

Aprimoramentos de performance de modelos LLM para aplicações de transcrição de texto em sistemas elétricos de potência

Victor H. Yoshizumi*. Sofia M. A. Lopes*. Danilo H. Spatti**. Rogério A. Flauzino*.
Ivan N. da Silva*. Ivan G. Ricci***. Alexandre G. C. Latorre***. Celso M. L. Junior***

*Departamento de Engenharia Elétrica e de Computação, Escola de Engenharia de São Carlos, Universidade de São Paulo (EESC-USP), São Carlos, São Paulo, Brasil (e-mail: yoshizumi@usp.br, sofia.moreira.lopes@usp.br, raflauzino@usp.br, insilva@sc.usp.br).

** Departamento de Sistemas de Computação, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (ICMC-USP), São Carlos, São Paulo, Brasil (e-mail: spatti@icmc.usp.br).

***ARGO Transmissão de Energia S/A, São Paulo, São Paulo, Brasil
(e-mail: ivan.ricci@argoenergia.com.br, alexandre.latorre@argoenergia.com.br).

Abstract: The use of speech transcribers in everyday operations of electrical power systems has grown in recent years thanks to the increasingly efficiency in models. Such transcribers can also be combined with Generative Artificial Intelligence language models, such as ChatGPT, further increasing their performance. However, in certain cases, given the particularity of the words and sentences to be transcribed, improvements with personalized vocabularies are necessary, allowing specific instructions to be directed to the ChatGPT and thus improving transcription results. This process is known as Prompt Engineering and was used in this work to considerably improve the transcription of telephone conversations between the ONS and the transmission operator.

Resumo: a utilização de transcritores de fala em operações cotidianas dos sistemas elétricos de potência tem crescido nos últimos anos graças ao surgimento de modelos de linguagem cada vez mais eficientes. Tais transcritores podem ser também combinados com modelos de Inteligência Artificial Generativa, tal como o ChatGPT, aumentando ainda mais a sua performance. Porém, em certos casos, tendo em vista a particularidade das palavras e sentenças a serem transcritas, faz-se necessário melhoramentos com vocabulários personalizados permitindo orientar instruções específicas ao ChatGPT e com isso melhorar os resultados de transcrição. Este processo é conhecido como Engenharia de *Prompt* e foi empregado neste trabalho para melhoria considerável da transcrição de conversas telefônicas entre a ONS e o operador de transmissão.

Keywords: Speech Transcription, Speech Recognition, ChatGPT, Prompt Engineering, Whisper Model, LLM Model.

Palavras-chaves: Transcrição de Fala, Reconhecimento de Fala, ChatGPT, Engenharia de Prompt, Modelo Whisper, Modelo LLM.

1. INTRODUÇÃO

A busca pelo refinamento de modelos de transcrição de texto tem crescido nos últimos anos, impulsionada pelo surgimento de diversas propostas inteligentes capazes de realizar de maneira muito consistente o processo de Reconhecimento Automático de Fala (Yoshizumi *et al.* 2023). Este tipo de reconhecimento é fundamental para que aplicações automatizadas sejam capazes de reconhecer palavras e sentenças em uma fala encadeada (Scart *et al.* 2022). Ainda, quando esse tipo de transcrição deve ser empregue em finalidades de segurança de operação ou auditoria, torna-se fundamental o funcionamento com elevadas taxas de conversão, destacando-se um alinhamento de pesquisas cada vez mais promissoras envolvendo a combinação de técnicas de processamento de sinais, extração de características e

ferramentas da área de inteligência, tal como apontado em (Li, 2022, Alharbi *et al.* 2021).

Neste contexto, é possível destacar a Engenharia de *Prompt* como um importante passo em direção a transcrições de texto cada vez mais eficientes quando da utilização de LLMs (*Large Language Models*), tais como o ChatGPT (White *et al.* 2023). A Engenharia de *Prompt* é o que permite que os modelos de Inteligência Artificial Generativa sejam capazes de serem direcionados para tarefas específicas, permitindo ajustes com bases nos erros dos modelos de transcrição (Ekin, 2023, Ye *et al.* 2023). Tal tarefa é necessária pois a maioria destes modelos de transcrição utilizam como dados de aprendizagem um vocabulário genérico e ao se empregar termos e sentenças mais específicas de determinadas áreas, a taxa de erros de transcrição tende a subir. A chamada fraseologia padrão para certas áreas de atuação consistem então em um grande desafio

para o refinamento dos sistemas de transcrição (Helmke *et al.* 2020).

Nos últimos anos, os modelos de Inteligência Artificial Generativa tem ganhado muita notoriedade nas mais diversas aplicações, no entanto, a integração sistêmica entre os LLMs e estas ferramentas não é uma tarefa trivial, sendo também uma área em pleno crescimento, pois, permite a utilização de modelos de transcrição de fala para texto já consagrados, tais como Whisper (Gong *et al.* 2023) em aplicações específicas, desde que a combinação dos LLMs ajustados com vocabulários particularizados atendam as demandadas taxas de assertividade (Min *et al.* 2023).

Apesar das diversas questões que podem ser abordadas com o uso de sistemas de reconhecimento automático de fala em Sistemas Elétricos de Potência (SEPs), há uma escassez de estudos nessa área, com a maioria das pesquisas sendo conduzidas por grupos internacionais, como Jiangping *et al.* (2021), Zhang *et al.* (2021), Yu *et al.* (2020) e Li *et al.* (2020). Isso evidencia uma lacuna significativa no desenvolvimento de aplicações voltadas para o processamento de fala em português do Brasil.

A combinação de Engenharia de *Prompt* para o aperfeiçoamento dos modelos de linguagem é o objeto principal deste trabalho, uma vez que foi criado um vocabulário personalizado para o contexto de operação de Sistemas Elétricos de Potência a fim de aprimorar a performance de modelos de LLM quando aplicados para a correção de transcrições de texto.

Para tais desenvolvimentos foram empregados os procedimentos apontados em (Giray, 2023) e os avanços obtidos com o processamento de áudios em (Yoshizumi *et al.* 2023), onde o cenário alvo destas ferramentas é a transcrição com elevadas taxas de acertos envolvendo os diálogos entre os operadores da ONS (Operador Nacional do Sistema Elétrico) e dos centros de operação de transmissão.

2. SISTEMA DE TRANSCRIÇÃO DE CHAMADAS TELEFÔNICAS

O sistema de Transcrição de Chamadas Telefônicas entre a ONS e os Operadores do Centro de Operações do Sistema Elétrico (COS), fonte de dados desta pesquisa, é o empregado em (Yoshizumi *et al.* 2023), e sua estrutura pode ser sintetizada a partir da Fig. 1.

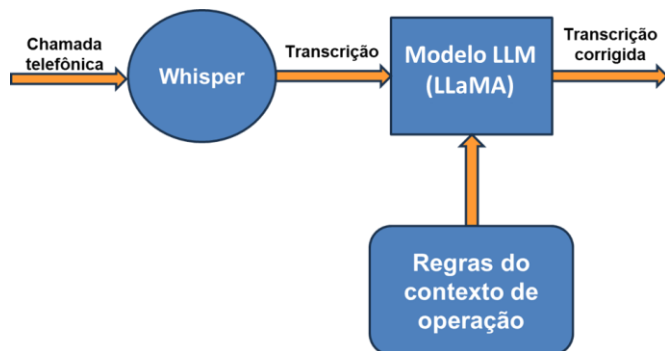


Fig. 1 Modelo de Transcrição de Chamadas Telefônicas.

Sendo assim, o modelo LLM é incorporado ao sistema, de forma que as transcrições fornecidas pelo algoritmo Whisper sejam corrigidas com base em regras pré-determinadas, que incorporam o contexto de operação em linhas de transmissão.

O modelo de linguagem LLaMA (*Large Language Model Meta AI*) foi escolhido para investigação por ser um modelo de código aberto, lançado pela Meta AI em fevereiro de 2023. Com capacidade de ser treinado com até 70 bilhões de parâmetros, o LLaMA permite realizar correções textuais baseadas em regras de contexto. Uma das principais vantagens dessa abordagem é a facilidade de modificar e incorporar novas regras, o que possibilita ao sistema se adaptar facilmente a mudanças (Thirunavukarasu *et al.* 2023). Para compor o *dataset*, chamadas disponibilizadas de dez ocorrências foram então rotuladas nas seguintes classes:

- Chamada Adequada – referente às chamadas nas quais não houve interrupções de fala, ou nas quais tais interrupções ocorreram no início ou no fim da chamada, momentos nos quais usualmente não são repassadas informações de extrema importância.
- Chamada Moderada – referente às chamadas nas quais houve interrupções de fala e estas ocorreram em momentos onde informações importantes estavam sendo repassadas, mas não houve perda da informação.
- Chamada Inadequada – referente às chamadas nas quais houve interrupções de fala, em momentos nos quais informações importantes eram repassadas, e que ocasionaram na perda da informação.

3. TABELA DE GROUNDTRUTH

Para se realizar a validação de resultados em sistemas de reconhecimento de texto, geralmente é empregada a tabela de *GroundTruth*, que tem como pontos principais a transcrição humana e a definição das entidades que estão presentes no áudio. Deste modo, é possível realizar essa validação com base na Taxa de Erros de Palavras (WER – *Word Error Rate*), que quanto menor, indica que o sistema é mais preciso. O WER é também a combinação de três tipos de erros de transcrição que podem ocorrer, sendo eles:

- Erro de inserção (I): palavras presentes na transcrição da hipótese que não estão presentes na verdade.
- Erros de substituição (S): palavras que estão presentes na hipótese e na verdade, mas não foram transcritas corretamente.
- Erros de exclusão (D): palavras que estão faltando na hipótese, mas presentes na verdade.

Um exemplo de como essa tabela de *GroundTruth* é estruturada, com todos os seus aspectos e características, é apresentado na Tabela 1.

Tabela 1. Exemplo de Tabela de *GroundTruth*.

#N	ID	Operador	Transcrição humana			
1	2000_2023_07_06_08_12_36_3334	*	SE Parnaíba Jhon bom dia Ô Jhon Juvenal do COS ARGO Oi Juvenal Ô Jhon agora às 8:13 vamos iniciar a SI PES 683/2023 referente a adequação das proteções de perda de sincronismo das linhas Bacabeira Parnaíba circuitos UNO e 2 e Parnaíba, dos Z Edges do RED 670 tá Ok 8:13 então iniciada a SI 683			
Entidades						
Nomes de interlocutores	Nomes de empresas	Instalação ou subestação envolvida	Equipamento envolvido	Documentos citados	Horário da tratativa	Classificação da manobra
Jhon; Juvenal	COS ARGO; SE Parnaíba; ONS	Linhas Bacabeira Parnaíba	*	SI PES 683/2023	08:13	Proteções de perda de sincronismo

O sistema será então validado com base nesta tabela de *GroundTruth*, utilizando 95 arquivos áudios, provenientes das chamadas telefônicas envolvendo as conversas entre o ONS e os operadores do COS.

4. METODOLOGIA

4.1 Construção de Vocabulário Customizado

A implementação de sistemas inteligentes de reconhecimento automático de fala é fortemente influenciada pela qualidade dos conjuntos de dados utilizados no treinamento desses sistemas. Certas deficiências se tornam evidentes quando esses sistemas são aplicados a contextos ou idiomas não abordados nos conjuntos de dados utilizados para treinamento. No contexto específico da língua portuguesa e dos sistemas elétricos de potência, é observada uma lacuna na capacidade desses sistemas em reconhecer termos comuns do dia a dia da operação dos SEPs.

Por esse motivo, foi realizada a construção de um vocabulário customizado que abrange os principais termos utilizados na operação do SEP, para os quais os sistemas de reconhecimento de fala frequentemente apresentam dificuldades de interpretação.

Para isso, utilizou-se como base uma extensa lista de termos específicos da operação do SEP. Essa lista foi lida e gravada por sete funcionários diferentes da Companhia ARGO Energia. Ressalta-se a importância de gravar a lista de termos com mais de uma pessoa, a fim de avaliar a influência de fatores como gênero, sotaque e cadência da fala na interpretação do vocabulário.

A partir dessas gravações, os arquivos de áudio obtidos foram então submetidos ao sistema de transcrição Whisper. A avaliação da performance do Whisper frente às gravações foi realizada a partir da análise da métrica de WER. Foi desenvolvido um *script* para avaliar a transcrição obtida

pelo Whisper em contraste com uma transcrição realizada por um ser humano. A métrica WER é então calculada avaliando o número de inserções, remoções e substituições verificadas entre os textos.

Para cada áudio foi possível avaliar termo a termo a versão correta e a versão interpretada pelo Whisper. Assim, foi possível identificar os termos que apresentavam maior dificuldade para o sistema. Os erros de transcrição cometidos pelo Whisper não são padronizados, deste modo, um termo não compreendido, pode ser transcrito erroneamente de diferentes formas. Um exemplo é o termo “Acarau”, que apresentou três diferentes escritas pelo Whisper, uma para cada áudio, sendo essas: **acara u**, **a cara u**, **a carau**. Assim, as três variações foram identificadas para a construção da lista de vocabulário.

A tabela de vocabulário é composta de duas colunas, a primeira indica o termo com a escrita correta (esperada), enquanto a segunda coluna apresenta possíveis variações que podem ser utilizadas pelo Whisper. Um exemplo desta tabela pode ser encontrado na Tabela 2.

Tabela 2. Exemplo de Tabela de vocabulário criada.

Termos Corretos	Variações
rele	relogio, relegio
acarau 3	acara u3, a cara u3, a carau 3
act	alfa charlie tango, a ct
acu 3	asu3, a3, a4-3
aes	as
a	alfa
aq	aqu, alfa quebec
ard	alfa romeo delta
argo	argon
arinos 2	arinos 2, arimos 2, arilo 2
ariquemis	arichemes, arigemes
aterramentos	aterramento
aut	alfa uniforme tango

impedimento	alimento
autotrafo	auto trafo, autografo, auto-traff
autotransformador	auto transformador, auto-transformador
bacabeira	baça beira
barramento	barra-mede
bay	bei, b, bi, bain
bcb	bravo charlie bravo

4.2 Instruções para o modelo de LLM

Para realizar a correção dos textos, instruções baseadas no vocabulário personalizado criado foram fornecidas para o modelo de LLM, sendo elas:

1. O idioma do texto é PT-BR;
2. Letras do alfabeto fonético no texto e termos similares devem ser substituídos pelas letras do alfabeto romano. O padrão do texto são as letras do alfabeto latino. Exemplo: "alfa" deve ser escrito como "a", "bravo" como "b", "charlie" como "c", "delta" como "d", "foxtrot" como "f", "Delta, Juliette, Foxtrot 30" deve ser escrito como "DJF30", "FOX30" deve ser escrito como "F30", entre outras variações;
3. Para citar nomes de documentos, seguir o padrão numérico: "239 traço 23" deve ser escrito como "239-23"; "03 UNO ponto 289" como "031.289"; "30 e UNO 28923" como "3128923";
4. As horas devem seguir o formato HH:MM: "8 horas" deve ser escrito como "08:00"; "agora 7 e 5" como "agora 07:05"; "as 10 e 20" como "as 10:20"; "5 horas e 28 minutos" como "05:28";
5. Verificar a lista de nomes frequentemente utilizados;
6. Substituir no texto as variações incorretas encontradas de termos técnicos, nomes de documentos, nomes de empresas citadas, nomes de locais e outras similares, pelos termos corretos da lista.

Para que as análises sejam possíveis, os textos devem ser normalizados, tendo sua pontuação removida e deve também ignorar letras maiúsculas. Além disso, o texto proveniente da transcrição humana deve ainda empregar um código para identificar trechos incompreensíveis como, por exemplo, #\$.#.

Ainda, algumas diretrizes devem ser definidas para os textos. Além do alfabeto padrão, é possível utilizar dígitos de 0 a 9, e que não serão também utilizados símbolos. Palavras podem ser adotadas para números inferiores a 10.

Por fim, deve-se utilizar números para medidas, moeda e grandes numerais, e padronizar o formato das horas. Assim, novas regras podem ser adicionadas futuramente a essa tabela de *GroundTruth*, para que o desempenho do sistema seja aprimorado.

5. RESULTADOS

A Tabela 3 sintetiza o resultado das 10 primeiras amostras de transcrições, quando comparamos o resultado utilizando somente o Whisper (WER_W) com o Whisper juntamente com o vocabulário personalizado (WER_W&V).

Tabela 3. Comparativo WER.

Amostras	WER_W	WER_W&V	Diferença
Amostra01	0,6875	0,4531	-0,2344
Amostra02	0,4201	0,3018	-0,1183
Amostra03	0,2291	0,2123	-0,0168
Amostra04	0,4393	0,2897	-0,1495
Amostra05	0,4935	0,3766	-0,1169
Amostra06	0,4180	0,3934	-0,0246
Amostra07	0,3623	0,2415	-0,1208
Amostra08	0,3941	0,3529	-0,0412
Amostra09	0,5000	0,4836	-0,0164
Amostra10	0,5833	0,5463	-0,0370

Ao considerar todas as 95 amostras de transcrições, 52 amostras apresentaram redução de erros de palavras WER quando se emprega a combinação de Whisper e vocabulário personalizado, com uma redução máxima de aproximadamente 40%, 19 amostras não apresentaram alteração e 24 tiveram aumento do WER. Além disso, a média de WER utilizando somente o Whisper foi de 0.5195, já para a combinação de Whisper com vocabulário personalizado foi de 0.4867, o que representa uma redução global superior a 6%.

Tendo em vista a presença de *outliers* em ambos os casos, pode-se confeccionar um histograma, tal como apresentado na Fig. 2, onde é possível verificar uma elevação do número de amostras no primeiro compartimento, o qual contém as menores taxas de WER.

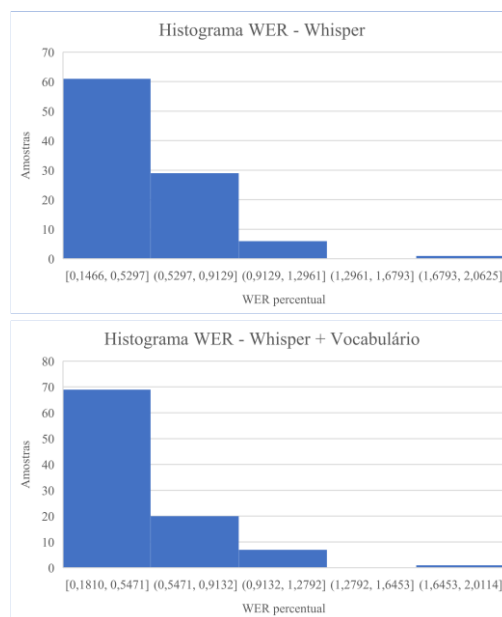


Fig. 2 Histogramas envolvendo o comparativo entre a transcrição usando Whisper com e sem a adição de vocabulário.

Tomando-se como exemplo a Amostra01 da Tabela 3, com WER para Whisper de 0,6875 e Whisper + Vocabulário com 0,4531, é possível construir a Tabela 4 com a comparação da transcrição Humana (H), Whisper (W) e Whisper + vocabulário (WV). De maneira análoga a Tabela 5 foi construída, contendo as mesmas análises para a Amostra09 da Tabela 3.

É possível observar que foram escolhidos os casos com maior e menor redução de WER, entre os apresentados na Tabela 3 para posteriores análises.

Tabela 4. Detalhes das transcrições Amostra01.

Resultado	Transcritor
COS ARGO Fábio É, Gustavo ONS Norte Centro Oeste. Oi Gustavo A Eletronorte então interrompeu a etapa diária da Intervenção 6594, pra linha de Bacabeira Minas 2, circuito 2, então agora 17:01, você pode fechar o disjuntor D30, na subestação de Bacabeira. Positivo, então agora 17:01, a gente vai fechar o disjuntor D30, com o ## da Estação Bacabeira. ## Obrigado. Valeu Ronaldo Obrigado	H
Jair, é o Thiago, Fábio? É, Gustavo, é a Cota de Troés. E o Fábio? A Letronorte então, entreu um pretapa diária da Intervenção 6594, da linha de Bacabeira Minas 2, circuito 2, então agora 17 horas e 1 minuto, você pode fechar o Junitor Delta 30, na subestação de Bacabeira. Ok, então agora 17 horas e 1 minuto, já fechou o Junitor Delta 30, com o entramão da Estação Bacabeira. Com ele em meio dado. Obrigado. Todo mundo, obrigado.	W
Jair, é o Thiago, Fábio? É, Gustavo, é a Equatorial. E o Fábio? A Eletronorte então, entrou com uma solicitação diária da Intervenção 6594, da linha de Bacabeira Minas 2, circuito 2, então agora 17:01, você pode fechar o Disjuntor DJ30, na subestação de Bacabeira. Ok, então agora 17:01, já fechou o Disjuntor DJ30, com o retorno da Estação Bacabeira. Com ele em serviço. Obrigado. Todo mundo, obrigado.	WV

Tabela 5. Detalhes das transcrições Amostra09.

Resultado	Transcritor
COS ARGO Donavan bom tarde O Donavan, é o Ramon na SE Parnaíba 3, tudo bom? O, Ramon, bom demais, filho. Concluimos aqui o ##, ok? E estou entregando para você a SE Parnaíba 3, a LT 500 KV, Parnaíba 3 Tianguá, e o disjuntor 7DJF30 livre para operação, ok? Positivo, então 16:50, entregue pela SE Parnaíba 3, a LT 500 KV Parnaíba 3 Tianguá 2, circuito 1, e o disjuntor F30 . Positivo, Donavan, positivo. Tá, então a gente confirma também a conclusão da SI239. 16:50, conclusão da SI239, ok?	H

Positivo, eu vou disponibilizar para o ONS lá, tá bom? Ok, você me dá um retorno, Donavan? Quando entregar o ONS, por favor? Positivo, eu te informo aí. Tá bom, obrigado. Nada.	
Seu Esquerda, bom ano, bom tarde. Bom dia, meu irmão. Eu, Ramon, nascer, Parna e Batre, tudo bom? Bom, Ramon, bom demais, filho. Concluimos aqui o presidente, ok? E estou entregando para você ser Parna e Batre, ser ele ser 500 KV, Parna e Batre, Xangua, e juntar os 7 Beltas de Juliette, Fox 3, então livre para operação, ok? Positivo, então 1650, entregue pela S.E. Parna e Batre, a LT 500 KV Parna e Batre, Xangua, do circuito 1, e o Juntor Fox 30. Positivo, Domão, positivo. Tá, então a gente confirma também a conclusão da SI239. 1650, conclusão da SI239, ok? Positivo, eu vou descorrer-nos a FOM INES, tá bom? Ok, você me dá um retorno, não é? Quando eu entregar o INES, por favor? Positivo, no time, parna aí. Tá bom, obrigado. Na hora.	W
Seu Esquerda, bom ano, boa tarde. Bom dia, meuirmão. Eu, Ramon, nascer, Parna e Batre, tudo bom? Bom, Ramon, bom demais, filho. Concluimos aqui o presidente, ok? E estou entregando para você ser Parna e Batre, ser ele ser 500 KV, Parna e Batre, Xangua, e juntar os 7 Beltas de Juliette, F30, então livre para operação, ok? Positivo, então 16:50, entregue pela SE Parna e Batre, LT 500 KV Parna e Batre, Xangua, do circuito 1, e o Disjuntor F30. Positivo, Domão, positivo. Tá, então a gente confirma também a conclusão da SGI-239. 16:50, conclusão da SGI-239, ok? Positivo, eu vou descorrer-nos a SGI NOVE TRÊS, tá bom? Ok, você me dá um retorno, não é? Quando eu entregar o INES, por favor? Positivo, no time, Parna aí. Tá bom, obrigado. Na hora.	WV

O comparativo entre as Tabelas 4 e 5 envolvendo as amostras 01 e 09, mostra que ainda são necessárias melhorias em termos da definição do vocabulário, bem como com relação a Engenharia de *Prompt* para que o ChatGPT consiga aperfeiçoar o processo de transcrição com as palavras particularizadas para o caso de Sistemas Elétricos de Potência, porém, é possível verificar que existe uma melhora entre a utilizar apenas do modelo Whisper e com os ajustes propostos neste artigo, objetivando-se um transcritor mais fiel possível ao humano.

6. CONCLUSÕES

Diante dos resultados obtidos pode-se observar que a abordagem proposta contribuiu positivamente para redução do erro de transcrição fornecido pelo sistema. Deste modo, a avaliação do sistema baseada na métrica WER se mostrou eficiente para ajustar um conjunto adequado de regras de correção da transcrição.

No entanto, o conjunto que compõe o vocabulário personalizado, juntamente com a tabela de *GroundTruth*, podem ser aprimorados para que todos os erros sistemáticos

de transcrição observados nos diálogos sejam considerados e, esse ganho de desempenho possa ser verificado para a totalidade das amostras.

Tais melhorias são de grande importância para os avanços do sistema, uma vez que a construção de uma tabela completa de *GroundTruth*, apesar de custosa, agrega um nível elevado de informação ao trabalho, que pode ser utilizado posteriormente pela companhia para padronização da fraseologia de comunicação envolvendo a ONS e os operadores, ou quaisquer outros setores que necessitam de um serviço confirmado.

AGRADECIMENTOS

Este trabalho foi desenvolvido no âmbito do Programa de P&D regulado pela ANEEL, contando também com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Processo 88887.636079/2021-00).

REFERÊNCIAS

- Alharbi, S. et al. (2021). “Automatic speech recognition: Systematic literature review”. *IEEE Access*, 9, 131858-131876.
- Ekin, S. (2023). “Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices”. *Authorea Preprints*.
- Helmke, H. et al. (2020). “Machine learning of air traffic controller command extraction models for speech recognition applications”. In *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)* (pp. 1-9). IEEE.
- Giray, L. (2023). “Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*”, 51(12), 2629-2633.
- Gong, Y. et al. (2023). “Whisper-at: Noise-robust automatic speech recognizers are also strong general audio event taggers”. *arXiv preprint arXiv:2307.03183*.
- Jiangping, J. et al. (2021). “Analysis of power grid dispatching instructions based on Bert-Bigru mode. In: *IEEE Conference on Telecommunications, Optics and Computer Science*, 2021, Shenyang, China.
- Li, C. et al. (2020). “A novel speech recognition algorithm for substation safety notification based on IPASSLSTM”. In: *International Conference on Smart Grids and Energy Systems*, 2020, Perth, Australia.
- Li, J. (2022). “Recent advances in end-to-end automatic speech recognition”. *APSIPA Transactions on Signal and Information Processing*, 11(1).
- Min, Z. e Wang, J. (2023). “Exploring the integration of large language models into automatic speech recognition systems: An empirical study”. In *International Conference on Neural Information Processing* (pp. 69-84). Singapore: Springer Nature Singapore.
- Scart, L. G., Vassallo, R. F. e Samatelo, J. L. A. (2022). “Aplicação de um Modelo Neural para Reconhecimento de Fala em Áudios com Características de Comunicação via Rádio”. In *Congresso Brasileiro de Automática-CBA* (Vol. 3, No. 1).
- Thirunavukarasu, A. J., et al. (2023). “Large language models in medicine”. *Nature medicine*, 29(8), 1930-1940.
- White, J. et al. (2023). “A prompt pattern catalog to enhance prompt engineering with chatgpt”. *arXiv preprint arXiv:2302.11382*.
- Ye, Q. et al. (2023). “Prompt engineering a prompt engineer”. *arXiv preprint arXiv:2311.05661*.
- Yoshizumi, V. H. et al. (2023). “Análise de Interrupções de Áudio em Processos de Reconhecimento de Fala Para Aplicações em Sistemas Elétricos de Potência”. In: *XVI Simpósio Brasileiro de Automação Inteligente (SBAI 2023)*, 2023, Manaus. *Simpósio Brasileiro de Sistemas Elétricos (SBSE)*, 2023.
- Yu, Y. et al. (2020). “Intelligent classification and automatic annotation of violations based on neural network language model”. In: *International Joint Conference on Neural Networks*, 2020, Glasgow, UK.
- Zhang, H. et al. (2021). “Research on speech recognition of power grid dispatching based on big data and deep learning”. In: *International Conference on Power System Technology*, 2021, Haikou, China.