. Em cenários onde a linguagem é cuidadosamente elaborada e apresentada, as ferramentas de IA tendem a ter um desempenho melhor, como evidenciado na Tabela 3.

### **5.5.4 AVALIAÇÃO DAS FERRAMENTAS DE TRANSCRIÇÃO ESPECÍFICAS**

As ferramentas especializadas em transcrição, como a Trint, a Sonix e a solução desenvolvida para este estudo, mostraram níveis altos de acurácia, ultrapassando a marca de 90%. A Trint se destacou por sua capacidade de identificar falantes individuais, o que pode ser um recurso valioso para transcrições de reuniões com múltiplos participantes. No entanto, seu modelo de assinatura pode representar um custo significativo para usuários com demandas de transcrição volumosas.

A Sonix e a solução proposta neste artigo apresentaram uma combinação ideal de custo-benefício, precisão e velocidade de transcrição, sendo ambas soluções viáveis para usuários frequentes. Contudo, a Sonix opera com um modelo de cobrança por hora de vídeo transcrita, o que pode se tornar oneroso para usuários com grandes volumes de conteúdo.

A análise detalhada de acurácia revela que, embora existam diversas ferramentas de IA capazes de oferecer transcrições automatizadas, há variações consideráveis em termos de precisão e custo. A ferramenta desenvolvida neste estudo se mostrou notavelmente precisa e econômica, representando um avanço significativo na acessibilidade e eficiência da transcrição automatizada.

Tabela 2 - Acurácia x Ferramenta

	<b>Vídeo 01 - Quantidade de Palavras (144)</b>	<b>Vídeo 02 - Quantidade de Palavras (154)</b>	<b>Vídeo 01 - Quantidade de Palavras (442)</b>
<b>Ferramenta</b>	Palavras corretas	Palavras corretas	Palavras corretas
Google Meet	102	-	-
Microsoft Teams	107	-	-
Trint	132	152	427
Happy Scribe	-	150	-
Sonix	138	152	426
Solução Nova	140	153	438

Fonte: O autor (2023)

Tabela 3 - Porcentagem de Acurácia

<b>Ferramenta</b>	<b>Acurácia (%) - V1</b>	<b>Acurácia (%) - V1</b>	<b>Acurácia (%) - V1</b>
Google Meet	71%	-	-
Microsoft Teams	74%	-	-
Trint	92%	98%	97%
Happy Scribe	0	97%	0
Sonix	95%	98%	96%
Solução Nova	97%	99%	99%

Fonte: O autor (2023)

## 6. CONSIDERAÇÕES FINAIS

### 6.1 SÍNTESE DE ESTUDO

Este estudo abordou uma necessidade emergente resultante da transição global para o trabalho remoto e o ensino a distância, exacerbada pela pandemia de COVID-19. A exigência de documentar eficientemente o conteúdo de reuniões e aulas online levou ao desenvolvimento de uma ferramenta de transcrição automatizada baseada em Inteligência Artificial (IA), projetada para superar as limitações das soluções existentes. Ao longo deste trabalho, investigamos várias ferramentas de transcrição, avaliando-as em termos de acurácia, eficiência de custo e usabilidade, com o objetivo de oferecer uma alternativa superior.

## **6.2 AVALIAÇÃO CRÍTICA DAS FERRAMENTAS**

As ferramentas de transcrição existentes no mercado, como revelado pelas análises, apresentam um espectro variado de eficácia e custo. As plataformas de videoconferência populares, apesar de úteis, mostraram-se insuficientes em termos de funcionalidades de transcrição para propósitos de documentação completa. Por outro lado, as ferramentas especializadas exibiram uma maior precisão, mas muitas vezes a um custo proibitivo, destacando a complexidade de equilibrar acessibilidade e qualidade.

## **6.3 CONTRIBUIÇÕES DO DESENVOLVIMENTO PROPOSTO**

A solução desenvolvida neste trabalho se destacou em todos os critérios estabelecidos. A integração de tecnologias de ponta em IA e um modelo de custo efetivo resultou em uma ferramenta que não apenas atende, mas também excede as expectativas de desempenho e acessibilidade. A capacidade de transcrever com alta precisão e rapidez, juntamente com uma interface amigável, confere a esta solução um potencial significativo para mudar o panorama da transcrição automatizada.

## **6.4 LIMITAÇÕES E DIREÇÕES FUTURAS**

Reconhecemos as limitações do estudo, incluindo a necessidade de testes adicionais em um conjunto mais diversificado de condições de fala e ambientes acústicos. As direções futuras sugeridas incluem a expansão da funcionalidade da ferramenta para suportar múltiplos idiomas e dialetos, bem como a integração de recursos avançados, como a identificação de falantes e a capacidade de gerar resumos automáticos.

## **6.5 CONCLUSÃO SOBRE A INOVAÇÃO E IMPACTO**

Este projeto ilustra a inovação no campo da IA e seu impacto na transcrição de vídeos. A ferramenta proposta oferece não apenas uma solução técnica para um problema imediato, mas também uma contribuição significativa para o campo da IA aplicada ao processamento de linguagem natural. As implicações são vastas, sugerindo melhorias na eficiência, comunicação e gestão de conhecimento em diversos setores.

Por fim, a ferramenta desenvolvida representa um passo significativo em direção a um futuro onde a IA é integral na facilitação de comunicações digitais. As considerações finais deste estudo enfatizam a importância da inovação contínua, da pesquisa aplicada e do

desenvolvimento de soluções que atendam às necessidades reais dos usuários em um mundo em rápida transformação.

## 7. REFERÊNCIAS

YOUTUBE. **Amostra 1**. Disponível em: <https://www.youtube.com/watch?v=NRTMJ72cJ8w>. Acesso em: 26 out. 2023

YOUTUBE. **Número de mortos entre israelenses e palestinos passa de 650**. Disponível em: <https://youtu.be/8hOYkktnJaI?si=HtUIAFwbjSeE3SbY>. Acesso em: 28 out. 2023

YOUTUBE. **Assembleia Geral Ordinária**. Disponível em: [https://youtu.be/7Pr6BzY\\_38I?si=obowF7W1gcP\\_GC9c](https://youtu.be/7Pr6BzY_38I?si=obowF7W1gcP_GC9c). Acesso em: 30 out. 2023

BARÓN BIRCHENALL, L. F. **El juego de imitación de Turing y el pensamiento humano**. Avances en Psicología Latinoamericana, [S. l.], v. 26, n. 2, p. 195-210, 2009. Disponível em: <https://revistas.urosario.edu.co/index.php/apl/article/view/61>. Acesso em: 30 nov. 2023.

CARDENUTO, Giovana. **Métricas de custo e esforço de software**. 2020. TCC (Graduação) - Curso Bacharelado em Engenharia de Software, UniCesumar. Disponível em: [https://www.youtube.com/watch?v=yu4m0hB\\_f3Y](https://www.youtube.com/watch?v=yu4m0hB_f3Y). Acesso em: 30 out. 2023.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Pesquisa sobre acessibilidade das ferramentas de videoconferência em plataforma web**. 2020. Disponível em: <https://acervo.ceweb.br/acervos/conteudo/55a7a288-3bd5-466a-8628-6054d7c81b5c>. Acesso em: 26 out. 2023.

HODGES, Andrew. **Alan Turing: The Enigma**. New York, Simon and Schuster, 1983.  
LEE, Kai-Fu. **Inteligência artificial**. Globo Livros, 2019.

PASSOS, Mauro Romero Leal; JÚNIOR, José Eleutério. **Chatbot, ChatGPT: inteligência artificial e/ou inteligência comercial e/ou inverdades robotizadas, por enquanto**. Brazilian Journal of Sexually Transmitted Diseases, v. 35, 2023.

PEDRO, João. **Melhores softwares de transcrição de vídeo e áudio automática**. 2023. Disponível em: <https://filmora.wondershare.com.br/duio-editing/best-automatic-transcription-software.html>. Acesso em: 11. nov. 2023.

RIBEIRO, Denise; WELLS, Anthony. **Com pandemia, demanda por videoconferências dispara em empresas brasileiras**. 2021. Disponível em: <https://www.cnnbrasil.com.br/economia/com-pandemia-demanda-por-videoconferencias-dispara-em-empresas-brasileiras/>. Acesso: 14. nov. 2023.

TAVARES, Josafá. Conecte-se sem custo: **Descubra as 8 melhores ferramentas gratuitas para videochamada.** 2020. Disponível em: <https://www.mindtek.com.br/2020/03/8-melhores-ferramentas-gratuitas-videochamadas/>. Acesso em: 31 out. 2023.

VIEIRA, Renata; LOPES, Lucelene. **Processamento de linguagem natural e o tratamento computacional de linguagens científicas.** Em corpora, p. 183, 2010.

APÊNDICE

```
igorflpaula, 2 months ago | 1 author (igorflpaula)
generator client {
  provider = "prisma-client-js"
}

igorflpaula, 2 months ago | 1 author (igorflpaula)
datasource db {
  provider = "sqlite"
  url      = env("DATABASE_URL")
}

igorflpaula, 2 months ago | 1 author (igorflpaula)
model Video {
  id          String @id @default(uuid())
  name        String
  path        String
  transcription String?
  createdAt   DateTime @default(now())
}

igorflpaula, 2 months ago | 1 author (igorflpaula)
model Prompt {
  id      String @id @default(uuid())
  title   String
  template String
}
```

APENDICE A - Configuração do arquivo schema.prisma

Video x +					
Filters None Fields All Showing 34 of 34 Add record					
	id A	name A	path A	transcription A? ▼	createdAt ▢
	db03be38-f0ae-45d4-9f...	audio.mp3	/home/igor/Documents/...	abertura, abertura, a...	2023-09-28T18:49:54.8...
	4c15c222-c06f-45ed-8f...	audio.mp3	/home/igor/Documents/...	Olá, vamos retomar o ...	2023-10-05T13:39:39.5...
	ad6eb420-01d2-4aba-a4...	audio.mp3	/home/igor/Documents/...	Olá, meu nome é Carol...	2023-10-05T17:54:07.5...
	3cf5d7a4-a009-4802-8e...	audio.mp3	/home/igor/Documents/...	Olá pessoal, sejam be...	2023-09-28T18:21:52.0...
	262be632-f895-4c36-a4...	audio.mp3	/home/igor/Documents/...	Olá pessoal, sejam be...	2023-09-25T13:46:28.2...
	8a22fb77-58a9-4132-90...	audio.mp3	/home/igor/Documents/...	Olá pessoal, sejam be...	2023-09-25T13:49:38.1...
	5855cd62-e5fe-423e-a1...	audio.mp3	/home/igor/Documents/...	Olá pessoal, sejam be...	2023-09-28T17:43:13.5...
	1c8cadf2-1c71-4360-91...	audio.mp3	/home/igor/Documents/...	Olá pessoal, sejam be...	2023-10-05T13:25:43.5...
	2986020c-0756-46b5-a0...	vintage.mp3	/home/igor/Documents/...	O que é que você está...	2023-09-14T17:53:55.0...
	a5ceef46-f037-4c92-97...	audio.mp3	/home/igor/Documents/...	O que você está fazen...	2023-09-21T20:38:51.8...

APENDICE B - Estrutura da tabela VIDEO

```

app.post('/videos', async (request, reply) => {
  const data = await request.file()

  if (!data) {
    return reply.status(400).send({ error: 'Sem arquivo' })
  }

  const extension = path.extname(data.filename)

  if (extension !== '.mp3') {
    return reply.status(400).send({ error: 'Tipo inválido, a aplicação só aceita arquivos MP3' })
  }

  const fileName = path.basename(data.filename, extension)
  const fileUploadName = `${fileName}-${randomUUID()}${extension}`

  const uploadDestination = path.resolve(__dirname, '../tmp', fileUploadName)

  await pump(data.file, fs.createWriteStream(uploadDestination))

  const video = await prisma.video.create({
    data: {
      name: data.filename,
      path: uploadDestination,
    }
  })

  return { video }
})

```

APENDICE C - Conteúdo da rota uploadVideo