



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS ARARANGUÁ
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE
TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

Gabriel Miranda Cruz da Silva

**Explorando o Potencial do Processamento de Linguagem Natural e
Visualização de Dados em Transcrições de Áudio**

Araranguá
2024

Gabriel Miranda Cruz da Silva

**Explorando o Potencial do Processamento de Linguagem Natural e
Visualização de Dados em Transcrições de Áudio**

Trabalho de Conclusão de Curso submetido ao curso de Graduação em Tecnologias da Informação e Comunicação do Centro de Ciências, Tecnologias e Saúde da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Bacharel em Tecnologias da Informação e Comunicação.

Orientadora: Prof.^a Dra. Marina Carradore Sérgio

Araranguá

2024

Da Silva, Gabriel Miranda Cruz
Explorando o Potencial do Processamento de Linguagem
Natural e Visualização de Dados em Transcrições de Áudio /
Gabriel Miranda Cruz Da Silva ; orientadora, Marina
Carradore Sérgio, 2024.
88 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Campus Araranguá,
Graduação em Tecnologias da Informação e Comunicação,
Araranguá, 2024.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2.
Transcrições de Áudio. 3. Processamento de Linguagem
Natural. 4. Perguntas e Respostas. 5. Modelagem de
Tópicos. I. Sérgio, Marina Carradore. II. Universidade
Federal de Santa Catarina. Graduação em Tecnologias da
Informação e Comunicação. III. Título.

Gabriel Miranda Cruz da Silva

**Explorando o Potencial do Processamento de Linguagem Natural e Visualização de
Dados em Transcrições de Áudio**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de bacharel e aprovado em sua forma final pelo Curso de Tecnologias da Informação e Comunicação.

Araranguá, 17 de dezembro de 2024.

Prof. Fabrício Herpich, Dr.
Coordenação do Curso

Banca examinadora

Prof.^a Marina Carradore Sérgio, Dra.
Orientadora

Prof. Alexandre Leopoldo Gonçalves, Dr.
Universidade Federal de Santa Catarina

Prof. Cristian Cechinel, Dr.
Universidade Federal de Santa Catarina

Araranguá, 2024.

Dedico este trabalho à minha família por todo amor, carinho e apoio que me proporcionaram ao longo de todos esses anos. Sou imensamente grato por tê-los em minha vida.

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, à minha orientadora, Prof.^a Dra. Marina Carradore Sérgio, por todo o apoio, incentivo, paciência, perseverança, e prestatividade ao longo de todo o processo de orientação. Sem o seu apoio, eu certamente não teria conseguido. Desde que a conheci, você sempre se empenhou em inspirar o melhor em mim e nas pessoas ao seu redor. E por isso, sou grato e lhe desejo sucesso na sua jornada como professora e pesquisadora.

Agradeço também a todos os amigos que estiveram presentes ao longo dessa jornada, em especial, ao meu amigo e colega Thiago da Silva Fialho, por sua companhia, apoio, e conselhos ao longo do desenvolvimento deste trabalho. Suas orientações e sugestões me auxiliaram muito a evoluir na escrita científica, algo que foi fundamental para este projeto. Desejo a você muito sucesso e que você realize todos os seus sonhos, em virtude da sua dedicação, sinceridade, e otimismo.

Deixo também meus agradecimentos à minha psicóloga Karmel Nardi por todos os conselhos e conversas que tivemos. Através de algumas delas, tomei decisões que mudaram o rumo da minha vida acadêmica e profissional, e por isso, sou imensamente grato por ter conhecido você.

Por fim, agradeço a cada uma das pessoas que contribuíram, diretamente ou indiretamente, para o sucesso deste projeto, e que me apoiaram ao longo dessa trajetória. A todos, muito obrigado.

RESUMO

Nos últimos anos, o advento do *big data* e a ampla difusão de aplicações que utilizam dados de fala criaram desafios e oportunidades na análise de dados. Nesse contexto, os avanços no processamento de linguagem natural (NLP) e a utilização de técnicas de visualização de dados abriram novas possibilidades de extrair conhecimentos, levando à melhor compreensão de conteúdos falados e a obtenção de *insights*. Sendo assim, este estudo propõe um método para análise de transcrições de áudio, utilizando essas tecnologias para identificar as temáticas presentes e obter respostas para perguntas específicas. Para isso, foram empregados diferentes modelos de linguagem de grande escala (LLMs) em três módulos distintos: (1) transcrição de áudio, (2) análise de perguntas e respostas (Q&A) e (3) modelagem e visualização de tópicos. O primeiro módulo utiliza um modelo para a geração de transcrições de áudio, armazenando-as em um arquivo. O segundo módulo emprega a técnica de geração aumentada por recuperação (RAG) para obter respostas contextualizadas de um LLM, baseadas no conteúdo das transcrições. Por fim, o terceiro módulo emprega um *framework* adaptável, que emprega modelos baseados na arquitetura *transformer*, algoritmos de agrupamento, técnicas de modelagem estatística, e visualização de dados para a obtenção e exploração de prováveis tópicos no conteúdo analisado. Para demonstração e avaliação, o método proposto foi aplicado em um conjunto de 12 horas de áudio de um curso on-line. Os resultados apontam que a abordagem se mostrou eficaz ao utilizar técnicas NLP e visualização de dados para a condução das análises, possibilitando a exploração dos dados tanto em uma perspectiva específica quanto abrangente. Além disso, este estudo forneceu perspectivas em relação à aplicação dessas tecnologias em dados de áudio, destacando algumas de suas vantagens, desvantagens e possibilidades.

Palavras-chave: transcrição de áudio; processamento de linguagem natural; perguntas e respostas; modelagem de tópicos.

ABSTRACT

In recent years, the advent of big data and the widespread adoption of applications that utilize speech data have created challenges and opportunities in data analysis. In this context, advances in natural language processing (NLP) and the use of data visualization techniques have opened new possibilities for knowledge extraction, leading to a better understanding of spoken content and the generation of insights. Therefore, this study proposes a method for analyzing audio transcriptions, using these technologies to identify present themes and obtain answers to specific questions. To achieve this, different large language models (LLMs) were employed in three distinct modules: (1) audio transcription, (2) question-and-answer analysis (Q&A), and (3) topic modeling and visualization. The first module utilizes a model for generating audio transcriptions, storing them in a file. The second module applies the retrieval-augmented generation (RAG) technique to obtain contextualized answers from an LLM based on the transcription content. Finally, the third module employs an adaptable framework that uses transformer-based models, clustering algorithms, statistical modeling techniques, and data visualization to extract and explore probable topics within the analyzed content. For demonstration and evaluation, the proposed method was applied to a dataset of 12 hours of audio from an online course. The results indicate that the approach effectively utilized NLP techniques and data visualization for conducting the analyses, enabling the exploration of data from both specific and broader perspectives. Furthermore, this study provided insights into the application of these technologies to audio data, highlighting some of their advantages, disadvantages, and possibilities.

Keywords: speech transcription; natural language processing; question answering; topic modeling.

LISTA DE FIGURAS

Figura 1 – Visão Geral do Processo de KDD	25
Figura 2 – Estrutura Básica de uma Rede Neural Profunda	29
Figura 3 – Mapa Conceitual de Tarefas de NLP	34
Figura 4 – Evolução dos Modelos de Linguagem	37
Figura 5 – Arquitetura Transformer	39
Figura 6 – Modelo BERT	41
Figura 7 – Funcionamento do RAG	44
Figura 8 – Visão Geral do Método	55
Figura 9 – “Alucinação” em uma transcrição	61
Figura 10 – Resultado da Consulta ao Banco Vetorial	63
Figura 11 – Resposta do Modelo Gemini com RAG	64
Figura 12 – Resposta do Modelo Gemini e Contexto	65
Figura 13 – Modularidade do Framework BERTopic	67
Figura 14 – Tópicos Gerados pelo BERTopic	69
Figura 15 – Distribuição de Fragmentos por Tópico	71
Figura 16 – Proporção de Fragmentos Atribuídos a um Tópico	72
Figura 17 – Pontuação de Termos por Tópico	73
Figura 18 – Matriz de Similaridade Entre Tópicos	74
Figura 19 – Visão Geral do Gráfico de Dispersão	76
Figura 20 – Visão Aproximada de um Tópico	77

LISTA DE QUADROS

Quadro 1 – Metodologia DSRM	48
Quadro 2 – Implementação do Módulo I	56
Quadro 3 – Comparativo das Transcrições Geradas	58
Quadro 4 – Implementação do Módulo II	62
Quadro 5 – Respostas a Perguntas Específicas	65
Quadro 6 – Implementação do Módulo III	67

LISTA DE ABREVIATURAS E SIGLAS

AI	<i>Artificial Intelligence</i>
API	<i>Application Programming Interface</i>
ASR	<i>Automatic Speech Recognition</i>
BERT	<i>Bidirectional Encoder Representations From Transformers</i>
DNN	<i>Deep Neural Networks</i>
DSRM	<i>Design Science Research Methodology</i>
DTM	<i>Document-Term Matrix</i>
GMM	<i>Gaussian Mixture Models</i>
HDBSCAN	<i>Hierarchical Density-Based Spatial Clustering of Applications with Noise</i>
HMM	<i>Hidden Markov Models</i>
IA	<i>Inteligência Artificial</i>
IoT	<i>Internet of Things</i>
IR	<i>Information Retrieval</i>
KDD	<i>Knowledge Discovery in Databases</i>
LLM	<i>Large Language Model</i>
LM	<i>Language Modelling</i>
ML	<i>Machine Learning</i>
NER	<i>Named Entity Recognition</i>
NLM	<i>Neural Language Model</i>
NN	<i>Neural Network</i>
NLP	<i>Natural Language Processing</i>
PLM	<i>Pre-Trained Language Model</i>
RAG	<i>Retrieval Augmented Generation</i>
SLM	<i>Statistical Language Model</i>
SBERT	<i>Sentence-Bidirectional Encoder Representations From Transformers</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>

SUMÁRIO

1 INTRODUÇÃO	16
1.1 OBJETIVO	18
1.1.1 Objetivo geral	18
1.1.2 Objetivos Específicos	18
1.2 JUSTIFICATIVA	18
1.3 ESTRUTURA DO TRABALHO	20
2 FUNDAMENTAÇÃO TEÓRICA	21
2.1 PROCESSAMENTO DE ÁUDIO E RECONHECIMENTO DE FALA	22
2.1.1 Conceitos de som e fala	22
2.1.2 Reconhecimento automático de fala (ASR)	23
2.2 BIG DATA E DESCOBERTA DO CONHECIMENTO	23
2.2.1 Processo de descoberta do conhecimento em bases de dados	24
2.2.2 Visualização de Dados	26
2.3 INTELIGÊNCIA ARTIFICIAL	27
2.3.1 Machine Learning	28
2.3.2 Redes Neurais	28
2.3.3 Deep Learning	30
2.3.4 Métodos de Aprendizado de Máquina	30
2.3.5 Algoritmos de Agrupamento: HDBSCAN	31
2.4 MINERAÇÃO DE TEXTO E PROCESSAMENTO DE LINGUAGEM NATURAL	32
2.4.1 Processamento de Linguagem Natural	33
2.4.2 Representação de dados textuais	34
2.4.3 Word Embeddings	36
2.4.4 Modelos de Linguagem de Grande Escala	37
2.4.5 Arquitetura Transformer	38
2.4.6 Modelo BERT	40

2.4.7 SBERT	41
2.4.8 Questions and Answers (Q&A)	42
2.4.9 Geração Aumentada por Recuperação	43
3 METODOLOGIA	46
3.1 CARACTERIZAÇÃO DA PESQUISA	46
3.2 METODOLOGIA DESIGN SCIENCE RESEARCH METHODOLOGY	46
3.3 DESENVOLVIMENTO DA PESQUISA	47
3.4.1 População e Amostra	49
3.4.2 Coleta de Dados e Procedimentos	50
3.4.3 Análise de dados	51
3.4.4 Resultados esperados	52
3.4.4.1 Módulo 1: Transcrição e Criação de um JSON de Transcrições	52
3.4.4.2 Módulo 2: Q&A com Modelo Gemini e RAG	53
3.4.4.3 Módulo 3: Modelagem de Tópicos com BERTopic e Visualização de Dados	53
3.4.4.4 Resultados gerais esperados	54
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	55
4.1 MÓDULO I: TRANSCRIÇÃO	56
4.1.1 Teste Inicial	57
4.1.2 Análise das Transcrições	60
4.2 MÓDULO II: Q&A	61
4.2.1 Análise dos Componentes do Q&A	63
4.2.2 Perguntas e Respostas (Q&A)	65
4.3 MÓDULO III: MODELAGEM DE TÓPICOS	66
4.3.1 Análise Preliminar dos Resultados	69
4.3.2 Visualização de Tópicos	70
4.3.2.1 Distribuição de Fragmentos por Tópico	70
4.3.2.2 Proporção de Outliers	71
4.3.2.3 Distribuição de Palavras por Tópico	72

4.3.2.4 Matriz de Similaridade Entre Tópicos	74
4.3.2.5 Gráfico de Dispersão	75
5 CONSIDERAÇÕES FINAIS	79
5.1 LIMITAÇÕES DA PESQUISA	79
5.2 TRABALHOS FUTUROS	80
REFERÊNCIAS	81

1 INTRODUÇÃO

Nas últimas décadas, a evolução tecnológica impulsionou a produção e o armazenamento de grandes volumes de dados de forma contínua. Mahmoudian *et al.* (2023) destacam que o surgimento de aplicações web e *mobile*, combinado com o avanço da Internet das Coisas (IoT) e da computação em nuvem, intensificaram esse fluxo de informações. Na visão dos autores, esse fenômeno, conhecido como *big data*, tem atraído a atenção de pesquisadores e organizações, que buscam maneiras eficazes de interpretar e analisar esses dados massivos.

Segundo Cook, Lee e Majumder (2016), os dados são um recurso fundamental para entender o mundo, reunindo informações oriundas de diversas fontes. Essa diversidade caracteriza uma ampla gama de formas de dados e representa uma oportunidade significativa para extrair informações valiosas que auxiliam na tomada de decisões (Keim; Qu; Ma, 2013). Assim, os dados são considerados uma fonte essencial de conhecimento, impulsionando avanços na ciência de dados, além de se configurarem como um diferencial competitivo no mercado (Mauro; Greco; Grimaldi, 2015).

Dentre esse vasto volume de dados gerados, os dados de fala representam uma fração significativa, sendo encontrados em diversos contextos. Mehrish *et al.* (2023) ressaltam que o campo do processamento de fala tem adquirido crescente importância, com aplicações nas áreas de telecomunicações, entretenimento e saúde. Desse modo, aplicações que envolvem dados de fala têm se tornado cada vez mais presentes, destacando o papel dos dados de áudio em análises.

Nesse cenário, Chen *et al.* (2020) destacam a visualização de dados como um componente central e indispensável na descoberta de conhecimento em diversas disciplinas. De acordo com os autores, esse processo possui o potencial de revelar padrões relevantes, relações ocultas e novas conexões, utilizando de forma eficaz os mecanismos de percepção visual humana. Dessa forma, a visualização de dados facilita a interpretação de resultados e fomenta novas descobertas, o que evidencia sua relevância na análise de grandes volumes de dados (Ali *et al.*, 2016).

Além disso, Tucker, Capps e Shamir (2020) apontam que os avanços nas ferramentas tecnológicas viabilizaram estudos e análises antes impraticáveis por métodos manuais, ampliando significativamente as oportunidades no campo da

descoberta de conhecimento. No contexto desses avanços, a combinação de técnicas de Processamento de Linguagem Natural (do inglês *Natural Language Processing* – NLP) e visualização de dados aplicada a transcrições de áudio surge como uma abordagem promissora, capaz de identificar padrões complexos e gerar *insights* em diversas áreas do conhecimento.

O NLP, um ramo da inteligência artificial, inclui diversas tarefas voltadas para a interação entre máquinas e linguagem humana. Entre as tarefas mais relevantes para este trabalho, destacam-se:

- Transcrição de áudio para texto: Conversão de gravações de voz em texto estruturado;
- Perguntas e Respostas (do inglês *Questions and Answers* – Q&A): Sistemas que respondem a perguntas baseadas em bases de dados textuais;
- Modelagem de Tópicos: Identificação de temas recorrentes em grandes volumes de dados textuais.

Nesse cenário marcado pelo *big data* e pelos avanços tecnológicos, torna-se essencial desenvolver métodos e técnicas que transformam dados em conhecimento. Paralelamente, as tecnologias emergentes abrem novas possibilidades para análises inovadoras. Assim, este trabalho propõe o desenvolvimento de um método para a transcrição automatizada de arquivos de áudio em texto. Essas transcrições serão utilizadas como base para a aplicação de técnicas de processamento de linguagem natural (NLP) e visualização de dados, permitindo análises mais aprofundadas e facilitando a interpretação dos resultados obtidos.

Com base nessa proposta, a pesquisa busca responder à seguinte questão: Como técnicas de processamento de linguagem natural e visualização de dados podem contribuir para explorar, analisar e extrair conhecimentos presentes em transcrições de áudio?

1.1 OBJETIVO

Para melhor compreensão dos objetivos deste projeto, estes foram divididos em objetivo geral e objetivos específicos, delimitados a seguir.

1.1.1 Objetivo geral

Propor um método para análise de transcrições de áudio, capaz de responder a perguntas sobre as transcrições.

1.1.2 Objetivos Específicos

- Desenvolver um módulo de transcrição para conversão de áudios em texto;
- Criar um mecanismo de perguntas e respostas que possibilite obter informações específicas acerca do conteúdo das transcrições;
- Utilizar algoritmos de processamento de linguagem natural para identificar os principais tópicos abordados nas transcrições.

1.2 JUSTIFICATIVA

Keim, Qu e Ma (2013) definem a era contemporânea como uma era orientada por dados, na qual a coleta e análise de informações desempenham um papel essencial na fundamentação de decisões rápidas e eficazes em setores como negócios, saúde e segurança. Esse cenário foi potencializado pelo fenômeno do *big data*, que, conforme destacado por Ali *et al.* (2016), vem transformando diversos setores, incluindo governo, academia e indústria, impulsionado pelo crescimento da Internet das Coisas (IoT) e pela digitalização massiva de registros físicos.

Auber *et al.* (2024) apontam que um dos maiores desafios atuais é oferecer suporte para a análise de grandes volumes de dados heterogêneos, especialmente para usuários sem formação técnica específica, como jornalistas de dados, cientistas sociais e formuladores de políticas públicas. Nesse contexto, a visualização de dados se apresenta como uma ferramenta poderosa, permitindo que

mesmo usuários não especializados compreendam e extraiam conhecimento de grandes volumes de informações, viabilizando inferências que seriam inviáveis apenas por métodos tradicionais.

Cook, Lee e Majumder (2016) enfatizam que a capacidade de utilizar visualizações de dados de forma eficaz tornou-se uma habilidade essencial no mundo contemporâneo. Segundo os autores, a aplicação adequada de visualizações é fundamental para lidar com os desafios impostos pelo *big data*, transformando dados brutos em recursos acionáveis para estudos científicos e decisões estratégicas. Ward, Grimstein e Keim (2015) corroboram essa visão, destacando a necessidade de desenvolver técnicas e ferramentas que comuniquem informações de maneira eficiente, dado o crescimento exponencial no volume de dados gerados diariamente.

No domínio específico dos dados de áudio de fala, Mehrish *et al.* (2023) destacam que os avanços em *deep learning* têm revolucionado o campo do processamento de fala, eliminando a necessidade de engenharia manual de características. A capacidade de extrair automaticamente informações valiosas de sinais de áudio brutos não apenas melhorou o desempenho e a precisão, mas também abriu novas possibilidades para resolver desafios práticos em áreas como educação, saúde e mídia.

Considerando o crescente volume de informações textuais e não textuais derivadas de dados de áudio, os avanços no processamento de linguagem natural (NLP), aliados ao uso de modelos como GPT, Gemini e BERT, têm transformado a maneira como essas informações são analisadas e interpretadas. Essas tecnologias, integradas a ferramentas de visualização dinâmica e interativa, criam oportunidades para a descoberta de conhecimento e para a comunicação de *insights*.

Ao integrar técnicas de transcrição automatizada, perguntas e respostas (Q&A), modelagem de tópicos e visualização de dados, este projeto busca atender à demanda crescente por ferramentas que combinem eficiência e facilidade de acesso, promovendo avanços em áreas interdisciplinares e orientadas por dados.

1.3 ESTRUTURA DO TRABALHO

O primeiro capítulo introduz o tema do estudo, delineando os objetivos gerais e específicos, além de apresentar a justificativa para a realização da pesquisa.

O segundo capítulo desenvolve a fundamentação teórica, explorando conceitos fundamentais relacionados ao processamento de fala e transcrição de áudio, *big data*, inteligência artificial e subáreas relacionadas, visualização de dados, mineração de texto e processamento de linguagem natural.

No terceiro capítulo, é descrita a metodologia adotada, detalhando os processos de transcrição de áudio, Q&A e *Retrieval Augmented Generation* (RAG), modelagem de tópicos, e aplicação de técnicas de visualização de dados.

O quarto capítulo apresenta os resultados obtidos a partir das análises realizadas, com ênfase nos *insights* revelados pelas respostas geradas pelos modelos de linguagem de grande escala (LLMs), pelas visualizações de dados e pelo agrupamento semântico.

Por fim, o quinto capítulo traz as considerações finais, refletindo sobre as contribuições deste trabalho e propondo possíveis direções para futuras pesquisas no campo de processamento de linguagem natural, visualização de dados e processamento de áudio.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, serão abordados os conceitos teóricos essenciais para a compreensão dos pilares que sustentam este trabalho. A era da informação, caracterizada pela explosão na geração e complexidade dos dados, reforça a relevância de tecnologias como o processamento de fala, o processamento de linguagem natural e a visualização de dados para a análise e interpretação de informações extraídas de dados sonoros.

O capítulo inicia com uma explanação sobre os fundamentos do processamento de áudio e fala, incluindo tópicos como a representação digital do som, a extração de características acústicas e o reconhecimento automático de fala. Na sequência, são apresentados os principais conceitos de *big data*, descoberta do conhecimento em bases de dados e visualização de dados, destacando o papel dessas técnicas na interpretação de grandes volumes de informações. Também será introduzido o campo da inteligência artificial, com foco em suas aplicações no reconhecimento de padrões em dados de fala.

No contexto do processamento de linguagem natural (NLP), será discutido o potencial de tecnologias avançadas, como o *Retrieval-Augmented Generation* (RAG) e os algoritmos de modelagem de tópicos, utilizando modelos como Gemini e BERT, para compreender o significado contextual por trás das palavras e responder a questões específicas em transcrições de áudio. Além disso, serão exploradas técnicas de agrupamento, como o HDBSCAN, que permitem a identificação de padrões e a classificação de informações relevantes a partir de grandes conjuntos de dados textuais derivados de transcrições.

Esses conceitos interligados formam a base para as análises e metodologia desenvolvidas nos capítulos subsequentes. Ao integrar tecnologias de processamento de áudio, processamento de linguagem natural e visualização de dados, este trabalho busca explorar as oportunidades tecnológicas oferecidas para a extração de conhecimento a partir de dados sonoros, contribuindo para avanços em áreas como educação, saúde, e mídia.

2.1 PROCESSAMENTO DE ÁUDIO E RECONHECIMENTO DE FALA

2.1.1 Conceitos de som e fala

O som é um fenômeno ondulatório que origina da ação vibratória de algum corpo, caracterizado por ondas de pressão que se propagam pelo meio de forma longitudinal e transversal (Christensen, 2019). Essas ondas são captadas pelo sistema auditivo humano e interpretadas como som, desempenhando um importante papel na comunicação, na orientação espacial, na detecção de eventos, dentre outros aspectos. De acordo com Li, Drew e Liu (2021), as variações de pressão que compõem o som podem ser convertidas em sinais elétricos por meio de transdutores, o que dá origem a um sinal de som contínuo transformável em um sinal discreto.

Desse modo, os autores destacam a amplitude e a frequência como propriedades físicas do som, associadas à percepção de intensidade e tom, respectivamente. Normalmente, as etapas de amostragem e quantização são utilizadas para extrair essas propriedades. Após a conversão para o formato digital, é possível reproduzir, armazenar, transmitir e modificar sinais de áudio por meio de softwares e aplicações.

Nesse contexto, entende-se como sinal de fala um sinal de áudio que contém dados de fala, cujas características captam diferentes informações (Mehrish *et al.*, 2023). Essas características são classificadas de acordo com o domínio do tempo e da frequência, e podem ser usadas em conjunto para realizar determinadas tarefas, como síntese, análise, e reconhecimento de fala. Adalmarki *et al.* (2022) complementam essa ideia ao citar determinados padrões reconhecíveis na fala, como o idioma, as emoções, a identidade do falante e a transcrição textual do conteúdo falado.

Sendo assim, entende-se que dados de fala apresentam aspectos variados que são reconhecidos e representados por técnicas de processamento de fala, abrangendo não só o conteúdo verbal, mas também informações sobre o falante. Neste estudo, o conteúdo verbal será o foco da análise.

2.1.2 Reconhecimento automático de fala (ASR)

De acordo com Chen *et al.* (2021), o campo de processamento de fala é vasto e engloba diversas tarefas, que abrangem desde a extração de informações de sinais de fala até a geração desses sinais. Desse modo, Gabler *et al.* (2023) definem o reconhecimento automático de fala (do inglês *Automatic Speech Recognition* – ASR) como uma área consolidada e em constante desenvolvimento, cuja principal aplicação é converter sinais de fala em texto. Linke *et al.* (2024) reforçam essa ideia ao afirmar que o objetivo geral de uma arquitetura de ASR é prever uma sequência de palavras a partir de um sinal de fala.

Além disso, os autores apontam que o campo de ASR desenvolveu diversas abordagens e metodologias ao longo do tempo, com vantagens e desafios próprios. Nesse sentido, Linke *et al.* (2024) destacam três abordagens influentes: Modelos Ocultos de Markov com Modelos de Mistura Gaussiana (HMM-GMM), Modelos Ocultos de Markov com Redes Neurais Profundas (HMM-DNN) e arquiteturas de ASR baseadas em *transformer*.

Mehrish *et al.* (2023) afirmam que, recentemente, o campo de processamento de fala tem experienciado significativos avanços por conta do advento do *deep learning*, cuja capacidade de reconhecer características relevantes em sinais de fala de maneira automática dispensa a engenharia manual de características. O estudo destaca que esse fato tem contribuído significativamente para melhorias na performance, robustez, flexibilidade e precisão dos sistemas de processamento de fala, especialmente em condições desafiadoras, como ruídos e variações linguísticas.

Portanto, sistemas modernos de ASR podem ser utilizados para criar aplicações que transformam arquivos de áudio em fontes de dados textuais de maneira mais confiável e eficaz.

2.2 BIG DATA E DESCOBERTA DO CONHECIMENTO

Na era atual, grandes volumes de dados são produzidos a todo momento. Entretanto, observam-se diferentes entendimentos acerca do que é “*big data*”. Segundo Gandomi e Haider (2015), embora a rápida evolução do termo tenha ocasionado certa confusão a respeito do seu significado, a definição de Laney

(2001) tem sido um referencial comum para descrevê-lo, conforme pontuam os autores ao citar Chen *et al.* (2012) e Kwon *et al.* (2014).

Laney (2001) estabelece três dimensões de desafios no gerenciamento de dados, conhecidas como os 3 Vs do *big data*: “Volume”, “Velocidade”, e “Variedade”. Entretanto, L'Heureux *et al.* (2017) afirmam que a definição de Ohlhorst (2012) tem sido mais bem aceita, a qual acrescenta um quarto V às dimensões de Laney: a “Veracidade”. De acordo com os autores:

- “Volume” refere-se à grande quantidade de dados a serem gerenciados.
- “Velocidade” refere-se ao ritmo acelerado em que dados são gerados;
- “Variedade” refere-se à ampla diversidade de formatos, estruturas e semânticas em que dados se apresentam.
- “Veracidade”, por sua vez, aborda desafios relacionados à qualidade, procedência, incertezas, e ruídos nos dados.

Por outro lado, L'Heureux *et al.* (2017) ressaltam que outros Vs podem ser encontrados na literatura, como nos estudos de Fan e Bifet (2012) e Demchenko *et al.* (2013), que incluem um quinto V, o “valor”. Desse modo, entende-se que o termo “*big data*” se refere a um fluxo de dados caracterizado por grandes proporções, ritmo acelerado de geração, alta variedade de formatos e estruturas, além de diferentes procedências e graus de qualidade, mas que podem abranger outros aspectos representados por outros Vs. Essas características demandam tecnologias e estruturas específicas para gerenciá-los, o que na definição dos autores, evidencia os desafios decorrentes de sua complexidade.

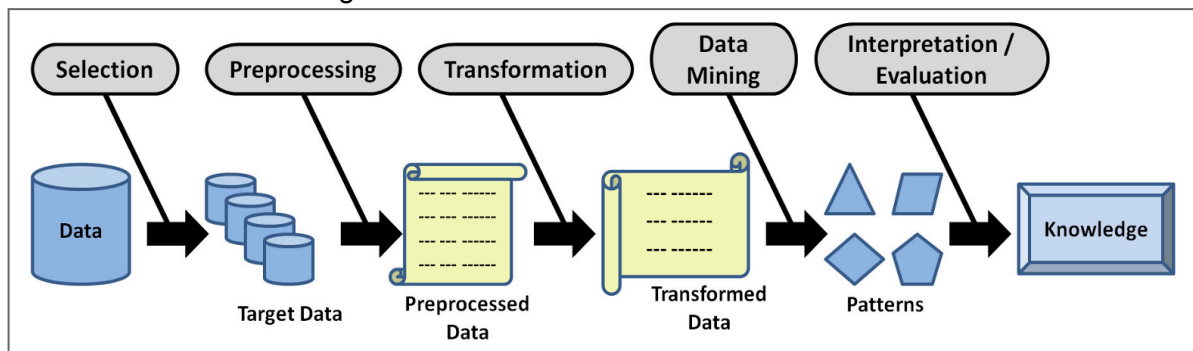
2.2.1 Processo de descoberta do conhecimento em bases de dados

Diante desse entendimento, os 4 Vs do *big data* destacam os desafios no seu gerenciamento, enquanto a capacidade de superá-los representa oportunidades de obter conhecimentos úteis a partir de informações aparentemente desconexas. Nesse contexto, a descoberta do conhecimento em bases de dados (*Knowledge Discovery in Databases - “KDD”*) emerge como um campo que busca dar sentido aos dados através de métodos e técnicas (Fayyad *et al.*, 1996). De acordo com os autores, a KDD é definida como um processo de identificar nos dados determinados padrões, que são:

- Não-triviais: Envolvem certa busca ou inferência. Ou seja, não resultam da computação direta de valores pré-definidos;
- Válidos: Aplicam-se a novos dados com certo nível de certeza;
- Novos: Não eram conhecidos até o momento da análise;
- Potencialmente úteis: Proporcionam algum benefício ou vantagem ao usuário ou a tarefas subsequentes;
- Compreensíveis: Seja de imediato ou após novas etapas de processamento.

O processo de KDD, ilustrado na Figura 1, é interativo e iterativo, dividido em cinco etapas principais: (i) seleção, (ii) pré-processamento, (iii) transformação, (iv) mineração de dados e (v) interpretação/avaliação.

Figura 1 – Visão Geral do Processo de KDD



Fonte: Gullo (2015).

Gullo (2015) descreve o processo de Fayyad *et al.* de acordo com suas etapas. Segundo o autor, o KDD começa com a fase de seleção, na qual o objetivo é obter, a partir de um conjunto de dados, um subconjunto de variáveis ou amostras pertinentes à descoberta. Depois disso, no pré-processamento, são realizadas operações de limpeza dos dados, visando remover ruídos e lidar com dados faltantes, levando em consideração informações de tempo e sequência. Na terceira fase, de transformação, busca-se reduzir e transformar os dados para obter uma representação adequada para a tarefa a ser realizada.

Em seguida, a quarta fase, de mineração de dados, envolve a extração de padrões por meio de métodos específicos, algoritmos adequados para a tarefa e representações apropriadas das saídas resultantes. As principais tarefas de mineração de dados incluem classificação, predição (tarefas preditivas), agrupamento, associação, e sumarização (tarefas descritivas) (Safhi; Frikh; Ouhbi,

2018). Na definição de Fayyad *et al.* (1996), a mineração de dados não é um processo à parte, mas uma etapa do processo de descoberta do conhecimento em bases de dados.

Ainda segundo Gullo (2015), na quinta e última fase, de interpretação/avaliação, o usuário explora os padrões minerados para extrair e interpretar conhecimentos, utilizando visualizações para facilitar a compreensão dos padrões, modelos ou dados. Desse modo, a visualização de dados demonstra um papel importante no processo de descoberta do conhecimento, provendo uma interface entre o usuário e os padrões obtidos. A seguir, serão abordados seus principais conceitos.

2.2.2 Visualização de Dados

Chen *et al.* (2020) definem a visualização de dados como a transformação de dados provenientes de modelos, simulações ou medições em representações gráficas para apresentação, exploração ou análise. Para os autores, um dos principais pontos fortes dessa abordagem é o uso eficiente da visão e cognição, permitindo a rápida compreensão e exploração de conceitos e informações, além da observação de fenômenos, correlações, tendências, e conexões que complementam conhecimentos e hipóteses.

Schmitt (2020) reforça essa ideia ao destacar a importância da visualização de dados na análise exploratória, apresentando-a como um meio não apenas de demonstrar e comunicar resultados, mas também de descobrir padrões ocultos nos números, tornando-os mais inteligíveis. Assim, a visualização de dados se configura como um componente essencial na compreensão e descoberta de conhecimento em grandes volumes de dados, evidenciando, por meio de diferentes representações gráficas, informações que, de outra forma, passariam despercebidas.

No contexto organizacional, Balusamy *et al.* (2021) afirmam que a visualização de dados é um complemento essencial no ciclo de vida de análise do *big data*, uma vez que, sem ela, a interpretação dos resultados se restringe aos analistas, limitando seu potencial na tomada de decisões informadas. Desse modo, a visualização de dados é uma ferramenta que torna o conhecimento extraído no processo de análise de dados acessível a partes interessadas em todos os níveis organizacionais.

Sendo assim, o uso adequado de visualizações de dados não apenas melhora a compreensão de informações, mas também viabiliza a interpretação e avaliação dos padrões identificados no processo de descoberta do conhecimento. Existem diferentes tipos de representações gráficas aplicáveis a dados textuais, como nuvem de palavras (*word clouds*), gráfico de barras, e gráfico de dispersão. Entretanto, a escolha de quais utilizar varia de acordo com fatores como número e tipos de variáveis e objetivos da análise. Mais adiante, na seção 2.4, serão apresentados determinados tipos de representações textuais, a partir das quais são geradas as visualizações que revelam os padrões descobertos.

2.3 INTELIGÊNCIA ARTIFICIAL

Russell e Norvig (2021) definem a inteligência artificial (do inglês *Artificial Intelligence – AI*) como uma área multidisciplinar que busca perceber, compreender, prever e manipular o mundo. Mais do que isso, visa criar entidades inteligentes que possam calcular como agir de maneira segura e eficaz em diversas situações. Segundo os autores, existem quatro métodos distintos para descrever e modelar a inteligência artificial, que se dividem em duas principais dimensões: humano *versus* racional, e pensamento *versus* comportamento.

Na abordagem humana, explica-se que o objetivo é desenvolver uma inteligência semelhante à do ser humano, envolvendo observações e hipóteses sobre comportamento e processos de pensamento, configurando-se, em parte, como uma ciência empírica ligada à psicologia. Por outro lado, a abordagem racionalista utiliza conceitos de matemática e engenharia, ligando-se à estatística, economia e teoria de controle. Embora distintas, essas abordagens estabeleceram princípios fundamentais que orientam o desenvolvimento da inteligência artificial.

Na era contemporânea, Satyam e Geetha (2023) descrevem a inteligência artificial como um campo em constante evolução que tem transformado a maneira como se vive e trabalha, apresentando uma influência crescente em áreas como saúde, transporte, finanças e educação. Esse potencial inovador abrange desde a melhoria da eficiência e precisão dos sistemas existentes até a criação de novos sistemas, possibilitando descobertas e avanços.

Atualmente, existem diversas aplicações de AI, divididas em subcampos especializados em tarefas específicas. Entre as suas principais disciplinas, pode-se

mencionar o processamento de linguagem natural, a representação do conhecimento, o raciocínio automatizado, o aprendizado de máquina, a visão computacional, e a robótica (Russell; Norvig, 2021). Na seção 2.4, serão abordados os conceitos fundamentais de processamento de linguagem natural, subcampo destacado neste estudo.

2.3.1 Machine Learning

O *machine learning* (aprendizado de máquina) é um ramo da inteligência artificial que utiliza dados e algoritmos para replicar o processo de aprendizado humano, aprimorando sua precisão de forma contínua (IBM, 2024b). Em outras palavras, o ML possibilita que máquinas aprendam com a experiência de maneira similar ao ser humano, utilizando algoritmos de aprendizagem baseados em dados para obter modelos capazes de realizar previsões em novas observações (Zhou, 2021). Esses modelos analíticos são treinados com dados de um problema ou questão específicos, visando resolver tarefas relacionadas (Janiesch; Zschech; Heinrich, 2021).

LeCun, Bengio e Hinton (2015) apontam que as técnicas tradicionais de *machine learning* possuíam limitações significativas ao processar dados brutos. Segundo os autores, a construção de um sistema de *machine learning* que pudesse extrair e representar adequadamente as características dos dados, utilizadas para tarefas como reconhecimento ou classificação de padrões, exigia uma engenharia meticulosa e um alto nível de expertise no domínio. No entanto, os avanços no campo mudaram significativamente esse cenário.

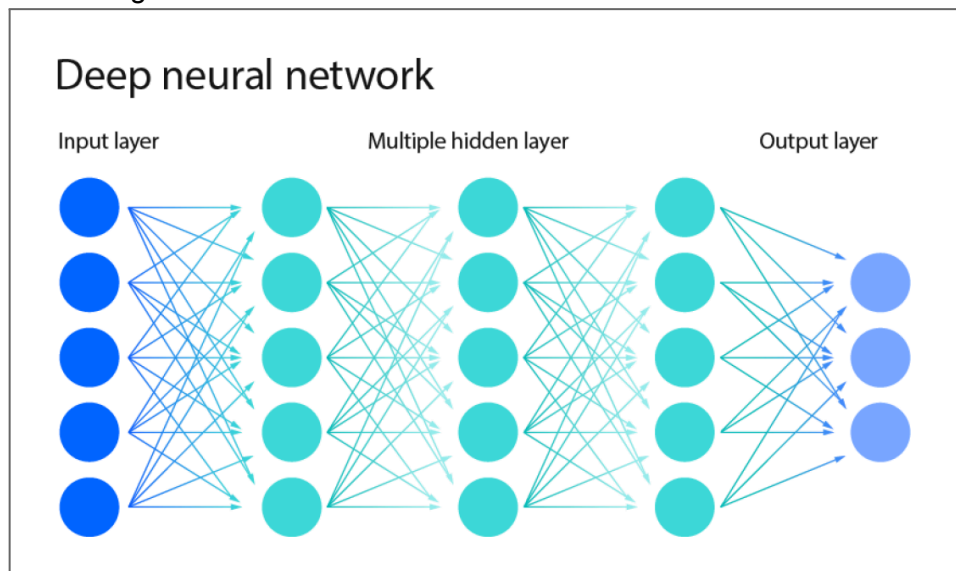
2.3.2 Redes Neurais

Nesse contexto, as redes neurais (NNs) se mostram como um subcampo do *machine learning*, cuja popularidade tem aumentado graças ao seu desempenho de ponta em uma ampla gama de tarefas (DeVore; Hanin; Petrova, 2020). De acordo com Uhrig (1996), as redes neurais são sistemas de processamento de dados que utilizam uma arquitetura composta por múltiplos elementos básicos interconectados, inspirada no córtex cerebral. Dessa forma, elas permitem que computadores

realizem tarefas em que os humanos geralmente têm bom desempenho, mas que são desafiadoras para as máquinas.

Na definição do autor, uma rede neural é composta por elementos de processamento interconectados, com funções de ponderação ajustáveis para cada entrada. Esses elementos, semelhantes a neurônios, são ativados quando as entradas atingem um limiar definido por uma função de ativação. Além disso, eles geralmente são organizados em três ou mais camadas: uma camada de entrada, onde os dados são introduzidos; uma ou mais camadas intermediárias, chamadas de camadas ocultas; e uma camada de saída, onde ocorre a resposta para a entrada fornecida. *Buffers* são utilizados tanto na camada de entrada quanto na de saída.

Figura 2 – Estrutura Básica de uma Rede Neural Profunda



Fonte: IBM (2024a).

Por fim, Uhrig (1996) destaca dois processos fundamentais na operação de uma rede neural artificial: aprendizado (*learning*) e recuperação (*recall*). Durante o aprendizado, os pesos das conexões são ajustados com base nos exemplos fornecidos no *buffer* de entrada, seguindo uma regra que define como esse ajuste deve ocorrer. Na recuperação, a resposta para uma entrada é determinada pelo conhecimento previamente adquirido pela rede. Desse modo, as redes neurais são capazes de se adaptar e fornecer saídas específicas com base no conhecimento adquirido durante o treinamento.

2.3.3 Deep Learning

Zhou (2021) afirma que os avanços no poder computacional e o *big data* possibilitaram a criação de redes neurais com múltiplas camadas densas, característica que define o *deep learning*. Sendo assim, entende-se como *deep learning* um tipo específico de rede neural que possui mais de três camadas — redes neurais com essa característica podem ser consideradas redes neurais profundas ou algoritmos de aprendizado profundo (IBM, 2024a). Dessa forma, o *deep learning* aumenta a complexidade da rede, possibilitando a resolução de problemas que envolvem várias camadas de processamento.

LeCun, Bengio e Hinton (2015) descrevem os métodos de *deep learning* como técnicas de aprendizado de representação que, a partir de dados brutos, obtêm automaticamente as representações necessárias para classificar ou detectar padrões. Nesse sentido, o *deep learning* permite a utilização de múltiplas camadas de abstração, possibilitando o aprendizado de funções complexas quando um número suficiente de abstrações é alcançado.

Entretanto, os autores destacam que o diferencial do *deep learning* é que essas camadas de características não são projetadas por humanos, mas aprendidas diretamente dos dados por meio de um procedimento de aprendizado de propósito geral. Devido a essas características, a pesquisa indica que o *deep learning* tem se mostrado extremamente eficiente na identificação de estruturas complexas em dados de alta dimensionalidade.

2.3.4 Métodos de Aprendizado de Máquina

De acordo com Nadkarni (2016), os métodos de aprendizado de máquina são classificados como supervisionados, não supervisionados, e semissupervisionados. Segundo a autora, no aprendizado supervisionado, os valores das saídas do programa são previamente conhecidos na fase de treinamento. No aprendizado não supervisionado, o software organiza as entradas, mas os valores de saída são desconhecidos ou inexistentes. Já o aprendizado

semisupervisionado, utiliza uma abordagem mista; algumas saídas são previamente conhecidas, mas outras não.

Na definição de Saravanan e Sujatha (2018), o aprendizado supervisionado é um método que utiliza dados rotulados durante o processo de treinamento. Ao fazer uma previsão, o algoritmo compara os resultados obtidos com os esperados, identificando erros e ajustando o modelo conforme necessário. Em contraste, o aprendizado não supervisionado dispensa o uso de dados rotulados, sendo aplicado quando os dados não possuem classificação ou descrição. Nesse tipo de abordagem, os autores ressaltam que o sistema não identifica a saída correta, mas é capaz de descobrir padrões ocultos nos dados ao deduzir funções para explicá-los.

Sendo assim, a escolha da abordagem para o treinamento do modelo varia conforme a disponibilidade de conhecimentos prévios sobre os dados, levando em consideração suas vantagens, desvantagens e adequação às tarefas a serem desempenhadas.

2.3.5 Algoritmos de Agrupamento: HDBSCAN

No processo de análise, após a obtenção dos dados, é necessário estabelecer estratégias para lidar com eles, de forma que o conteúdo que representam faça sentido. Uma das maneiras de compreender melhor um conjunto de elementos é a classificação segundo características que possuem em comum.

Xu e Wunsch (2008) destacam que uma das tarefas mais importantes na análise de dados é o agrupamento, que organiza objetos em grupos com base em suas similaridades. Nesse contexto, os algoritmos de agrupamento são métodos de aprendizado não supervisionado que agrupam os dados com base em características comuns, permitindo a extração de padrões em cenários em que não há informações rotuladas (Dalal, 2020). Um dos algoritmos utilizados para o agrupamento de dados é o *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN), desenvolvido por Campello, Moulavi e Sander (2013).

McInnes, Healy e Astels (2024) descrevem o HDBSCAN como uma extensão do DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), que identifica *clusters* em dados com base na densidade. Segundo os autores, o

HDBSCAN aprimora o DBSCAN ao transformá-lo em um algoritmo de agrupamento hierárquico. Desse modo, uma hierarquia de *clusters* é criada, na qual os dados são organizados em diferentes níveis de granularidade. Em seguida, uma técnica é empregada para extrair um agrupamento plano (ou seja, um único nível de *clusters*) a partir dessa hierarquia, com base na estabilidade dos *clusters*. Tal estabilidade, por sua vez, é determinada pela persistência dos *clusters* ao longo dos diferentes níveis da hierarquia.

Desse modo, o HDBSCAN é útil em situações em que os dados não possuem uma estrutura clara e com ruídos. Considerando que transcrições de áudios são suscetíveis a falhas e variações que afetam a fidelidade do conteúdo textual transcrito, cuja natureza é não estruturada, o algoritmo HDBSCAN se mostra como uma alternativa viável para análises de agrupamento de dados de áudio.

2.4 MINERAÇÃO DE TEXTO E PROCESSAMENTO DE LINGUAGEM NATURAL

Na era contemporânea, dados textuais estão presentes em todos os lugares. Eles podem ser encontrados em documentos, artigos, livros, e-mails, mensagens de texto, postagens em redes sociais e outras fontes, como legendas de vídeos e transcrições de áudio. Diante dos desafios do *big data* destacados por Laney (2001), a extração de conhecimentos requer técnicas específicas que deem sentido a grandes volumes de dados textuais. Conforme apontam Gharehchopogh e Khalifelu (2011), o texto é o meio mais comum para a troca formal de informações, mas também é desestruturado, sem forma e difícil de lidar.

Nesse contexto, os autores definem a mineração de texto como o processo de analisar textos com o objetivo de extrair informações úteis para fins específicos e identificar padrões. Em contraste com a mineração de dados, que lida com dados estruturados, a mineração de texto lida com dados não estruturados. Dessa forma, ela pode ser aplicada em cenários que envolvem dados textuais, possibilitando a obtenção de conhecimentos mesmo em conjuntos sem um formato definido.

De acordo com Durga *et al.* (2023), o objetivo principal da mineração de texto é obter conhecimentos úteis a partir de textos, identificando fatos ou padrões relevantes que não seriam visíveis apenas com abordagens manuais. Para isso, os autores afirmam que são utilizadas tecnologias como o processamento de linguagem natural e algoritmos de aprendizado, que possibilitam a compreensão das

relações entre palavras em conjuntos de textos. Exemplos de tarefas de mineração de texto mencionados no estudo incluem *clusterização*, extração de conceitos ou entidades, análise de sentimentos, sumarização de documentos, entre outras.

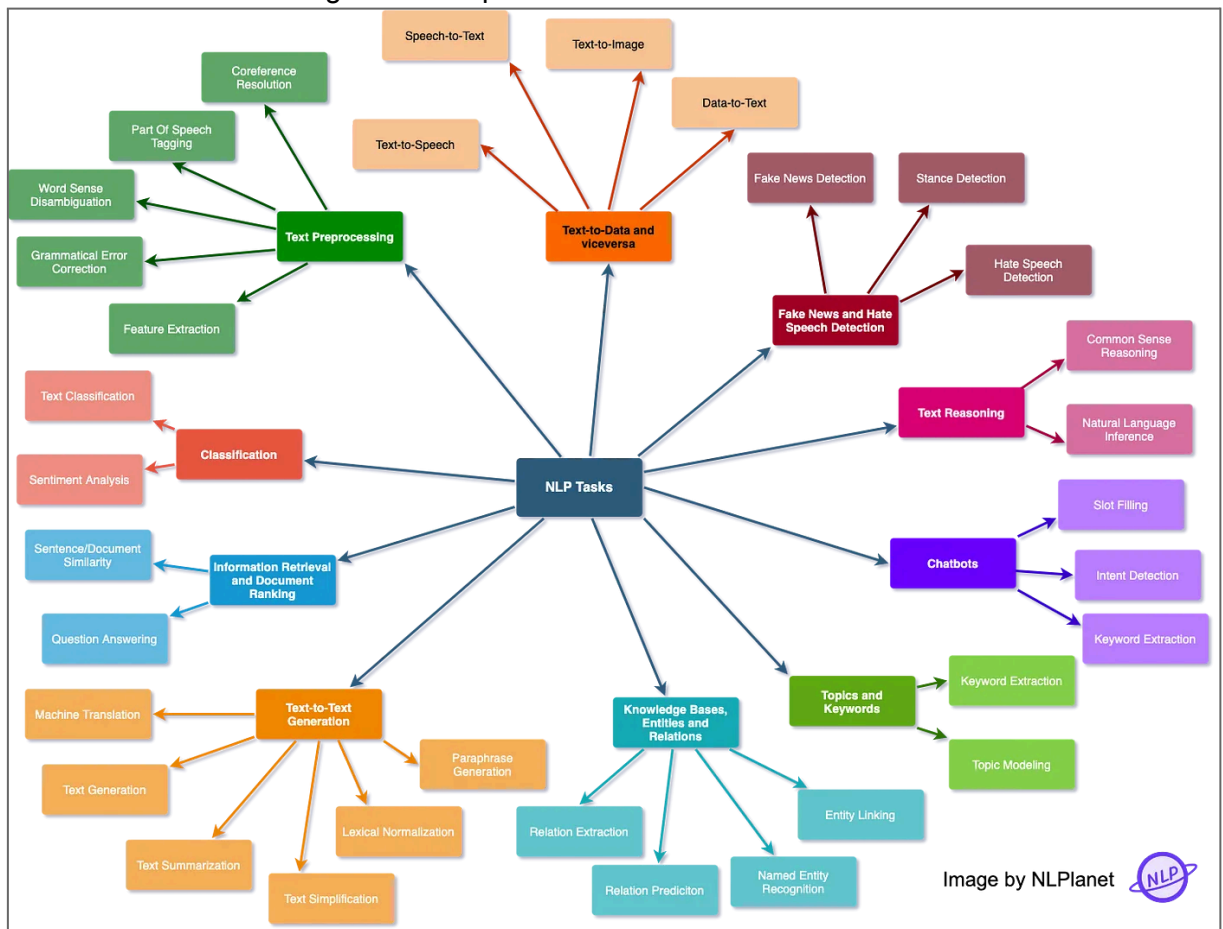
2.4.1 Processamento de Linguagem Natural

Omar *et al.* (2022) definem o processamento de linguagem natural (do inglês *Natural Language Processing* – NLP) como um subcampo da inteligência artificial que possibilita que máquinas leiam, compreendam, e adquiram significados a partir de amplos artefatos de linguagem. Nesse contexto, seus principais benefícios residem na capacidade de ensinar computadores a analisar grandes volumes de dados textuais. O estudo indica que esses benefícios têm feito o NLP ser amplamente utilizado para prover interpretações relevantes para os dados e resolver desafios, além de fazer parte de aplicações cotidianas, como tradutores e assistentes virtuais.

Na definição de Gonzalez-Gomez *et al.* (2024), o processamento de linguagem natural possibilita que computadores compreendam e interpretem a linguagem humana ao combinar inteligência artificial, ciência da computação e linguística, tornando-se fundamental para analisar grandes quantidades de dados textuais. Desse modo, o NLP configura-se como um campo multidisciplinar, que abrange diferentes áreas e tecnologias.

Para melhor compreensão do NLP, Chiusano (2021) fornece uma visão abrangente de suas principais tarefas (Figura 3).

Figura 3 – Mapa Conceitual de Tarefas de NLP



Fonte: Chiusano (2021).

Segundo Al-Sarori *et al.* (2023), as técnicas de NLP têm sido utilizadas em diversas áreas, como categorização de textos, análise de sentimentos, reconhecimento de entidades nomeadas (do inglês *Named Entity Recognition* – NER), perguntas e respostas (Q&A), *clustering*, e modelagem de tópicos. Sendo assim, o processamento de linguagem natural pode ser empregado em uma ampla gama de tarefas, tornando análises textuais mais eficientes ao utilizar técnicas que melhor capturam os significados presentes em dados textuais e transcrições de áudio.

2.4.2 Representação de dados textuais

Dados costumam se apresentar em diferentes níveis de estruturação, que são descritos por Durga *et al.* (2023). Segundo os autores, dados estruturados são organizados no formato de uma tabela, o que otimiza seu armazenamento e

processamento. Em contraste, dados não estruturados não possuem um formato pré-definido e abrangem fontes diversas, como redes sociais, arquivos de áudio e vídeo, e avaliações de produtos. Dados semiestruturados, por sua vez, representam uma mistura de ambos os formatos.

Na definição de Ward, Grimstein e Keim (2015), os dados textuais são componentes de um conjunto de documentos denominado *corpus*, podendo conter metadados do documento de origem. Esses dados costumam ser minimamente estruturados e permitem consultas que retornam informações específicas, como número de parágrafos e palavras, frequência ou distribuição de palavras, e relacionamentos entre parágrafos ou documentos de um corpus. Nesse contexto, as representações textuais são divididas em três níveis: léxico, sintático e semântico. Segundo os autores:

- No nível léxico, ocorre a transformação de sequências de caracteres em entidades chamadas *tokens*, que podem ser compostos por caracteres, n-gramas, palavras, frases, entre outros;
- No nível sintático, a principal tarefa é identificar e anotar a função de cada *token*, definindo características como classe gramatical, posição na frase, data, local, entre outras;
- No nível semântico, é realizada a extração de relacionamentos e significados entre conhecimentos, derivados das estruturas identificadas no nível anterior. A função principal deste nível é definir uma interpretação analítica do texto, que pode ou não estar relacionada a algum contexto.

De modo geral, as representações textuais têm como objetivo capturar informações essenciais dos dados textuais e suas relações. Um exemplo desse tipo de estrutura é o modelo espaço-vetorial, citado na obra dos autores.

Após etapas de pré-processamento, como *stemming* e remoção de *stopwords*, uma representação vetorial é criada para descrever um conjunto de documentos. Isso é feito através da construção de uma matriz documento-termo (DTM), na qual as linhas representam documentos, e as colunas, termos, que podem ser palavras ou frases. Desse modo, cada célula possui um valor numérico, que representa a importância ou grau de ocorrência de um termo no documento (Van Otten, 2023).

Em seguida, pesos são atribuídos às palavras de um texto por meio de cálculos estatísticos, o TF-IDF, que confere maior importância a termos que

aparecem frequentemente no documento de origem, mas pouco no *corpus*. Dessa forma, representa-se numericamente um conjunto de relações entre palavras, possibilitando operações que determinam, por exemplo, qual documento mais se aproxima de outro através do cálculo de similaridade do cosseno (Ward; Grimstein; Keim, 2015).

2.4.3 Word Embeddings

De acordo com Kang e Bissyandé (2019), "*word embeddings*" são definidos como representações de palavras em um espaço vetorial, no qual busca-se atribuir valores próximos a termos semanticamente similares. Nesse contexto, Ethayarajh (2019) afirma que a representação de palavras em um vetor contínuo de baixa dimensionalidade possibilitou a aplicação de métodos de aprendizado profundo no campo de NLP. Entretanto, esses vetores representavam as palavras de forma estática; ou seja, cada palavra possuía uma única representação, independentemente do seu significado no contexto em questão.

Embora as representações textuais tradicionais tenham vantagens como simplicidade e facilidade de implementação, suas limitações impulsionaram o desenvolvimento de novos modelos, capazes de compreender melhor o significado das palavras. Conforme apontam Qiu *et al.* (2024), modelos tradicionais, como *one-hot encoding* e *bag-of-words*, enfrentavam barreiras significativas, como alta dimensionalidade e capacidade limitada de representar correlações semânticas.

No entanto, essas barreiras foram superadas com novos métodos e modelos, permitindo representações de texto mais enxutas, robustas e precisas. Segundo os autores, a evolução das representações de texto resultou em maior velocidade e eficiência na seleção de características, representações contínuas, densas e de baixa dimensionalidade, e níveis de representação com múltiplos graus de granularidade, abrangendo não só itens lexicais, mas também frases, parágrafos e capítulos inteiros.

Nesse cenário, Ethayarajh (2019) ressalta o advento dos modelos de linguagem pré-treinados, como ELMO e BERT, como um importante fator na evolução das representações textuais, marcada pela transição dos *word embeddings* tradicionais para representações sensíveis ao contexto. Essas representações proporcionaram avanços significativos no processamento de

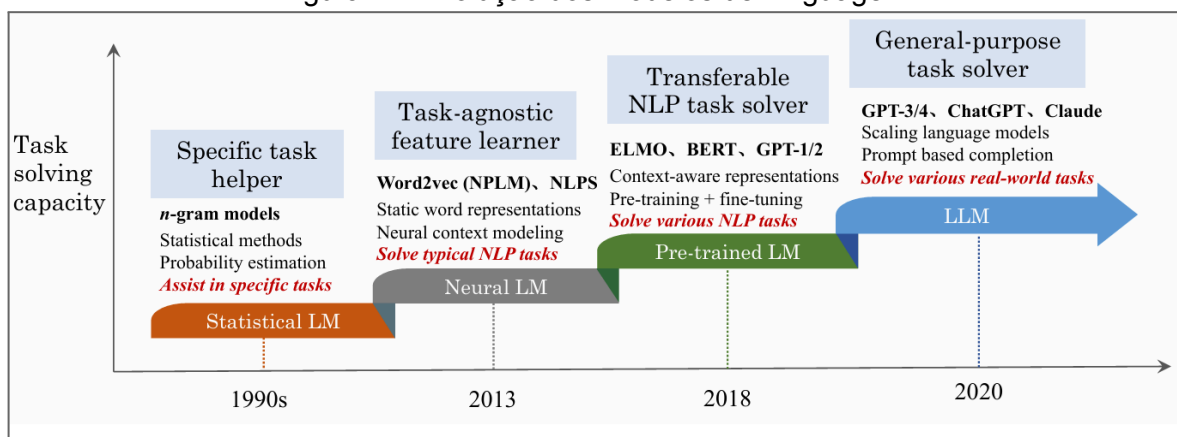
linguagem natural, destacando o impacto desses modelos no campo e áreas relacionadas.

2.4.4 Modelos de Linguagem de Grande Escala

Segundo Raiaan *et al.* (2024), os LLMs demonstraram uma capacidade extraordinária em diversas tarefas de NLP, tais como tradução de idiomas, geração de texto e perguntas e respostas (Q&A). Nesse sentido, os autores os destacam como uma parte nova e essencial do NLP, devido à sua capacidade de compreender padrões verbais complexos, e gerar respostas de forma coerente e apropriada com base em um contexto específico.

Na definição de Zhao *et al.* (2024), o objetivo geral da modelagem de linguagem (LM) é modelar a probabilidade generativa de sequências de palavras, com o intuito de prever a probabilidade de *tokens* ausentes ou futuros. Como área de pesquisa, os autores apontam que a modelagem de linguagem divide-se em quatro estágios de desenvolvimento principais: modelos de linguagem estatísticos (SLMs), modelos de linguagem neurais (NLMs), modelos de linguagem pré-treinados (PLMs) e modelos de linguagem de grande escala (LLMs). Na Figura 4, a evolução dos modelos de linguagem é representada.

Figura 4 – Evolução dos Modelos de Linguagem



Fonte: Zhao *et al.* (2024).

Nesse cenário, o estudo aponta que modelos de linguagem recentes utilizam arquiteturas que geram representações de palavras sensíveis ao contexto por meio de um processo de pré-treinamento, que são altamente eficientes como características semânticas de propósito geral. Em seguida, o modelo em questão

pode ser ajustado para tarefas ou contextos específicos, num processo conhecido como *fine-tuning*. Segundo os autores, estudos posteriores estabeleceram esses paradigmas de aprendizado e novos modelos foram propostos.

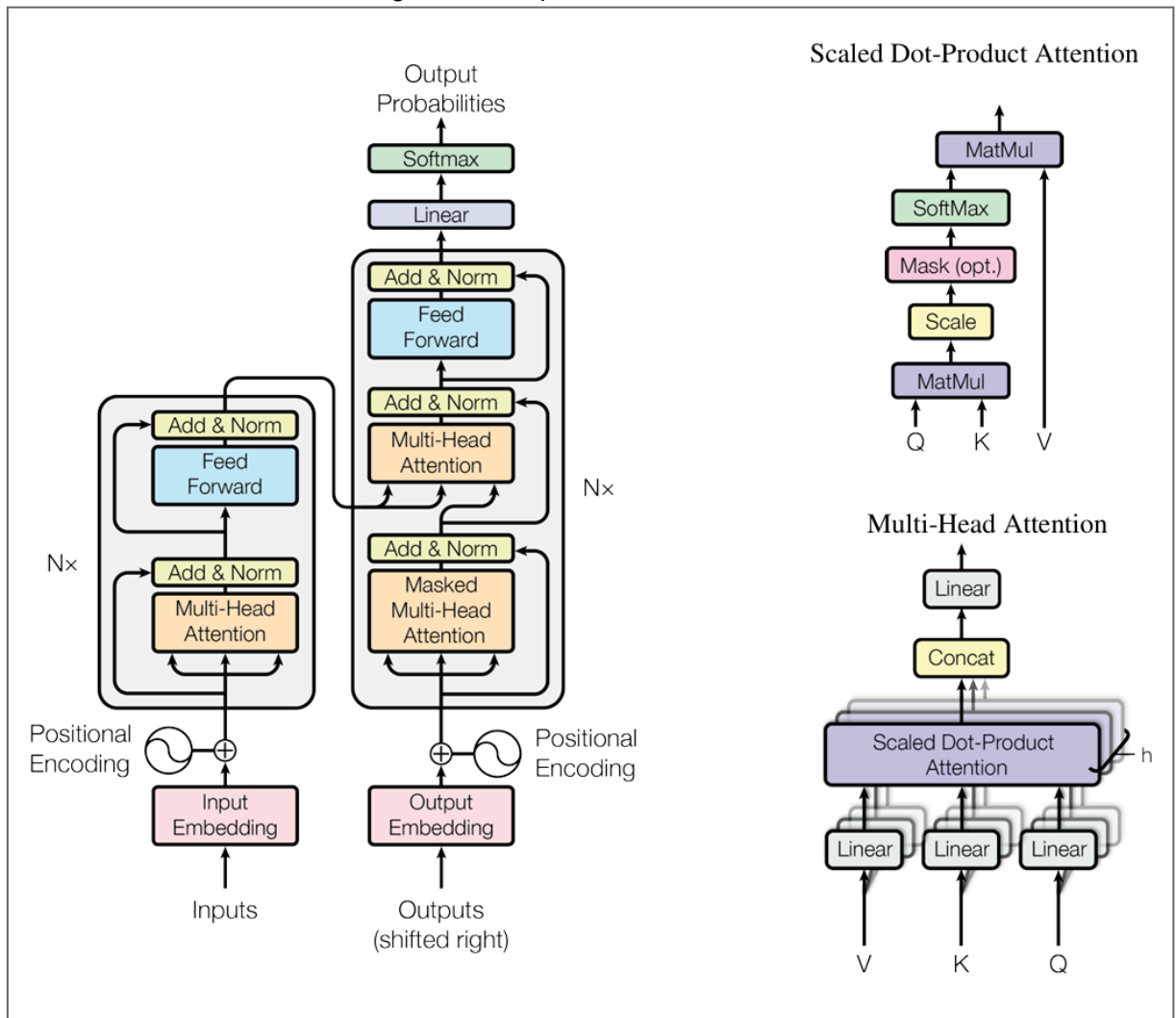
Nesse contexto, Zhao *et al.* (2024) destacam que modelos de linguagem de grande escala, como GPT-4 e Llama, demonstraram uma capacidade surpreendente em resolver tarefas complexas, comportando-se de maneira distinta de modelos pré-treinados menores. Isso se deve não apenas às suas tecnologias subjacentes, mas também ao grande número de parâmetros utilizados, o que expandiu significativamente a capacidade desses modelos de resolverem tarefas e reconhecerem contextos. Desse modo, o emprego de um grande número de parâmetros tem sido um importante fator no desempenho dos modelos de linguagem atuais.

No entanto, essa característica também impõe limitações. De acordo com Rostam *et al.* (2024), os LLMs exigem uma quantidade extrema de parâmetros para atingir alto desempenho, o que aumenta consideravelmente os custos computacionais e de memória. Essas barreiras dificultam que pesquisadores acessem os recursos necessários para treiná-los ou aplicá-los. Sendo assim, os autores ressaltam que desenvolver *frameworks*, bibliotecas e propor novas técnicas para superar tais barreiras é essencial para tornar o uso dos LLMs mais acessível e eficiente.

2.4.5 Arquitetura *Transformer*

Nesse cenário de avanços impulsionados por LLMs poderosos, a arquitetura *Transformer* se encontra no cerne de modelos como BERT, GPT-3 e T5. De acordo com Lin *et al.* (2021), o *Transformer* é um modelo de aprendizado profundo notável que tem sido muito utilizado em vários campos, como NLP, processamento de fala e visão computacional. Segundo os autores, estudos recentes evidenciam que modelos pré-treinados baseados em *Transformer* demonstraram desempenho de ponta em várias tarefas, fato que tem contribuído para a sua vasta aplicação. Na Figura 5, a arquitetura *transformer* é representada.

Figura 5 – Arquitetura Transformer



Fonte: Vaswani *et al.* (2021).

De acordo com Lauren (2022), tal conquista tecnológica pode ser atribuída ao mecanismo de atenção, um aspecto primário do modelo que possibilita que redes neurais aprendam representações contextualizadas de palavras. Conforme aponta Nazeri (2024), esse mecanismo possibilita que LLMs consigam acompanhar e compreender melhor o conteúdo de conversas, auxiliando o modelo a atuar nas porções relevantes da sequência de entrada ao gerar as saídas. Segundo o autor:

- O mecanismo de atenção compara cada item de uma sequência consigo mesmo e com os demais, um a um. Essa comparação é feita através do cálculo do produto escalar desses itens.
- Os números resultantes dessas operações são chamados de *attention scores*, que são normalizados, diminuindo suas dimensões;

- No passo seguinte, a função *softmax* é aplicada nas pontuações de atenção. Assim, obtém-se a sua distribuição de probabilidade, que indica a quais partes da sequência o foco deve ser direcionado. Após a aplicação dessa função, a soma dos números resultantes é igual a um.
- Por fim, a saída para cada palavra é a soma ponderada de todas as palavras na frase, cujos pesos são determinados pelo grau de atenção que a palavra em questão deve ter em relação às demais.

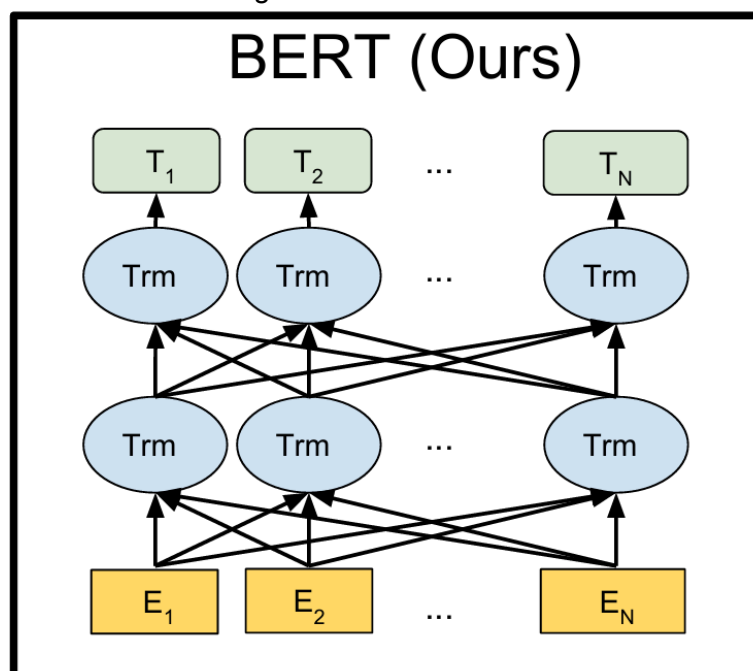
Esses passos permitem que o modelo capture as relações entre as palavras de uma mesma frase. Além do mecanismo de atenção, outros aspectos relevantes da arquitetura proposta por Vaswani *et al.* (2021) incluem o *multi-head attention*, que calcula a atenção de múltiplas partes da sequência em paralelo, capturando diversas relações contextuais. O uso das saídas do codificador como contexto adicional para a geração das saídas do decodificador é outro aspecto importante, pois fornece informações essenciais para a geração das saídas. A codificação posicional, por sua vez, incorpora as informações de posição das palavras em uma frase aos *embeddings*, permitindo que o modelo entenda a ordem das palavras.

Para garantir a estabilidade e a preservação das informações originais, as operações “*add & norm*” são aplicadas após cada transformação. Por fim, as subcamadas “*feed forward*” aplicam uma transformação não linear aos dados processados, permitindo que o modelo aprenda representações mais complexas. Desse modo, a arquitetura reforça o contexto das sequências ao longo de todo o processo de codificação e decodificação (Vaswani *et al.*, 2021).

2.4.6 Modelo BERT

O BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de linguagem baseado na arquitetura *Transformer*, que utiliza mecanismos de atenção para processar sequências de palavras e compreender as relações contextuais entre elas (MD, 2022). Ele opera por meio de uma “pilha de *encoders*”, que analisa o contexto em ambas as direções. Isso permite ao modelo captar o significado semântico real do texto, característica essencial para diversas tarefas de NLP. A estrutura básica do modelo BERT é representada na Figura 6.

Figura 6 – Modelo BERT



Fonte: Devlin *et al.* (2019).

Devlin *et al.* (2019) ressaltam que, além de sua natureza bidirecional, o BERT pode ser facilmente adaptado a tarefas específicas com mudanças mínimas em sua arquitetura por meio de *fine-tuning*. Essas características resultaram em avanços notáveis em comparação aos modelos unidirecionais, proporcionando maior desempenho e flexibilidade.

2.4.7 SBERT

De acordo com Zhao *et al.* (2024), o modelo BERT marcou um avanço significativo na evolução dos modelos de linguagem, destacando-se por sua capacidade de gerar representações de palavras sensíveis ao contexto. Por outro lado, a criação de representações em nível de frase (*sentence embeddings*) pode ser necessária para certas tarefas de NLP. Nesse cenário, Reimers e Gurevych (2019) introduzem o SBERT (Sentence-BERT), uma adaptação do BERT projetada para gerar *embeddings* de frases.

Na definição dos autores, o modelo utiliza arquiteturas de redes siamesas e *triplet*¹ para derivar representações de frases com base nos *embeddings* gerados

¹ DEEP Learning Proteins Using a Triplet-Bert Network. 2021. Disponível em: <https://openreview.net/pdf?id=hga6dk7nxFB>. Acesso em: 26 dez. 2024.

pelo BERT. Além disso, uma camada de *pooling* é empregada, permitindo representar a frase por meio do *token* “CLS” ou pela combinação dos *embeddings* de cada palavra em uma única representação. Essa combinação pode ser feita calculando a média dos *embeddings* (*mean pooling*) ou aplicando o cálculo *max-over-time* nos vetores de saída.

Dessa forma, é possível comparar frases utilizando métricas como distância euclidiana e similaridade do cosseno, ao mapear seus *embeddings* em um espaço vetorial, o que reduz significativamente a carga de processamento. Essa abordagem se mostra especialmente útil em tarefas como busca semântica e agrupamento (*clustering*), visto que basta aplicar o BERT apenas uma vez para cada frase (Reimers e Gurevych, 2019).

2.4.8 Questions and Answers (Q&A)

A capacidade dos LLMs de compreender contextos e resolver tarefas torna a sua utilização útil em uma ampla variedade de contextos. Nesse sentido, uma das tarefas de NLP que podem ser utilizadas para obter respostas e informações específicas é o *question answering* ou *questions and answers* (Q&A). Na definição de Mishra, Mishra e Sharma (2012), o Q&A é um processo que surgiu em meados dos anos 60 e 70, cujo objetivo é extrair respostas para perguntas realizadas em linguagem natural. De acordo com os autores, os sistemas de Q&A podem ser divididos em três principais abordagens:

- Q&A baseado em NLP: As perguntas do usuário são mapeadas em um modelo formal de palavras, que visa fornecer as respostas mais confiáveis;
- Q&A baseado NLP e recuperação da informação (IR): Para a extração de fatos de grandes volumes de texto, a recuperação da informação é utilizada em conjunto com o processamento de linguagem natural;
- Q&A baseado em *templates*: Nessa abordagem, o modelo faz a correspondência entre a consulta do usuário e um conjunto de *templates*, que abrangem as porções mais consultadas de áreas de conhecimento.

Frente ao crescimento na quantidade de dados presente em diversos contextos, Patil *et al.* (2023) destacam a importância do Q&A como um ferramenta necessária para recuperar informações de forma eficiente e precisa. No contexto de

transcrições de áudio, isso é particularmente útil para automatizar a recuperação de informações em áudios longos ou em conjuntos de áudios, que podem conter informações específicas relevantes para o usuário.

Apesar de suas vantagens, o Q&A também possui limitações que devem ser consideradas. De acordo com Pandey e Roy (2024), embora a inteligência artificial e o processamento de linguagem natural tenham revolucionado os sistemas de Q&A, o campo ainda enfrenta desafios consideráveis. A geração de respostas para perguntas complexas continua difícil, pois envolve a busca por dados com diferentes graus de estruturação de fontes diversas. Desse modo, são necessárias etapas cautelosas de coleta, limpeza e integração de dados, além de técnicas avançadas de NLP, para extrair respostas de grandes volumes de dados de forma eficiente.

Nesse cenário, os autores destacam que o uso de modelos pré-treinados, como o BERT, é uma abordagem eficaz para criar sistemas capazes de responder a perguntas em contextos específicos. Isso permite aproveitar a capacidade avançada desses modelos de compreender frases e contextos, fornecendo respostas mais precisas.

2.4.9 Geração Aumentada por Recuperação

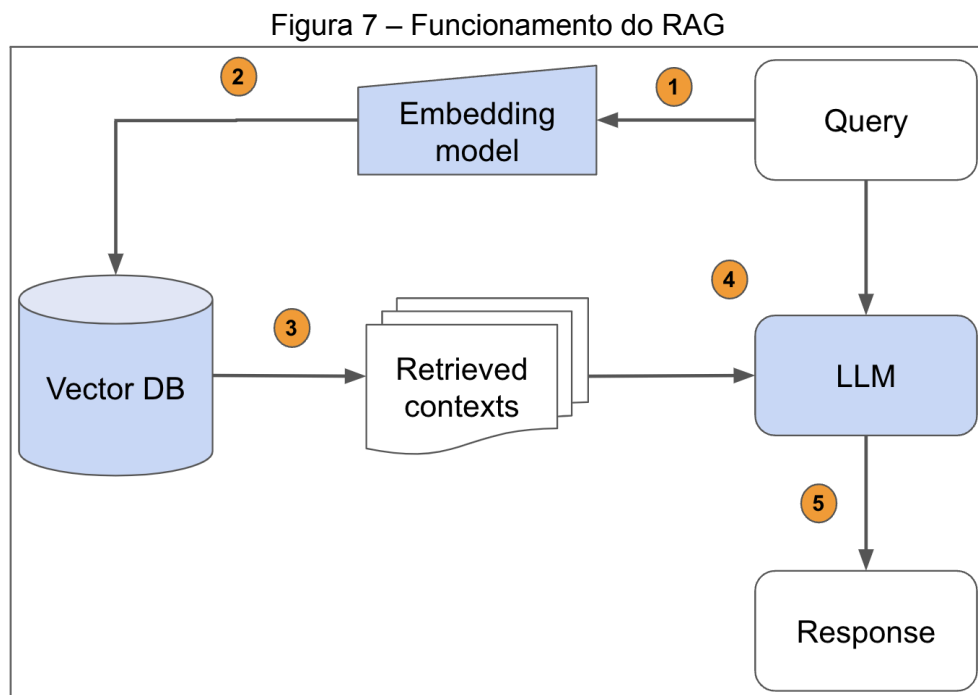
Em sistemas de Q&A, identificar os fragmentos (*chunks*) de texto no qual uma resposta se encontra não só auxilia a encontrar informações, mas também proporciona mais clareza quanto aos fatores que levaram o modelo a fornecer a saída em questão. Entretanto, em aplicações modernas, pode ser desejável gerar uma resposta de forma mais elaborada, utilizando a vasta base de conhecimentos presente em modelos como o GPT-4 e Gemini. Desse modo, modelos *encoder-only* como o BERT podem não ser suficientes para prover respostas satisfatórias, uma vez que, devido à sua arquitetura (Devlin *et al.*, 2019), o modelo não gera textos.

Por outro lado, modelos generativos também enfrentam obstáculos. De acordo com Zhu *et al.* (2024), uma das limitações que LLMs apresentam é a dependência de grandes volumes de texto para o processo de treinamento, o que resulta em performance inferior ao lidar com tarefas ligadas a um domínio específico. Desse modo, os autores afirmam que ao se deparar com uma pergunta cuja resposta não se encontra em sua base de dados, o modelo pode sofrer de um

fenômeno chamado “alucinação”, que refere-se à geração de respostas sem sentido ou inconsistentes em relação ao material fornecido.

Uma das possíveis abordagens para contornar essas limitações é submeter o modelo de linguagem a um processo *fine-tuning*, que ajusta seus parâmetros utilizando uma base de dados específica, aprimorando as respostas para o domínio em questão. Contudo, esse processo pode nem sempre ser viável, uma vez que demanda submeter a rede a um processo de treinamento. Outra limitação é a flexibilidade exigida em cenários em que as informações tornam-se obsoletas rapidamente, uma vez que a base de conhecimentos utilizada pelo modelo é fixa (Amazon, 2024).

Nesse cenário, a Geração Aumentada por Recuperação (do inglês *Retrieval Augmented Generation* – RAG) emerge como uma alternativa que possibilita fazer uso de conhecimentos externos para fornecer um contexto a modelos generativos, que utilizam esse contexto para gerar respostas mais precisas e relevantes (Wu *et al.*, 2024). Para isso, as informações externas relativas à pergunta, representadas como dados vetoriais, são recuperadas através da similaridade semântica (Zhu *et al.*, 2024). Desse modo, os autores afirmam que o RAG não só reduz a chance de ocorrência de alucinações, mas também melhora a performance do modelo ao produzir respostas mais precisas. Na Figura 7, as etapas do RAG são representadas.



Fonte: Mohandas e Moritz (2023).

Entretanto, apesar de suas vantagens, o RAG possui alguns pontos a serem observados. Edge *et al.* (2024), citados por Zhu *et al.* (2024), ressaltam que, ao converter os dados em vetores, a conexão contextual entre as informações pode se perder, tornando a relação entre diferentes partes dos dados menos clara e dificultando a interpretação correta. Além disso, os autores apontam que métodos de RAG dependem muito da fragmentação (*chunking*) dos documentos; se a fragmentação não for feita adequadamente, a eficácia pode ser comprometida, resultando em respostas menos precisas e relevantes.

Outra barreira importante diz respeito às informações recuperadas. Wu *et al.* (2024) apontam que, caso documentos não sejam recuperados ou documentos que não correspondem à consulta sejam enviados como contexto, a resposta pode ser comprometida, levando a respostas incorretas. Sendo assim, é fundamental considerar não apenas os parâmetros dos métodos de RAG utilizados, mas também a base de dados usada como contexto e as respostas esperadas, para obter melhores resultados.

3 METODOLOGIA

Este capítulo tem como objetivo fornecer uma descrição da metodologia empregada na condução da pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

Os autores Gerhardt e Silveira (2009) destacam que o processo de pesquisa tem início a partir da definição de uma problemática ou questão de interesse, que demanda a utilização de um método ou caminho específico para alcançar resultados e, assim, gerar conhecimento. Nesse sentido, este estudo adota uma abordagem metodológica que combina pesquisa aplicada e exploratória, com o objetivo de extrair *insights* relevantes de um conjunto de dados.

A pesquisa aplicada, conforme discutido por Vieira, Leite e Kuhn (2023), tem como objetivo gerar conhecimento por meio da resolução de um problema específico, contribuindo para o avanço científico e inovação dentro de um determinado campo. Por outro lado, a pesquisa exploratória, segundo Gil (2008), é amplamente utilizada quando se trata de temas pouco explorados ou compreendidos, e visa proporcionar uma visão abrangente sobre o tópico, possibilitando aos pesquisadores uma maior familiaridade com suas particularidades.

Este trabalho integra ambos os tipos de pesquisa, ao aplicar algoritmos de modelos de linguagem de grande escala (LLMs) sobre uma base de dados formada a partir da transcrição de áudio, com o intuito de explorar os *insights* e tendências que possam surgir. A próxima seção detalha a metodologia adotada neste estudo.

3.2 METODOLOGIA DESIGN SCIENCE RESEARCH METHODOLOGY

A metodologia Design Science Research Methodology (DSRM), proposta por Peffers *et al.* (2007), fornece uma estrutura para a condução de pesquisas em design voltadas à área de sistemas da informação. Segundo os autores, seu foco está na criação de artefatos bem-sucedidos para resolver problemas ou alcançar objetivos específicos.

Conforme apontam Abdurrahman e Mulyana (2020), embora a DSRM siga um fluxo de etapas definido, ela permite adaptações conforme os objetivos e resultados esperados. A seguir, apresenta-se uma breve descrição de cada passo dessa metodologia (Peffer *et al.*, 2007).

- Passo 1 - Identificação do problema e motivação: Consiste na definição e descrição do problema de pesquisa, incluindo a justificativa da solução proposta e o valor agregado por essa solução;
- Passo 2 - Expressão dos objetivos para o resultado: Foca na explicitação do raciocínio e das ações necessárias a partir da problemática definida, com o objetivo de orientar as etapas subsequentes da pesquisa;
- Passo 3 - Criação e desenvolvimento: Trata da criação de um artefato, que pode ser um *framework*, método, arquitetura, entre outros, e define as funcionalidades que o artefato deverá ter para responder à problemática apresentada;
- Passo 4 - Demonstração: Consiste na apresentação do artefato desenvolvido, demonstrando como ele pode resolver, parcial ou totalmente, o problema dentro de um cenário de teste;
- Passo 5 - Avaliação: Envolve a observação e análise do uso do artefato para verificar sua eficácia na resolução do problema proposto;
- Passo 6 - Comunicação: Este passo envolve a divulgação do problema abordado, da modelagem realizada e dos resultados obtidos com o artefato, incluindo sua avaliação e eficácia;

Esses passos estruturam o processo da *Design Science Research Methodology*, proporcionando um caminho para a pesquisa e desenvolvimento de soluções inovadoras dentro do contexto da investigação científica.

3.3 DESENVOLVIMENTO DA PESQUISA

O processo de transcrição de áudio, essencial para o desenvolvimento deste trabalho, foi realizado utilizando o modelo Whisper, que é um Modelo de Linguagem de Grande Escala (LLM), especializado em converter áudios em texto de forma

eficiente. Após a transcrição, os modelos Gemini e uma variante do BERT² foram utilizados para criar *embeddings*, que possibilitaram uma representação semântica dos dados textuais. Em seguida, a técnica de modelagem de tópicos, utilizando o BERTopic, foi empregada para identificar e elucidar os tópicos presentes nas transcrições, ajudando na organização e compreensão dos conteúdos extraídos. Além disso, um módulo de perguntas e respostas (Q&A) foi implementado para facilitar a recuperação de informações específicas a partir das transcrições geradas.

Com o objetivo de dar seguimento à pesquisa, a implementação da metodologia *Design Science Research Methodology* (DSRM) foi adaptada para este contexto, conforme esquematizado no Quadro 1 a seguir:

Quadro 1 – Metodologia DSRM

Identificação do problema e motivação	Explorar os dados coletados por meio de transcrições de áudio, identificando a necessidade de ferramentas que facilitem a análise automatizada de grandes volumes de dados textuais provenientes de áudios, com foco na identificação de padrões e tendências inovadoras e obtenção de respostas a perguntas (Q&A).
Definição dos requisitos	Estabelecer os critérios para a análise exploratória das transcrições, incluindo a identificação de grupos temáticos e padrões semânticos e a recuperação de informações, a fim de compreender os tópicos relevantes discutidos nas transcrições e obter informações específicas a partir do conteúdo transcrito.
Projeto e desenvolvimento	Desenvolver e implementar um método para análise de dados textuais provenientes de áudios, dividido em três módulos: Transcrição, Q&A, e Modelagem de Tópicos. Para isso, utilizar os modelos de linguagem para transcrever e representar os dados textuais de áudio de forma eficiente. A partir dessas representações, implementar um algoritmo para perguntas e respostas e modelagem de tópicos, com o objetivo de organizar as transcrições de áudio em tópicos semanticamente e contextualmente semelhantes e possibilitar a realização de perguntas a respeito de conteúdos específicos, facilitando a análise de grandes volumes de dados textuais em diferentes escalas.

² HUGGING FACE. sentence-transformers/all-distilroberta-v1. Disponível em: <https://huggingface.co/sentence-transformers/all-distilroberta-v1>. Acesso em: 15 dez. 2024.

Demonstração	Aplicar o modelo Whisper para transcrição dos áudios, seguido pela análise das transcrições utilizando o framework BERTopic, gerando visualizações dinâmicas que representem os agrupamentos identificados e os principais <i>insights</i> extraídos das transcrições. Além disso, implementar um método de Q&A com RAG para possibilitar a utilização do conteúdo de bases de áudio como uma fonte de informações, facilitando a obtenção de informações específicas.
Avaliação	Verificar a relevância dos tópicos identificados nas transcrições e a eficácia da ferramenta de Q&A na provisão de respostas contextualizadas ao usuário.
Comunicação	Apresentar os resultados obtidos em relatórios e visualizações que destaquem os <i>insights</i> , padrões e tendências identificadas, bem como demonstrar o funcionamento do módulo de Q&A para realizar perguntas relativas a um contexto, discutindo o impacto desses achados na gestão e análise de dados transcritos.

Fonte: Autor (2024).

3.4.1 População e Amostra

Para viabilizar o estudo proposto, optou-se por utilizar uma fonte de dados composta por áudios transcritos do curso “*The Surveillance State: Big Data, Freedom, and You*”, de Paul Rosenzweig (2016). Os áudios abordam a tecnologia como um mecanismo de vigilância, explorando questões políticas, geopolíticas e éticas, como privacidade e o impacto da tecnologia na sociedade.

A amostra para análise foi composta por aproximadamente 12 horas de áudios transcritos. A análise buscou, por meio de técnicas de processamento de linguagem natural (NLP), identificar padrões semânticos e tópicos em evidência nas discussões. Além disso, foi proposto um método baseado em perguntas e respostas (Q&A), no qual os *embeddings* foram gerados com o modelo Gemini, para aprofundar a compreensão sobre os temas abordados e facilitar a extração de *insights* relevantes das transcrições.

3.4.2 Coleta de Dados e Procedimentos

A coleta de dados foi realizada por meio da transcrição automatizada de áudios utilizando o modelo Whisper. Após a *transcrição*, o texto foi armazenado em um arquivo JSON e utilizado como base para a construção de um módulo de Q&A. Em seguida, foi utilizado o BERTopic para a identificação de tópicos predominantes nas transcrições. O processo de coleta de dados seguiu as seguintes etapas:

- **Seleção dos Áudios:** Inicialmente, foi realizada a seleção de áudios de um curso, composto por 24 áudios, com temas relacionados à tecnologia como um mecanismo de vigilância. Todos os áudios foram utilizados.
- **Transcrição Automática com Whisper:** Utilizou-se o modelo Whisper para transcrever os áudios, convertendo as gravações em texto de forma precisa. Esses textos foram organizados em um arquivo JSON;
- **Geração de *Embeddings* com Gemini e SBERT:** Após a transcrição, os modelos Gemini e SBERT foram utilizados para criar *embeddings* que representassem semanticamente as transcrições em seus respectivos módulos, permitindo a análise semântica do conteúdo representado por diferentes modelos de linguagem;
- **Modelo de Perguntas e Respostas (Q&A):** Como parte do processo de análise, um modelo de Q&A foi implementado, utilizando o *framework* LangChain para RAG e os *embeddings* gerados pelo modelo Gemini, permitindo a extração de informações específicas das transcrições. Essas informações foram então enviadas como contexto ao modelo Gemini para geração das respostas, facilitando a obtenção de informações relevantes.
- **Identificação de Tópicos com BERTopic:** O BERTopic foi aplicado para organizar as transcrições em tópicos, agrupando os tópicos de acordo com o conteúdo discutido nos áudios. Como parâmetros, foram utilizados os algoritmos UMAP para redução de dimensionalidade e HDBSCAN para agrupamento, bem como c-TF-IDF para a extração dos tópicos. Por fim, visualizações utilizando bibliotecas como Plotly e Matplotlib foram geradas.

3.4.3 Análise de dados

A fase de análise de dados é fundamental para a extração de *insights* e padrões a partir das transcrições de áudio. Neste estudo, foi adotada uma abordagem multifacetada, que combina análise de agrupamento, técnicas de processamento de linguagem natural (NLP) e visualização de dados, com o objetivo de identificar, interpretar e explorar os tópicos discutidos nas transcrições.

Os principais procedimentos adotados na análise de dados foram:

- Transcrição Automática com Whisper: Inicialmente, os áudios foram transcritos para texto utilizando o modelo Whisper, um sistema automatizado de transcrição altamente preciso. Essa etapa garantiu que o conteúdo oral fosse convertido com qualidade, permitindo que as análises subsequentes fossem realizadas com base em dados textuais precisos;
- Geração de *Embeddings* com Gemini e SBERT: Após a transcrição dos áudios, utilizou-se os modelos Gemini e SBERT (*Sentence-Bidirectional Encoder Representations from Transformers*) para gerar *embeddings*, ou representações vetoriais, para cada trecho transcrito. Esses modelos foram escolhidos por sua capacidade de capturar o significado semântico de palavras e frases, permitindo uma análise mais profunda das transcrições. Dessa forma, representa-se o contexto semântico das transcrições, facilitando a análise e o agrupamento de informações relevantes;
- Modelagem de Tópicos com BERTopic: O *framework* BERTopic foi aplicado para identificar e organizar os tópicos discutidos nas transcrições. O BERTopic é uma técnica de modelagem de tópicos que utiliza *embeddings* para agrupar palavras e frases semelhantes, formando tópicos coesos. Esta etapa permitiu a identificação das principais áreas de discussão, como as questões de privacidade, vigilância e ética, abordadas nos áudios.
- Agrupamento com o Algoritmo HDBSCAN: Para agrupar as transcrições com base na semelhança semântica, foi utilizado o algoritmo HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*). Este algoritmo foi escolhido por sua capacidade de identificar *clusters* de densidade variável, sendo especialmente útil em dados de alta dimensão,

como os *embeddings* gerados pelo SBERT. O HDBSCAN permite a criação de agrupamentos de tópicos relacionados, ajudando a identificar padrões nas discussões dos áudios;

- Modelo de Perguntas e Respostas (Q&A): Para facilitar a extração de informações específicas e relevantes, foi implementado um modelo de perguntas e respostas (Q&A). Utilizando os *embeddings* gerados pelo modelo Gemini, o módulo de Q&A permitiu que questões específicas sobre os tópicos dos áudios fossem respondidas de forma precisa, extraindo *insights* e aprofundando a compreensão dos temas tratados nas transcrições;
- Visualização de Dados: A última etapa da análise envolveu a criação de gráficos e representações visuais para ilustrar os *clusters* e tópicos identificados, bem como os padrões emergentes. Técnicas de visualização dinâmicas foram utilizadas para facilitar a interpretação dos resultados, permitindo uma análise mais intuitiva e acessível dos dados. As visualizações ajudaram a identificar a distribuição dos tópicos, o relacionamento entre os diferentes temas e a intensidade das discussões nos áudios.

Dessa forma, a análise de dados foi realizada de maneira integrada, utilizando ferramentas avançadas de NLP e *clustering* para fornecer uma visão detalhada e semântica dos temas discutidos nas transcrições de áudio. O uso do Gemini, Q&A, BERTopic, e visualizações de dados possibilitou a extração de *insights* e a organização das informações de maneira clara e significativa.

3.4.4 Resultados esperados

O método foi desenvolvido a partir de três módulos interligados, os resultados esperados incluem a obtenção de *insights* abrangentes tanto da perspectiva "micro" (individual) quanto da perspectiva "macro" (global) dos dados. A seguir, descrevem-se os principais resultados esperados de cada módulo:

3.4.4.1 Módulo 1: Transcrição e Criação de um JSON de Transcrições

No módulo 1 são esperados os seguintes resultados:

- Transcrição precisa: A transcrição dos áudios será realizada usando o modelo Whisper Turbo. A expectativa é que cada áudio seja corretamente transcrito, com metadados como título e autor associados a cada transcrição, permitindo uma organização estruturada das informações;
- Armazenamento em formato JSON: As transcrições serão armazenadas em um formato JSON, o que permitirá uma integração fácil com os módulos subsequentes, garantindo que todas as informações sejam acessíveis e estruturadas para análise posterior.

3.4.4.2 Módulo 2: Q&A com Modelo Gemini e RAG

No módulo 2 são esperados os seguintes resultados:

- Extração de respostas: O modelo Gemini, combinado com a técnica de Geração Aumentada por Recuperação (RAG), permitirá a geração de respostas para perguntas específicas sobre as transcrições. A divisão das transcrições em fragmentos (*chunks*) de 1000 caracteres, com sobreposição de 200 caracteres, possibilitará que o modelo possa lidar com transcrições longas, melhorando a eficiência da recuperação de informações;
- Geração contextualizada de respostas: Através da análise semântica dos *embeddings*, o modelo Gemini será capaz de gerar respostas baseadas no contexto mais relevante extraído das transcrições e fornecido via *prompt*, oferecendo respostas para perguntas sobre temas como vigilância digital, privacidade e questões éticas;

3.4.4.3 Módulo 3: Modelagem de Tópicos com BERTopic e Visualização de Dados

No módulo 3 são esperados os seguintes resultados:

- Identificação e classificação de tópicos: O uso do BERTopic permitirá a identificação de tópicos predominantes nas transcrições, agrupando frases de forma coerente de acordo com o conteúdo discutido. Espera-se que o modelo identifique tópicos relacionados a questões-chave, como a vigilância digital, a privacidade e as implicações políticas e éticas dessas tecnologias;

- Distribuição de tópicos ao longo das transcrições: As transcrições poderão ser analisadas tanto a nível de frase quanto de documento, permitindo identificar a distribuição e a evolução dos tópicos ao longo dos diferentes áudios. Isso permitirá uma análise mais granular do conteúdo e uma melhor compreensão de como os temas se inter-relacionam;
- Visualização gráfica dos tópicos: A visualização de dados através de gráficos gerados pelo BERTopic, Matplotlib e Plotly fornecerá uma forma intuitiva de explorar a distribuição dos tópicos nas transcrições. Espera-se que as visualizações permitam identificar quais tópicos predominam e como os temas estão organizados, facilitando a interpretação dos dados.
- Flexibilidade na análise: A flexibilidade do Módulo 3, permitindo a análise tanto de frases isoladas quanto de documentos completos, possibilitará que a metodologia possa ser aplicada a conjuntos de dados de diferentes tamanhos e complexidades, oferecendo *insights* tanto de forma granular quanto global.

3.4.4.4 Resultados gerais esperados

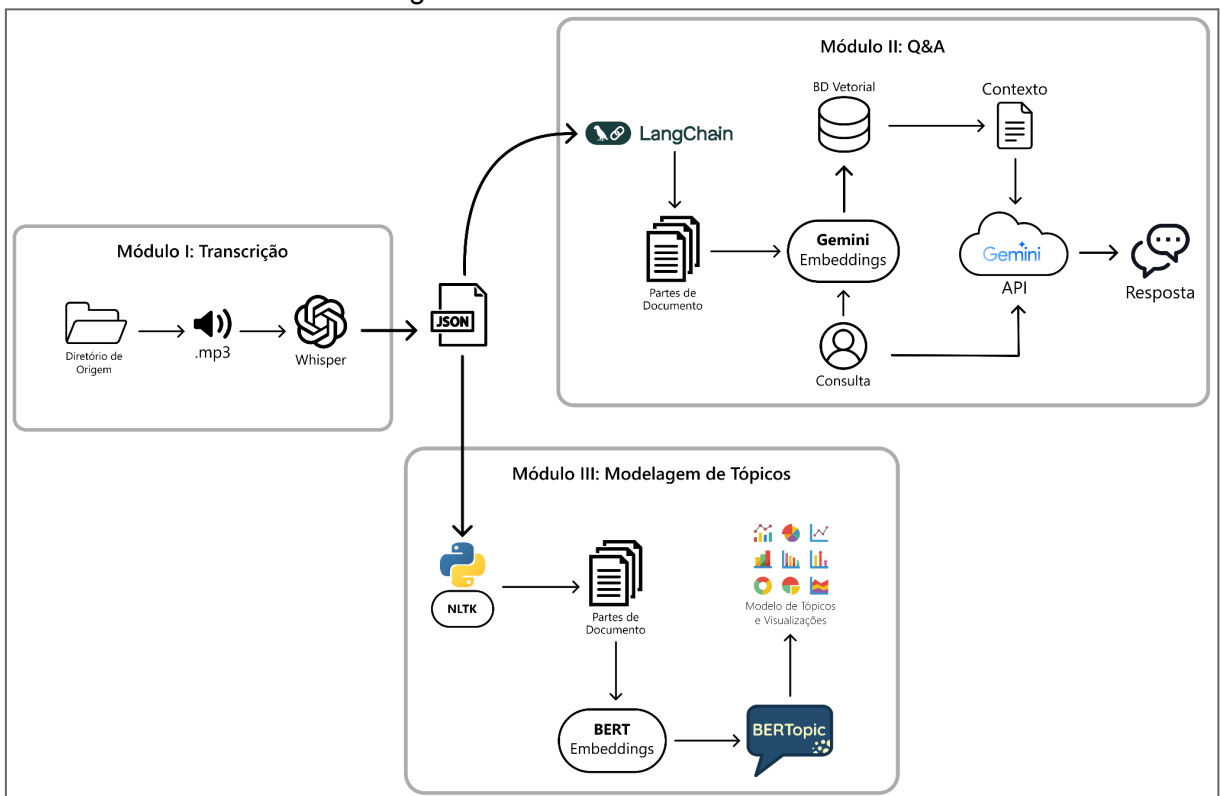
A combinação de modelagem de tópicos, análise de Q&A e visualizações de dados ajudará a entender melhor os principais assuntos das transcrições. Os resultados devem trazer *insights* sobre os áudios e seus temas. O uso dos três módulos permitirá analisar os dados de diferentes formas, tornando mais fácil e acessível extrair informações. O Módulo 3, em especial, poderá ser aplicado a áudios de tamanhos e formatos variados, oferecendo uma visão completa do conteúdo e facilitando a compreensão do seu contexto.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção, serão apresentados os resultados observados ao longo do processo de desenvolvimento do método de transcrição, Q&A, e modelagem de tópicos de dados de áudio.

O código utilizado foi implementado no ambiente de execução Google Colab, o qual provê um plano gratuito de computação em nuvem para o desenvolvimento colaborativo de algoritmos em linguagem Python. Cada módulo foi implementado em um *notebook* próprio, para melhor organização e modularidade. Na Figura 8, uma visão de alto nível do funcionamento do método proposto é representada.

Figura 8 – Visão Geral do Método



Fonte: Autor (2024).

O *pipeline* do método inicia-se no módulo I, responsável por realizar as transcrições dos áudios presentes em uma pasta especificada pelo usuário. Como saída, um arquivo JSON contendo as transcrições e metadados dos áudios de origem é criado, utilizado como entrada para os módulos II e III.

No módulo II, o objetivo é explorar informações específicas acerca do conteúdo das transcrições através de RAG e Q&A. Desse modo, o JSON é carregado para acessar as transcrições, que são convertidas em *embeddings* e utilizadas como contexto para uma pergunta específica através do *framework* LangChain. Em seguida, um *prompt* contendo a pergunta e o contexto é enviado à API do modelo Gemini, que fornece uma resposta personalizada.

No módulo III, busca-se obter uma visão geral dos tópicos abordados nos áudios. Assim como no módulo II, o JSON fornece as transcrições a serem analisadas, que são divididas em partes menores, convertidas em *embeddings*, e armazenadas em um *dataframe*. Em seguida, o BERTopic é aplicado para a modelagem de tópicos, identificando os principais tópicos e possibilitando a geração de visualizações.

A seguir, serão abordados os detalhes da implementação dos módulos e os resultados observados.

4.1 MÓDULO I: TRANSCRIÇÃO

O primeiro passo para a análise de dados textuais provenientes de áudios é a transcrição do conteúdo falado. Para isso, iniciou-se o processo utilizando o sistema de reconhecimento de fala OpenAI Whisper, devido à sua capacidade de transcrever áudios com precisão e eficácia. No Quadro 2, são descritos os passos para a construção do primeiro módulo.

Quadro 2 – Implementação do Módulo I

Principais recursos utilizados	Google Colab; JSON; OpenAI Whisper; Google Drive; Pydub; TinyTag;
Passo 1	A implementação do módulo se inicia com a instalação das dependências e bibliotecas necessárias para ler, manipular e transcrever arquivos de áudio. Nesse passo, também é realizada a conexão com o Google Drive, onde a pasta com os arquivos de áudio foi armazenada.
Passo 2	Em seguida, são definidas as funções para listar os áudios e extrair seus metadados. A função "list_audios" recebe como parâmetro o caminho de um arquivo ou pasta, no formato de <i>string</i> . Caso o

	<p>caminho seja de uma pasta, a função busca por áudios dentro da pasta, armazenando em um dicionário o ID do áudio como chave e seus metadados (incluindo o caminho do arquivo) como valor. Caso o caminho seja de um arquivo, o procedimento é feito apenas para o arquivo em questão. Desse modo, obtém-se um dicionário que contém as informações de todos os arquivos de áudio de um caminho especificado.</p>
Passo 3	<p>No terceiro passo, o Whisper é carregado e uma função simples para gerar as transcrições é declarada, possibilitando a definição de certos parâmetros para o modelo, como idioma, tarefa, e uso de timestamps. Neste estudo, foi utilizado o modelo "large-v3-turbo", que oferece performance similar ao modelo "large-v3", mas exigindo consideravelmente menos recursos.</p>
Passo 4	<p>No quarto passo, a função "create_transcripts" é declarada, recebendo como parâmetro o dicionário de áudios gerado pela função "list_audios". Ao ser executada, a função cria uma pasta no mesmo diretório que o primeiro áudio do dicionário, caso ela não exista, para armazenar os arquivos de transcrição gerados. Em seguida, o algoritmo percorre o dicionário com as informações dos áudios, abrindo o arquivo de áudio em questão por meio do caminho extraído, e gerando as transcrições. Para cada transcrição concluída, o resultado da transcrição é armazenado como valor no dicionário fornecido e em um arquivo de texto separado, gerado para facilitar consultas posteriores (caso necessário). Após realizar a transcrição de todos os áudios, a função salva o dicionário como um arquivo JSON.</p>

Fonte: Autor (2024).

4.1.1 Teste Inicial

Após a implementação, as primeiras execuções do módulo foram realizadas para explorar a capacidade de transcrição de diferentes variações multilíngues do modelo Whisper: "base", "medium" e "large-v3-turbo". Mais informações sobre o Whisper podem ser encontradas na página oficial do modelo no GitHub³.

³ OPENAI. Whisper. Disponível em: <https://github.com/openai/whisper>. Acesso em: 8 dez. 2024.

Para isso, foram realizadas transcrições de dois áudios gravados pelo autor: um de um minuto e sete segundos e outro de um minuto e 41 segundos. O primeiro áudio é a leitura de dois parágrafos de uma página em inglês da Wikipedia⁴, com sotaque brasileiro, alguns ruídos no microfone, e pequenos erros de pronúncia.

O segundo áudio é a leitura de um texto em português gerado pelo modelo Copilot, com pronúncia clara e levemente acelerada. Essas características foram mantidas para explorar, superficialmente, a capacidade do modelo de lidar com situações adversas, comuns em áudios gravados no cotidiano.

Após a transcrição do primeiro áudio, observou-se que os três modelos tiveram desempenho similar, mas com algumas diferenças. O modelo “base” apresentou o maior número de erros dos três modelos, como esperado, transcrevendo “*these abilities*” como “*disabilities*” e “*corpora*” como “*corpore*”, entre outros erros. No entanto, apesar de ser um dos menores modelos, seu desempenho foi muito bom, transcrevendo quase todo o texto corretamente.

Quadro 3 – Comparativo das Transcrições Geradas

<p>Texto lido pelo autor</p>	<p>A large language model is a type of computational model designed for natural language processing tasks such as language generation. As language models, LLMs acquire these abilities by learning statistical relationships from vast amounts of text during a self-supervised and semi-supervised training process. The largest and most capable LLMs are artificial neural networks built with a decoder-only transformer-based architecture, enabling efficient processing and generation of large-scale text data. Modern models can be fine-tuned for specific tasks or guided by prompt engineering. These models acquire predictive power regarding syntax, semantics, and ontologies inherent in human language corpora, but they also inherit inaccuracies and biases present in the data on which they are trained.</p>
<p>Transcrição gerada pelo modelo “base”</p>	<p>A large language model is a type of computational model designed for natural language processing tasks such as language generation. As language models, LLMs acquire disabilities by learning statistical relationships from vast amounts of text during a self-supervised and semi-supervised training process. The largest and most capable LLMs are artificial neural networks built with a decoder only,</p>

⁴ WIKIPEDIA. Large language model. Disponível em: https://en.wikipedia.org/wiki/Large_language_model. Acesso em: 9 dez. 2024.

	transformer-based architecture enabling efficient processing and generation of large-scale text data. Modern models can be fine tuned for specific tasks or be guided by prompt engineering. These models acquire predictive power regarding syntax, semantics and ontologies inherent in human language corpore, but also inherent inaccuracies and biases present in the data on which they are trained.
Transcrição gerada pelo modelo "medium"	A large language model is a type of computational model designed for natural language processing tasks, such as language generation. As language models, LLMs acquire disabilities by learning statistical relationships from vast amounts of text during a self-supervised and semi-supervised training process. The largest and most capable LLMs are artificial neural networks built with a decoder-only transformer-based architecture, enabling efficient processing and generation of large-scale text data. Modern models can be fine-tuned for specific tasks or be guided by prompt engineering. These models acquire predictive power regarding syntax, semantics, and ontologies inherent in human language corpora, but they also inherit inaccuracies and biases present in the data on which they are trained.
Transcrição gerada pelo modelo "large-v3-turbo"	A large language model is a type of computational model designed for natural language processing tasks such as language generation. As language models, LLMs acquire these abilities by learning statistical relationships from vast amounts of text during a self-supervised and semi-supervised training process. The largest and most capable LLMs are artificial neural networks built with a decoder-only, transformer-based architecture enabling efficient processing and generation of large scale text data. Modern models can be fine-tuned for specific tasks or be guided by prompt engineering. These models acquire predictive power regarding syntax, semantics and ontologies inherent in human language corpora, but they also inherit inaccuracies and biases present in the data on which they are trained.

Fonte: Autor (2024).

O modelo "medium" apresentou algumas melhorias em relação ao modelo "base", especialmente na pontuação, captando melhor as pausas entre palavras. A transcrição da palavra "*corpora*" foi corrigida, mas o erro de transcrição para as palavras "*these abilities*" permaneceu. Assim como o modelo "base", o modelo "medium" se mostrou eficaz no exemplo aplicado. O modelo "large-v3-turbo",

conforme esperado, apresentou a melhor performance entre os três modelos, transcrevendo corretamente todo o texto.

Em seguida, o teste foi repetido com o segundo áudio, lido em português. Nesse caso, os modelos demonstraram diferenças mais acentuadas entre si. O modelo "base" apresentou o maior número de erros, transcrevendo incorretamente algumas palavras e errando a escrita de outras. Já o modelo "medium" mostrou ganhos significativos na precisão da transcrição em relação ao modelo "base", mas ainda com erros perceptíveis. O modelo "large-v3-turbo", por fim, demonstrou performance consideravelmente superior em relação aos outros dois modelos; após a leitura do texto gerado, não foram encontrados erros perceptíveis na transcrição do segundo áudio de teste.

Contudo, vale ressaltar que o fenômeno observado é esperado devido às diferenças no número de parâmetros entre os modelos testados, que oferecem opções com diferentes custos computacionais para diversas aplicações. Radford *et al.* (2022) apresentam mais detalhes sobre o funcionamento do Whisper, abordando aspectos da arquitetura do sistema e sua performance em relação a outros modelos.

Com base nessas considerações iniciais, decidiu-se utilizar o modelo "large-v3-turbo" para conduzir as análises, devido à sua performance superior em ambos os idiomas e custo computacional similar ao do modelo "medium". Dessa forma, foi possível gerar transcrições para todos os áudios do curso analisado de forma automatizada, possibilitando o acesso ao conteúdo falado.

4.1.2 Análise das Transcrições

Após a seleção do modelo, o algoritmo foi aplicado na série de áudios do curso "*The Surveillance State: Big Data, Freedom, and You*", de Paul Rosenzweig (2016). Ao todo, 24 arquivos de áudio, de aproximadamente 30 minutos cada, foram transcritos, em um processo que durou cerca de 36 minutos, utilizando o recurso GPU T4 do ambiente de execução do Colab. O resultado das transcrições foi armazenado em um arquivo JSON, contendo todas as transcrições.

De modo geral, ao ler o arquivo, observou-se que a maioria das transcrições se mostrou precisa e coerente na maior parte dos casos. Entretanto, um fenômeno específico foi observado; ao final de determinadas transcrições, o modelo utilizado apresentou "alucinações", incorporando frases e palavras que não foram ditas, logo

após as frases finais devidamente transcritas. Na Figura 9, uma captura de tela mostra uma das ocorrências do fenômeno, destacada em azul.

Figura 9 – “Alucinação” em uma transcrição

state residents said the greater the surveillance was of them, the more negative they found the work environment. If it was you at the dinner table making a joke about taping a friendly conversation, you might be the one loudest to laugh. And if it was the person next to you, you might instinctively respond suspiciously, or at the least very cautiously. The parable of the panopticon, the circular jail, is a provocative stage on which the tension between the observer and the observed plays out. And how you feel might very well depend on whether you are the one observing or being observed. This willselling the scene of Carpenter support, five hours on the safety привил Clemset stage ofzenima care plans, and episode velocity out of the air of the Od запись triggered that by the Timar te sicphet of the poop knight on the broadcast. So it turns way to theんな actors that want to aware that action is in the debate. Things willampoo take four hours and take four hours"

},
"audio 4": f

Fonte: Autor (2024).

Desse modo, ao considerar que nos casos observados apenas uma pequena porção do texto foi afetada (especificamente após o final do áudio transcrito), decidiu-se manter o artefato. No entanto, o resultado destaca a necessidade de implementar mecanismos para minimizar e tratar alucinações ao utilizar modelos generativos para transcrição de áudio, pois estas podem impactar negativamente análises sensíveis a ruídos e falhas.

Portanto, é essencial desenvolver soluções que garantam a precisão e a confiabilidade das transcrições, especialmente em contextos em que a integridade dos dados é crítica.

4.2 MÓDULO II: Q&A

Uma vez geradas as transcrições, conteúdos falados de áudios podem ser utilizados como uma fonte de informações úteis. Uma das tarefas de NLP aplicáveis a dados textuais são as perguntas e respostas (Q&A), que utiliza modelos para identificar porções de texto relevantes a uma consulta. Desse modo, o segundo módulo do método visa possibilitar que o usuário faça perguntas a um modelo e obtenha respostas contextualizadas, facilitando a obtenção de informações específicas. No Quadro 4, descreve-se o processo de implementação do módulo II.

Quadro 4 – Implementação do Módulo II

Principais recursos utilizados	Google Colab; Google Drive; JSON; Langchain; LangGraph; Pathlib; Google Vertex AI; Gemini; Chroma DB.
Passo 1	A implementação do módulo II, assim como no módulo I, inicia-se instalando as dependências e bibliotecas necessárias. A conexão com o Google Drive é realizada para acessar o arquivo JSON contendo as transcrições geradas. Além disso, efetua-se também a conexão com a API do LangChain, que facilita a criação de aplicações que utilizam LLMs.
Passo 2	O segundo passo é o carregamento do arquivo JSON que contém as transcrições por um objeto do tipo DocumentLoader, que o armazena como uma lista de documentos.
Passo 3	No terceiro passo, os documentos são divididos em fragmentos (<i>chunks</i>) de 1000 caracteres, com 200 caracteres de sobreposição. Esse passo é necessário para não exceder a janela de contexto do modelo de linguagem. Além disso, a divisão em fragmentos facilita a recuperação das porções mais relevantes do documento.
Passo 4	No quarto passo, os fragmentos são convertidos em <i>embeddings</i> utilizando o modelo Gemini. Em seguida, os dados são armazenados em um banco de dados vetorial (Chroma), que possibilita a recuperação de informações pertinentes à consulta.
Passo 5	No quinto passo, é realizada a conexão com a API do Vertex AI, uma plataforma que permite a conexão com o modelo Gemini para envio de <i>prompts</i> e geração de texto. O processo é feito por meio de uma chave gerenciada no console do Google Cloud.
Passo 6	Em seguida, o RAG é executado, recuperando as informações específicas armazenadas no banco vetorial a partir de uma pergunta ou consulta do usuário. Em seguida, o contexto e a pergunta são enviados ao modelo Gemini, que elabora a resposta.
Passo 7	No último passo, são implementadas determinadas funcionalidades do LangGraph, que não só facilita o gerenciamento e utilização do LLM, mas também proporciona uma visão mais clara das funcionalidades implementadas anteriormente.

Fonte: Autor (2024).

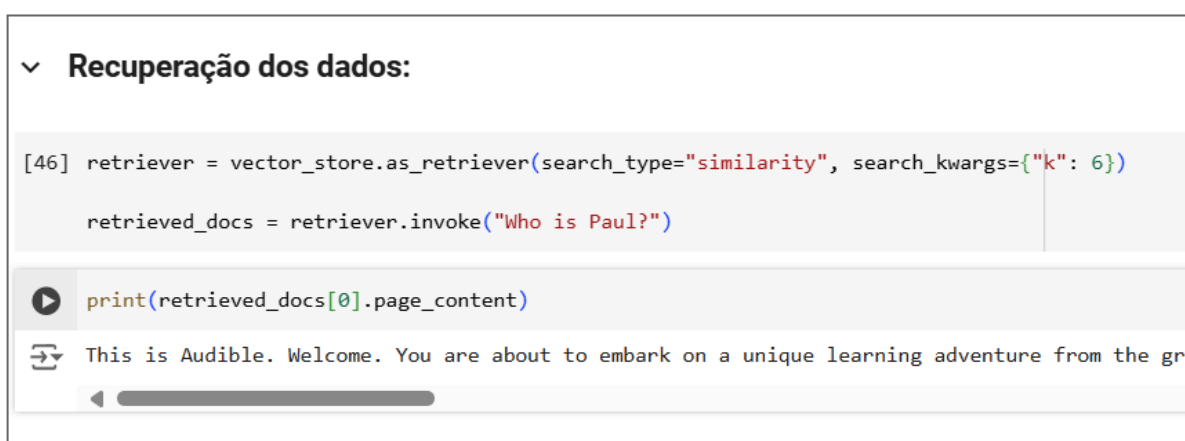
4.2.1 Análise dos Componentes do Q&A

Após a implementação dos componentes principais, o primeiro passo foi identificar se o arquivo JSON havia sido corretamente carregado. Para isso, foram utilizados comandos *print* para identificar os elementos da lista de documentos, que foram exibidos na tela. O mesmo procedimento foi realizado para os fragmentos, que exibiram informações a respeito do documento de origem, como metadados, número sequencial, e índice do início do documento.

Em seguida, os comandos para geração dos *embeddings* foram executados, e os dados, armazenados no banco de dados vetorial. A partir dessa etapa, é possível realizar consultas ao banco vetorial utilizando a interface *retriever* do LangChain, que retorna os fragmentos de documentos mais próximos aos termos da consulta, através da função "*retriever.invoke*".

Ao realizar a consulta "*Who is Paul?*", por exemplo, o *retriever* forneceu como resultado mais provável os fragmentos do primeiro áudio, que introduzem o curso e falam sobre o autor, Paul Rosenzweig. Desse modo, conforme a Figura 10, a geração dos *embeddings* demonstrou a eficácia esperada ao capturar o significado das palavras, cujos documentos correspondentes foram corretamente recuperados.

Figura 10 – Resultado da Consulta ao Banco Vetorial



```
▼ Recuperação dos dados:

[46] retriever = vector_store.as_retriever(search_type="similarity", search_kwargs={"k": 6})

    retrieved_docs = retriever.invoke("Who is Paul?")

▶ print(retrieved_docs[0].page_content)

↔ This is Audible. Welcome. You are about to embark on a unique learning adventure from the gre
```

Fonte: Autor (2024).

A partir do resultado observado, o próximo passo foi efetivar o RAG através da integração do LangChain com o modelo Gemini. O *prompt* enviado ao modelo possui o seguinte formato:

“You are an assistant for question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just say that you don't know. Use three sentences maximum and keep the answer concise.

Question: (question goes here)

Context: (context goes here)

Answer:”

Após a execução das linhas de código necessárias para incorporar o contexto e a consulta do usuário ao *prompt* a ser enviado, a conexão com a API do Gemini foi estabelecida e a pergunta realizada. Para fins de comparação, foi enviada a pergunta *“Who is Paul?”*, desta vez ao modelo Gemini junto ao contexto incorporado automaticamente. A Figura 11 apresenta o resultado da consulta.

Figura 11 – Resposta do Modelo Gemini com RAG



```

from langchain_core.output_parsers import StrOutputParser
from langchain_core.runnables import RunnablePassthrough

def format_docs(docs):
    return "\n\n".join(doc.page_content for doc in docs)

rag_chain = (
    {"context": retriever | format_docs, "question": RunnablePassthrough()}
    | prompt
    | llm
    | StrOutputParser()
)

for chunk in rag_chain.stream("Who is Paul?"):
    print(chunk, end="", flush=True)

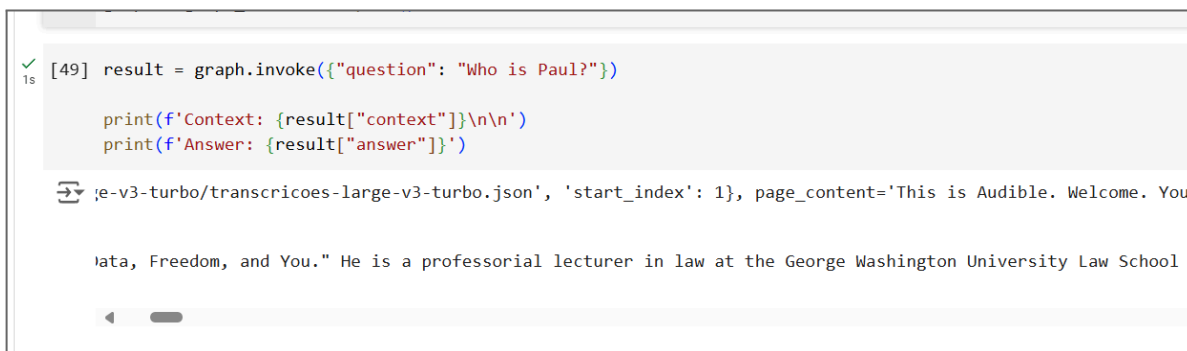
```

Paul Rosenzweig is the lecturer for the Great Courses series titled "The Surveillance State, Big Data, Freedom, and You."

Fonte: Autor (2024).

Na saída exibida na Figura 11, observou-se que a resposta foi respondida corretamente com base no conteúdo das transcrições. Além disso, o algoritmo proveu a resposta utilizando linguagem natural, o que demonstra o funcionamento esperado do RAG. Entretanto, a saída do programa não exibiu os fragmentos do documento utilizados como contexto. Desse modo, o último passo é executado para melhor observação dos resultados.

Figura 12 – Resposta do Modelo Gemini e Contexto



```

✓ [49] result = graph.invoke({"question": "Who is Paul?"})
      print(f'Context: {result["context"]}\n\n')
      print(f'Answer: {result["answer"]}')

↩ e-v3-turbo/transcricoes-large-v3-turbo.json', 'start_index': 1}, page_content='This is Audible. Welcome. You
ata, Freedom, and You." He is a professorial lecturer in law at the George Washington University Law School a

```

Fonte: Autor (2024).

Na Figura 12, a saída é exibida de forma mais clara, apresentando os fragmentos de documentos relacionados à consulta enviados como contexto, além da resposta do modelo. Sendo assim, é possível identificar quais documentos foram utilizados para responder à pergunta, proporcionando maior transparência no processo. Uma vez demonstrado o funcionamento do modelo, a análise exploratória das transcrições foi realizada por meio de três perguntas específicas.

4.2.2 Perguntas e Respostas (Q&A)

Para verificar o funcionamento do Módulo II, foram elaboradas três perguntas específicas, baseadas nos conteúdos das transcrições: “*What is privacy?*”, “*What is personal data?*”, e “*What is freedom?*”. Essas perguntas refletem conceitos e reflexões abordados ao longo dos áudios sob determinadas nuances. Desse modo, caso o modelo não utilize corretamente o contexto, é provável que as respostas apresentadas sejam mais genéricas e desconexas. No Quadro 5, as saídas do modelo são exibidas:

Quadro 5 – Respostas a Perguntas Específicas

Pergunta	Resposta do modelo Gemini + RAG
<i>What is privacy?</i>	<i>Answer: Privacy refers to the independence of personal activity and autonomy. We protect privacy through secrecy and by allowing individuals to exercise their choices, like in voting booths or freedom of religion. Privacy allows people to act freely in public.</i>

<p><i>What is personal data?</i></p>	<p><i>Answer: Personal data refers to information that can be used to identify an individual. This data is widely collected and available in our increasingly networked world. The collection and analysis of personal data has become increasingly prevalent, used by both governments and businesses for various purposes, from targeted advertising to identifying potential threats.</i></p>
<p><i>What is freedom?</i></p>	<p><i>Answer: The provided context focuses on the tension between privacy and security, particularly within the context of American government and its history. It doesn't offer a direct definition of freedom, but it suggests that freedom is closely tied to the concept of autonomy and the ability to make choices without undue government interference.</i></p>

Fonte: Autor (2024).

Portanto, ao analisar as respostas e compará-las com os contextos fornecidos, observou-se que o modelo baseou suas respostas no conteúdo das transcrições, destacando aspectos políticos, geopolíticos e éticos, e o uso da primeira pessoa ao fornecer as respostas também foi observado. Além disso, a limitação de três frases por resposta foi respeitada diante das perguntas realizadas. Desse modo, o modelo apresentou consistência e aderência aos procedimentos estabelecidos no *prompt*, fornecendo respostas curtas e contextualizadas.

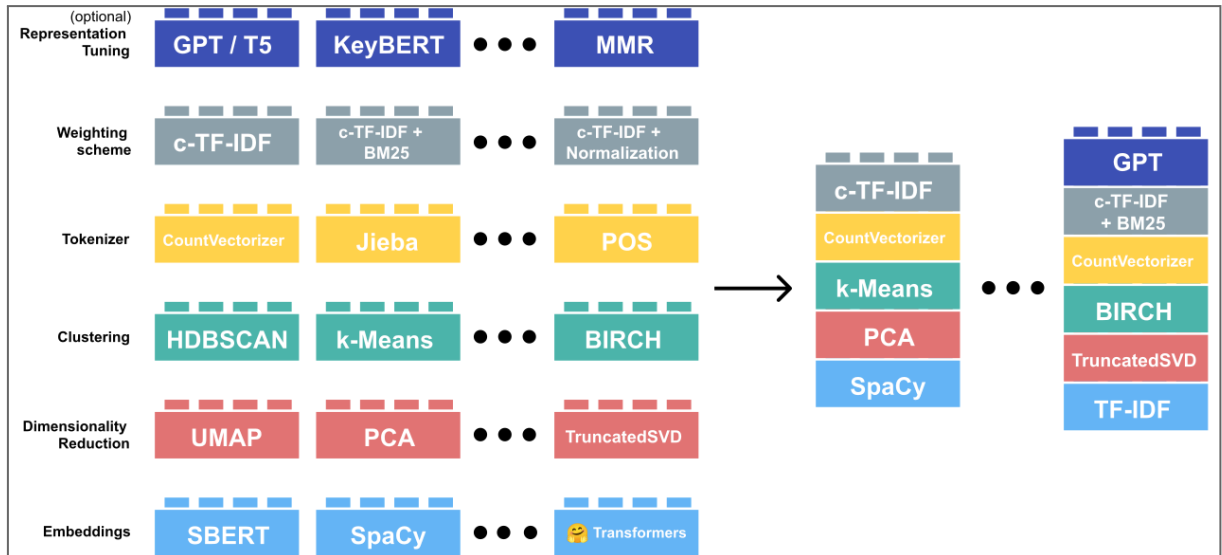
4.3 MÓDULO III: MODELAGEM DE TÓPICOS

O Módulo II, de perguntas e respostas, fornece uma visão próxima dos dados, possibilitando a obtenção de informações específicas. Por outro lado, observá-los a partir de uma visão geral ajuda não só a compreender o todo, mas também proporciona *insights* que levam a melhores perguntas. Sob essa perspectiva, o Módulo III do método foi desenvolvido com o objetivo de melhor compreender as temáticas, padrões e tendências em grandes volumes de informações presentes em transcrições de áudio.

Para isso, o BERTopic foi utilizado. O BERTopic é um *framework* que integra diferentes técnicas e modelos com o objetivo de criar representações textuais eficientes, e utilizá-las para a extração de tópicos. Isso é feito através do agrupamento de dados textuais no formato de *embeddings*, que passam por uma

etapa de redução de dimensionalidade. Em seguida, os tópicos são obtidos com uma variante do TF-IDF baseada em classes (Grootendorst, 2022).

Figura 13 – Modularidade do *Framework* BERTopic



Fonte: Grootendorst (2024).

Uma das características notáveis do BERTopic é a sua modularidade, que possibilita a construção de um modelo de tópicos personalizado. Desse modo, é possível ajustá-lo de forma a melhor representar diversos tipos de dados textuais e comparar, de maneira simplificada, o desempenho de diferentes técnicas. Na Figura 13, a estrutura do BERTopic é representada.

Tendo em vista as suas vantagens, o módulo III foi construído com base no *framework* BERTopic, não só com o objetivo identificar temas e tópicos relevantes, mas também para explorar o potencial da visualização de dados na observação desses dados, atribuindo um significado aos *clusters* identificados pelo algoritmo de agrupamento. A seguir, os passos para a construção do terceiro módulo são descritos.

Quadro 6 – Implementação do Módulo III

Principais recursos utilizados	Google Colab; Google Drive; BERTopic; JSON; Pandas; NLTK; SentenceTransformer; UMAP; HDBSCAN; CountVectorizer; Matplotlib; Plotly.
--------------------------------	--

Passo 1	No primeiro passo, as dependências são instaladas e as bibliotecas instaladas, assim como nos módulos anteriores. A conexão com o Google Drive é realizada para acessar o conteúdo das transcrições.
Passo 2	No segundo passo, ocorre o carregamento dos dados, através da leitura do arquivo JSON. Em seguida, o conteúdo é armazenado em um <i>dataframe</i> , que contém não apenas a transcrição, mas também alguns metadados do áudio de origem.
Passo 3	Devido ao baixo número de amostras no <i>dataframe</i> utilizado, a análise foi conduzida a nível de frases, o que possibilita a identificação de tópicos em diferentes partes de um mesmo áudio. Para isso, é utilizado o método “sent_tokenize”, armazenando os fragmentos em uma lista. Uma segunda lista é utilizada para que a informação do título do áudio não seja perdida, associando cada frase a um áudio.
Passo 4	Nessa etapa, o modelo “all-distilroberta-v1” é carregado e inicializado para a geração dos <i>embeddings</i> , aplicando-o em seguida nos fragmentos das transcrições. Para isso, é utilizada a biblioteca Sentence Transformers (SBERT), que obtém o modelo através da plataforma Hugging Face. Vale ressaltar que o modelo em questão foi treinado para tarefas de NLP em inglês. Contudo, modelos multilíngues também podem ser facilmente carregados e utilizados.
Passo 5	Na quinta etapa, o BERTopic é inicializado, definindo seus parâmetros e gerando o modelo de tópicos a partir do conjunto de <i>embeddings</i> das transcrições. Para essa análise, com exceção do gerador de <i>embeddings</i> , especificado anteriormente, foram utilizados os parâmetros e algoritmos padrão do BERTopic (UMAP, HDBSCAN, CountVectorizer, e c-TF-IDF + MMR). Com isso, obtém-se uma tabela, com os tópicos identificados, e a distribuição de termos e frases relevantes. Para a remoção de <i>stopwords</i> , a função integrada do vetorizador é utilizada.
Passo 6	Por fim, na última etapa, as visualizações de dados são geradas utilizando as funções do BERTopic e as bibliotecas Matplotlib e Plotly.

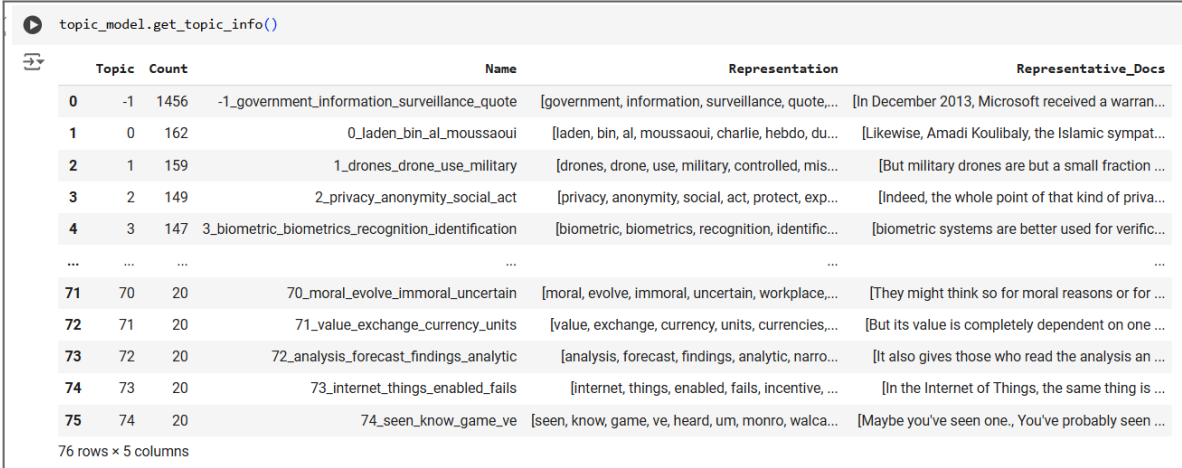
Fonte: Autor (2024).

4.3.1 Análise Preliminar dos Resultados

O primeiro aspecto analisado na execução do algoritmo foi o Passo 3. Como o conjunto de transcrições de áudio foi segmentado em partes menores para facilitar a identificação da distribuição de tópicos em um mesmo áudio, tornou-se necessário verificar se esses fragmentos estavam corretamente associados aos documentos de origem. Para isso, foi realizada uma verificação preliminar, imprimindo as listas de fragmentos e seus títulos correspondentes, garantindo a manutenção da relação entre os dados e sua fonte. Essa verificação confirmou que os títulos estavam devidamente associados aos fragmentos, permitindo o avanço para a próxima etapa da análise.

Ao aplicar os passos do framework BERTopic, foi gerada uma tabela contendo os tópicos identificados, os agrupamentos correspondentes e as frases pertencentes a cada tópico. No total, o algoritmo identificou 74 tópicos potenciais em um conjunto de 5.306 fragmentos de transcrições. Desses, 1.456 fragmentos não se encaixaram em nenhum tópico específico (designados como tópico -1) e foram classificados como *outliers* pelo algoritmo. A Figura 14 ilustra algumas linhas da tabela gerada.

Figura 14 – Tópicos Gerados pelo BERTopic



Topic	Count	Name	Representation	Representative Docs	
0	-1	1456	-1_government_information_surveillance_quote	[government, information, surveillance, quote,...	[In December 2013, Microsoft received a warran...
1	0	162	0_laden_bin_al_moussaoui	[laden, bin, al, moussaoui, charlie, hebdo, du...	[Likewise, Amadi Koulibaly, the Islamic sympat...
2	1	159	1_drones_drone_use_military	[drones, drone, use, military, controlled, mis...	[But military drones are but a small fraction ...
3	2	149	2_privacy_anonymity_social_act	[privacy, anonymity, social, act, protect, exp...	[Indeed, the whole point of that kind of priva...
4	3	147	3_biometric_biometrics_recognition_identification	[biometric, biometrics, recognition, identific...	[biometric systems are better used for verific...
...
71	70	20	70_moral_evolve_immoral_uncertain	[moral, evolve, immoral, uncertain, workplace,...	[They might think so for moral reasons or for ...
72	71	20	71_value_exchange_currency_units	[value, exchange, currency, units, currencies,...	[But its value is completely dependent on one ...
73	72	20	72_analysis_forecast_findings_analytic	[analysis, forecast, findings, analytic, narro...	[It also gives those who read the analysis an ...
74	73	20	73_internet_things_enabled_fails	[internet, things, enabled, fails, incentive, ...	[In the Internet of Things, the same thing is ...
75	74	20	74_seen_know_game_ve	[seen, know, game, ve, heard, um, monro, walca...	[Maybe you've seen one., You've probably seen ...

76 rows x 5 columns

Fonte: Autor (2024).

A partir dos dados obtidos, foi possível aplicar diferentes visualizações para explorar o conteúdo de forma mais eficiente. Para isso, foram utilizadas tanto as

funções de visualização integradas do BERTopic quanto gráficos personalizados, gerados com as bibliotecas Matplotlib e Plotly. A seguir, são apresentados os gráficos resultantes, que oferecem uma compreensão mais clara e detalhada dos dados analisados.

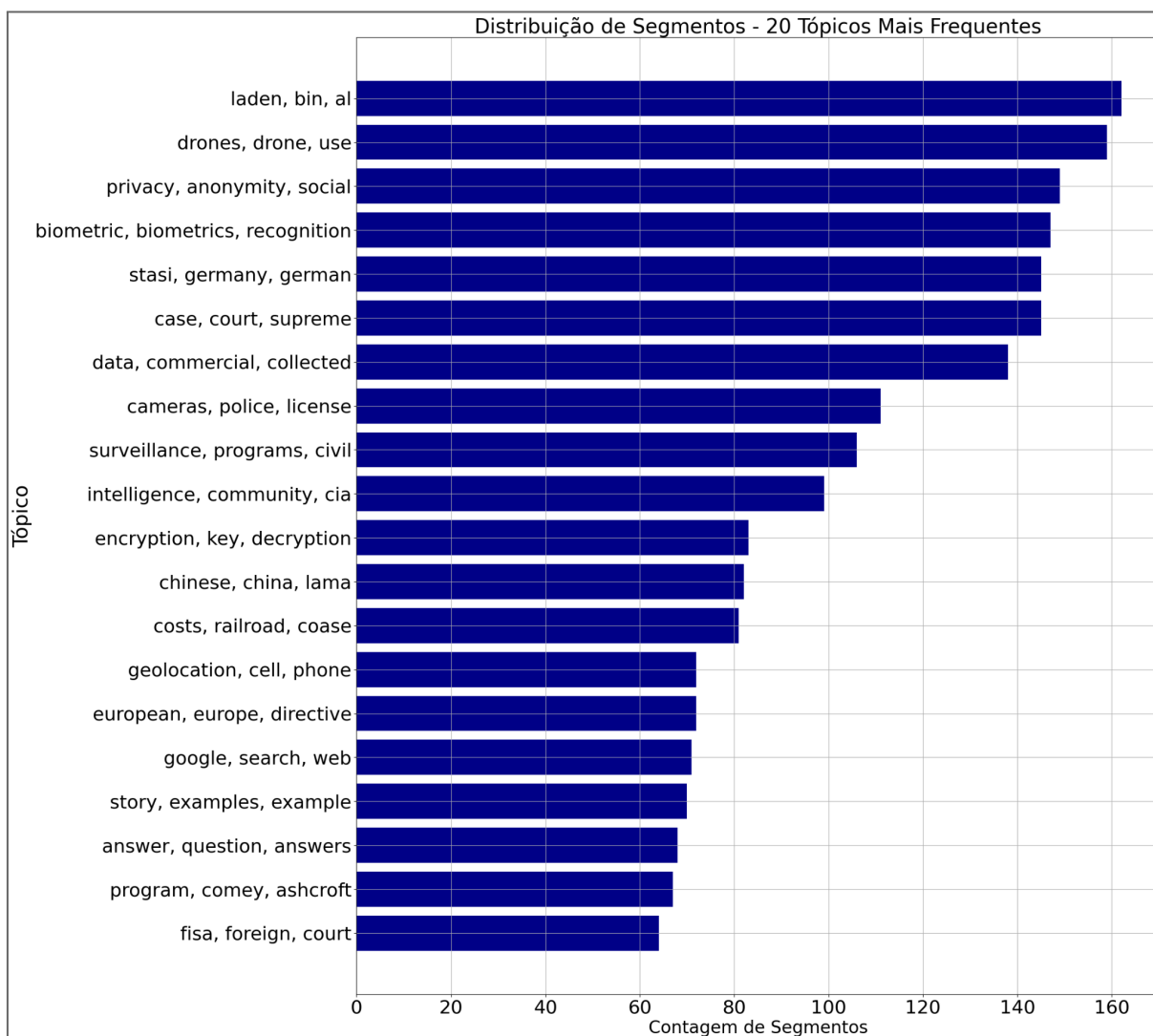
4.3.2 Visualização de Tópicos

4.3.2.1 Distribuição de Fragmentos por Tópico

Diante da grande quantidade de tópicos gerados pelo algoritmo, uma maneira eficaz de compreendê-los é visualizar a distribuição de fragmentos por tópico. Assim, a análise exploratória de dados iniciou-se com a criação de um gráfico de barras destacando os 20 tópicos mais frequentes, conforme apresentado na Figura 15.

Ao analisar o gráfico, observa-se que os tópicos “laden, bin, al” e “drones, drone, use” possuem o maior número de fragmentos associados. Em seguida, destacam-se tópicos relacionados à privacidade, biometria, Stasi e Alemanha, casos jurídicos e coleta de dados. Além disso, temas como câmeras, programas de vigilância e serviços de inteligência também aparecem com frequência significativa.

Figura 15 – Distribuição de Fragmentos por Tópico



Fonte: Autor (2024).

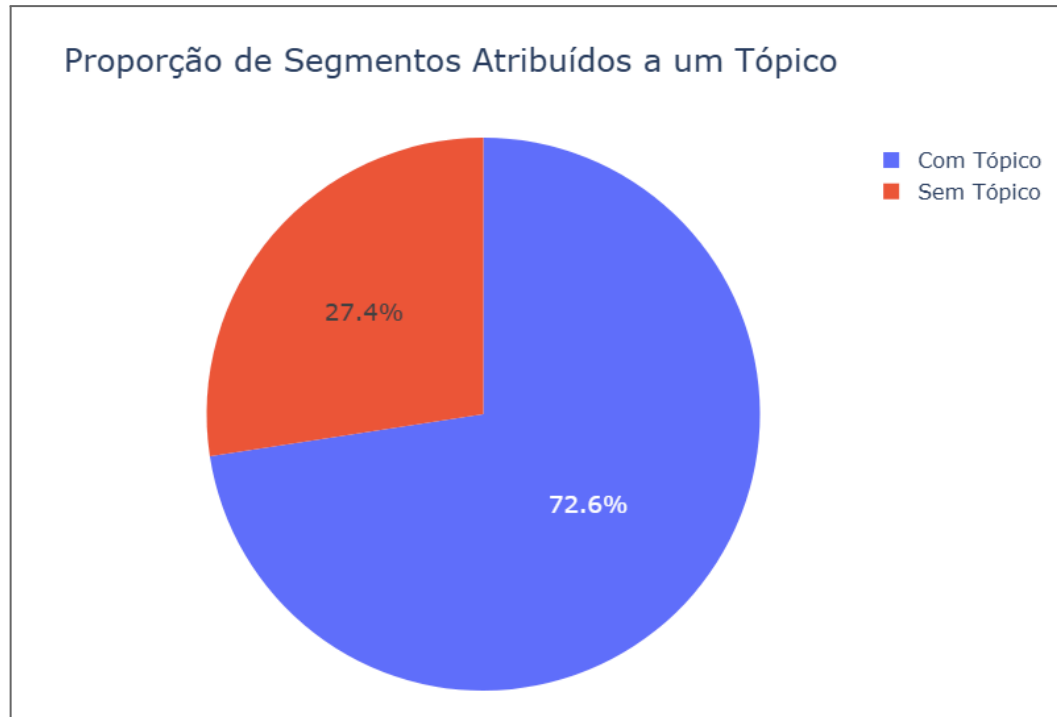
Nessa análise, a visualização ressalta a ênfase em aspectos geopolíticos e sociais, alinhando-se à temática central do curso. Além disso, o gráfico evidencia a frequência com que os tópicos são abordados, incentivando análises mais aprofundadas sobre temas específicos. Por outro lado, alguns fragmentos não foram associados a nenhum tópico definido, destacando a necessidade de explorar melhor esse cenário.

4.3.2.2 Proporção de Outliers

Para tornar esse fenômeno mais claro, foi elaborada uma visualização em formato de gráfico de pizza, ilustrando a proporção entre fragmentos com tópicos

definidos e aqueles classificados como *outliers*. O resultado pode ser observado na Figura 16.

Figura 16 – Proporção de Fragmentos Atribuídos a um Tópico



Fonte: Autor (2024).

Ao gerar o gráfico, observou-se que a maioria dos fragmentos (72,6%) foi atribuída a tópicos específicos. No entanto, aproximadamente um quarto dos fragmentos foi classificado como *outlier*, sem associação a nenhum tópico definido. Assim, o gráfico de pizza oferece uma visão geral do total de fragmentos, complementando o primeiro gráfico e fornecendo insights sobre os dados e aspectos do processo de modelagem de tópicos.

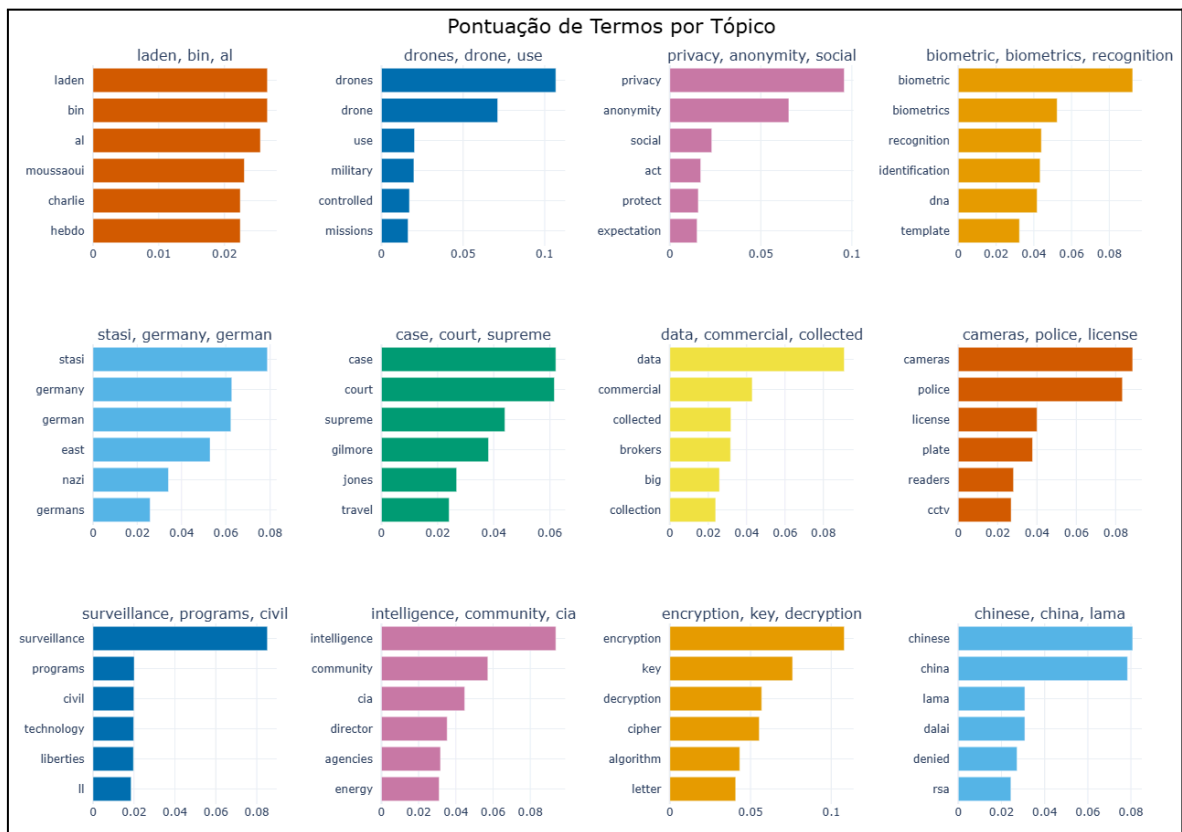
4.3.2.3 Distribuição de Palavras por Tópico

Com base na análise da distribuição dos fragmentos por tópico, buscou-se identificar os termos mais relevantes utilizados pelo algoritmo na modelagem dos tópicos mais comuns. Para isso, foram selecionados os 12 tópicos mais frequentes, e gráficos de barras foram gerados para ilustrar a pontuação dos termos, conforme mostrado na Figura 17.

A inspeção dos gráficos revela que o primeiro tópico, “laden, bin, al”, faz referência a nomes associados ao 11 de Setembro e a outros incidentes envolvendo conflitos ideológicos ou políticos. Já o segundo tópico, “drones, drone, use”, apresenta uma alta pontuação para termos relacionados ao dispositivo, sugerindo também possíveis aplicações militares. No tópico “privacy, anonymity, social”, destacam-se termos associados à privacidade e ao anonimato, incluindo referências a legislações e questões sociais. Por sua vez, o quarto tópico abrange termos relacionados à biometria e ao reconhecimento de identidade.

Assim, de maneira semelhante aos tópicos mais frequentes, os demais tópicos abrangem termos com diferentes graus de relação ao tema atribuído. Essas relações, representadas nos gráficos, indicam que certos tópicos são mais bem definidos e consistentes do que outros.

Figura 17 – Pontuação de Termos por Tópico



Fonte: Autor (2024).

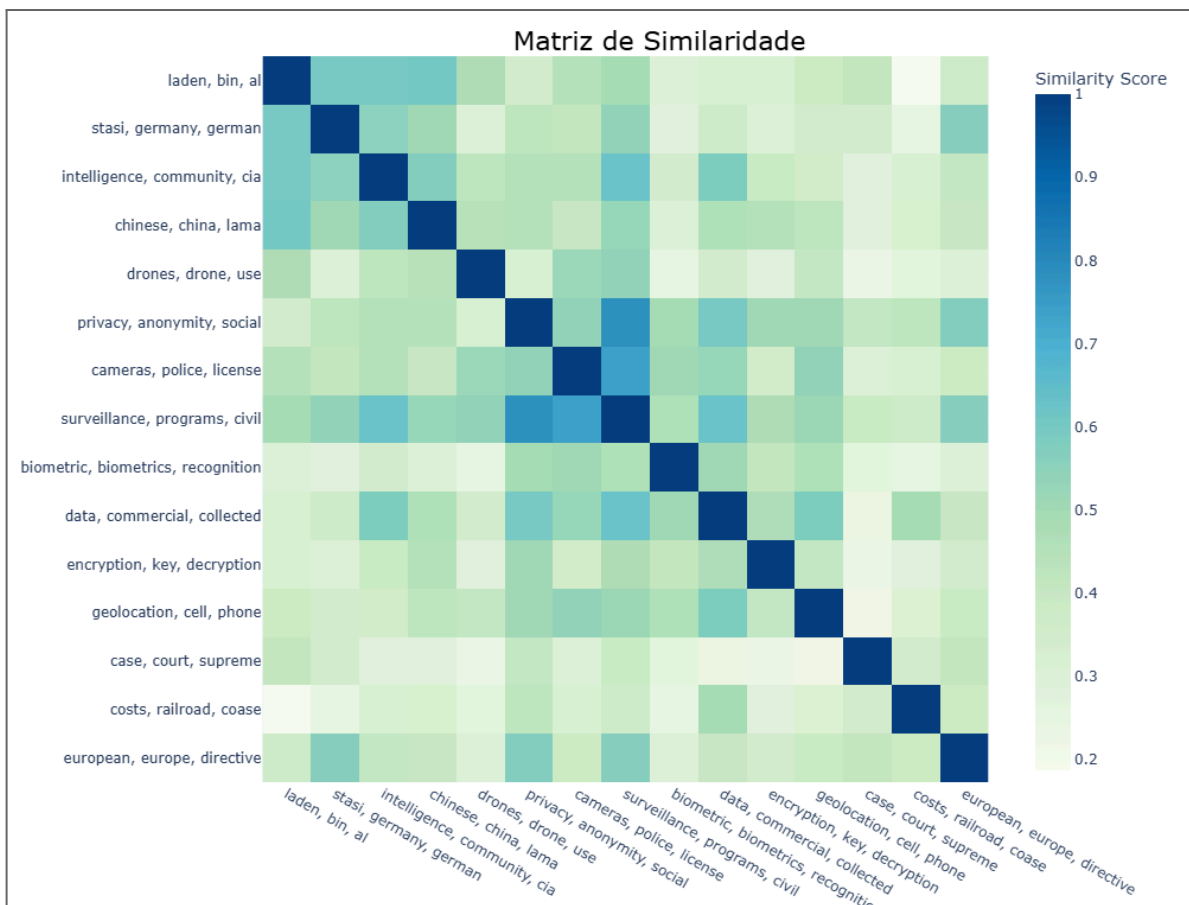
Ainda assim, os gráficos demonstram que o modelo foi capaz de identificar as temáticas centrais de determinados áudios, abordando aspectos políticos,

ideológicos, éticos e de privacidade. Isso pode estar relacionado ao uso de *embeddings* para a representação dos dados textuais, o que contribui para o agrupamento de termos semanticamente e contextualmente similares.

4.3.2.4 Matriz de Similaridade Entre Tópicos

Depois de explorar as relações entre os termos e os tópicos, foi gerado um mapa de calor para observar a similaridade entre tópicos. Para isso, foram selecionados os quinze tópicos mais frequentes, exibidos na Figura 18. Ao utilizar o parâmetro “*n_clusters=5*”, foi possível visualizar melhor a similaridade entre determinados termos.

Figura 18 – Matriz de Similaridade Entre Tópicos



Fonte: Autor (2024).

O gráfico denota que a maioria dos tópicos demonstra similaridade baixa ou mediana, com valores entre 20% e 50%. Entretanto, nota-se também um grau de

similaridade alto entre determinados tópicos, representado em azul claro. Os tópicos “surveillance, programs, civil” e “privacy, anonymity, social” destacam-se em relação aos demais, com grau de similaridade próximo a 78%. Entre os tópicos “cameras, police, license” e “surveillance, programs, civil”, o mesmo ocorre, com grau de similaridade de aproximadamente 74%. Estas informações sugerem que, no contexto dos áudios de origem, há uma possível correlação entre esses tópicos.

Nos demais tópicos, não foram observados casos que particularmente se destacam, porém é possível identificar algumas zonas um pouco mais escuras no gráfico, com graus de similaridade próximos a 60%. Entre “surveillance, programs, civil” e “intelligence, community, cia”, por exemplo, o grau de similaridade é razoável, com cerca de 63%. O mesmo ocorre entre os termos “surveillance, programs, civil” e “data, commercial, collected”.

Entre “data, commercial, collected” e “privacy, anonymity, social”, a similaridade é próxima a 60%. No topo do gráfico, o tópico “laden, bin, al”, apresenta um grau de similaridade de cerca de 60% entre os tópicos “stasi, germany, german”, “intelligence, community, cia” e “chinese, china, lama”. O tópico “intelligence, community, cia”, por sua vez, apresenta cerca de 55% de similaridade entre “stasi, germany, german” e “chinese, china, lama”.

Portanto, nota-se uma relação de similaridade acima da média entre determinados tópicos, o que pode indicar uma abordagem temática consistente ao longo do conteúdo falado.

4.3.2.5 Gráfico de Dispersão

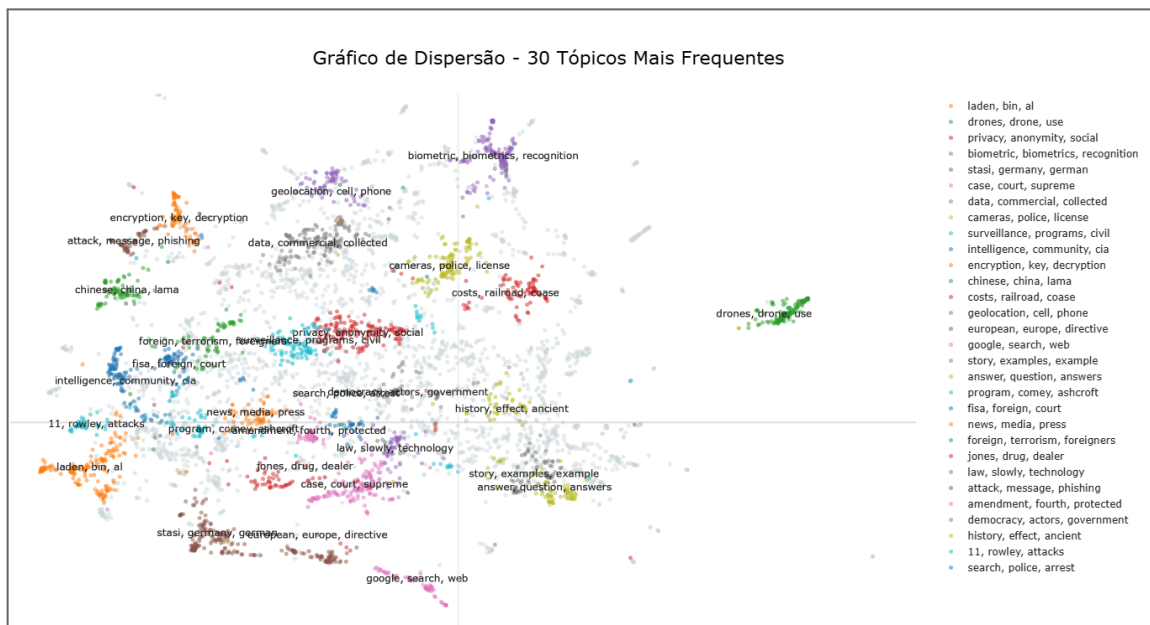
Após a observação da distribuição de tópicos em diferentes níveis e algumas de suas relações, a geração de uma visualização que sintetiza essas informações em um gráfico interativo possibilita melhor compreender o panorama dos tópicos descobertos. Sendo assim, utilizando as funções do BERTopic, um gráfico de dispersão foi gerado, possibilitando a identificação dos *clusters* criados pelo algoritmo de agrupamento.

Na fase inicial de implementação, observou-se que o BERTopic aborda, por padrão, cada item no conjunto de dados como um documento. Por essa razão, as primeiras visualizações geradas mostraram pouquíssimos pontos, refletindo a quantidade de transcrições presentes no arquivo JSON. Tendo isso em vista, a

biblioteca NLTK foi utilizada para dividir as transcrições em fragmentos, o que aumentou a granularidade da análise.

Desse modo, o gráfico representa a distribuição de tópicos ao longo das transcrições, em que cada ponto é um fragmento de transcrição. Na Figura 19, o gráfico com os 30 tópicos mais frequentes é apresentado. Essa seleção foi necessária para evitar a sobreposição excessiva de informações, dificultando a visualização do gráfico e sobrecarregando o modelo. Ao analisá-lo, foram realizadas determinadas observações.

Figura 19 – Visão Geral do Gráfico de Dispersão



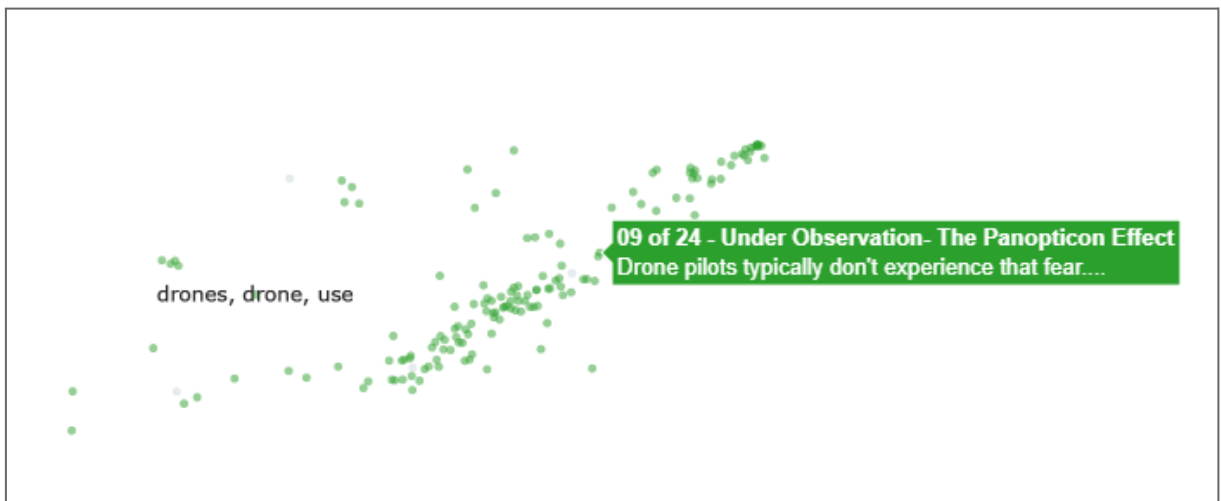
A primeira delas é que, apesar do baixo número de transcrições, o gráfico apresentou uma quantidade considerável de tópicos e fragmentos, sendo possível identificar com clareza alguns agrupamentos. Os tópicos “drones, drone, use” e “google, search, web”, por exemplo, apresentaram *clusters* bem definidos, sem se sobrepor a outros tópicos. Em outros casos, contudo, observou-se um agrupamento pouco definido, com pontos dispersos e misturados em meio a outros grupos.

Por outro lado, a distribuição dos pontos no gráfico remeteu ao potencial de compreensão semântica do modelo de linguagem pré-treinado utilizado, sugerindo a sua eficácia na análise de agrupamento e modelagem de tópicos. Ao explorar o gráfico, observou-se que o *framework* foi capaz de identificar, em diferentes casos,

as correlações entre fragmentos de um mesmo áudio e entre fragmentos de áudios distintos.

Além disso, uma das vantagens observadas na visualização ao longo do processo de análise exploratória foram as possibilidades de interação com o gráfico. A biblioteca Plotly, utilizada pelo BERTopic para a geração dos gráficos, possibilita destacar tópicos específicos, ocultando os demais. Além disso, também é possível aproximar-se de pontos específicos do gráfico ou identificar o fragmento em questão ao passar o cursor sobre os pontos, facilitando a compreensão da distribuição dos documentos ou fragmentos. Na Figura 20, essa funcionalidade é demonstrada.

Figura 20 – Visão Aproximada de um Tópico



Fonte: Autor (2024).

Ao utilizar o recurso para explorar os pontos do tópico “drones, drone, use”, observou-se que a maioria dos fragmentos correspondem ao áudio “10 of 24 - Drones, Drones Everywhere”, que aborda este tópico em específico, fenômeno comum observado também em outros tópicos. Entretanto, fragmentos de outros áudios, como o exibido na Figura 20, também foram identificados no *cluster* apesar da sua distância dos demais pontos.

Nessa perspectiva, o resultado sugere a eficácia da fragmentação como estratégia para aumentar a granularidade da análise, uma vez que ela pode proporcionar a identificação de correlações que talvez não ocorreriam a nível de documento.

Sendo assim, os resultados do módulo III apontam que o *framework* BERTopic possibilitou melhor compreensão das temáticas e fenômenos identificados no conjunto de transcrições, provendo ferramentas diversas para explorar os dados sob diferentes métodos e ângulos e destacando o papel da visualização de dados como uma ferramenta essencial na análise de dados.

5 CONSIDERAÇÕES FINAIS

As considerações finais deste estudo refletem os principais achados e a contribuição do método proposto para a análise de transcrições de áudio, destacando sua aplicabilidade, flexibilidade e escalabilidade. A combinação de técnicas de processamento de linguagem natural (NLP) e visualização de dados proporcionou uma abordagem sólida para compreender, explorar e analisar de forma eficiente os temas abordados nos áudios.

O método, composto por três módulos interconectados, demonstrou ser eficaz na transcrição de áudio, na recuperação de respostas contextuais por meio de Q&A, e na modelagem e visualização de tópicos, facilitando a análise de conjuntos de áudios. A flexibilidade do Módulo 3, que permite a análise tanto a nível de frases quanto de documentos inteiros, amplia a aplicabilidade da abordagem a diferentes cenários de análise.

O uso de Modelos de Linguagem de Grande Escala (LLMs), como o modelo Gemini, foi fundamental para a implementação da análise de Q&A e a Geração Aumentada por Recuperação (RAG), permitindo a extração de informações específicas de maneira eficiente e precisa. Este módulo possibilitou que as perguntas fossem respondidas com base em fragmentos específicos das transcrições, melhorando a precisão e a relevância das respostas.

Por outro lado, a modelagem de tópicos, realizada com o BERTopic, forneceu uma visão geral sobre os principais temas discutidos nas transcrições, organizando e agrupando as informações em tópicos relevantes. Essa análise semântica mais abrangente permitiu uma exploração detalhada dos tópicos abordados nos áudios, como vigilância digital, privacidade e segurança dos usuários, entre outros. A visualização de dados, complementada por gráficos interativos, facilitou a interpretação dos resultados, tornando a análise mais acessível e compreensível.

5.1 LIMITAÇÕES DA PESQUISA

Entretanto, este estudo enfrentou algumas limitações no processamento das transcrições com o modelo Whisper. A variante “large-v3-turbo” apresentou

fenômenos de alucinação ao final de determinadas transcrições, evidenciando a necessidade de estratégias para lidar com saídas inesperadas do modelo.

Além disso, o número limitado de áudios no *dataset* fez com que os padrões e informações encontrados refletissem especificamente o contexto do curso analisado. Recomenda-se, portanto, a reaplicação dos métodos em uma base de dados maior para avaliar como o algoritmo se comporta com grandes volumes de dados.

A eficiência do método também pode ser aprimorada. Durante o desenvolvimento, o processo de carregamento de dados e geração de *embeddings* foi repetido várias vezes. Abordagens recentes buscam reduzir o consumo de recursos integrando melhor os componentes das soluções, criando soluções de ponta a ponta.

Por fim, devido ao caráter exploratório da análise, métricas de avaliação como acurácia e F1 score não foram utilizadas para mensurar o desempenho dos modelos. Recomenda-se, portanto, a aplicação dessas métricas em análises futuras para obter resultados mais precisos e relevantes.

5.2 TRABALHOS FUTUROS

Embora os resultados obtidos sejam promissores, futuras pesquisas podem explorar a aplicação dessa abordagem em contextos mais complexos e com áudios de maior diversidade. Além disso, novos estudos podem analisar os seguintes aspectos:

- Aprimorar as técnicas de modelagem de tópicos e integrar novas ferramentas de NLP;
- Desenvolver análises temporais, permitindo a investigação de padrões ao longo do tempo;
- Utilizar LLMs para extração de tópicos com maior precisão;
- Empregar representações que tornem a relação entre os tópicos mais clara, como grafos ou redes;
- Realizar a *clusterização* com base nos *embeddings* dos áudios e, para cada *cluster*, utilizar os textos originais correspondentes para interagir com o LLM, possibilitando a extração de tópicos, a criação de redes, a geração de *insights* e outras análises potenciais.

REFERÊNCIAS

- ABDURRAHMAN, L; MULYANA, T. Parallel Construction of Information Technology Value Model: Design-Science Research Methodology. In: 8th International Conference on Information and Communication Technology (ICICT), Yogyakarta, Indonesia, 2020.
- ALDARMAKI, Hanan; ULLAH, Asad; RAM, Sreepratha; ZAKI, Nazar. Unsupervised Automatic Speech Recognition: A review. *Speech Communication*, v. 139, p. 76-91, 2022. ISSN 0167-6393. DOI: 10.1016/j.specom.2022.02.005.
- ALI, S. M.; GUPTA, N.; NAYAK, G. K.; LENKA, R. K. Big data visualization: Tools and challenges. In: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 2016. p. 656-660. DOI: 10.1109/IC3I.2016.7918044.
- AL-SARORI, Mokhtar Hussein; PUND, Sachin S.; CHARITHA, Bondili Bhavya; KAUSHAL, Ashish; LAL, Bechoo; CHATURVEDI, Sudhir Kumar. An Overview of Natural Language Processing Techniques for Information Analysis. In: INTERNATIONAL CONFERENCE ON SMART GENERATION COMPUTING, COMMUNICATION AND NETWORKING, 2023, Bangalore. Anais [...]. Bangalore: IEEE, 2023. DOI: 10.1109/SMARTGENCON60755.2023.10442702.
- AMAZON. O que é RAG (Retrieval-Augmented Generation)? Disponível em: <https://aws.amazon.com/what-is/retrieval-augmented-generation/>. Acesso em: 07 dez. 2024.
- AUBER, David; BIKAKIS, Nikos; CHRYSANTHIS, Panos K.; PAPASTEFANATOS, George; SHARAF, Mohamed. Interactive big data visualization and analytics. *Big Data Research*, v. 36, 2024, 100445. ISSN 2214-5796. DOI: 10.1016/j.bdr.2024.100445.
- BALUSAMY, Balamurugan; ABIRAMI, R. Nandhini; KADRY, Seifedine; GANDOMI, Amir H. *Big Data: Concepts, Technology, and Architecture*. 1. ed. John Wiley & Sons, 2021. ISBN 9781119701828. DOI: 10.1002/9781119701859.
- BECKS, A.; TOEBERMANN, J.-C. Mining Textual Project Documentation in Process Engineering. *Computer Aided Chemical Engineering*, v. 10, p. 835-840, 2002. DOI: 10.1016/S1570-7946(02)80167-5.
- CAMPELLO, Ricardo J. G. B.; MOULAVI, Davoud; SANDER, Joerg. Density-Based Clustering Based on Hierarchical Density Estimates. In: PEI, Jian; TSENG, Vincent S.; CAO, Longbing; MOTODA, Hiroshi; XU, Guandong (eds). *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science*, vol. 7819. Springer, Berlin, Heidelberg, 2013. DOI: 10.1007/978-3-642-37456-2_14.
- CHEN, Min; HAUSER, Helwig; RHEINGANS, Penny; SCHEUERMANN, Gerik (Eds.). *Foundations of Data Visualization*. Cham: Springer Nature Switzerland AG, 2020. DOI: 10.1007/978-3-030-34444-3.

CHEN, Yi-Chen; CHI, Po-Han; YANG, Shu-wen; CHANG, Kai-Wei; LIN, Jheng-hao; HUANG, Sung-Feng; LIU, Da-Rong; LIU, Chi-Liang; LEE, Cheng-Kuang; LEE, Hung-yi. SpeechNet: A Universal Modularized Model for Speech Processing Tasks. arXiv, 2021. DOI: 10.48550/arXiv.2105.03070.

CHIUSANO, Fabio. Two minutes NLP — 33 important NLP tasks explained. Medium, 7 dez. 2021. Disponível em: <https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b>. Acesso em: 6 dez. 2024.

CHRISTENSEN, M. G. Introduction to Audio Processing. Cham: Springer, 2019. DOI: 10.1007/978-3-030-11781-8.

COOK, Dianne; LEE, Eun-Kyung; MAJUMDER, Mahbubul. Data Visualization and Statistical Graphics in Big Data Analysis. Annual Review of Statistics and Its Application, v. 3, p. 133-159, 2016. DOI: 10.1146/annurev-statistics-041715-033420.

DALAL, Kushal Rashmikant. Analysing the Role of Supervised and Unsupervised Machine Learning in IoT. In: INTERNATIONAL CONFERENCE ON ELECTRONICS AND SUSTAINABLE COMMUNICATION SYSTEMS (ICESC), 2., 2020, Coimbatore. Proceedings [...]. Coimbatore: IEEE, 2020. DOI: 10.1109/ICESC48915.2020.9155761.

DE MAURO, Andrea; GRECO, Marco; GRIMALDI, Michele. What is big data? A consensual definition and a review of key research topics. AIP Conf. Proc., v. 1644, n. 1, p. 97-104, 9 fev. 2015. DOI: 10.1063/1.4907823.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv, 2018. DOI: 10.48550/arXiv.1810.04805.

DEVORE, Ronald; HANIN, Boris; PETROVA, Guergana. Neural Network Approximation. arXiv, 2020. DOI: 10.48550/arXiv.2012.14501

DURGA, B. H. S.; SANJANA, K. S.; BAIG, Y.; TENDULKAR, N. V. R.; MOTHUKURI, R.; VIGNESH, T. Information Extraction From Text Messages Using Natural Language Processing. In: INTERNATIONAL CONFERENCE ON COMPUTER COMMUNICATION AND INFORMATICS, 2023, Coimbatore. Anais [...]. Coimbatore: IEEE, 2023. p. 1-6. DOI: 10.1109/ICCCI56745.2023.10128641.

ETHAYARAJH, Kawin. How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. arXiv, 2019. Disponível em: <https://arxiv.org/pdf/1909.00512>. Acesso em: 6 dez. 2024.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. AI Magazine, v. 17, n. 3, p. 37-54, 1996. DOI: 10.1609/aimag.v17i3.1230.

GABLER, Philipp; GEIGER, Bernhard C.; SCHUPPLER, Barbara; KERN, Roman. Reconsidering Read and Spontaneous Speech: Causal Perspectives on the Generation of Training Data for Automatic Speech Recognition. Information, v. 14, n. 2, p. 137, 2023. DOI: 10.3390/info14020137.

GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, v. 35, n. 2, p. 137-144, 2015. DOI: 10.1016/j.ijinfomgt.2014.10.007.

GERHARDT, T. E; SILVEIRA, D. T. Métodos de pesquisa. Porto Alegre: Editora da UFRGS, 2009.

GHAREHCHOPOGH, F. S.; KHALIFELU, Z. A. Analysis and evaluation of unstructured data: text mining versus natural language processing. In: INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGIES, 5., 2011, Baku. Anais [...]. Baku: IEEE, 2011. p. 1-4. DOI: 10.1109/ICAICT.2011.6111017.

GIL, Cristina. Clustering. 2008. Disponível em: https://rpubs.com/cristina_gil/clustering. Acesso em: 07 dez. 2024.

GONZALEZ-GOMEZ, Luis Jose; HERNANDEZ-MUNOZ, Sofia Margarita; BORJA, Abiel; AZOFEIFA, Jose Daniel; NOGUEZ, Julieta; CARATOZZOLO, Patricia. Analyzing Natural Language Processing Techniques to Extract Meaningful Information on Skills Acquisition From Textual Content. *IEEE Access*, v. 12, p. 86038-86056, 2024. DOI: 10.1109/ACCESS.2024.3465409.

GROOTENDORST, Maarten. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794, 2022.

GROOTENDORST, Maarten. The Algorithm - BERTopic. Disponível em: <https://maartengr.github.io/BERTopic/algorithm/algorithm.html>. Acesso em: 9 dez. 2024.

GULLO, Francesco. From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. *Physics Procedia*, v. 62, p. 18-22, 2015. DOI: 10.1016/j.phpro.2015.02.005.

IBM. What is a Neural Network? Disponível em: <https://www.ibm.com/topics/neural-networks>. Acesso em: 03 dez. 2024a.

IBM. What is Machine Learning (ML)? Disponível em: <https://www.ibm.com/topics/machine-learning>. Acesso em: 02 dez. 2024b.

JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. *Electronic Markets*, v. 31, p. 685-695, 2021. DOI: 10.1007/s12525-021-00475-2.

KAMATH, Uday; LIU, John; WHITAKER, James. Deep Learning for NLP and Speech Recognition. Cham: Springer, 2019. DOI: 10.1007/978-3-030-14596-5.

KANG, H. J.; BISSYANDÉ, T. F.; LO, D. Assessing the Generalizability of Code2vec Token Embeddings. In: IEEE/ACM INTERNATIONAL CONFERENCE ON AUTOMATED SOFTWARE ENGINEERING (ASE), 34., 2019, San Diego. Proceedings [...]. San Diego: IEEE, 2019. p. 1-12. DOI: 10.1109/ASE.2019.00011.

KARUNA, E. N.; SOKOLOV, P. V. Comparative Analysis of Text Information Clustering Methods. In: INTERNATIONAL CONFERENCE ON SOFT COMPUTING AND MEASUREMENTS (SCM), 2021, St. Petersburg, Russia. Anais [...]. IEEE, 2021. p. 109-112. DOI: 10.1109/SCM52931.2021.9507189.

KEIM, D.; QU, H.; MA, K.-L. Big-Data Visualization. IEEE Computer Graphics and Applications, v. 33, n. 4, p. 20-21, jul./ago. 2013. DOI: 10.1109/MCG.2013.54.

KINRA, Aseem; BEHESHTI-KASHI, Samaneh; BUCH, Rasmus; NIELSEN, Thomas Alexander Sick; PEREIRA, Francisco. Examining the potential of textual big data analytics for public policy decision-making: A case study with driverless cars in Denmark. Transport Policy, v. 98, p. 68-78, 2020. DOI: 10.1016/j.tranpol.2020.05.026.

L'HEUREUX, Alexandra; GROLINGER, Katarina; ELYAMANY, Hany F.; CAPRETZ, Miriam A. M. Machine Learning With Big Data: Challenges and Approaches. IEEE Access, v. 5, p. 7776-7797, 2017. DOI: 10.1109/ACCESS.2017.2696365.

LANEY, Doug. 3D Data Management: Controlling Data Volume, Velocity, and Variety. 2001. Disponível em: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--and-variety>. Acesso em: 25 nov. 2024.

LAUREN, P. Reconstructing Word Representations from Pre-trained Subword Embeddings. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND COMPUTATIONAL INTELLIGENCE (CSCI), 2022, Las Vegas. Proceedings [...]. Las Vegas: IEEE, 2022. p. 35-40. DOI: 10.1109/CSCI58124.2022.00013.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. Nature, v. 521, p. 436-444, 2015. DOI: 10.1038/nature14539.

LEE, In. Big data: Dimensions, evolution, impacts, and challenges. Business Horizons, v. 60, n. 3, p. 293-303, 2017. DOI: 10.1016/j.bushor.2017.01.004.

LI, J.; LIU, W.; LIU, M.; HUANG, M. Study on Chinese Text Clustering Algorithm Based on K-mean and Evaluation Method on Effect of Clustering for Software-intensive System. In: INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING AND APPLICATION (ICCEA), 2020, Guangzhou, China. Anais [...]. IEEE, 2020. p. 513-519. DOI: 10.1109/ICCEA50009.2020.00114.

LIN, Tianyang; WANG, Yuxin; LIU, Xiangyang; QIU, Xipeng. A Survey of Transformers. arXiv, 2021. Disponível em: <https://arxiv.org/abs/2106.04554>. Acesso em: 6 dez. 2024.

LINKE, Julian; GEIGER, Bernhard C.; KUBIN, Gernot; SCHUPPLER, Barbara. What's so complex about conversational speech? A comparison of HMM-based and transformer-based ASR architectures. Computer Speech & Language, v. 90, p. 101738, 2025. DOI: 10.1016/j.csl.2024.101738.

MAHMOUDIAN, M.; ZANJANI, S. M.; SHAHINZADEH, H.; KABALCI, Y.; KABALCI, E.; EBRAHIMI, F. An Overview of Big Data Concepts, Methods, and Analytics:

Challenges, Issues, and Opportunities. In: 2023 5th Global Power, Energy and Communication Conference (GPECOM), Nevsehir, Turquia, 2023. p. 554-559. DOI: 10.1109/GPECOM58364.2023.10175760.

MCINNES, Leland; HEALY, John; ASTELS, Steve. How HDBSCAN Works. Disponível em: https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html. Acesso em: 05 dez. 2024.

MD. Blending Weighted TF-IDF & BERT for Improving Semantic Search. In: INTERNATIONAL CONFERENCE ON ADVANCED RESEARCH IN COMPUTING (ICARC), 2., 2022, Belihuloya, Sri Lanka. Anais [...]. Belihuloya: IEEE, 2022. p. 154-159. DOI: 10.1109/ICARC54489.2022.9753875.

MEHRISH, Ambuj; MAJUMDER, Navonil; BHARADWAJ, Rishabh; MIHALCEA, Rada; PORIA, Soujanya. A review of deep learning techniques for speech processing. Information Fusion, [S.l.], v. 99, 2023. ISSN 1566-2535.

MISHRA, M.; MISHRA, V. K.; SHARMA, H. R. Performance measurement for the quality of question answering approaches in natural language. In: NATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND SIGNAL PROCESSING (CISP), 2., 2012, Guwahati. Proceedings [...]. Guwahati: IEEE, 2012. p. 95-98. DOI: 10.1109/NCCISP.2012.6189685.

MOHANDAS, Goku; MORITZ, Philipp. Building RAG-based LLM Applications for Production. 25 out. 2023. Disponível em: <https://www.anyscale.com/blog/a-comprehensive-guide-for-building-rag-based-llm-applications-part-1>. Acesso em: 07 dez. 2024.

NADKARNI, Prakash. Clinical Research Computing: A Practitioner's Handbook. 1. ed. Iowa City: Academic Press, 2016. DOI: <https://doi.org/10.1016/C2014-0-03836-1>.

NAZERI, Sina. The Power of Paying Attention: How ChatGPT Understands Conversations. Medium, 5 jan. 2024. Disponível em: <https://medium.com/@sina.nazeri/the-power-of-paying-attention-how-chatgpt-understands-conversations-eb774c3599be>. Acesso em: 7 dez. 2024.

OLIVEIRA, Roger Alves de; BOLLEN, Math H.J. Deep learning for power quality. Electric Power Systems Research, v. 214, Part A, 2023. DOI: 10.1016/j.epsr.2022.108887.

OMAR, M.; CHOI, S.; NYANG, D.; MOHAISEN, D. Robust Natural Language Processing: Recent Advances, Challenges, and Future Directions. IEEE Access, v. 10, p. 86038-86056, 2022. DOI: 10.1109/ACCESS.2022.3197769.

OPENAI. Whisper. Disponível em: <https://github.com/openai/whisper>. Acesso em: 8 dez. 2024.

PANDEY, A. Kumar; ROY, S. Sekhar. Extractive Question Answering Over Ancient Scriptures Texts Using Generative AI and Natural Language Processing Techniques. IEEE Access, v. 12, p. 101197-101209, 2024. DOI: 10.1109/ACCESS.2024.3431282.

PASSRICHA, Vishal; AGGARWAL, Rajesh Kumar. Chapter 2 - End-to-End Acoustic Modeling Using Convolutional Neural Networks. In: DEY, Nilanjan (Ed.). Intelligent Speech Signal Processing. Academic Press, 2019. p. 5-37. ISBN 9780128181300. DOI: 10.1016/B978-0-12-818130-0.00002-7.

PATIL, R.; PATIL, P. D.; JOSHI, Y.; KHANDELWAL, S.; NALAWADE, S.; PALVE, B. NLP Based Question Answering System. In: INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION, CONTROL AND AUTOMATION (ICCUBEA), 7., 2023, Pune. Proceedings [...]. Pune: IEEE, 2023. p. 1-5. DOI: 10.1109/ICCUBEA58933.2023.10392202.

PEFFERS, K; TUUNANEN, T; ROTHENBERGER, M. A; CHATTERJEE, S. A Design Science Research Methodology For Information Systems Research. Journal of management information systems : JMIS, v. 24, n. 3, p. 45–77, 2007.

QIU, Qinjun; TIAN, Miao; TAO, Liufeng; XIE, Zhong; MA, Kai. Semantic information extraction and search of mineral exploration data using text mining and deep learning methods. Ore Geology Reviews, v. 165, p. 105863, 2024. DOI: 10.1016/j.oregeorev.2023.105863.

RADFORD, Alec; KIM, Jong Wook; XU, Tao; BROCKMAN, Greg; McLEAVEY, Christine; SUTSKEVER, Ilya. Robust Speech Recognition via Large-Scale Weak Supervision. 2022. Disponível em: <https://arxiv.org/pdf/2212.04356>. Acesso em: 09 dez. 2024.

RAIAAN, M. A. K.; MUKTA, Md. Saddam Hossain; FATEMA, Kaniz; FAHAD, Nur Mohammad; SAKIB, Sadman; MIM, Most Marufatul Jannat. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. IEEE Access, v. 12, p. 26839-26874, 2024. DOI: 10.1109/ACCESS.2024.3365742.

RAJATH, S.; KUMAR, A.; AGARWAL, M.; SHEKAR, S.; PRASAD, V. B. Data Mining Tool To Help The Scientific Community Develop Answers To Covid-19 Queries. In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES (ICDS), 5., 2021, Fez, Morocco. Anais [...]. Fez: IEEE, 2021. p. 1-5. DOI: 10.1109/ICDS53782.2021.9626771.

REIMERS, Nils; GUREVYCH, Iryna. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 2019. DOI: 10.48550/arXiv.1908.10084.

ROSENZWEIG, Paul. The Surveillance State: Big Data, Freedom, and You. The Great Courses, 2016. Disponível em: <https://www.thegreatcourses.com/courses/the-surveillance-state-big-data-freedom-and-you>. Acesso em: 9 dez. 2024.

ROSTAM, Z. R. K.; SZÉNÁSI, S.; KERTÉSZ, G. Achieving Peak Performance for Large Language Models: A Systematic Review. IEEE Access, v. 12, p. 96017-96050, 2024. DOI: 10.1109/ACCESS.2024.3424945.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 4. ed. Global Edition. Pearson Education, 2021.

SAFHI, Hicham Moad; FRIKH, Bouchra; OUHBI, Brahim. Assessing reliability of Big Data Knowledge Discovery process. *Procedia Computer Science*, v. 148, p. 30-36, 2019. DOI: 10.1016/j.procs.2019.01.005.

SARAVANAN, R.; SUJATHA, P. A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND CONTROL SYSTEMS (ICICCS), 2., 2018, Madurai. Proceedings [...]. Madurai: IEEE, 2018. DOI: 10.1109/ICCONS.2018.8663155.

SATYAM; GEETHA, P. Comprehensive Overview of the Opportunities and Challenges in AI. In: INTERNATIONAL CONFERENCE ON SUSTAINABLE COMPUTING AND SMART SYSTEMS (ICSCSS), 2023, Coimbatore. Anais [...]. IEEE, 2023. DOI: 10.1109/ICSCSS57650.2023.10169722

SCHMITT, Églantine. *Big Data: An Art of Decision Making*. 1. ed. Londres: ISTE Ltd, 2020. ISBN 9781786305558. DOI: 10.1002/9781119777014.

SÖNMEZ, Yeşim Ülgen; VAROL, Asaf. In-depth investigation of speech emotion recognition studies from past to present –The importance of emotion recognition from speech signal for AI–. *Intelligent Systems with Applications*, [S.I.], v. 22, 200351, 2024. ISSN 2667-3053. DOI: 10.1016/j.iswa.2024.200351.

TUCKER, Ethan C.; CAPPS, Colton J.; SHAMIR, Lior. A data science approach to 138 years of congressional speeches. *Heliyon*, [S.I.], v. 6, n. 8, e04417, 2020. ISSN 2405-8440. DOI: 10.1016/j.heliyon.2020.e04417.

UHRIG, R. E. Introduction to artificial neural networks. In: PROCEEDINGS OF IECON '95 - 21st Annual Conference on IEEE Industrial Electronics, 1995, Orlando, FL, USA. Anais [...]. Orlando: IEEE, 1995. p. 33-37. DOI: 10.1109/IECON.1995.483329.

VAN OTTEN, Neri. *Vector Space Model Made Simple With Examples & Tutorial*. Spot Intelligence, 7 set. 2023. Disponível em: <https://spotintelligence.com/2023/09/07/vector-space-model/>. Acesso em: 6 dez. 2024.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUHKIN, Illia. Attention Is All You Need. *arXiv*, 2017. Disponível em: <https://arxiv.org/pdf/1706.03762>. Acesso em: 7 dez. 2024.

VIEIRA, J. A; LEITE, A. R; KUHN, A. S. Perspectivas da Produção de Pesquisa Aplicada, Inovação e Desenvolvimento Científico e Tecnológico nos Institutos Federais. *Revista Valore*. v. 8, 2023.

WARD, Matthew O.; GRINSTEIN, Georges; KEIM, Daniel. *Interactive Data Visualization: Foundations, Techniques, and Applications*. CRC Press, 2015.

WIKIPEDIA. Large language model. Disponível em: https://en.wikipedia.org/wiki/Large_language_model. Acesso em: 9 dez. 2024.

WU, R.; CHEN, S.; SU, X.; ZHU, Y.; LIAO, Y.; WU, J. A Multi-Source Retrieval Question Answering Framework Based on RAG. In: INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE, PARALLEL AND DISTRIBUTED SYSTEMS (ISPDS), 5., 2024, Guangzhou. Proceedings [...]. Guangzhou: IEEE, 2024. p. 644-647. DOI: 10.1109/ISPDS62779.2024.10667535.

XU, R.; WUNSCH, D. Clustering. IEEE Press Series on Computational Intelligence. Wiley, 2008. ISBN 978-0-470-38278-3.

ZHAO, Wayne Xin; ZHOU, Kun; LI, Junyi; TANG, Tianyi; WANG, Xiaolei; HOU, Yupeng; MIN, Yingqian; ZHANG, Beichen; ZHANG, Junjie; DONG, Zican; DU, Yifan; YANG, Chen; CHEN, Yushuo; CHEN, Zhipeng; JIANG, Jinhao; REN, Ruiyang; LI, Yifan; TANG, Xinyu; LIU, Zikang; LIU, Peiyu; NIE, Jian-Yun; WEN, Ji-Rong. A Survey of Large Language Models. 2024. Disponível em: <https://arxiv.org/abs/2303.18223v15>. Acesso em: 15 dez. 2024.

ZHOU, Zhi-Hua. Machine Learning. Springer, 2021. DOI: 10.1007/978-981-15-1967-3.

ZHU, Z.; QI, G.; SHANG, G.; HE, Q.; ZHANG, W.; LI, N. Enhancing Large Language Models with Knowledge Graphs for Robust Question Answering. In: IEEE INTERNATIONAL CONFERENCE ON PARALLEL AND DISTRIBUTED SYSTEMS (ICPADS), 30., 2024, Belgrade. Proceedings [...]. Belgrade: IEEE, 2024. p. 262-269. DOI: 10.1109/ICPADS63350.2024.00042.