



# Extração, Classificação e Estruturação de Discussões e Argumentos em Redes Sociais

**RICARDO JORGE ALVES COSTA BARROS**

Julho de 2025

**Extração, Classificação e Estruturação de  
Discussões e Argumentos em Redes Sociais:  
Uma Abordagem com Modelos de Linguagem de Grande  
Escala e Grafos de Conhecimento**

**Ricardo Jorge Alves Costa**

**Dissertação para obtenção do Grau de Mestre em  
Engenharia Informática, Área de Especialização em  
Sistemas de Informação e Conhecimento**

**Orientador: Luís Conceição**

**Co-orientador: João Carneiro**

Porto, junho 2025



# Declaração de Integridade

Declaro ter conduzido este trabalho académico com integridade.

Não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Portanto, o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P.PORTO.

ISEP, Porto, 29 de junho de 2025



# Resumo

Plataformas como o Reddit concentram diariamente milhares de discussões públicas sobre temas políticos e sociais. No entanto, a informalidade da linguagem utilizada, a dimensão das conversas e a redundância argumentativa tornam difícil a identificação dos principais pontos de vista defendidos pelos participantes. Esta dissertação aborda o desafio de simplificar estes debates, propondo uma solução que recorre a *Large Language Models* (LLMs) e Grafos de Conhecimento para estruturar, analisar, classificar e sintetizar os argumentos utilizados, de forma automática e acessível. A abordagem desenvolvida assenta numa *pipeline* modular que extrai e processa todos os dados relevantes de uma discussão, utilizando modelos generativos nas diversas fases e tarefas que a compõem. A solução procura identificar os argumentos mais relevantes, classificando-os individualmente quanto à sua posição (a favor, contra ou neutro) em relação ao tópico do debate, e agrupando-os semântica e visualmente através de um grafo de conhecimento. O protótipo permite ainda a geração de sumários expositivos, análises detalhadas e avaliações quantitativas e qualitativas da *performance* do sistema.

O projeto dá especial atenção a discussões de natureza política e social, tendo em conta a forma como os argumentos ideológicos são formulados, disseminados e contrapostos em espaços digitais neste contexto. Esta perspetiva, permite, simultaneamente, testar a robustez do sistema e explorar a dinâmica discursiva e polarização associada a tópicos politicamente sensíveis. O objetivo final também passa por tentar contribuir para a melhoria da qualidade do discurso público e político em ambiente digital, reduzindo a desinformação e promovendo decisões mais informadas.

Este trabalho foi desenvolvido com base na metodologia *Action Research* e inclui uma fase experimental de avaliação do sistema utilizando os próprios LLMs para validação automática. Foram ainda respeitadas as diretrizes e normas éticas definidas pelo RGPD e pelo AI Act, com especial atenção à anonimização dos dados e à mitigação de possíveis vieses dos modelos. Os resultados demonstram o potencial interpretativo e generativo dos LLMs combinados com Grafos de Conhecimento para promover uma compreensão clara, estruturada e crítica dos debates públicos e políticos ocorridos em redes sociais.

**Palavras-chave:** *Argument Mining*, Sintetização de Argumentos, Redes Sociais, *Large Language Models*, Grafos de Conhecimento



# Abstract

Platforms such as Reddit host thousands of public discussions every day, particularly on political and social issues. However, the informal nature of the language used, the scale of the conversations, and the redundancy of arguments make it difficult to identify the main points of view defended by participants. This dissertation addresses the challenge of simplifying these debates by proposing a solution that leverages Large Language Models and Knowledge Graphs to structure, analyze, classify, and synthesize the arguments used, in an automated and accessible way. The proposed approach is based on a modular pipeline that extracts and processes all relevant data from a discussion, using generative models throughout the different phases and tasks involved. The system identifies relevant arguments, classifies them individually according to their stance (in favor, against, or neutral) toward the topic under discussion, and semantically and visually groups them through a knowledge graph. It also enables the generation of expository summaries, detailed analyses, and qualitative evaluations of the system's performance.

The project pays particular attention to political and social discussions, taking into account how ideological arguments are formulated, disseminated, and countered in digital spaces within this context. This perspective allows, simultaneously, for testing the system's robustness and exploring the discursive dynamics and polarization associated with politically sensitive topics. The ultimate goal is also to contribute to improving the quality of public and political discourse in digital environments, reducing misinformation, and promoting more informed decision-making.

This work follows the Action Research methodology and includes an experimental evaluation phase using the LLMs themselves for automatic validation. The ethical standards of the GDPR and the AI Act were followed, with special attention to data anonymization and the mitigation of potential model biases. The results demonstrate the interpretative and generative potential of LLMs combined with Knowledge Graphs in promoting a clear, structured, and critical understanding of public political debates taking place on social media platforms.

**Keywords:** Argument Analysis, Argument Summarization, Social Media, Large Language Models, Knowledge Graphs



# Agradecimentos

Dezoito anos de ensino (e seis de ensino superior) culminam no desenvolvimento deste projeto e eu gostava de deixar alguns agradecimentos especiais às pessoas que fizeram com que este ciclo corresse da melhor forma possível. Primeiramente, agradecer de coração à minha avó e à minha mãe que foram os dois pilares neste percurso e as pessoas que me providenciaram todas as ferramentas e recursos necessários durante todos estes anos; à minha namorada que me acompanhou em todo este longo processo desde o início; à minha irmã, que começa agora o ciclo de ensino superior.

Gostava de deixar um agradecimento sentido aos dois orientadores que me acompanharam no processo de desenvolvimento deste projeto: ao João Carneiro, que me deu a oportunidade de abraçar este desafio e integrar a equipa da Devoteam, que me ensinou muito e acompanhou semanalmente; e ao Professor Luís Conceição que, de igual forma, acompanhou o processo e mostrou-se sempre disponível para ajudar no que fosse necessário.

Gostava ainda de agradecer à Devoteam, que me acolheu e providenciou todas as condições necessárias para a conclusão com sucesso deste projeto.

Finalmente, gostava de agradecer ao meu colega Hugo que me acompanhou e ajudou nestes dois últimos anos de ensino superior.



# Índice

1	Introdução.....	1
1.1	Contexto e Enquadramento.....	1
1.1.1	A Influência das Redes Sociais na Posição Ideológica dos Utilizadores.....	2
1.1.2	A Influência das Redes Sociais na Posição Política dos Utilizadores.....	3
1.2	Apresentação do Problema.....	5
1.3	Descrição dos Objetivos.....	6
1.4	Considerações Éticas.....	8
1.4.1	RGPD: Regulamento Geral sobre a Proteção de Dados.....	8
1.4.2	AI Act: Regulamento da Inteligência Artificial.....	9
1.5	Metodologia.....	11
1.6	Estrutura do Documento.....	12
1.7	Nota Terminológica.....	13
2	Estado da Arte.....	15
2.1	Argumentação e Negociação baseada em Argumentos.....	15
2.1.1	Argumentos e Argumentação.....	16
2.1.2	Negociação Baseada em Argumentos.....	18
2.1.3	Argumentação e Negociação Baseada em Argumentos em Contexto Digital.....	20
2.2	<i>Large Language Models</i> (Modelos de Linguagem de Larga Escala).....	21
2.2.1	<i>Generative AI</i> .....	21
2.2.2	<i>Large Language Models</i> : O conceito.....	22
2.2.3	<i>Natural Language Processing</i> - NLP.....	23
2.2.4	<i>Tokens</i> e <i>Tokenização</i> .....	24
2.2.5	Parâmetros.....	25
2.2.6	<i>Transformers</i> .....	26
2.2.7	Modelos <i>Open-Source</i> e <i>Closed Source</i> .....	28
2.2.8	<i>Prompts</i> e <i>Prompt Engineering</i> .....	29
2.2.9	Treino e <i>Fine-Tuning</i> de Modelos.....	30
2.2.10	Desafios da Utilização de LLMs.....	31
2.3	Grafos de Conhecimento.....	32
2.3.1	Grafos de Conhecimento Como Complementos aos <i>Large Language Models</i> .....	33
2.3.2	Utilização de Grafos de Conhecimento na Análise de Política Social.....	33
2.4	Análise de Argumentos em Redes Sociais.....	34
2.4.1	<i>Stance Detection</i> em Redes Sociais: Estado da Arte.....	35
3	Design e Conceção da Solução.....	41
3.1	Requisitos do Sistema.....	41
3.1.1	Requisitos Funcionais.....	42

3.1.2	Requisitos Não Funcionais .....	43
3.2	Tecnologias e Ferramentas Utilizadas .....	45
3.2.1	Reddit API .....	45
3.2.2	OpenAI GPT (via API) .....	47
3.2.3	LangChain .....	47
3.2.4	Neo4j .....	48
3.2.5	FastAPI .....	48
3.2.6	Streamlit .....	48
3.2.7	Bibliotecas de Suporte .....	48
3.3	Arquitetura Geral do Sistema .....	49
3.3.1	Descrição dos Componentes .....	50
3.4	Funcionamento da <i>Pipeline</i> de Processamento .....	52
3.4.1	Avaliação da <i>Pipeline</i> .....	55
4	Implementação da Solução .....	57
4.1	Abordagens e Tentativas Experimentais .....	57
4.1.1	Modelos <i>Open-Source</i> : Tentativas Iniciais com Modelos <i>Pre-trained</i> .....	58
4.1.2	Tentativas de Construção Automática do Grafo via LLM .....	60
4.1.3	Classificação de Comentários e Respostas ao Invés de Argumentos .....	61
4.1.4	Agrupamento de Argumentos – Primeiras Abordagens .....	61
4.1.5	Reflexão Final .....	62
4.2	Metodologia de Desenvolvimento .....	62
4.2.1	Considerações Éticas e Conformidade Durante o Desenvolvimento .....	63
4.3	Desenvolvimento dos Diferentes Módulos .....	64
4.3.1	Arquitetura Base com FastAPI .....	64
4.3.2	Extração e Recolha de Dados - <i>Reddit Scraper</i> .....	65
4.3.3	Identificação do Tópico de Discussão .....	67
4.3.4	Classificação do Posicionamento dos Comentários e Respostas – <i>Stance Classification</i> .....	69
4.3.5	Sumarização das Posições – <i>Stance Summarization</i> .....	70
4.3.6	Extração, Classificação e Agrupamento de Argumentos .....	72
4.3.7	Criação dos Grafos de Conhecimento .....	76
4.3.8	Exploração dos Grafos de Conhecimento .....	77
4.3.9	<i>Framework</i> de Avaliação .....	79
5	Resultados e Discussão .....	85
5.1	Demonstração Funcional .....	86
5.1.1	<i>Header</i> da Discussão – Tópico, Descrição, <i>Stances</i> e Distribuições .....	87
5.1.2	Resumo dos Argumentos por <i>Stance</i> .....	88
5.1.3	Criação do Grafo de Conhecimento .....	89
5.1.4	Grafo de Conhecimento Criado .....	90

5.1.5	Exploração Inicial do Grafo de Conhecimento .....	92
5.1.6	Exploração de Cada <i>Post</i> e Respetivos Argumentos Extraídos.....	92
5.1.7	Exploração do Grafo através de <i>Queries</i> Pré-Definidas .....	93
5.2	Avaliação Crítica da Solução.....	95
5.2.1	Identificação do Tópico de Discussão – Avaliação .....	96
5.2.2	Extração de Argumentos – Avaliação .....	98
5.2.3	Classificação de <i>Stances</i> dos Argumentos – Avaliação.....	103
5.2.4	<i>Clustering</i> dos Argumentos – Avaliação.....	107
5.2.5	Considerações Finais da Avaliação Global da Solução .....	110
6	Conclusões .....	113
6.1	Objetivos Cumpridos.....	113
6.2	Resultados Alcançados.....	117
6.3	Limitações e Trabalho Futuro .....	118
6.3.1	Limitações Identificadas .....	118
6.3.2	Trabalho Futuro .....	119
6.4	Impacto Científico .....	120
	Referências .....	121
	Apêndices.....	127
	Apêndice A – Excertos de Código .....	127
	Apêndice A1 – Definição dos <i>Endpoints</i> do Sistema.....	127
	Apêndice A2 – Função Principal do <i>Reddit Scraper</i> .....	127
	Apêndice A3 – <i>Helper Function</i> “ <i>process_text</i> ” .....	128
	Apêndice A4 - Objeto “ <i>thread_data</i> ” .....	128
	Apêndice B – Demonstração da Solução .....	129
	Apêndice C – Avaliação: <i>Clusters</i> Criados .....	133
	Anexos.....	135
	Anexo A – Constituição de um Argumento segundo Alpert (2016).....	135



# Lista de Figuras

Figura 1 - Comparação entre Democratas e Republicanos Sobre a Importância das Redes Sociais na Atividade Política (Clancy, 2022). .....	5
Figura 2 - Estrutura de Um Argumento (Toulmin, 2003).....	17
Figura 3 - Exemplo Prático da Estrutura de Um Argumento (Toulmin, 2003). .....	17
Figura 4 - Ordem Cronológica do Lançamento de Diferentes LLMs Relevantes. Naveed et al., 2024 .....	22
Figura 5 - Processo de <i>Tokenização</i> (Databricks, 2024).....	25
Figura 6 - Exemplo da Estrutura de uma <i>Thread</i> do Reddit.....	46
Figura 7 – Diagrama de Componentes do Sistema.....	50
Figura 8 - <i>Pipeline</i> de Processamento .....	55
Figura 9 - Exemplo de URL e Thread do Reddit .....	65
Figura 10 - Exemplo de URL e Thread do Reddit em Formato JSON.....	66
Figura 11 - <i>Prompt</i> Utilizada para a Identificação do Tópico da Discussão.....	68
Figura 12 - Definição do Modelo para Identificação do Tópico de Discussão .....	68
Figura 13 - Definição do Modelo para Stance Classification.....	69
Figura 14 - <i>Prompt</i> Utilizada para <i>Stance Classification</i> .....	70
Figura 15 - Definição do Modelo para Sumarização .....	71
Figura 16 - <i>Prompt</i> Utilizada para Sumarização Argumentativa .....	71
Figura 17 - Definição do Modelo para Extração, Classificação e Agrupamento de Argumentos .....	72
Figura 18 - <i>Prompt</i> Utilizada para Extração de Argumentos.....	73
Figura 19 - <i>Prompt</i> Utilizada para Classificação de Argumentos .....	74
Figura 20 - <i>Prompt</i> Utilizada para o <i>Clustering</i> de Argumentos .....	75
Figura 21 - Exemplos das funções Cypher criadas .....	77
Figura 22 - Estrutura das <i>Queries</i> Pré-definidas para Exploração do Grafo de Conhecimento..	78
Figura 23 - <i>Prompt</i> Utilizada para Avaliação da Extração de Argumentos.....	80
Figura 24 - <i>Prompt</i> Utilizada para Avaliação das <i>Stances</i> Atribuídas .....	81
Figura 25 - <i>Prompt</i> Utilizada para Avaliação dos <i>Clusters</i> Criados.....	82
Figura 26 - <i>Prompt</i> de Avaliação Geral da Extração de Argumentos .....	83
Figura 27 - <i>Prompt</i> de Avaliação Geral do <i>Clustering</i> .....	83
Figura 28 - <i>Prompt</i> de Avaliação Geral da Classificação de <i>Stances</i> .....	83
Figura 29 – Demonstração: <i>Thread</i> a Analisar .....	86
Figura 30 – Demonstração: <i>Header</i> da Discussão .....	87
Figura 31 - Demonstração: Resumos dos Argumentos por <i>Stance</i> .....	88
Figura 32 – Demonstração: Comentários Originais e Respetivas <i>Replies</i> Classificadas .....	89
Figura 33 - Demonstração: Criação do Grafo de Conhecimento .....	90

Figura 34 - Demonstração: Grafo de Conhecimento Criado.....	90
Figura 35 - Demonstração: <i>Overview</i> do Grafo .....	91
Figura 36 - Demonstração: Contagem de Argumentos Identificados por Posição .....	92
Figura 37 - Demonstração: Análise dos Argumentos Extraídos de Um <i>Post</i> (1) .....	92
Figura 38 - Demonstração: Análise dos Argumentos Extraídos de Um <i>Post</i> (2) .....	93
Figura 39 - Demonstração: Consultas disponíveis .....	94
Figura 40 - Demonstração: Execução da Consulta " <i>List of Arguments by Stance</i> " .....	94
Figura 41 - Demonstração: Execução da Consulta " <i>Argument Groups by Popularity</i> ". .....	95
Figura 42 - Avaliação: <i>Thread</i> Utilizada .....	97
Figura 43 - Avaliação: Tópico Identificado e Explicações das Respetivas <i>Stances</i> .....	98
Figura 44 - Avaliação: Grafo de Conhecimento Criado .....	99
Figura 45 - Avaliação: Argumentos Extraídos por Classificação.....	99
Figura 46 - Avaliação: Estatísticas Gerais da Discussão .....	100
Figura 47 - Avaliação: Classificação Geral da Extração de Argumentos.....	101
Figura 48 - Avaliação: Classificação Individual de Extração (1) .....	102
Figura 49 - Avaliação: Classificação Individual de Extração (2) .....	103
Figura 50 - Avaliação: Média Geral da Classificação de <i>Stances</i> .....	104
Figura 51 - Avaliação: Classificação Individual da Atribuição de <i>Stances</i> (1) .....	105
Figura 52 - Avaliação: Classificação Individual da Atribuição de <i>Stances</i> (2) .....	106
Figura 53 - Avaliação: Classificação Geral do <i>Clustering</i> .....	107
Figura 54 - Avaliação: Classificação Individual do <i>Clustering</i> (1) .....	108
Figura 55 - Avaliação: Classificação Individual do <i>Clustering</i> (2) .....	109
Figura 56 - Avaliação: Classificação Global dos Módulos .....	111
Figura 57 - Estrutura do ficheiro "main.py" .....	127
Figura 58 - Função Principal do <i>Reddit Scraper</i> .....	127
Figura 59 - <i>Helper Function</i> "process_text" .....	128
Figura 60 - Objeto "thread_data" .....	128
Figura 61 - Comentários Originais e Respetivas <i>Replies</i> Classificadas .....	129
Figura 62 - Resumo dos Argumentos Neutros.....	129
Figura 63 - Demonstração: Execução da Consulta " <i>Argument Groups by Popularity – Full Detail</i> " .....	130
Figura 64 - Demonstração: Execução da Consulta " <i>Replies to Supporting Comments</i> " .....	130
Figura 65 - Demonstração: Execução da Consulta " <i>Popular Comments by Score</i> " .....	130
Figura 66 - Demonstração: Execução da Consulta " <i>Discussion url</i> " .....	131
Figura 67 - Detalhes do Grafo de Conhecimento .....	131
Figura 68 - Detalhes de Nós de Tópico, Comentário e <i>Reply</i> .....	131
Figura 69 - Detalhes de Nós de Argumento e <i>Cluster</i> de Argumentos .....	132
Figura 70 - Base de Dados de Grafos com Várias Discussões Analisadas .....	132

# Lista de Tabelas

Tabela 1 - Comparação entre Democratas e Republicanos Sobre a Importância das Redes Sociais na Atividade Política - 2014 (Smith, 2014).....	4
Tabela 2 - Requisitos Funcionais do Sistema .....	42
Tabela 3 - Requisitos Não Funcionais do Sistema.....	44



# Acrónimos e Símbolos

## Lista de Acrónimos

<b>AI/IA</b>	Inteligência Artificial ( <i>Artificial Intelligence</i> )
<b>API</b>	<i>Application Programming Interface</i>
<b>BART</b>	<i>Bidirectional and Autoregressive Transformer</i>
<b>BERT</b>	<i>Bidirectional Encoder Representations from Transformers</i>
<b>CNN</b>	<i>Convolutional Neural Network</i>
<b>GPT</b>	<i>Generative Pre-Training Transformer</i>
<b>RGPD</b>	Regulamento Geral sobre a Proteção de Dados
<b>LLMs</b>	<i>Large Language Models</i>
<b>LSGDM</b>	<i>Large Scale Group Decision Making</i>
<b>LSTM</b>	<i>Long Short-Term Memory</i>
<b>NLP</b>	<i>Natural Language Processing</i>
<b>RAG</b>	<i>Retrieval-Augmented Generation</i>
<b>RNN</b>	<i>Recurrent Neural Network</i>
<b>SVM</b>	<i>Support Vector Machine</i>



# 1 Introdução

O presente capítulo estabelece as bases fundamentais da dissertação, contextualizando e enquadrando o tópico em análise e apresentando o problema que motivou o desenvolvimento da solução proposta. São ainda definidos os vários objetivos do projeto, enquadradas as principais preocupações éticas e metodológicas, e apresentada a estrutura global do documento. Adicionalmente, é apresentada uma nota terminológica que justifica o uso de determinados termos técnicos ao longo do documento.

## 1.1 Contexto e Enquadramento

De acordo com dados do Statista, uma plataforma global de dados e inteligência empresarial, a fevereiro de 2025 existiam 5.56 mil milhões de utilizadores de internet em todo o mundo, representando 67.9% da população global. Destes, cerca de 5.24 mil milhões, ou 63.9% da população mundial, eram utilizadores de redes sociais (Statista, 2025).

O advento das plataformas de comunicação social veio possibilitar níveis de conectividade e partilha informacional sem precedentes, tendo, simultaneamente, originado um novo panorama de discurso público caracterizado pela rápida permutação de ideias, pela amplificação de diferentes perspetivas e pela emergência inevitável de divergências e discussões (Rauniar et al., 2013). Estas ocorrem diariamente, em larga escala, nas mais variadas

plataformas, tanto nas mais conhecidas como o Twitter/X<sup>1</sup>, Facebook<sup>2</sup> ou o Reddit<sup>3</sup>, como noutras especialmente focadas nessa vertente, como o Politiq.net<sup>4</sup>, debatepolitics.com<sup>5</sup> ou a Kialo<sup>6</sup>, abrangendo tópicos que vão desde questões políticas e sociais até debates sobre inovações tecnológicas e desporto.

No entanto, à medida que o volume de interações aumenta, surgem também desafios significativos relacionados com a sobrecarga de informação e a dificuldade de identificar os argumentos mais relevantes dentro destas discussões.

### **1.1.1 A Influência das Redes Sociais na Posição Ideológica dos Utilizadores**

A crescente ubiquidade das redes sociais tem transformado significativamente a forma como os indivíduos interagem com a informação e entre si. As plataformas digitais tornaram-se espaços predominantes de troca de ideias, onde os utilizadores não apenas consomem conteúdo, mas também constroem e partilham as suas próprias narrativas. Este fenómeno tem impacto direto na formação e reformulação das posições ideológicas dos utilizadores, com implicações significativas a nível político e social.

Kuhn & Crowell (2011), desenvolveram um estudo de nome *“Dialogic Argumentation as a Vehicle for Developing Young Adolescents’ Thinking”*, em que chegaram à conclusão que a argumentação dialógica em contextos digitais pode desempenhar um papel central no desenvolvimento do pensamento crítico, particularmente entre adolescentes. Os autores conduziram uma investigação que utilizou diálogos digitais sobre questões sociais como meio de fomentar habilidades de raciocínio argumentativo, e referem que o uso de plataformas digitais oferece benefícios únicos, incluindo criação de um registo escrito das interações, que facilita a reflexão e o desenvolvimento de argumentos mais elaborados. Os resultados demonstram que a argumentação dialógica em ambientes digitais promove uma transição do raciocínio social para o individual, permitindo que os participantes utilizem as habilidades adquiridas em contextos sociais, em tarefas individuais. Além disso, a abordagem dialógica aumenta a consciência sobre a relevância de evidências no processo argumentativo,

---

<sup>1</sup> [www.x.com](http://www.x.com)

<sup>2</sup> [www.facebook.com](http://www.facebook.com)

<sup>3</sup> [www.reddit.com](http://www.reddit.com)

<sup>4</sup> [www.politiq.net](http://www.politiq.net)

<sup>5</sup> [www.debatepolitics.net](http://www.debatepolitics.net)

<sup>6</sup> [www.kialo.com](http://www.kialo.com)

contribuindo para a formação de opiniões mais fundamentadas. Os autores sublinham ainda que o debate em redes digitais contribui não só para o desenvolvimento cognitivo, mas também para uma maior sensibilidade à complexidade dos problemas sociais contemporâneos (Kuhn & Crowell, 2011).

### **1.1.2 A Influência das Redes Sociais na Posição Política dos Utilizadores**

As redes sociais tornaram-se, igualmente, em espaços estratégicos de comunicação política. Desde as eleições presidenciais dos EUA de 2008, em que Barack Obama utilizou o Twitter como ferramenta de campanha e como um meio para criar e manter proximidade com os eleitores, é visível a tentativa de instrumentalização destas plataformas por figuras políticas para moldar a opinião pública. A diplomacia digital transformou-se, tornando-se mais imediata, pública e personalizada (Antila, 2024). Com a amplificação destas estratégias por parte de líderes políticos de todo o mundo, são facilmente identificáveis os que mais se aproveitam e beneficiam das mesmas aos dias de hoje. O sucessor de Barack Obama, Donald Trump, amplificou esta tendência ao adotar o X/Twitter como principal canal de comunicação política. A sua estratégia, que utilizou durante as duas campanhas eleitorais em que participou como candidato, marcada por imprevisibilidade e decisões não convencionais, exemplifica uma nova forma de discurso político que privilegia o espetáculo e a comunicação direta (Antila, 2024). É neste mesmo contexto que se têm popularizado as discussões e argumentações políticas nas redes sociais, onde os utilizadores, expostos a uma multiplicidade de perspetivas, interagem com argumentos que tentam moldar ou desafiar as suas convicções ideológicas.

Num estudo feito pelo *Pew Research Center* em 2014, em que inquiriram americanos sobre o impacto das redes sociais na política do país, 35% dos Democratas e 25% dos Republicanos referem que estas são importantes para o recrutamento de pessoas. Uma parte significativa dos inquiridos — 34% dos Democratas e 23% dos Republicanos — destaca a importância das redes sociais para encontrar outras pessoas com visões políticas semelhantes e, 32% dos Democratas e 24% dos Republicanos referem a sua importância para debates políticos. Importa destacar que os resultados deste estudo, obtidos em 2014, apresentam, à data da elaboração desta dissertação, um intervalo temporal de 11 anos, o que pode implicar alterações, ligeiras ou significativas, nos dados apresentados. No entanto, é possível observar a relevância que as redes sociais já detinham na vida política dos Estados Unidos nesse período (Tabela 1).

Tabela 1 - Comparação entre Democratas e Republicanos Sobre a Importância das Redes Sociais na Atividade Política - 2014 (Smith, 2014).

	<b>Democratas</b>	<b>Republicanos</b>
Acompanhar atividade política	48%	34%
Recrutar pessoas para se envolverem	35%	25%
Encontrar outros com os mesmos pontos de vista	34%	23%
Debater questões políticas	32%	24%

Noutro estudo mais recente (Clancy, 2022), também levado a cabo pelo *Pew Research Center*, e com o mesmo foco, os republicanos continuam a ter avaliações mais negativas sobre o impacto das redes sociais na sociedade americana em comparação com os democratas, especialmente em relação à eficácia das redes sociais na conscientização pública, criação de movimentos sociais e na criação de atenção dos políticos a questões importantes. Além disso, consideram que as redes sociais não tornam as pessoas mais informadas ou mais tolerantes, e que contribuem para a divisão política. Ambos os grupos concordam que as redes sociais facilitam a manipulação com informações falsas, mas os democratas veem isso como uma maior ameaça. Dentro de cada partido, existem diferenças ideológicas significativas nas opiniões dos americanos sobre o impacto das redes sociais (Clancy, 2022). Os republicanos conservadores e os democratas liberais são mais propensos do que os moderados, nos seus partidos, a considerar as redes sociais divisivas politicamente e a afirmar que estas têm tornado as discussões políticas menos civis. Apesar de preocupados com o impacto das redes sociais, os republicanos conservadores e os democratas liberais são os mais propensos a divulgar mensagens políticas ou sociais. Embora apenas 23% dos americanos que usam redes sociais façam publicações sobre questões políticas ou sociais, a maioria desses utilizadores (55%) identifica-se como republicano conservador ou democrata liberal (Clancy, 2022). Na Figura 1 é possível observar os resultados descritos.

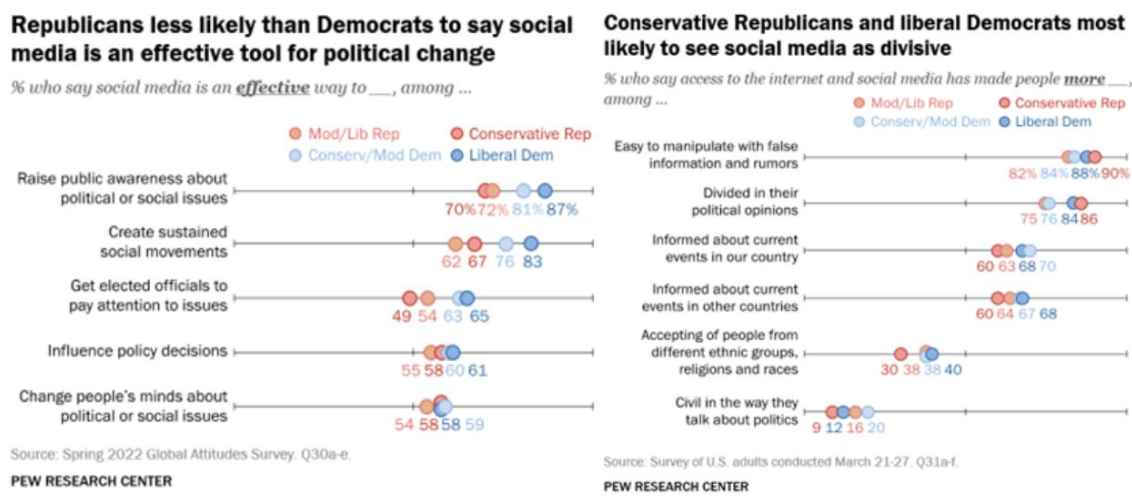


Figura 1 - Comparação entre Democratas e Republicanos Sobre a Importância das Redes Sociais na Atividade Política (Clancy, 2022).

O estudo realizado por Diehl et al. (2015) destaca que mesmo interações sociais não políticas podem levar à persuasão política. A manutenção de redes diversas em plataformas digitais promove a exposição a opiniões diferentes, o que estimula a reflexão sobre as próprias crenças e a reconsideração de posições políticas.

A exposição a redes heterogêneas promove a reflexão e a reconsideração de posições, favorecendo atitudes mais flexíveis e tolerantes. O desejo de preservar relações interpessoais e reputações digitais também influencia os utilizadores a ajustar ou moderar as suas posições, reforçando o papel social e psicológico das redes na formação política individual (Diehl et al., 2015).

## 1.2 Apresentação do Problema

O principal problema enfrentado nas discussões realizadas em redes sociais é a dispersão das múltiplas ideias e a falta de estrutura nos argumentos apresentados, o que é natural dado o caráter informal e desregulado destas plataformas. Apesar de facilitarem a partilha de opiniões, estas discussões decorrem frequentemente de forma desorganizada, com sobreposição de ideias, ausência de estrutura argumentativa clara e presença de comentários redundantes, pouco informativos ou até difamatórios. Tal cenário pode contribuir para a fragmentação dos debates, o aprofundamento de bolhas ideológicas e a disseminação de desinformação.

Como demonstrado por Diehl et al. (2015), as redes sociais influenciam significativamente as visões políticas dos utilizadores, tanto diretamente, através do consumo de notícias, quanto indiretamente, por meio da interação social e da exposição a opiniões divergentes. Esta influência torna ainda mais relevante a necessidade de promover estruturas discursivas claras e acessíveis. A elevada participação de utilizadores, por sua vez, cria um ambiente ruidoso onde se torna difícil distinguir argumentos bem fundamentados de opiniões superficiais ou manipulatórias. Este desafio afeta não só os utilizadores comuns, mas também investigadores, moderadores e analistas que pretendam compreender os pontos centrais de uma discussão pública.

No entanto, apesar da importância destes debates, ainda são escassas as ferramentas automatizadas capazes de estruturar e sintetizar discussões online de forma coerente, especialmente em ambientes altamente dinâmicos e polarizados como os de teor político. Neste contexto, torna-se então urgente o desenvolvimento de soluções que facilitem a análise, compreensão e simplificação destas discussões complexas.

### **1.3 Descrição dos Objetivos**

Tendo em consideração o problema apresentado, o objetivo principal deste projeto consiste no desenvolvimento de uma solução modular capaz de analisar e simplificar discussões extraídas de redes sociais, com foco na classificação e organização de argumentos expressos pelos utilizadores. Pretende-se, em particular, responder ao desafio da desinformação, polarização e fragmentação de debates em temas políticos e sociais, onde a clareza argumentativa é essencial para a tomada de decisão informada. Para tal, tem-se como alvo a criação de uma *pipeline* baseada em *Large Language Models* (LLMs), complementada com uma base de dados de grafos (*Knowledge Graphs*), que permita representar e explorar as discussões e elementos argumentativos de forma estruturada e detalhada.

De forma concreta, procura-se criar uma ferramenta capaz de identificar automaticamente os tópicos centrais dos debates, extrair argumentos relevantes a partir dos comentários e respostas, classificá-los de acordo com os seus posicionamentos argumentativos (*stance*) como “a favor”, “contra” ou “neutro” no contexto do tópico identificado, e organizá-los de forma

estruturada. Este processo inclui ainda o agrupamento de argumentos semelhantes, com o objetivo de evidenciar padrões discursivos, consensos e polarizações.

Outro objetivo essencial, mas não menos importante, consiste na síntese e exposição dos argumentos mais relevantes, apresentando os pontos principais feitos pelos utilizadores de forma clara, objetiva e imparcial. A componente de síntese tem como objetivo proporcionar uma compreensão mais eficiente e descomplicada dos diferentes pontos de vista presentes e reduzir a sobrecarga cognitiva associada ao excesso de informação.

Para facilitar a compreensão, os objetivos do projeto são listados abaixo:

1. **Realizar uma revisão abrangente da literatura e estado da arte** sobre argumentos e negociação baseada em argumentação, *Large Language Models*, Grafos de Conhecimento e análise de argumentos em redes sociais.
2. **Conceber e implementar uma *pipeline* de processamento baseada em LLMs**, capaz de sumarizar uma discussão realizada numa rede social, identificando o tópico central de uma discussão e extraindo, classificando e agrupando os argumentos centrais utilizados.
3. **Integrar um sistema de dupla classificação de posicionamento**, atribuindo uma primeira classificação aos comentários e uma segunda, mais granular, aos argumentos extraídos individualmente, de forma a atingir a maior precisão de classificação possível e utilidade analítica da representação discursiva.
4. **Agrupar argumentos semelhantes através de *clustering* semântico com LLMs**, garantindo a coerência temática e a preservação da orientação argumentativa (a favor, contra, neutro).
5. **Construir e armazenar um grafo de conhecimento estruturado**, onde se representem explicitamente as entidades da discussão (tópico, comentários, respostas, argumentos e grupos de argumentos), bem como as relações entre estas.
6. **Permitir a análise detalhada da discussão analisada**, através de consultas ao grafo de conhecimento criado.
7. **Avaliar a qualidade da extração, classificação e agrupamento de argumentos**, utilizando *prompts* específicas para análise automática da performance de cada componente do sistema com base em modelos generativos e critérios qualitativos.

8. **Analisar as limitações, falhas e enviesamentos do sistema**, identificando pontos de melhoria futura e discutindo o equilíbrio entre desempenho, controlo e interpretabilidade na utilização de modelos generativos em contextos argumentativos.

## 1.4 Considerações Éticas

O desenvolvimento de uma solução de Inteligência Artificial (IA) que cumpra os objetivos propostos levanta várias questões éticas que devem ser cuidadosamente consideradas. Tais preocupações abrangem tanto o cumprimento de normas legais como a garantia de imparcialidade e justiça no tratamento dos dados utilizados. Duas normas fundamentais que regem a utilização de dados pessoais e o uso de Inteligência Artificial são o Regulamento Geral sobre a Proteção de Dados (RGPD) e o Regulamento da Inteligência Artificial (*AI Act*).

### 1.4.1 RGPD: Regulamento Geral sobre a Proteção de Dados

O RGPD (Regulamento (UE) 2016/679) é uma legislação da União Europeia que estabelece regras rigorosas para a proteção de dados pessoais. A sua implementação visa garantir que as empresas e organizações que processam dados de cidadãos da UE o façam de forma transparente, responsável e com respeito pela privacidade dos indivíduos.

#### Implicações para o desenvolvimento de IA

A solução de IA proposta pode envolver o processamento de grandes quantidades de dados pessoais, o que implica a necessidade de garantir conformidade com as disposições do RGPD, já que este estabelece vários princípios fundamentais que devem ser seguidos durante o desenvolvimento da solução:

##### 1. Artigo 5.º – Princípios relativos ao tratamento de dados pessoais

**Alínea c) – Minimização dos dados:** “Os dados pessoais devem ser adequados, pertinentes e limitados ao que é necessário relativamente às finalidades para as quais são tratados.”

O princípio da minimização de dados estabelece que apenas os dados necessários para atingir os objetivos do projeto devem ser recolhidos. No contexto deste projeto, é crucial garantir que os modelos apenas processam dados estritamente necessários para atingir o objetivo final.

## **2. Artigo 6.º – Licitude do tratamento**

**Alínea 1.a) – Consentimento:** “O tratamento só é lícito se [...] o titular dos dados tiver dado o seu consentimento para o tratamento dos seus dados pessoais para uma ou mais finalidades específicas.”

## **3. Artigo 25.º – Proteção de dados desde a conceção e por defeito**

“O responsável pelo tratamento aplica [...] medidas técnicas e organizativas adequadas [...] destinadas a aplicar de forma eficaz os princípios de proteção de dados [...] como a minimização dos dados.”

O consentimento de indivíduos cujos dados estão a ser processados é uma exigência central do RGPD. Para garantir a conformidade com este ponto, todos os dados extraídos e processados serão anonimizados.

### **1.4.2 AI Act: Regulamento da Inteligência Artificial**

O *AI Act* é um regulamento proposto pela União Europeia para regulamentar o uso da IA em diversos contextos, incluindo a análise de dados pessoais. O *AI Act* visa garantir que a IA seja desenvolvida e utilizada de forma ética e segura, com foco na transparência e na proteção dos direitos fundamentais dos cidadãos.

#### **Obrigações Legais do AI Act (UE 2024/1689)**

##### **1. Artigo 15.º – Transparência**

“Os sistemas de IA de alto risco devem ser concebidos e desenvolvidos de forma que as suas operações sejam suficientemente transparentes [...] para que os utilizadores possam interpretar corretamente a produção do sistema.”

##### **2. Artigo 16.º – Obrigações dos fornecedores de sistemas de IA**

“Assegurar que os seus sistemas de IA foram concebidos e desenvolvidos em conformidade com os requisitos de: segurança, transparência, vigilância humana e ausência de viés.”

O *AI Act* reforça a necessidade de responsabilidade pelos impactos da IA. Neste projeto, embora os dados sejam anonimizados, ainda existe uma responsabilidade sobre os resultados e

potenciais efeitos da análise, como a propagação de desinformação ou a amplificação de argumentos polarizados. Isto implica a necessidade de manter uma documentação rigorosa das decisões e ações tomadas pelos modelos, incluindo a metodologia de treino e as decisões tomadas no processo de análise. Caso o sistema produza efeitos inesperados ou prejudiciais, a responsabilidade deve ser claramente atribuída ao desenvolvedor do sistema, garantindo mecanismos de revisão e correção.

### **Vieses nos Modelos**

Ao utilizar *Large Language Models* para tentar atingir o objetivo proposto, é fundamental garantir que estes modelos respeitem os princípios do *AI Act*. Um dos principais desafios é garantir que os modelos não perpetuem ou amplifiquem vieses, como o viés contra minorias ou a preferência por certos tipos de argumentos.

Os modelos podem, inadvertidamente, refletir ou amplificar vieses presentes nos dados com os quais são treinados. No caso da análise de argumentos em redes sociais, isto pode resultar em discriminação contra certos grupos sociais, políticos ou étnicos, ou favorecer argumentos de maiorias, tornando-os mais visíveis ou predominantes. A presença de viés pode comprometer a imparcialidade e a justiça dos resultados, resultando em discriminação e reforço de divisões sociais.

### **Tentativas de Mitigação de Vieses**

#### **1. Artigo 10.º – Qualidade dos conjuntos de dados**

“Os sistemas de IA de alto risco devem ser desenvolvidos com base em conjuntos de dados que sejam relevantes, representativos, isentos de erros e completos.”

Um passo essencial é garantir que os possíveis dados utilizados para treinar os modelos sejam diversos e representem diferentes perspetivas e grupos sociais, já que pode ajudar a evitar que favoreçam ou discriminem certos grupos.

#### **2. Artigo 29.º – Supervisão humana**

“Os sistemas de IA de alto risco devem ser concebidos para permitir a supervisão humana ao longo do seu ciclo de vida.”

A implementação de práticas de monitorização contínua é necessária porque permite avaliar a equidade e imparcialidade das decisões tomadas pela IA.

## 1.5 Metodologia

Para atingir os objetivos delineados, recorreu-se a uma metodologia conhecida como *Action-Research*, abordada por Somekh (2006). Esta abordagem caracteriza-se por um conjunto de etapas estruturadas, baseadas num desenvolvimento contínuo e orientadas para a resolução do problema em questão. Diferencia-se das metodologias mais convencionais pelo foco no estudo científico. No contexto desta metodologia, realizou-se uma análise sistemática do problema a resolver e tentou-se fundamentar as futuras intervenções em princípios teóricos.

O processo iniciou-se com a identificação do problema, permitindo a formulação de uma possível hipótese de resolução. Subsequentemente, a informação foi continuamente recolhida, organizada e analisada, de forma a tentar conceber uma proposta capaz de solucionar o problema em questão. Por fim, os resultados obtidos ao longo da investigação fornecem a base para a formulação de conclusões. Para assegurar a correta aplicação desta metodologia e alcançar os objetivos propostos, são definidas cinco etapas interligadas, descritas em baixo:

- 1 **Definição do Problema e das suas Características:** Nesta fase, procede-se à identificação do problema, analisam-se as razões que justificam a sua existência, reúnem-se as suas características principais e formula-se a hipótese de resolução.
- 2 **Revisão Contínua e Incremental do Estado da Arte:** Corresponde ao estudo das abordagens relacionadas com o objeto de investigação.
- 3 **Desenvolvimento da Solução:** Com base na informação recolhida nas etapas anteriores, é concebida uma solução que procura atingir os objetivos definidos, possibilitando a validação ou refutação das hipóteses formuladas.
- 4 **Experimentação e Implementação da Solução por Meio de um Protótipo:** Consiste na elaboração de um protótipo que integre os recursos definidos na solução proposta, observando-se o seu comportamento para avaliar a sua eficácia.
- 5 **Análise e Avaliação de Resultados e Elaboração de Conclusões:** Compreende a validação do protótipo, verificando se a implementação atinge os objetivos

previamente estabelecidos, o que permite retirar conclusões relativamente à hipótese de investigação.

## 1.6 Estrutura do Documento

Após este capítulo introdutório, o documento apresenta a seguinte estrutura:

- **Capítulo 2 – Estado da Arte:** Reúne e analisa a literatura relevante sobre os principais pilares da investigação: Teoria da Argumentação, *Large Language Models*, Grafos de Conhecimento e Análise de Argumentos em Redes Sociais. Serve como base teórica para a solução proposta.
- **Capítulo 3 – Design e Conceção da Solução:** Descreve os requisitos funcionais e não funcionais do sistema, as tecnologias utilizadas e as decisões de design arquitetónico. Apresenta o funcionamento da estrutura modular do sistema e o racional por trás da escolha de ferramentas e abordagens.
- **Capítulo 4 – Implementação da Solução:** Detalha o desenvolvimento da *pipeline* desenvolvida e dos módulos que a compõem. Explica o funcionamento por trás de cada componente, desde a recolha de dados até à criação de grafos de conhecimento.
- **Capítulo 5 – Resultados e Discussão:** Apresenta os resultados obtidos através da execução da solução, incluindo demonstrações práticas e avaliações críticas. Analisa o desempenho dos módulos, a utilidade das representações geradas e a qualidade das classificações e agrupamentos criados.
- **Capítulo 6 – Conclusões:** Resume o trabalho realizado e as principais contribuições. Avalia o cumprimento dos objetivos e requisitos definidos, destaca os resultados alcançados, discute limitações identificadas e propõe direções para trabalho futuro. Inclui ainda uma reflexão sobre o impacto científico do projeto, incluindo a menção a um artigo científico escrito e submetido para publicação durante o desenvolvimento do projeto.
- **Referências, Apêndices e Anexos:** Reúnem todos os materiais complementares que sustentam o trabalho desenvolvido, incluindo código, exemplos de execução da solução, excertos de dados e visualizações adicionais.

## 1.7 Nota Terminológica

No âmbito desta dissertação, opta-se por manter em inglês determinados termos técnicos e conceitos-chave amplamente utilizados na literatura científica e na prática profissional na área da Inteligência Artificial e Processamento de Linguagem Natural. Esta decisão prende-se com a ausência de traduções consensuais ou equivalentes exatos em português, bem como com a frequência e familiaridade com que estes termos surgem em publicações especializadas, *frameworks* e documentação técnica. Assim, expressões como *Large Language Models*, *Natural Language Processing*, *prompt*, *fine-tuning*, *embedding*, entre outras, serão utilizadas ao longo do texto na sua forma original em inglês. Esta abordagem visa assegurar clareza, precisão e alinhamento com a terminologia internacionalmente estabelecida, facilitando também a consulta de fontes e a articulação com trabalhos relacionados.



## 2 Estado da Arte

O presente capítulo tem como principal objetivo apresentar uma análise crítica e sistemática da literatura existente nos domínios centrais desta dissertação, de modo a fundamentar teoricamente o trabalho desenvolvido. A revisão do estado da arte permitiu identificar as principais abordagens, tecnologias e desafios associados à análise e sintetização de argumentos em redes sociais, com foco na utilização de *Large Language Models* e Grafos de Conhecimento.

Os tópicos abordados neste capítulo foram selecionados com base numa análise exploratória inicial da literatura, articulada com os objetivos definidos para este projeto. Procurou-se garantir uma cobertura abrangente dos conceitos e ferramentas essenciais para a concretização da solução proposta, dando particular destaque às contribuições mais recentes e relevantes em cada área. Para isto, o capítulo encontra-se organizado em quatro subcapítulos: (1) Argumentação e Negociação Baseada em Argumentos, (2) *Large Language Models*, (3) Grafos de Conhecimento, e (4) Análise de Argumentos em Redes Sociais.

A análise realizada neste capítulo constitui a base teórica que sustenta as decisões metodológicas e tecnológicas adotadas na fase de desenvolvimento da solução proposta.

### 2.1 Argumentação e Negociação baseada em Argumentos

Atualmente, com o incentivo de plataformas como o Facebook, Twitter, Reddit e outras redes sociais já mencionadas, encontram-se, frequentemente, opiniões divergentes que são muitas

vezes tratadas como discussões. O confronto de opiniões distintas não se limita ao ambiente digital. Este fenômeno manifesta-se em vários contextos, seja na ciência, em conversas íntimas com amigos e familiares, na política ou até mesmo diante da televisão, quando o conteúdo apresentado gera forte discordância (Feger, 2021).

### **2.1.1 Argumentos e Argumentação**

Antes de avançar para a designação do conceito composto da “Argumentação Online”, é importante introduzir o conceito de “Argumentação”. De acordo com o *Cambridge Dictionary*, um Argumento pode ser definido como “a razão ou conjunto de razões pelas quais se apoia ou se opõe a uma ideia ou sugestão, ou o processo de explicar essas razões.” Acrescentando a isto, pode ainda ser definido como “as razões para a opinião sobre a veracidade de algo ou uma explicação de porque se acredita que algo deve ser feito.” Krauthoff et al. (2016), define a Argumentação como a troca racional de posições, razões e justificativas, e como sendo uma ferramenta vital sempre que um grupo de duas ou mais pessoas necessita de decidir um curso de ação, determinar o que aceitar como verdade, concordar com um conjunto de valores compartilhados ou, simplesmente, alcançar um entendimento comum sobre as posições dos membros do grupo.

Van Eemeren et al. (1996), no livro “*Fundamentals of Argumentation Theory*”, apresenta a sua visão sobre o conceito de Argumentação:

*Argumentação é, por fim, uma atividade racional, o que indica que o argumentador refletiu sobre o assunto. Apresentar um argumento significa que o argumentador tenta demonstrar que uma explicação racional pode ser dada para sua posição sobre o tema. Isso não significa que emoções não possam desempenhar um papel na adoção de uma posição, mas sim que esses motivos internos, embora assimilados no discurso, não são diretamente relevantes como tais. Quando as pessoas apresentam seus argumentos em uma argumentação, colocam suas considerações no domínio da razão.*

Alpert (2016), refere que na filosofia um argumento não representa um conflito ou um debate, mas sim uma defesa estruturada de uma afirmação sobre determinado tema. Ao construir um argumento, uma pessoa não se limita a apresentar uma opinião ou resposta pessoal a uma questão filosófica específica, já que isso seria meramente a expressão de um ponto de vista. Em vez disso, explica-se porque essa resposta é válida e fundamentada. Idealmente, também se

demonstra porque essa resposta é superior às visões alternativas sobre a mesma questão (Alpert, 2016). Feger (2021), também alega que o tópico da argumentação remonta à antiguidade, estando particularmente associado a filósofos como Aristóteles.

Toulmin (2003), compara um argumento a um organismo que possui tanto uma estrutura anatômica geral, quanto uma estrutura mais refinada, quase como uma fisiologia. Dizendo que quando reduzido ao detalhe, um argumento pode ocupar várias páginas ou levar cerca de um quarto de hora para ser apresentado.

A estrutura de um argumento também é objeto de estudo e pode definir a adequação e qualidade do mesmo. Toulmin (2003) apresenta várias estruturas de argumentos rudimentares, antes de chegar à estrutura final que considera ser a mais adequada e que engloba todas as variáveis e cenários possíveis. Na Figura 2 é representada a estrutura de um argumento segundo Toulmin, em que o D representa os Dados (*Data*), o Q representa o Qualificador (*Qualifier*), o C representa a Afirmação (*Claim*), o W representa a Justificação ou Fundamento Lógico (*Warrant*), o B representa a Corroboração (*Backing*), e o R representa a refutação (*Rebuttal*). Na Figura 3, um exemplo prático criado pelo próprio autor.

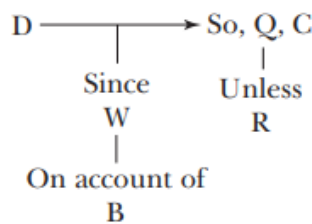


Figura 2 - Estrutura de Um Argumento (Toulmin, 2003).

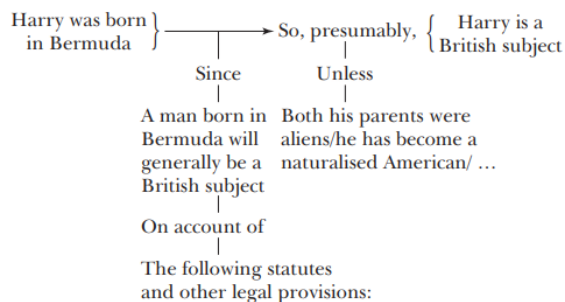


Figura 3 - Exemplo Prático da Estrutura de Um Argumento (Toulmin, 2003).

Alpert (2016) refere que os argumentos têm uma estrutura essencial que distingue uma simples afirmação de um argumento propriamente dito. Essa estrutura consiste em pelo menos uma

premissa e uma conclusão. Quando se apresenta uma razão na forma de uma proposição (uma frase que pode existir de forma independente), ela é chamada de premissa. A alegação que se está a defender é chamada de conclusão do argumento, pois supõe-se que seja o resultado do raciocínio apresentado em sua defesa. Assim, um argumento deve incluir pelo menos uma premissa e uma conclusão. Alpert (2016), faz ainda uma simples demonstração prática dos fundamentos teóricos apresentados, que pode ser encontrada nos Anexos (Anexo A – Constituição de um Argumento segundo Alpert (2016)).

### **2.1.2 Negociação Baseada em Argumentos**

Dentro do tema da Argumentação, encontra-se o subtema da Negociação Baseada em argumentos. Rahwan et al. (2003) define Negociação como “uma forma de interação em que um grupo de agentes, com interesses conflitantes e um desejo de cooperação, procura alcançar um acordo mutuamente aceitável quanto à divisão de recursos escassos.” El-Sisi & Mousa (2014) definem o termo como “o processo até se alcançar um acordo sobre os termos de uma transação, como preço e quantidade, envolvendo duas ou mais partes em sistemas multiagente, como no caso do Comércio Eletrônico. Tenta maximizar os benefícios para todas as partes envolvidas”. Já Carneiro (2017), define negociação como “o processo entre duas ou mais partes que procuram alcançar um acordo”.

Das três abordagens de negociação existentes – baseadas na teoria dos jogos, em heurísticas e em argumentação – o presente trabalho centrará a sua atenção na última: as abordagens baseadas em argumentação.

Rahwan et al. (2003), refere que a negociação baseada em argumentos procura superar as limitações dos modelos tradicionais de negociação, como os baseados na teoria dos jogos e em heurísticas. Essas limitações incluem a troca restrita de informações, a suposição de que as preferências ou utilidades dos agentes são totalmente conhecidas e fixas, e a incapacidade de os agentes influenciarem diretamente as preferências ou atitudes mentais de outros agentes.

Na abordagem baseada em argumentos, os agentes podem trocar informações adicionais, justificar as suas propostas e influenciar as crenças, objetivos ou preferências dos seus interlocutores. Desta maneira, os argumentos podem ser usados para (Rahwan et al., 2003):

- Justificar posições de negociação: Um agente pode explicar porque faz uma proposta ou porque não pode aceitar outra, facilitando a compreensão mútua.
- Influenciar preferências: A troca de argumentos pode modificar as áreas de aceitação de um agente ou reconfigurar o espaço de negociação, tornando propostas inicialmente inaceitáveis mais atraentes.
- Reformular objetos de negociação: Os agentes podem introduzir novos atributos ao objeto negociado, ampliando as possibilidades de acordo.
- Exercer pressão: Podem ser feitas ameaças ou promessas para influenciar o comportamento do interlocutor.

O autor ilustra essas ideias com exemplos, como disputas trabalhistas e negociações salariais, onde argumentos e justificações podem levar a resultados mais eficientes e mutuamente aceitáveis. Dessa forma, a negociação baseada em argumentos não apenas resolve conflitos, mas também enriquece o processo ao incorporar elementos racionais e persuasivos (Rawhan et al., 2003).

Segundo Carneiro (2017), a negociação baseada em argumentos distingue-se pela possibilidade de sustentar propostas por meio de justificações e explicações, as quais assumem um papel central no contexto negocial. Esta abordagem permite que os intervenientes não se limitem à troca de propostas, mas também apresentem razões e fundamentações que as sustentem, exercendo influência mútua nas preferências dos participantes e, conseqüentemente, no desfecho do diálogo.

A persuasão através de argumentos já era objeto de estudo de Aristóteles, filósofo já mencionado anteriormente, que definia o conceito como o processo de provar, ou de aparentar provar, uma determinada posição, destacando que este método se diferencia por ser probativo, ou seja, por estabelecer uma conexão inferencial entre a conclusão sugerida e fatos evidentes ou convicções prévias do público (Rapp, 2023). Aristóteles também distinguia entre argumentos reais e aparentes, sendo que os últimos, ainda que falaciosos, podem exercer um efeito persuasivo quando a falácia não é percebida pelo público. O estudo dos argumentos falaciosos, segundo Aristóteles, é essencial não apenas para refutá-los quando utilizados por opositores, mas também para compreender a sua aplicação estratégica. (Rapp, 2023).

### **2.1.3 Argumentação e Negociação Baseada em Argumentos em Contexto Digital**

A argumentação e a negociação baseada em argumentos em contextos digitais, especialmente nas redes sociais, têm recebido atenção crescente na literatura acadêmica devido à relevância dessas plataformas na mediação de interações, tomadas de decisão e resolução de conflitos.

No que concerne a este tema, Feger (2021), define o conceito como a apresentação sistemática, assistida por computador, e transparente de declarações geradas por utilizadores, bem como das suas derivações e relações entre si. (Walton, 2007) destaca que a argumentação em ambientes digitais se caracteriza pela utilização de raciocínio estruturado para persuadir audiências, muitas vezes acompanhada de explicações e evidências. Em redes sociais, a argumentação é amplamente influenciada pelo contexto interativo e pela natureza pública das discussões, que podem atrair participantes com diferentes perspectivas e níveis de conhecimento. Van Eemeren et al. (1996) reforçam que o uso de argumentação em ambientes digitais é moldado pela acessibilidade e pela rapidez das interações, fatores que favorecem tanto a colaboração quanto a polarização. Argumentos podem ser utilizados não apenas para convencer, mas também para contestar informações falsas, reforçar identidades de grupo e mobilizar ações coletivas.

Quanto à negociação argumentativa em contextos digitais, Sycara (1992) destaca que esta permite que os agentes justifiquem as suas preferências e influenciem o modelo mental dos outros participantes, superando as limitações de abordagens tradicionais. Isto é particularmente relevante em redes sociais, onde informações adicionais e justificações são frequentemente utilizadas para moldar a percepção pública sobre determinados temas.

A Internet, e mais especificamente as redes sociais, forneceram a infraestrutura básica para viabilizar a argumentação em todos os tipos de grupos, independentemente do tamanho desses grupos, da localização dos seus membros ou do horário em que escolhem participar. Habermas (1984) identifica que estas plataformas podem atuar como espaços de deliberação pública, facilitando a troca de argumentos que contribuem para o entendimento mútuo e a resolução de conflitos. No entanto, Krauthoff et al. (2016) apresenta um sentimento de insatisfação sobre a situação atual do tema referindo que essa infraestrutura básica ainda não resultou na propagação esperada da troca racional de argumentos. Pelo contrário, o cenário parece apontar para um desfecho oposto. As secções de discussão de portais de notícias online, blogs e redes sociais carecem de estrutura e clareza, sofrem com a repetição frequente de

argumentos semelhantes, confundem diferentes aspectos de um mesmo tema ou apresentam-se enviesadas, irrelevantes, emocionalmente exaltadas e mal fundamentadas. Além disso, essas discussões promovem a fragmentação dos participantes e não conseguem escalar para grandes volumes de utilizadores (Krauthoff et al., 2016).

## **2.2 *Large Language Models* (Modelos de Linguagem de Larga Escala)**

A linguagem desempenha um papel fundamental na facilitação da comunicação e da expressão pessoal dos seres humanos, bem como na sua interação com máquinas. A necessidade de modelos generalizados decorre da crescente exigência de que as máquinas sejam capazes de lidar com tarefas linguísticas complexas, como tradução, sumarização, recuperação de informação, interações conversacionais, entre outras (Naveed et al., 2024).

Dentro do grande tema dos *Large Language Models*, emergem diversos subtemas de relevância que merecem análise e compreensão aprofundadas. Contudo, é imperativo começar pela definição do conceito central de modo a estabelecer uma base sólida que permita uma compreensão integrada e abrangente do tema.

### **2.2.1 *Generative AI***

Começando pelo princípio, antes de se tentar perceber o conceito de LLMs, convém definir o conceito de *Generative AI*. A *Gen AI*, em diminutivo, é uma parte do leque de aplicações e usos da Inteligência Artificial, que contém uma classe de algoritmos e modelos capaz de criar conteúdos, como imagens, texto ou música. A geração de conteúdo baseia-se em solicitações dos utilizadores, padrões existentes e exemplos provenientes de dados previamente disponíveis (Christensen, 2024). A *Gen AI* tem como ponto de partida um *foundation model* — um modelo de *deep learning* que serve de base para diversos tipos de aplicações generativas. Os modelos mais comuns atualmente são os *Large Language Models*, concebidos para aplicações de geração de texto. Contudo, também existem *foundation models* voltados para a geração de imagens, vídeos, sons e música, bem como modelos multimodais, capazes de suportar a geração de vários tipos de conteúdo em simultâneo (Stryker & Scapicchio, 2024).

### 2.2.2 Large Language Models: O conceito

Os modelos generativos surgiram como sistemas de inteligência artificial de ponta, capazes de processar e gerar texto com comunicação coerente, além de se generalizarem para múltiplas tarefas. Recentemente, avanços na área trouxeram uma transformação revolucionária, permitindo a criação de LLMs capazes de se aproximarem ao desempenho humano em diversas tarefas (Naveed et al., 2024). Na Figura 4 é possível observar uma linha cronológica do lançamento de vários LLMs relevantes, destacando o crescimento substancial no final do ano de 2021.

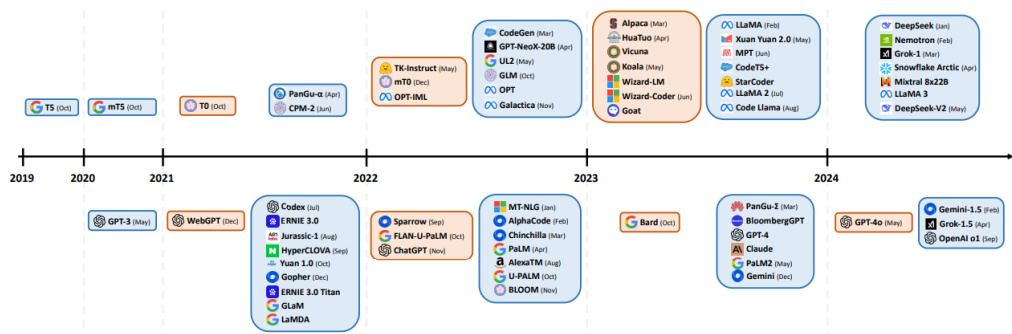


Figura 4 - Ordem Cronológica do Lançamento de Diferentes LLMs Relevantes. Naveed et al., 2024

### O que são, então, Large Language Models?

Os LLMs são estruturas de aprendizagem automática extremamente complexas, com uma capacidade inata de processar e gerar texto de forma incrivelmente semelhante à linguagem humana. A sua proficiência resulta de um treino extensivo em um vasto corpus de dados textuais recolhidos de diversos domínios, sendo treinados para identificar como os padrões presentes nesses textos criam contexto e significado (Broadhead, 2023).

Shanahan (2024) define LLMs como modelos matemáticos generativos que representam a distribuição estatística de *tokens* no vasto corpo público de textos gerados por humanos. Devlin et al. (2019) define o conceito como um modelo de linguagem que utiliza redes neurais profundas com bilhões de parâmetros, treinado para prever palavras subsequentes em uma sequência textual. Este processo, conhecido como *language modeling*, é essencial para tarefas como tradução automática, sumarização e respostas a perguntas. Kunz & Kuhlmann (2024), por sua vez, alega que os LLMs são redes neurais de grande escala, com milhões ou até bilhões de parâmetros, que aprendem a representar a linguagem ao prever palavras com base no seu

contexto. Ao serem treinados em grandes volumes de texto e realizarem essas previsões de palavras bilhões de vezes, os LLMs conseguem capturar informações estatísticas detalhadas sobre a linguagem. Estas características têm-se revelado extremamente úteis para quase todos os tipos de aplicações secundárias atualmente relevantes no campo do Processamento de Linguagem Natural (NLP).

Em todas as anteriores definições de diferentes autores, surgem vários termos cujos conceitos ainda não foram abordados e que necessitam de ser dissecados para total compreensão das definições.

### **2.2.3 *Natural Language Processing* - NLP**

O Processamento de Linguagem Natural ou *Natural Language Processing* – NLP, pode ser definido, de acordo com Copestake (2004), como o processamento automático (ou semiautomático) da linguagem humana. No âmbito da ciência da computação, o NLP está relacionado à teoria de linguagens formais, técnicas de compiladores, demonstração de teoremas, aprendizagem automática e interação humano-computador. O autor alega ainda que o termo NLP é, por vezes, utilizado de forma mais restrita, excluindo frequentemente a recuperação de informação e, em alguns casos, até a tradução automática Copestake (2004). Amaratunga (2023) define NLP como a tecnologia que tem como objetivo principal reduzir a lacuna entre a linguagem humana e a compreensão computacional, sendo que o tem sido aplicada, ao longo da sua história, a diversas tarefas relacionadas com a linguagem.

O Processamento de Linguagem Natural (NLP) abrange uma ampla gama de tarefas, cada uma com objetivos específicos que buscam melhorar a interação entre humanos e máquinas. Entre as tarefas mais comuns, Amaratunga (2023) destaca:

- **Classificação de Texto**, que consiste em atribuir categorias ou etiquetas a segmentos de texto, como na classificação de e-mails em “spam” ou “não spam”, na análise de sentimentos (positivo, negativo ou neutro) ou na categorização por tópicos.
- **Tradução Automática**, que permite converter conteúdo textual de um idioma para outro, automatizando processos de tradução linguística.
- **Geração de Texto**, tarefa em que o sistema produz texto com aparência humana, útil na criação de conteúdos, respostas automáticas e resumos.

- **Reconhecimento de Fala**, que converte linguagem falada em texto, sendo essencial em aplicações como assistentes virtuais ou legendas automáticas.
- **Sumarização de Texto**, que gera versões condensadas e coerentes de textos mais longos, mantendo os pontos-chave da informação.
- **Resposta a Perguntas**, que permite ao sistema interpretar perguntas em linguagem natural e fornecer respostas diretas com base no seu conhecimento ou no conteúdo analisado.
- **Modelagem de Linguagem**, uma tarefa central que consiste em prever a próxima palavra de uma sequência, sendo a base para muitas das outras tarefas referidas.

#### 2.2.4 Tokens e Tokenização

Outro dos conceitos de grande importância que estão presentes nas definições de LLMs acima, é o conceito de *token* e *tokenização*. A *tokenização* é uma etapa essencial de pré-processamento no treino de LLMs, responsável por dividir o texto em unidades indivisíveis denominadas *tokens*. Os *tokens* podem corresponder a caracteres, subpalavras, símbolos ou palavras, dependendo do processo de tokenização (Naveed et al., 2024). Amaratunga (2023) define tokenização como o processo de dividir um texto em unidades menores, tipicamente palavras ou subpalavras. Essas unidades menores são denominadas *tokens*.

A OpenAI (2024), referência mundial a nível de Inteligência Artificial, providencia a explicação perfeita do conceito de *token* no seu website referindo que estes podem ser entendidos como fragmentos de palavras. Esses *tokens* nem sempre coincidem exatamente com os limites das palavras, podendo incluir espaços finais e até mesmo subpalavras.

Para se compreender a dimensão dos *tokens*, a OpenAI também oferece algumas regras práticas que podem ser úteis:

- 1 *token*  $\approx$  4 caracteres em inglês
- 1 *token*  $\approx$   $\frac{3}{4}$  de uma palavra
- 100 *tokens*  $\approx$  75 palavras
- 1-2 frases  $\approx$  30 *tokens*
- 1 parágrafo  $\approx$  100 *tokens*
- 1.500 palavras  $\approx$  2.048 *tokens*

A célebre frase de Wayne Gretzky, "You miss 100% of the shots you don't take", contém 11 tokens. A forma como as palavras são divididas em *tokens* também varia conforme o idioma. Por exemplo, a frase em espanhol "Cómo estás" ("How are you", em inglês) contém 5 *tokens* para 10 caracteres (OpenAI, 2024).

É possível visualizar o processo de tokenização na imagem seguinte (Figura 5) providenciada pela Databricks (2024):

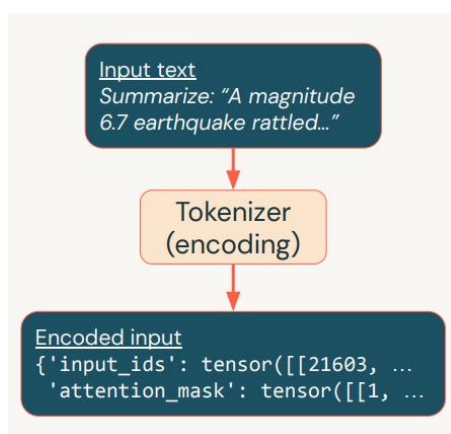


Figura 5 - Processo de *Tokenização* (Databricks, 2024)

### 2.2.5 Parâmetros

Também é importante perceber o que são os parâmetros de um modelo e Broadhead (2023) define-os como “os condutores que orquestram a forma como o modelo manipula e compreende os dados de entrada para gerar as saídas. Estes manifestam-se como pesos ou vieses, influenciando a relevância de características específicas dos dados de entrada no resultado gerado”. Um maior número de parâmetros geralmente corresponde a um modelo com maior complexidade e adaptabilidade (embora isto não seja estritamente verdade em todas as arquiteturas, é geralmente válido no contexto de arquiteturas baseadas em *transformers*). Um modelo de linguagem de grande escala com um maior número de parâmetros tem a capacidade de discernir padrões mais complexos nos dados, permitindo a geração de saídas mais ricas e precisas (Broadhead, 2023).

#### Temperatura

A temperatura é um hiperparâmetro central nos modelos generativos, que controla a aleatoriedade da seleção de *tokens* pelo modelo. Quando a temperatura é baixa (próxima de

0), o modelo tende a escolher os *tokens* com maior probabilidade, gerando texto mais determinístico e coerente; já valores mais elevados promovem variações e diversidade na saída, gerando respostas mais criativas, mas potencialmente menos consistentes (Peeperkorn et al., 2024).

No estudo "*Is Temperature the Creativity Parameter of Large Language Models?*" (Peeperkorn et al., 2024), os autores investigaram esta relação através da geração de narrativas usando diferentes valores de temperatura e avaliando quatro critérios de criatividade: originalidade, tipicidade, coesão e coerência. Concluíram que:

- A temperatura tem correlação fraca com a originalidade (mais diversidade).
- Existe uma relação moderada com a incoerência (alta temperatura aumenta inconsistência).
- Não há relação estatística significativa com coesão ou tipicidade.

Assim, embora temperatura influencie ligeiramente a criatividade em termos de diversidade, os autores demonstram que não deve ser encarada como o parâmetro definitivo da criatividade nos LLMs. Em vez disso, propõem que este conceito seja visto como o grau de abrangência com que o modelo recorre ao texto dos dados em que foi treinado (Murel & Noble, 2024).

### **2.2.6 Transformers**

Em 2017, colaboradores do *Google Brain* e *Google Research* propuseram uma arquitetura revolucionária para as redes neurais, voltada para o processamento de linguagem natural (NLP) e outras tarefas de sequência a sequência. No artigo "*Attention Is All You Need*" de Vaswani et al., (2017), os autores apresentaram uma abordagem inovadora baseada no uso intensivo de mecanismos de atenção, permitindo a paralelização do processamento, treino eficiente e a capacidade de capturar dependências de longo alcance nos dados (Amaratunga, 2023).

Os *transformers* são um tipo de arquitetura de redes neuronais que transforma ou converte uma sequência de entrada numa sequência de saída. Este processo é realizado através da aprendizagem do contexto e da análise das relações entre os componentes da sequência. Por exemplo, pode-se considerar a sequência de entrada: "Qual é a cor do céu?". O *transformer* utiliza uma representação matemática interna que identifica a relevância e as relações entre as

palavras "cor", "céu" e "azul". A partir desse conhecimento, o modelo gera a sequência de saída: "O céu é azul" (AWS, 2024).

As redes neurais são modelos altamente eficazes para analisar tipos de dados complexos, como imagens, vídeos, áudio e texto. No entanto, existem diferentes tipos de redes neurais otimizadas para diferentes tipos de dados. Por exemplo, para a análise de imagens, as mais comuns são as redes neurais convolucionais (CNNs). No entanto, para análise de texto, o mais comum é a utilização de *Recurrent Neural Networks* (RNN). De forma geral, estas redes imitam a maneira como o cérebro humano processa a informação visual (Markowitz, 2021).

Ao contrário das RNNs, os *Transformers* podem ser paralelizados de uma forma muito eficiente, ou seja, dividir uma tarefa em sub-tarefas menores que podem ser processadas simultaneamente, acelerando assim o processo (Markowitz, 2021).

A complexidade dos *transformers* pode ser decomposta em três conceitos principais:

- **Codificações Posicionais:** As codificações posicionais permitem que os *Transformers* representem a ordem das palavras diretamente nos dados de entrada, eliminando a necessidade de processamento sequencial típico das redes neurais recorrentes. Utilizando funções trigonométricas para gerar essas codificações, os *Transformers* aprendem, durante o treino, a interpretar a posição relativa das palavras, tornando o modelo mais eficiente e adequado à paralelização (Markowitz, 2021).
- **Attention:** Este tipo de arquitetura usa mecanismos de atenção inspirados na forma como os seres humanos processam informações, priorizando elementos mais relevantes em um dado momento e permitindo que o modelo atribua diferentes pesos às partes de uma sequência, destacando as mais relevantes para uma tarefa específica, como tradução ou geração de texto (Vaswani et al., 2017). A atenção atribui pesos aos *tokens* de entrada com base na sua importância, de modo que o modelo dê mais ênfase aos *tokens* mais relevantes (Naveed et al., 2024).
- **Self-Attention:** A *self-attention*, por outro lado, refere-se à capacidade de um *transformer* se concentrar em diferentes partes da própria sequência de entrada ao realizar previsões. O nome deriva do fato de que, ao contrário da "atenção regular", a *self-attention* foca-se na mesma sequência que está a ser codificada no momento (Wydanski, 2022). A *self-attention* permite que uma rede neuronal compreenda uma palavra no contexto das palavras que a rodeiam. Este mecanismo auxilia na

desambiguação de palavras, na etiquetagem, na resolução de entidades, na aprendizagem de papéis semânticos e em diversas outras tarefas (Markowitz, 2021).

### 2.2.7 Modelos *Open-Source* e *Closed Source*

Os modelos *open-source*, disponibilizados por plataformas como a Hugging Face<sup>7</sup>, oferecem aos utilizadores a possibilidade de utilizar, modificar e adaptar os modelos às necessidades específicas dos seus projetos. Exemplos notáveis incluem o GPT, o Llama e variantes do BERT (Hugging Face, 2024). Estes modelos são particularmente atrativos para investigadores e desenvolvedores que procuram flexibilidade e controlo sobre os dados e parâmetros utilizados na fase de treino. Para além disso, o código aberto promove colaboração e inovação, permitindo que a comunidade científica contribua para avanços rápidos nesta área.

Por outro lado, os modelos *closed-source*, como os disponibilizados pela OpenAI (por exemplo, o GPT-4), proporcionam acessibilidade através de APIs públicas que são fáceis de integrar em aplicações. No entanto, o acesso limitado ao funcionamento interno desses modelos pode representar um desafio para projetos que necessitem de personalizações profundas. Apesar disso, a qualidade e a robustez desses modelos tornam-nos ideais para soluções prontas para produção (Brown et al., 2020). Ainda, modelos *closed-source* como os da OpenAI, exigem o pagamento de uma taxa de utilização por *tokens* de input e output (OpenAI, 2024).

O hardware do utilizador constitui um fator determinante na escolha entre modelos *open-source* e *closed-source*. Modelos *open-source* são executados localmente na máquina do utilizador, o que pode revelar-se inviável ou mesmo impossível, dependendo dos recursos disponíveis e das especificidades do modelo e do objetivo final. Por outro lado, os modelos *closed-source* são processados em servidores da empresa responsável, garantindo, em princípio, um desempenho superior devido à infraestrutura otimizada dessas plataformas. Os LLMs requerem grandes volumes de dados diversificados, cobrindo múltiplos idiomas e domínios. Além disso, o treino de LLMs é altamente intensivo em recursos computacionais, exigindo infraestruturas de alta performance (Vaswani et al., 2017). A escolha entre *open* e *closed-source* dependerá, portanto, das necessidades específicas e requisitos do projeto em questão, do hardware disponível, e do orçamento do utilizador.

---

<sup>7</sup> Hugging Face – [www.huggingface.co](http://www.huggingface.co)

## Estruturação de Projetos com LLMs

A implementação de LLMs em projetos pode exigir a definição de *pipelines* de processamento que integrem diferentes componentes para atingir os objetivos definidos.

Tais *pipelines* podem incluir (Abbas, 2024):

- Pré-processamento de Dados: Limpeza e transformação dos dados de texto de entrada para um formato que o modelo possa processar.
- Seleção do Modelo: Escolha do modelo (ou modelos) apropriado(s) para realizar inferências.
- Inferência: Execução do modelo (LLM) para gerar previsões ou respostas com base nos dados de entrada.
- Pós-processamento: Formatação ou processamento adicional da saída do modelo, incluindo refinamento do texto, aplicação de lógica empresarial ou integração com sistemas externos.
- *Feedback* e Monitorização: Avaliação do desempenho do modelo, recolha de *feedback* dos utilizadores e refinamento do sistema para obter melhores resultados.

Aos dias de hoje, existem ferramentas próprias construídas com o intuito de ajudar e facilitar o trabalho dos utilizadores que pretendam construir *pipelines* de LLMs. Exemplos destas ferramentas incluem o LangChain ou o LlamaIndex, e *pipelines* desenvolvidas com alguma destas, conectam múltiplas chamadas a LLMs e/ou a sistemas externos, como bases de dados vetoriais ou pesquisas na web (Databricks, 2024). Essas *pipelines* permitem o uso de LLMs em tarefas complexas, como perguntas e respostas baseadas em bases de conhecimento ou num conjunto de documentos específicos. O desenvolvimento de aplicações com LLMs concentra-se frequentemente na construção dessas cadeias de modelos, em vez da criação de novos modelos de linguagem (Databricks, 2024).

### 2.2.8 Prompts e Prompt Engineering

Marvin et al. (2024), define *prompt* como uma entrada baseada em texto que é fornecida a um modelo de linguagem com o objetivo de orientar a sua resposta. Uma *prompt* consegue guiar um modelo a gerar o *output* desejado pelo utilizador (Hugging Face, 2024).

*Prompt engineering* surge como uma técnica crucial para potenciar as capacidades de LLMs. Esta técnica consiste na conceção estratégica de instruções específicas para uma determinada tarefa (*prompts*) com o objetivo de orientar a saída do modelo, sem necessidade de alterar os seus parâmetros (Sahoo et al., 2024). Ekin (2023), define *Prompt engineering* como o processo de conceção, refinamento e otimização de *prompts* com o objetivo de comunicar de forma eficaz a intenção do utilizador a um modelo de linguagem.

A importância do *prompt engineering* torna-se especialmente evidente no seu impacto transformador sobre a adaptabilidade dos LLMs. Ao fornecer um mecanismo para ajustar as respostas dos modelos através de instruções cuidadosamente formuladas, *prompt engineering* permite que estes modelos atinjam um desempenho elevado em tarefas e domínios variados (Sahoo et al., 2024).

### **2.2.9 Treino e *Fine-Tuning* de Modelos**

O treino de LLMs representa um dos avanços mais significativos na área, possibilitando capacidades avançadas em tarefas de compreensão e geração de linguagem natural. Este processo pode envolver diversas etapas complexas e utilizar extensos recursos computacionais e conjuntos de dados massivos.

O processo de treino de LLMs é tipicamente dividido em duas fases principais: o pré-treino (*pre-training*) e o *fine-tuning*. Durante o pré-treino, o modelo é exposto a grandes volumes de dados não supervisionados para aprender padrões gerais da linguagem, como sintaxe, semântica e relações contextuais. Esta etapa utiliza tarefas auto-supervisionadas, como a predição de palavras mascaradas ou a geração de texto, conforme descrito por Devlin et al. (2019) no desenvolvimento do BERT (*Bidirectional Encoder Representations from Transformers*). O treino destes modelos exige uma ampla experiência prática no processamento de dados em larga escala e em treino paralelo. Para desenvolverem LLMs eficazes, os utilizadores precisam de resolver questões de engenharia complexas (Zhao et al., 2024).

#### **Métodos de Treino: *Zero-shot*, *Few-shot* e *Fully Supervised***

No contexto de LLMs, os métodos de treino variam de acordo com a quantidade de exemplos fornecidos. O treino *zero-shot* refere-se à capacidade do modelo de realizar tarefas sem exposição prévia a exemplos específicos, baseando-se apenas no contexto geral aprendido

durante o pré-treino (Brown et al., 2020). Já o treino *few-shot* utiliza um número limitado de exemplos para ajudar o modelo a compreender melhor as nuances da tarefa, como destacado por autores no desenvolvimento do GPT-3 (Brown et al., 2020). Por outro lado, o treino supervisionado utiliza conjuntos de dados substanciais e anotados, embora exija recursos mais extensos e possa limitar a versatilidade do modelo (Raffel et al., 2020).

A segunda fase, de *fine-tuning*, adapta o modelo a tarefas específicas utilizando conjuntos de dados menores e supervisionados. Este método permite que o modelo se especialize em aplicações concretas, como classificação de texto ou respostas a perguntas, enquanto aproveita os conhecimentos adquiridos no treino (Howard & Ruder, 2018). O conceito de *fine-tuning* baseia-se na utilização de um modelo previamente treinado e na continuação do seu treino com dados diferentes, alterando apenas os pesos nas últimas camadas durante o processo. Este método requer apenas uma fração dos recursos utilizados no treino inicial, permitindo que seja realizado de forma significativamente mais rápida. Além disso, as estruturas aprendidas pelo modelo durante o pré-treino permanecem codificadas nas camadas iniciais e podem ser aproveitadas (Drost, 2023).

#### **2.2.10 Desafios da Utilização de LLMs**

Os *Large Language Models* têm emergido como ferramentas fundamentais para a realização de tarefas em contextos acadêmicos, empresariais e tecnológicos em geral. Embora tenham alcançado sucessos notáveis, enfrentam desafios significativos que não podem ser ignorados e abordá-los é fundamental para o avanço contínuo e aplicação prática destes modelos. Um dos principais desafios é o fenômeno das alucinações, que comprometem a fiabilidade das respostas - um fenômeno em que o modelo gera informações que, embora pareçam plausíveis, são infielis ou desprovidas de sentido (Z. Ji et al., 2023). Segundo Jing et al. (2024) diversos fatores contribuem para este problema: em primeiro lugar, os LLMs carecem de conhecimento específico de domínio, devido, principalmente, ao treino realizado em conjuntos de dados públicos, o que limita a sua capacidade de responder a questões especializadas; além disso, a atualização em tempo real do conhecimento interno dos LLMs apresenta desafios significativos, resultando em informações desatualizadas num mundo em constante mudança.

Outro problema relevante está relacionado com os elevados recursos computacionais necessários para treinar e operar LLMs. Este aspeto levanta preocupações ambientais e de acessibilidade, destacando a necessidade de práticas de AI sustentáveis (Jing et al., 2024).

## 2.3 Grafos de Conhecimento

A investigação sobre grafos de conhecimento (*Knowledge Graphs*) tem acelerado nos últimos anos, resultando no aparecimento de uma variedade de definições e descrições, por vezes contraditórias (Ehrlinger & Wöß, 2016). S. Ji et al. (2022) definem grafos de conhecimento como “uma representação estruturada de factos, composta por entidades, relações e descrições semânticas.” As entidades podem ser objetos, eventos e situações do mundo real ou conceitos abstratos, enquanto as relações representam as conexões entre as entidades, e as descrições semânticas das entidades e suas relações incluem tipos e propriedades com significados bem definidos (Stegeman, 2024; S. Ji et al., 2022). Paulheim (2016), por sua vez, refere que um grafo de conhecimento “descreve principalmente entidades do mundo real e as suas inter-relações, organizadas em forma de grafo; define classes e relações possíveis entre entidades por meio de um esquema; e possibilita a interligação arbitrária entre entidades.”

Stegeman (2024) já introduz o conceito de bases de dados de grafos, referindo que geralmente, os grafos de conhecimento são armazenados numa base de dados de grafos, que preserva de forma nativa as relações entre as entidades de dados.

Grafos de conhecimento bem construídos e em larga escala podem ser valiosos para diversas aplicações subsequentes, além de capacitar modelos conscientes de conhecimento com raciocínio baseado em senso comum, abrindo assim caminho para avanços na Inteligência Artificial (S. Ji et al., 2022).

### **Mas como funcionam estes grafos?**

Os grafos de conhecimento são frequentemente mencionados no contexto dos motores de pesquisa, sendo o *Google Knowledge Graph* a ferramenta que transformou a forma como pesquisamos e encontramos informações na Web e um bom exemplo para demonstrar o funcionamento destes grafos. Este agrega factos sobre pessoas, lugares e objetos, organizando-os numa rede estruturada de entidades. Quando se realiza uma pesquisa no Google, o sistema

utiliza as conexões entre entidades para apresentar os resultados mais relevantes no seu contexto (Stegeman, 2024).

As entidades no Google *Knowledge Graph* representam o mundo tal como o conhecemos, marcando uma mudança de paradigma de "*strings to things*". Por detrás desta expressão simples reside o conceito profundo de tratar a informação na web como entidades interligadas, em vez de apenas texto (Stegeman, 2024).

### **2.3.1 Grafos de Conhecimento Como Complementos aos *Large Language Models***

A combinação de LLMs e KGs visa superar as limitações que os LLMs apresentam devido à dependência de parâmetros estáticos, treinados em grandes corpos de dados. Esta fusão das duas tecnologias explora as forças complementares de cada abordagem, já que os grafos de conhecimento podem fornecer conhecimento factual atualizado e específico aos LLMs, melhorando sua capacidade de raciocínio; enquanto os LLMs podem acelerar a construção, validação e atualização dos grafos de conhecimento, reduzindo a necessidade de processos manuais (Kau et al., 2024). Segundo Erdl (2023), "os grafos de conhecimento são uma opção confiável para fundamentar LLMs, devido à sua capacidade de representar dados estruturados e não estruturados, ao contrário das bases de dados vetoriais." Podendo então tirar partido da RAG já mencionada anteriormente (Erdl, 2023).

Os grafos de conhecimento podem ser usados para "injetar" conhecimento em LLMs, seja por meio de *prompts* ou mecanismos de raciocínio *multi-hop* (quando um modelo conecta múltiplas informações presentes num texto ou diversas fontes e sistemas para gerar uma resposta). Além disso, ao incorporar conhecimento semântico estruturado, KGs permitem que LLMs ofereçam respostas mais precisas e fundamentadas, aumentando sua interpretabilidade e confiança do utilizador (Kau et al., 2024).

### **2.3.2 Utilização de Grafos de Conhecimento na Análise de Política Social**

Num estudo de 2020 realizado por Abu-Salih et al., os autores utilizam grafos de conhecimento para analisar a política social através de um modelo de *embeddings* orientado à credibilidade e à ontologia do domínio específico. Implementam uma estrutura inovadora que integra dados de redes sociais, da Wikipedia e de notícias para construir um grafo de conhecimento no

domínio da política. Esse grafo é enriquecido semanticamente com ferramentas como Google KG, IBM Watson e WordNet, e a qualidade é assegurada por um módulo de credibilidade que filtra informações provenientes de fontes não confiáveis.

Os autores não utilizam explicitamente *Large Language Models* para a geração de conteúdo, mas empregam ferramentas avançadas de processamento de linguagem natural, incluindo o IBM Watson NLU, que oferece análise semântica e classificação de texto em domínios específicos. As técnicas de *embedding* aplicadas transformam os elementos do grafo em representações vetoriais de baixa dimensão, permitindo tarefas como previsão de ligações, *clustering* e visualização. Os autores testam múltiplos modelos de *embedding*, incluindo TransE, DistMult e ConvKB, destacando o desempenho superior deste último (Abu-Salih et al., 2020).

Esta abordagem alia metodologias de *embedding* com credibilidade e integração semântica, tendo sido um avanço na análise de políticas sociais ao permitir uma compreensão mais estruturada e confiável de dados complexos e heterogêneos (Abu-Salih et al., 2020).

## **2.4 Análise de Argumentos em Redes Sociais**

A análise da argumentação centra-se na compreensão do conteúdo das cadeias argumentativas, bem como na análise da estrutura linguística, na determinação da relação entre os argumentos precedentes e subsequentes, no reconhecimento das crenças conceituais subjacentes e na compreensão da coerência dos argumentos à luz do conhecimento geral sobre o tema específico. A análise da argumentação permite uma melhor compreensão do discurso, um conhecimento mais amplo das intenções e crenças do orador e destaca as interações entre os oradores e as suas diferentes visões do mundo (Mochales & Moens, 2011).

Os *Large Language Models* têm demonstrado um impacto significativo em diversos domínios, incluindo a análise e a síntese de argumentos em debates e discussões online. Esta tecnologia emergente oferece ferramentas poderosas para processar grandes volumes de texto, identificar padrões semânticos e extrair informações estruturadas de forma eficiente (Brown et al., 2020).

### 2.4.1 *Stance Detection* em Redes Sociais: Estado da Arte

*Stance detection*, ou identificação de posicionamento em português, é o nome que se dá à ação de deteção do posicionamento de um sujeito sobre uma entidade ou afirmação. Kùçük & Fazli (2020) definem o conceito como um problema de classificação em que se procura identificar o posicionamento do autor do texto sob a forma de categoria pertencente ao seguinte conjunto: {A Favor, Contra, Nenhum}. Ocasionalmente, a categoria Neutro também é adicionada ao conjunto de categorias de posicionamento. Hasan & Ng (2013) optam por uma abordagem mais prática à tentativa de definição do conceito referindo que “dada uma publicação sobre um tópico discutido num fórum de debate online (por exemplo, «Deve o aborto ser proibido?»), o objetivo da classificação de posicionamento no debate é identificar qual dos dois lados (ou seja, a favor ou contra) o autor está a apoiar.

ALDayel & Magdy (2021) dividem o conceito em duas definições diferentes: “*Target-specific stance detection*”, em que o texto ou o utilizador constituem a principal entrada para prever o posicionamento em relação a um alvo específico e previamente definido; e “*Claim-based stance detection*”, em que o alvo da análise não é uma entidade explícita, mas sim uma afirmação presente numa afirmação ou notícia, e o objetivo é identificar o posicionamento expresso nos comentários feitos sobre essa, avaliando se confirmam a sua veracidade ou se a desafiam.

Os trabalhos iniciais sobre *stance detection* centraram-se na análise de debates em fóruns online, distinguindo-se dos trabalhos mais recentes, que se focaram predominantemente em plataformas de redes sociais, especialmente no Twitter. Essa distinção deve-se ao facto de os fóruns online apresentarem um contexto único e bem definido, em comparação com as plataformas de redes sociais. Nos fóruns online, os utilizadores debatem sob a forma de discussões em *threads*, onde há um fluxo de informação geralmente direcionado para o tópico em questão enquanto as discussões nas redes sociais sobre um determinado tópico são mais dispersas (ALDayel & Magdy, 2021).

ALDayel & Magdy (2021) veem-se como os pioneiros a identificar e acoplar os vários estudos sobre o tema de deteção de posicionamento em redes sociais. No seu estudo de 2021, exploram os métodos utilizados para *stance detection* na bibliografia existente à data, descrevendo as principais abordagens e as suas aplicações.

## Métodos de *Machine Learning* utilizados (adaptado de ALDayel & Magdy (2021))

1. ***Supervised Learning***: Segundo os autores, este é o método mais comum, onde um conjunto de dados rotulado é utilizado para treinar algoritmos como SVMs, Naive Bayes e modelos de *deep learning* (como RNNs e LSTMs). Por exemplo, no SemEval 2016 (workshop anual de NLP), modelos baseados em LSTMs (Long Short-Term Memories) obtiveram resultados superiores, com o F1-score a alcançar 72.3% (Li & Caragea, 2019).
2. ***Unconstrained Supervised Learning***: Técnicas como *transfer learning* e *weak supervision* são usadas para lidar com a escassez de dados rotulados. Por exemplo, o uso de modelos baseados em BERT demonstrou ser altamente eficaz para tarefas de detecção de posicionamento em conjuntos de dados como o SemEval 2016 e artigos de notícias online (Ghosh et al., 2019).
3. ***Unsupervised Learning***: Métodos baseados em *clustering* têm sido utilizados para detetar posicionamentos, especialmente em cenários onde não há dados rotulados disponíveis. Técnicas como DBSCAN foram aplicadas para criar partições iniciais de posicionamento, mostrando desempenho competitivo em certos casos (Darwish et al., 2020).

Em estudos como o de Hasan & Ng (2013), os modelos utilizados incluem abordagens binárias, de sequência e de nível fino. Modelos binários, como o *Naive Bayes* e as *Support Vector Machines* (SVM) classificam cada publicação de forma independente, enquanto os modelos de sequência, como os *Hidden Markov Models* (HMM) e os *Conditional Random Fields* (CRF), capturam a dependência entre publicações sequenciais em debates. Modelos de nível fino vão além dos anteriores, atribuindo rótulos de posicionamento tanto ao nível do documento quanto ao nível da frase, permitindo uma análise mais granular.

No estudo de Anand et al. (2011), utilizam-se abordagens supervisionadas e características extraídas do texto e do contexto das interações. Os principais métodos incluem classificadores, como o *Naive Bayes* e o JRip (um classificador baseado em regras). As características exploradas dividem-se em diversas categorias. Entre as textuais, destacam-se os n-grams (unigramas e bigramas) e as *cue words*, como sequências iniciais de palavras que indicam o tipo de resposta. No âmbito das características sintáticas, foram utilizadas dependências gramaticais extraídas por meio do *Stanford Parser* e dependências generalizadas, que consideram classes gramaticais

e polaridade de palavras de opinião. As características semânticas incluem o uso do LIWC (*Linguistic Inquiry and Word Count*), que fornece categorias conceituais, como emoções positivas e negativas. Por fim, as características contextuais modelam as relações entre *posts*, aproveitando informações do *post* anterior para capturar a dinâmica dialógica do debate. Os resultados mostram uma precisão variando entre 54% e 69%, dependendo do tópico e das características utilizadas.

No estudo de Mohammad et al. (2016), os autores utilizaram diversos métodos e abordagens para a detecção de posicionamentos em *tweets*. A tarefa foi dividida em duas categorias principais: uma abordagem supervisionada e uma abordagem *weakly supervised*. Os métodos supervisionados basearam-se em classificadores como SVMs, que foram amplamente utilizados devido à sua eficácia na classificação binária. Na abordagem *weakly supervised*, foram utilizados métodos baseados em dados não rotulados relacionados a um alvo específico. Em alguns casos foram implementadas *pipelines* multi-etapas que combinaram anotações automáticas baseadas em palavras-chave com modelos preditivos.

### **"Stance Detection on Social Media with Fine-Tuned Large Language Models" (Gül, Lebret e Aberer, 2024)**

É evidente que todos os estudos apresentados anteriormente representam exemplos relevantes de trabalhos desenvolvidos em contextos semelhantes ao desta dissertação. No entanto, um artigo fundamental para a análise é o de Gül, Lebret e Aberer (2024), intitulado "*Stance Detection on Social Media with Fine-Tuned Large Language Models*", uma vez que representa um caso de estudo, em tudo, muito similar ao abordado nesta dissertação. A análise detalhada a seguir tem como objetivo compreender e assimilar a estrutura e os procedimentos realizados pelos autores, no processo de analisar e detetar as posições dos utilizadores em publicações nas redes sociais, utilizando LLMs.

No trabalho explorado, os autores adotaram uma abordagem abrangente, que inclui a utilização de métodos de *fine-tuning*, *prompting* e o emprego de diferentes arquiteturas de modelos. Este processo visou otimizar a capacidade dos modelos de detetar posições de forma precisa e contextualizada.

### **Modelos Utilizados**

Os autores exploraram três modelos principais:

- **ChatGPT-1106 (GPT-3.5-turbo-1106):** Um modelo proprietário da OpenAI, cuja adaptação foi realizada através de *fine-tuning*, utilizando as diretrizes específicas da OpenAI para tarefas personalizadas.
- **LLaMa-2 (7B e 13B):** Modelos *open-source* desenvolvidos pela Meta AI, ajustados para tarefas de *stance detection* em diferentes tamanhos, com 7 e 13 mil milhões de parâmetros.
- **Mistral-7B:** Outro modelo *open-source*, otimizado para a mesma tarefa, oferecendo uma solução eficiente com menor número de parâmetros.

Estes modelos foram escolhidos devido à sua capacidade de compreender o contexto semântico em textos, mesmo em cenários complexos como os encontrados nas redes sociais.

### Conjuntos de Dados e Pré-Processamento

Os autores utilizaram três conjuntos de dados principais para treino e avaliação:

- 1 **SemEval-2016:** Este *dataset* abrange tópicos sociais e políticos, com posições classificadas como *Favor* (a favor), *Against* (contra) e *None* (neutro ou sem posição). Inclui temas como figuras políticas e questões sociais amplas, como o movimento feminista e as alterações climáticas.
- 2 **P-Stance:** Um conjunto de dados focado no domínio político, com tweets que expressam posições *Favor* ou *Against* em relação a figuras políticas, como Bernie Sanders, Donald Trump e Joe Biden.
- 3 **Twitter Stance Election 2020:** *Dataset* derivado de tweets relacionados com a eleição presidencial dos Estados Unidos em 2020, com posições classificadas como *Favor*, *Against* e *None*.

Antes do treino, os dados foram submetidos a um pré-processamento, incluindo limpeza textual, integração de dados, normalização e padronização, de forma a garantir consistência e adequação ao contexto da tarefa.

### Métodos de Treino

Os autores ajustaram os modelos utilizando abordagens específicas:

- **Fine-Tuning:** O ajuste dos modelos foi realizado para otimizar o desempenho em tarefas de *stance detection*. Foram utilizadas três épocas de treino para os *datasets* menores (SemEval-2016 e *Twitter Stance Election 2020*) e uma época para o maior *dataset* (P-Stance), de forma a evitar *overfitting*.
- **Low-Rank Adaptation (LoRA):** Esta técnica foi utilizada para ajustar partes específicas dos modelos, otimizando as consultas e os valores de *auto-attention*. Este método, integrado na *framework* Lit-GPT, permitiu melhorar a precisão e eficiência na tarefa de detecção de posições.

### Estratégias de Avaliação

Os modelos foram avaliados em três configurações principais:

- **Zero-Shot Learning:** Sem exemplos de treino, avaliando a capacidade do modelo de generalizar e responder com base no seu conhecimento pré-treinado.
- **Few-Shot Learning:** Utilizando um número reduzido de exemplos de treino, testando a capacidade do modelo para aprender rapidamente.
- **Fine-Tuning Completo:** Aplicando conjuntos de dados completos para otimizar o desempenho dos modelos.

### Prompts Estruturadas

Os autores utilizaram *prompts* detalhados para orientar os modelos na tarefa de detecção de posições. Por exemplo, para o SemEval-2016, a *prompt* incluía:

“Analisa o seguinte *tweet* no contexto do tema [tópico]. Considera significados implícitos para determinar a posição em relação ao alvo «[alvo]». As opções são: FAVOR, AGAINST, NONE.  
Tweet: [tweet]”

Para o P-Stance, a *prompt* focava-se no domínio político:

“Analisa o seguinte *tweet*, relacionado com o domínio político, para determinar a posição do autor em relação ao alvo «[alvo]». Escolhe entre FAVOR ou AGAINST.”

### Resultados

Os resultados indicaram que os modelos *fine-tuned* superaram consistentemente os métodos *baseline*. Em particular, o ChatGPT destacou-se como o modelo mais preciso, alcançando os melhores resultados na maioria dos testes. Os modelos LLaMa-2 e Mistral-7B, embora de menor dimensão, também demonstraram elevada eficiência após ajustes.

Todo este trabalho é de autoria de (Gül et al., 2024), sendo que a análise apresentada acima constitui uma revisão detalhada, uma vez que as ferramentas, métodos e objetivos do estudo são, em grande parte, similares aos abordados nesta dissertação.

### ***A Survey of Stance Detection on Social Media: New Directions and Perspectives (Zhang et al., 2024)***

Mais recentemente, Zhang et al. (2024) realizaram uma revisão abrangente da literatura na área de *stance detection* em redes sociais, reunindo os avanços mais relevantes e delineando tendências emergentes. Os autores identificam a evolução progressiva das abordagens, desde os modelos supervisionados clássicos — como SVMs, RNNs ou arquiteturas baseadas em BERT — até à adoção crescente de LLMs, como o GPT. Mais recentemente, Zhang et al. (2024) realizaram uma revisão abrangente da literatura na área de *stance detection* em redes sociais, reunindo os avanços mais relevantes e delineando tendências emergentes. Os autores identificam a evolução progressiva das abordagens, desde os modelos supervisionados clássicos — como SVMs, RNNs ou arquiteturas baseadas em BERT — até à adoção crescente de LLMs, como o GPT. Segundo os autores, os LLMs oferecem uma flexibilidade sem precedentes, permitindo que tarefas como a detecção de posicionamento sejam resolvidas com abordagens baseadas em *prompting*, mesmo em cenários *zero-shot* ou *few-shot*, sem necessidade de grandes quantidades de dados rotulados. Apesar do seu potencial, o artigo também sublinha desafios como a instabilidade em tarefas ambíguas, a sensibilidade aos *prompts* e a falta de transparência na tomada de decisão. Ainda assim, os autores consideram os LLMs como uma das direções mais promissoras para o futuro da detecção de posicionamento, sobretudo em contextos complexos e pouco estruturados como as redes sociais (Zhang et al., 2024). Esta visão alinha-se com a abordagem desta dissertação, que aposta na utilização de modelos de linguagem avançados como núcleo da sua arquitetura de análise argumentativa.

## 3 Design e Conceção da Solução

Após a definição do problema e a análise do estado da arte, este capítulo apresenta a conceção da solução desenvolvida para tentar responder aos objetivos delineados. A abordagem proposta assenta na construção de um protótipo de um sistema modular e automatizado, capaz de recolher, analisar, sintetizar e representar discussões e argumentos expressos por utilizadores em debates digitais, com recurso a *Large Language Models* e Grafos de Conhecimento. A conceção da solução teve como principais linhas orientadoras a modularidade e integração entre componentes, permitindo um fluxo de dados contínuo desde a recolha de dados até à apresentação final dos resultados da análise.

Neste capítulo, são apresentados os requisitos funcionais e não funcionais definidos, bem como a arquitetura geral do protótipo. São ainda apresentadas e descritas as tecnologias e ferramentas utilizadas. Por fim, é apresentada a estrutura e funcionamento da *pipeline* de processamento desenvolvida. Esta visão global serve de base à descrição detalhada do processo de desenvolvimento, que será explorada no capítulo seguinte.

### 3.1 Requisitos do Sistema

Um dos primeiros passos da etapa de conceção da solução passa pela definição de um conjunto de requisitos que assegurem o alinhamento do sistema com os objetivos estabelecidos, bem como a sua viabilidade técnica, de usabilidade e de desempenho.

Os requisitos são organizados em duas categorias distintas: requisitos funcionais, que descrevem os comportamentos específicos que o sistema deve apresentar; e requisitos não funcionais, que especificam critérios de qualidade relacionados com o desempenho e a usabilidade, entre outros aspetos transversais ao funcionamento do sistema. Os próximos subcapítulos listam e descrevem os dois conjuntos de requisitos criados.

### 3.1.1 Requisitos Funcionais

Os requisitos funcionais focam-se nas capacidades e funcionalidades específicas que o sistema deve garantir para cumprir o seu propósito central: a análise automatizada de discussões online, com extração, classificação, síntese e visualização de argumentos. A tabela seguinte (Tabela 2) descreve de forma clara os requisitos funcionais da solução:

Tabela 2 - Requisitos Funcionais do Sistema

ID	Título	Descrição
RF1	Recolher dados de uma discussão pública online	O sistema deve permitir a extração automática de títulos, <i>headers</i> , comentários e respostas de <i>threads</i> públicas.
RF2	Identificar o tópico central da discussão	O sistema deve identificar, de forma automática, o tema ou tópico principal em torno do qual a discussão se desenvolve.
RF3	Classificar a posição argumentativa dos comentários ( <i>stance detection</i> )	Cada comentário deve ser analisado e classificado como sendo “a favor”, “contra” ou “neutro” no contexto do tópico identificado.
RF4	Agrupar comentários com base na posição	Os comentários devem ser agrupados de acordo com a posição identificada, de forma a estruturar os argumentos em blocos coerentes.
RF5	Criar sumários representativos dos grupos de posicionamento	O sistema deve gerar resumos representativos dos principais argumentos utilizados em cada grupo de posição (a favor / contra / neutro).

RF6	<p>Extrair e classificar argumentos individuais de cada comentário</p>	<p>O sistema deve ser capaz de identificar e classificar automaticamente as proposições principais expressas em cada comentário.</p>
RF7	<p>Criar <i>clusters</i> de argumentos semanticamente semelhantes</p>	<p>Os argumentos extraídos e classificados devem ser agrupados com base na sua similaridade semântica, de forma a eliminar redundância e destacar ideias-chave.</p>
RF8	<p>Criar um pequeno sumário de cada <i>cluster</i> de argumentos</p>	<p>Depois de criados os agrupamentos de argumentos, o sistema deve sumarizar as proposições utilizadas de forma clara, para que, lendo apenas o sumário, seja possível compreender todo o grupo argumentativo.</p>
RF9	<p>Construir um grafo de conhecimento</p>	<p>O sistema deve criar uma representação gráfica das discussões analisadas, contendo os tópicos, comentários, argumentos, grupos de argumentos e as relações entre os elementos.</p>
RF10	<p>Apresentar os resultados de forma clara ao utilizador</p>	<p>O sistema deve disponibilizar ao utilizador todos os resultados da análise efetuada, incluindo sumários, argumentos classificados, e <i>clusters</i>.</p>
RF11	<p>Avaliar o desempenho dos principais módulos da <i>pipeline</i></p>	<p>O sistema deve conseguir avaliar, através de um LLM, o desempenho dos três módulos principais da <i>pipeline</i>: Extração de Argumentos, Classificação de Argumentos, e Criação de <i>Clusters</i>.</p>

### 3.1.2 Requisitos Não Funcionais

Os requisitos não funcionais focam-se na qualidade da experiência de utilização do sistema e visam assegurar que este, para além de funcional, seja eficiente, robusto e utilizável em contextos reais. Tendo em conta a complexidade do processamento textual e a natureza desregrada das interações nas redes sociais, foi atribuída especial importância à modularidade

da arquitetura, à interpretabilidade dos resultados gerados e à competência da solução ao lidar com volumes significativos de dados. Estes requisitos (Tabela 3) asseguram que o sistema não só cumpre os seus objetivos técnicos, mas também responde às expectativas de usabilidade e fiabilidade.

Tabela 3 - Requisitos Não Funcionais do Sistema

ID	Título	Descrição
RNF1	Usabilidade	A interface do sistema e apresentação dos resultados devem ser intuitivas, claras e acessíveis a utilizadores com diferentes níveis de literacia digital.
RNF2	Desempenho	O sistema deve ser capaz de lidar com discussões de grande dimensão sem comprometer o desempenho.
RNF3	Modularidade	A arquitetura deve ser modular, permitindo a substituição reconfiguração ou atualização independente de cada componente (ex.: troca de LLM, mudança na estratégia de agrupamento de argumentos), sem comprometer o funcionamento global da <i>pipeline</i> .
RNF4	Integração entre componentes	Deve existir uma comunicação eficiente entre os diferentes módulos da <i>pipeline</i> , garantindo fluidez no processamento.
RNF5	Escalabilidade	O sistema deve ter a capacidade de se adaptar a um aumento de funções de forma eficiente. Deve ser fácil de manter e atualizar, sem afetar a funcionalidade.
RNF6	Utilização de LLMs	O sistema deve integrar <i>Large Language Models</i> , capazes de executar as tarefas necessárias com elevado grau de precisão e eficiência.
RNF7	Bases de Dados de Grafos	O sistema deve ter capacidade de armazenar e gerir grandes quantidades de informação relacionadas com cada discussão analisada.
RNF8	Ética e Conformidade	O sistema deve respeitar princípios éticos fundamentais na forma como recolhe, processa, analisa e apresenta os

dados provenientes de redes sociais. Isto inclui o compromisso com a transparência no funcionamento dos modelos utilizados, a preservação da privacidade dos utilizadores cujos conteúdos são analisados (mesmo que esses dados sejam publicamente acessíveis), a não distorção dos argumentos originalmente expressos e a promoção de uma análise imparcial e não tendenciosa. Além disso, a solução deve evitar a propagação de estereótipos, enviesamentos ou interpretações possivelmente manipuladoras dos dados.

---

## **3.2 Tecnologias e Ferramentas Utilizadas**

O desenvolvimento da solução baseou-se na integração de diversas ferramentas e tecnologias que, em conjunto, possibilitam a recolha, análise, síntese e visualização de argumentos extraídos de discussões ocorridas em redes sociais. A seleção destas tecnologias teve por base os requisitos funcionais e não funcionais definidos, com especial foco na modularidade, desempenho e facilidade de integração. Nesta secção são apresentadas e justificadas as principais tecnologias utilizadas.

### **3.2.1 Reddit API**

O processo de escolha da fonte de dados a utilizar passou por um estudo inicial de duas plataformas sociais de elevada relevância no contexto de discussões online: o X (antigo Twitter) e o Reddit. Ambas se destacam por permitir discursos irrestritos e são o exemplo perfeito de plataformas em que os utilizadores se envolvem em debates abertos, informais e frequentemente acalorados, sobre os mais variados tópicos - especialmente políticos - tornando-se, assim, fontes privilegiadas para a recolha de discursos argumentativos espontâneos. No entanto, a escolha recaiu sobre o Reddit, motivada por fatores tanto técnicos como qualitativos.

A API do Reddit acabou por ser utilizada como fonte principal de dados textuais, permitindo o acesso facilitado a *threads* públicas. A principal razão que motivou esta escolha prende-se com

a facilidade de extração de conteúdo, uma vez que o Reddit oferece formas diretas e simples de obter *threads* completas. Ainda assim, a diversidade temática global e o volume de conteúdo gerado nas discussões foram fatores determinantes na decisão de utilizar esta plataforma em detrimento de outras redes sociais potencialmente mais conhecidas pelo público geral. Para além destes fatores, as *threads* apresentam uma estrutura hierárquica clara, o que é deveras importante num debate com múltiplos intervenientes e opiniões. É relevante referir que se atribui a designação de “comentário” (*comment*) às respostas feitas diretamente ao *post* original e a designação de “resposta” (*reply*) às respostas aos comentários, mantendo evidente a estrutura hierárquica da *thread*. Abaixo (Figura 6) apresenta-se um exemplo ilustrativo da estrutura de uma *thread* do Reddit:

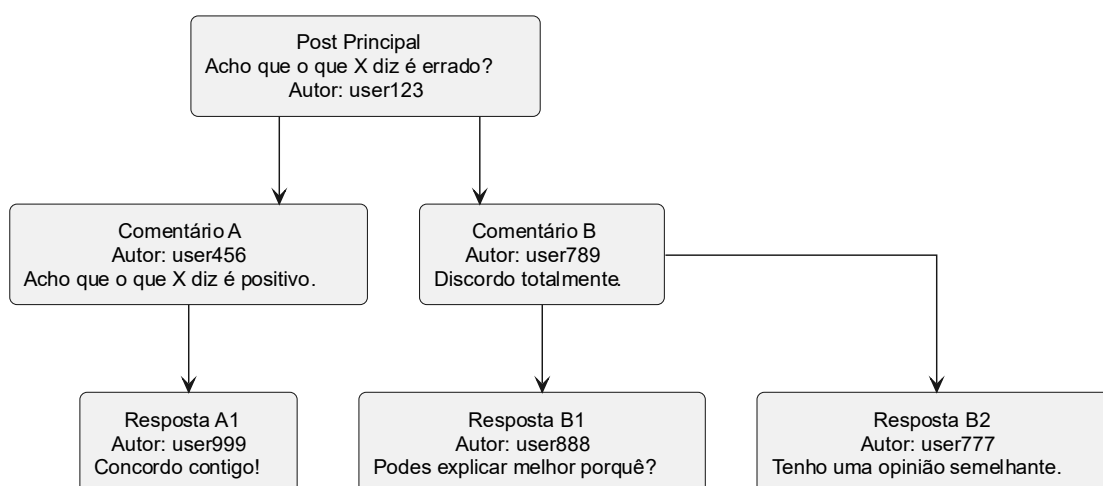


Figura 6 - Exemplo da Estrutura de uma *Thread* do Reddit

O Reddit também conta com um sistema de *upvotes*, semelhante ao sistema de *likes* adotado por algumas redes sociais mais comuns (Instagram, Facebook, X), que permite aos utilizadores expressar concordância ou relevância de publicações e comentários, influenciando a sua visibilidade ao ordenar os conteúdos com base na pontuação obtida.

No que toca a debates de teor político, o Reddit conta com dois *subreddits* (secções temáticas dedicadas a assuntos específicos dentro da plataforma) quase inteiramente dedicados a estas temáticas, com milhões de utilizadores e dezenas de novas discussões geradas todos os dias: o *r/PoliticalDiscussion* e o *r/ChangeMyView*. Estes dois espaços revelaram-se particularmente úteis, uma vez que todos os testes e validações realizados ao sistema foram conduzidos com base em dados extraídos a partir destas comunidades. No entanto, o *r/ChangeMyView* destacou-se como a melhor fonte de dados possível para os objetivos deste projeto, uma vez

que os *posts* publicados nesta comunidade seguem uma estrutura explícita: o autor do tópico apresenta, logo à partida, uma opinião pessoal bem definida sobre um determinado tema, abrindo espaço para que os restantes utilizadores respondam com argumentos contrários ou de apoio. Esta dinâmica binária, entre concordância e oposição fundamentada, revela-se particularmente adequada à lógica do sistema desenvolvido, que se baseia na identificação do tópico em debate e na classificação das posições dos comentários como “a favor”, “contra” ou “neutras”. Já no caso do *r/PoliticalDiscussion*, embora também se verifiquem interações ricas do ponto de vista discursivo, o formato das publicações é mais aberto e frequentemente orientado por perguntas ou temas genéricos, sem que o autor do tópico ou mesmo dos comentários assumam uma posição clara. Esta ausência de uma *stance* inicial dificulta a contextualização e classificação dos comentários subsequentes, tornando este subreddit menos apropriado para os objetivos específicos desta dissertação.

### **3.2.2 OpenAI GPT (via API)**

Para as tarefas de identificação de tópico, *stance detection*, criação de sumários, extração, classificação e agrupamento de argumentos, foram utilizados LLMs providenciados pela OpenAI, através da sua API oficial. A escolha destes modelos justifica-se pela sua elevada capacidade de compreensão de linguagem natural, contextualização semântica profunda, e flexibilidade na adaptação a tarefas específicas via *prompting*. O uso da API permite também garantir escalabilidade e melhor performance, já que não depende da capacidade de desempenho da máquina local.

### **3.2.3 LangChain**

A biblioteca LangChain foi utilizada para orquestrar chamadas à API dos LLMs, permitindo criar *pipelines* estruturadas de interações entre os modelos e os dados. Esta ferramenta simplifica a construção de cadeias de raciocínio, gestão de contexto, *templates* de *prompts* e integrações com outras fontes, como as bases de dados de grafos. A sua utilização foi essencial para garantir modularidade e reutilizabilidade na implementação das tarefas de processamento linguístico.

### 3.2.4 Neo4j

Para a representação estruturada dos dados processados, foi utilizada a base de dados de grafos Neo4j. Esta tecnologia permite modelar relações complexas entre entidades como tópicos, comentários, argumentos e agrupamentos, proporcionando consultas eficientes e visualizações interativas. A escolha do Neo4j justifica-se pela sua adequação à representação de dados relacionais, pelo suporte à linguagem *Cypher*, e pelo fornecimento de ferramentas de visualização integradas de fácil utilização (Neo4j Aura).

### 3.2.5 FastAPI

A *framework* FastAPI, foi utilizada para o desenvolvimento da API de *backend* responsável pela coordenação entre os módulos da *pipeline*. Esta tecnologia foi escolhida pela sua elevada simplicidade de implementação, performance e compatibilidade nativa com modelos assíncronos. A FastAPI permite, ainda, a construção de *endpoints* REST de forma modular, tornando o sistema mais escalável e fácil de manter.

### 3.2.6 Streamlit

A interface de apresentação e exploração dos resultados foi desenvolvida com recurso à biblioteca Streamlit, que permite a criação de *User Interfaces* (UI's) interativas em Python de forma rápida e intuitiva. O Streamlit foi escolhido pela sua simplicidade, integração direta com *pandas* e por não requerer conhecimentos avançados de desenvolvimento frontend.

### 3.2.7 Bibliotecas de Suporte

Adicionalmente, foram utilizadas bibliotecas auxiliares como:

- **Pandas** – Biblioteca Python utilizada para manipulação e análise de dados estruturados, como tabelas ou *dataframes*.
- **Pydantic** – Utilizada para validação e tipagem de dados nos modelos da API backend, em articulação com o FastAPI.
- **Uvicorn** – Servidor assíncrono utilizado para executar a aplicação FastAPI.

- **Requests** – Biblioteca de comunicações HTTP utilizada para aceder a dados de APIs externas.
- **re (expressões regulares)** – Utilizado para pré-processamento textual.

### 3.3 Arquitetura Geral do Sistema

A arquitetura geral do sistema proposto reflete uma abordagem modular, orientada à separação de responsabilidades entre os diferentes componentes que integram a solução. Esta organização visa garantir a escalabilidade, flexibilidade e facilidade de manutenção do sistema, tal como é referido acima na identificação dos requisitos não funcionais (RNF3, RNF4 e RNF5), enquanto permite a integração eficiente de tecnologias avançadas, como *Large Language Models* (RNF6) e bases de dados de grafos (RNF7).

A solução encontra-se dividida em duas grandes camadas principais: o *backend*, responsável pela extração, processamento e análise da informação; e o *frontend*, responsável pela visualização e interação com os resultados e grafos criados. Os componentes internos do *backend* foram concebidos na forma de uma *pipeline* sequencial e funcional, garantindo uma passagem de dados coerente e lógica entre todas as etapas. Adicionalmente, são utilizados componentes externos fundamentais ao funcionamento do sistema, como a API do Reddit, a API da OpenAI e a API do Neo4j. Na Figura 7 está representado o diagrama de componentes do sistema construído, ilustrando a sua arquitetura geral e representando os módulos principais e as interfaces utilizadas, bem como os fluxos de informação entre os diversos elementos do sistema.

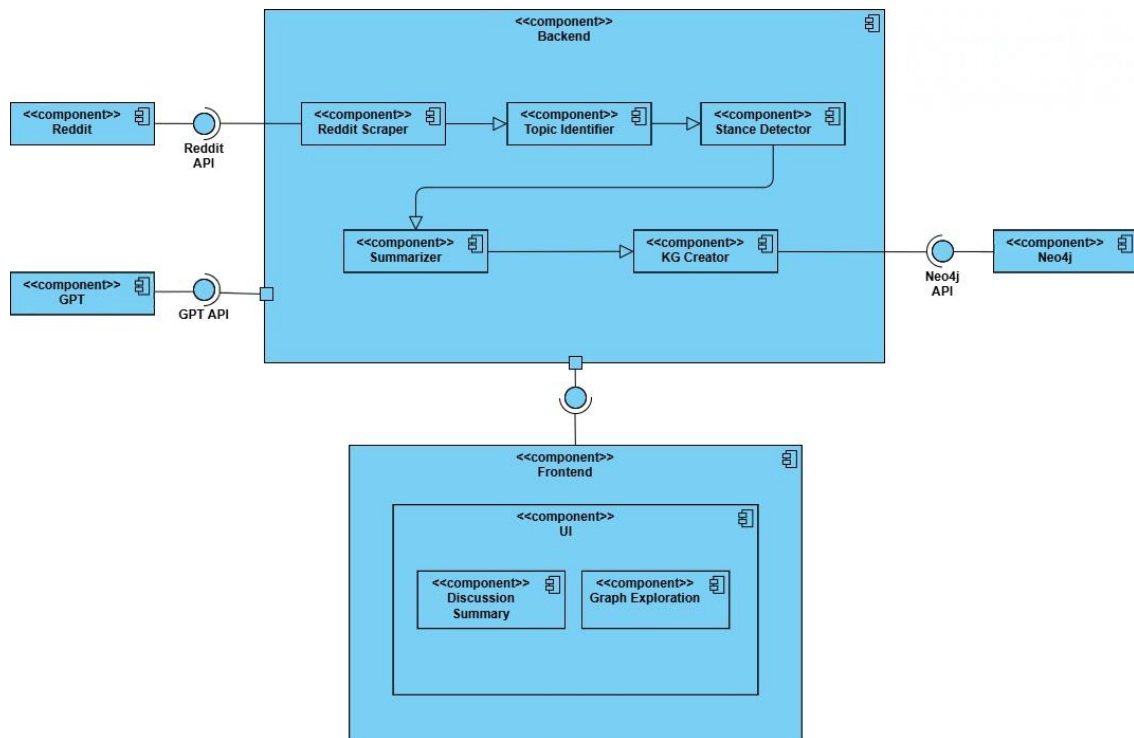


Figura 7 – Diagrama de Componentes do Sistema

### 3.3.1 Descrição dos Componentes

A arquitetura da solução encontra-se estruturada em torno de três grupos principais de componentes: componentes de *backend*, *frontend* e APIs externas. Cada um destes grupos desempenha um papel específico no fluxo de execução da *pipeline* e demonstração de resultados, assegurando todos os requisitos referidos anteriormente. Esta secção descreve detalhadamente cada grupo e os seus componentes, explicando a função de cada um no sistema.

#### Componentes de *Backend*

A camada de *backend* do sistema é responsável por processar e transformar os dados recolhidos, aplicando várias etapas de análise semântica e estrutural até à representação final em forma de grafo de conhecimento. Os seus principais componentes são os seguintes:

- **Recolha de Dados (*Reddit Scraper*):** Responsável por estabelecer ligação com a API pública do Reddit e extrair os dados relevantes de uma *thread* de discussão, incluindo o título, *header*, comentários e respetivas respostas. Esta componente constitui o ponto de entrada da *pipeline*.

- **Identificador de Tópico (*Topic Identifier*):** Responsável por determinar automaticamente o tópico central da discussão e gerar texto explicativo sobre o que significa ter uma posição “contra” ou “a favor” no contexto desse tema.
- **Classificador de Posição Argumentativa (*Stance Detector*):** Componente responsável por analisar cada comentário individualmente e classificar a sua posição face ao tópico como “a favor”, “contra” ou “neutro”.
- **Sumarizador (*Summarizer*):** Responsável por gerar sumários que representem de forma clara as principais proposições defendidas em cada grupo.
- **Criador de Grafos de Conhecimento (*KG Creator*):**

Responsável por quatro funções essenciais distintas:

1. Extrair os argumentos centrais expressos em cada comentário.
2. Classificar individualmente a posição (“a favor”, “contra” ou “neutro”) de cada argumento extraído.
3. Agrupar os argumentos semanticamente equivalentes, criando, para cada agrupamento, um pequeno sumário.
4. Construir e armazenar, a partir dos dados processados, um grafo de conhecimento em que se representam as entidades envolvidas (tópico, comentários, respostas, argumentos e grupos de argumentos). A base de dados de grafos Neo4j é utilizada para armazenar estes grafos.

### **Componentes de *Frontend***

O *frontend* do sistema é composto por duas interfaces interligadas que permitem ao utilizador aceder, explorar e interpretar os resultados da análise argumentativa. Para além da função de apresentação dos dados processados, que proporciona uma experiência interativa e informativa, o *frontend* serve também como ponto de entrada para a execução da *pipeline*.

- **Página inicial - *Discussion Summary*:**

Esta interface constitui o ponto de entrada da aplicação e, nesta fase, apenas é pedido ao utilizador que introduza o *link* da discussão do Reddit que pretenda analisar. Após a introdução da ligação à discussão, e uma vez concluído o processamento, é apresentado ao utilizador um sumário geral da mesma. Nesta página são exibidos:

1. O tópico principal identificado;

2. A explicação gerada sobre o significado de estar “a favor” ou “contra” relativamente a esse tópico;
3. Os sumários argumentativos correspondentes a cada grupo de posição. Esta interface fornece uma visão global e simplificada do debate, permitindo compreender os principais pontos em disputa sem necessidade de leitura extensiva.

- **Segunda página - *Graph Exploration*:**

Esta segunda interface destina-se a uma análise mais aprofundada da discussão. Através da integração com a API do Neo4j, o utilizador pode explorar consultas (*queries*) pré-definidas que permitem o acesso a informações detalhadas, tais como a listagem dos argumentos por *stance*, os grupos de argumentos, etc. Esta funcionalidade oferece uma perspetiva mais granular e estruturada das discussões processadas, permitindo análises minuciosas.

### **3.4 Funcionamento da *Pipeline* de Processamento**

A *pipeline* desenvolvida percorre todas as etapas do ciclo de vida da informação, desde a recolha de dados brutos até à apresentação visual dos argumentos estruturados e da rede de conhecimento que os interliga. Trata-se de um processo sequencial e modular, no qual cada componente da arquitetura contribui para transformar a linguagem natural informal presente nas discussões, em conhecimento estruturado e explorável. Do ponto de vista conceptual, a *pipeline* pode ser segmentada em duas grandes fases: uma primeira, centrada na criação de um sumário geral da discussão, e uma segunda, focada na geração do grafo de conhecimento, onde se analisam com maior profundidade os elementos argumentativos.

#### **1. Recolha e Pré-processamento dos Dados**

O ponto de partida é a extração de dados provenientes de uma discussão pública da plataforma Reddit, estruturada em formato de *thread*. Após a recolha dos dados da *thread*, o sistema realiza a identificação do tema central da discussão através de um *Large Language Model*. Para além disto, o modelo fornece ainda duas breves explicações do que seria, nesse contexto, estar "a favor" ou "contra" o tópico, detalhando o que significa estar de um lado ou de outro do debate. Esta etapa é crucial para contextualizar corretamente os comentários subsequentes e garantir que a análise interpretativa posterior é feita em função do tópico em causa.

## **2. Classificação do Posicionamento Argumentativo**

Nesta fase, os comentários e respostas são submetidos a outro LLM para uma etapa de classificação de posicionamento (*stance classification*), que distingue cada intervenção com base na sua orientação relativamente ao tópico identificado — “a favor”, “contra” ou “neutro”. Esta distinção permite dividir a discussão em blocos argumentativos coerentes, facilitando a organização posterior dos dados.

## **3. Sumarização das Posições**

Com base nesta categorização, outro modelo generativo elabora sumários expositivos dos argumentos defendidos por cada grupo de posicionamento. O objetivo é eliminar repetições, identificar os pontos principais de cada lado do debate e apresentar o conteúdo de forma mais clara, acessível e informativa. Esta fase é particularmente relevante para utilizadores que pretendam compreender rapidamente as ideias-chave sem precisar de ler todos os comentários.

## **4. Extração e Classificação de Argumentos**

Estas fases seguintes revelam-se as mais importantes de toda a *pipeline* de processamento já que levam à criação detalhada dos grafos de conhecimento, o que permite a análise minuciosa de cada discussão. Nesta etapa, cada comentário e resposta é analisado por um LLM para extrair os argumentos centrais de forma automática. Os argumentos extraídos consistem em proposições que expressam afirmações justificadas, críticas fundamentadas ou posições baseadas em factos. Em seguida, cada argumento é submetido a uma segunda instância do modelo, dedicada exclusivamente à classificação do seu posicionamento relativamente ao tópico, agora de forma individualizada para permitir a máxima veracidade nas classificações. Esta abordagem permite capturar com maior precisão a polaridade de cada proposição e evita os problemas da atribuição de um único rótulo a blocos textuais de maior dimensão (comentários e respostas) que, por vezes, contêm argumentos nos dois lados do espectro.

## **5. Agrupamento ou *Clustering* de Argumentos**

Os argumentos extraídos são então agrupados por outro modelo, com base na sua similaridade semântica. Este deteta ideias equivalentes ou complementares, ainda que formuladas de forma distinta, e organiza-as em grupos temáticos. Cada grupo recebe uma descrição detalhada e autoexplicativa que representa a ideia comum expressa pelos seus elementos. Esta estratégia

permite eliminar redundâncias, destacar os núcleos argumentativos principais de cada posição e oferecer uma visão condensada e inteligível do debate.

## **6. Construção do Grafo de Conhecimento**

Com todos estes elementos identificados e organizados, procede-se à criação de um grafo de conhecimento, onde se representam explicitamente as entidades envolvidas:

- Tópico da discussão
- Comentários
- Respostas
- Argumentos
- Grupos de Argumentos

Este conjunto de entidades, complementado com as relações que as ligam e os atributos que os compõem, criam uma estrutura gráfica que evidencia a lógica interna do debate, permitindo mapear relações de apoio, oposição ou neutralidade entre argumentos e explorar a dinâmica discursiva com maior profundidade. A escolha de utilização de uma base de dados de grafos, em detrimento de uma base de dados relacional tradicional, justifica-se pela natureza altamente interligada e contextual dos dados gerados. Ao tornar explícitas as ligações entre os vários elementos do discurso, o grafo não é apenas uma camada de armazenamento, mas sim uma ferramenta analítica fundamental para a exploração e compreensão da discussão.

## **7. Visualização e Exploração dos Resultados**

Por fim, todos os resultados são apresentados ao utilizador através de uma interface clara e interativa, que permite navegar pelos comentários, consultar os argumentos identificados e classificados, explorar os *clusters* criados, entre outros. A interface disponibiliza métricas, filtros e resumos que facilitam o acesso à informação mais relevante, promovendo uma leitura crítica e uma visão mais estruturada do debate em análise.

A Figura 8 apresenta uma visão geral da estrutura da *pipeline* desenvolvida, evidenciando as principais fases do processo e a forma como se articulam ao longo do sistema.

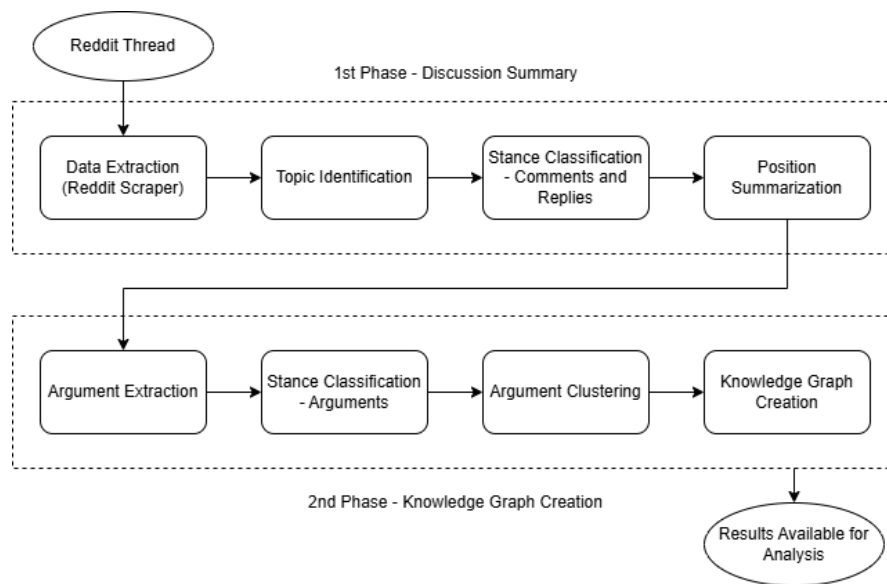


Figura 8 - Pipeline de Processamento

### 3.4.1 Avaliação da Pipeline

Para garantir a eficácia do sistema foi desenvolvida uma simples *framework* de avaliação automática, baseada num modelo generativo, que permite analisar criticamente o desempenho do sistema nas tarefas de extração, classificação de argumentos e *clustering*. Esta abordagem recorre ao mesmo tipo de modelo utilizado na geração inicial dos resultados, tirando partido da sua capacidade de realizar avaliações linguísticas, como a verificação da coerência, relevância e consistência argumentativa dos outputs produzidos.

Apesar de assentar na mesma arquitetura de linguagem, a utilização do mesmo modelo utilizado nas tarefas enquanto avaliador, justifica-se pela sua elevada competência na análise textual contextualizada, sendo capaz de emitir juízos críticos sobre conteúdos complexos quando devidamente instruído através de *prompt engineering*. A verdade é que para um humano conseguir avaliar tarefas como a extração de argumentos de texto informal, este necessita de um vasto conhecimento de causa e do tópico e contexto específico, e neste sentido, o modelo utilizado consegue espelhar e raciocinar como um humano, sendo competente e proficiente neste encargo.

Reconhece-se, contudo, que esta solução não é absolutamente fiável e poderá incorrer em enviesamentos ou limitações na objetividade da avaliação, dado que o modelo pode reproduzir falhas ou interpretações semelhantes às geradas durante o processo inicial. Por esse motivo, os

resultados obtidos devem ser encarados como indicadores qualitativos complementares, úteis para identificar padrões, inconsistências ou limitações, mas não como uma validação absoluta ou substituto total de avaliação humana.

Esta *framework* de avaliação aplica uma lógica de avaliação estruturada, fornecendo tanto um diagnóstico global como análises individuais por comentário, argumento e grupo de argumentos, permitindo detetar padrões de erro, inconsistências e limitações do modelo com maior precisão.

## 4 Implementação da Solução

Este capítulo descreve, de forma detalhada, todos os passos que fizeram parte do processo de desenvolvimento da *pipeline* e da solução proposta para a possível resolução do problema apresentado anteriormente. São ainda detalhadas as principais etapas do processo de implementação, abordando abordagens anteriores e tentativas experimentais, e com especial foco na lógica de funcionamento de cada módulo, nas decisões técnicas tomadas durante a construção do sistema, nos desafios encontrados e nas soluções adotadas para garantir a eficácia e robustez da aplicação. A articulação entre componentes e a gestão dos dados ao longo da *pipeline* são igualmente abordadas, destacando-se a forma como o sistema assegura a fluidez do processamento, desde a recolha dos dados até à visualização final dos argumentos e do grafo de conhecimento.

### 4.1 Abordagens e Tentativas Experimentais

O desenvolvimento do sistema passou por várias fases exploratórias e experimentais que foram fundamentais para refinar e atingir a abordagem final adotada. Antes de se alcançar uma arquitetura estável e eficaz, foram testadas diferentes estratégias para tarefas-chave como a identificação de tópicos, extração de argumentos, construção de grafos e agrupamento semântico. Esta secção descreve essas tentativas, os obstáculos enfrentados e as razões que motivaram as suas alterações.

#### 4.1.1 Modelos *Open-Source*: Tentativas Iniciais com Modelos *Pre-trained*

A primeira abordagem explorada consistiu na utilização de modelos já treinados, com o objetivo de realizar tarefas específicas, como a identificação de tópicos e *stance classification*. Estes dois módulos específicos, sendo introdutórios na *pipeline*, foram os que mais sofreram alterações e foram alvo de testes com modelos *open-source*, todos retirados da plataforma Hugging Face. Para a identificação do tópico da discussão foram testados modelos como o “cardiffnlp/tweet-topic-latest-multi”<sup>8</sup>, que se trata de um modelo “RoBERTa-base treinado com 168,86 milhões de tweets e posteriormente ajustado (*fine-tuned*) para classificação *multi-label* de tópicos, com base num corpus de 11.267 tweets” (Antypas et al., 2022). Este modelo, para além de ter a desvantagem de correr localmente, limitava os tópicos identificados a 19 *labels* diferentes que, apesar de variadas e abrangentes, não eram significativamente detalhadas e claras como pretendido.

No caso de *stance classification*, são mais facilmente encontrados modelos *fine-tuned* para esta tarefa, no entanto, modelos como o “classla/bcms-bertic-parlasent-bcs-bi”<sup>9</sup> fazem classificação binária – “negativo” ou “outro” – sendo que outros classificam como “a favor” ou “contra”. Como se pretendia classificar argumentos como neutros para filtrar proposições que não ocupem um lado específico, estes modelos foram excluídos. Quanto a modelos de classificação ternária como o “classla/bcms-bertic-parlasent-bcs-ter”<sup>10</sup> ou o “anhuu/argument\_classification\_stance\_topictexttarget\_bert”<sup>11</sup>, que classificam texto como “negativo”, “neutro” ou “positivo”, estes foram testados e excluídos devido à baixa performance das classificações.

Todos estes modelos testados apresentaram um desempenho claramente insatisfatório, com resultados irrelevantes e mal contextualizados. Ainda, a falta de flexibilidade na adaptação ao domínio específico das discussões e a dificuldade em controlar a forma de resposta tornaram-nos inviáveis para o propósito da solução.

---

<sup>8</sup> <https://huggingface.co/cardiffnlp/tweet-topic-latest-multi>

<sup>9</sup> <https://huggingface.co/classla/bcms-bertic-parlasent-bcs-bi>

<sup>10</sup> <https://huggingface.co/classla/bcms-bertic-parlasent-bcs-ter>

<sup>11</sup> [https://huggingface.co/anhuu/argument\\_classification\\_stance\\_topictexttarget\\_bert](https://huggingface.co/anhuu/argument_classification_stance_topictexttarget_bert)

## Modelos Open Source: Fine-tuning de Modelos Locais

Na sequência destes resultados, tentou-se outra alternativa mais elaborada. Primeiramente, foi realizada uma tentativa de *fine-tuning* de um modelo BART – “facebook/bart-large-mnli”<sup>12</sup> de 406 milhões de parâmetros – utilizando um conjunto de dados com cerca de 2500 exemplos. Segundo a descrição do modelo, este “revela-se particularmente eficaz quando é *fine-tuned* para tarefas de geração de texto (por exemplo, sumarização ou tradução), mas também apresenta um bom desempenho em tarefas de compreensão textual (como classificação de texto ou resposta a perguntas)” (Lewis et al., 2019). Por este motivo, considerou-se que seria uma escolha adequada para a tarefa de identificação de tópico.

O *dataset* utilizado - *SemEval-2016 Task 6*<sup>13</sup> - corresponde a um conjunto de dados anotados manualmente, construído para a tarefa de *stance detection* em debates online. Este inclui *tweets* relacionados com tópicos controversos, como por exemplo "Legalização da Marijuana" ou "Vacinação" e cada entrada é classificada consoante a sua posição relativamente ao tópico – “FAVOR”, “AGAINST” ou “NONE”.

Embora este primeiro processo de *fine-tuning* tenha corrido tecnicamente sem falhas, durando cerca de sete horas na máquina utilizada, os resultados obtidos foram medianos. O modelo não era capaz de produzir tópicos de forma limpa, e o seu formato de saída revelou-se difícil de controlar apenas por *prompt engineering*. Além disso, o tempo de inferência era desproporcionado face à simplicidade da tarefa.

Numa segunda tentativa, procedeu-se a um novo *fine-tuning* sobre o modelo já ajustado anteriormente, utilizando um novo *dataset* – IBM Argument Quality Ranking<sup>14</sup>. Este foi desenvolvido pela IBM no contexto de esforços para treinar modelos em tarefas relacionadas com extração e classificação de argumentos. Este conjunto de dados apresenta frases isoladas extraídas de fontes argumentativas como artigos, que são rotuladas consoante a sua posição relativamente a um tópico específico. As anotações possíveis são: 1 (*pro*), -1 (*con*) e 0 (*neutral*), representando, respetivamente, posições a favor, contra ou neutras em relação à proposição analisada. Esta segunda abordagem resultou numa degradação de desempenho ainda maior, servindo assim para chegar à conclusão de que múltiplos *fine-tunings* consecutivos sem uma

---

<sup>12</sup> <https://huggingface.co/facebook/bart-large-mnli>

<sup>13</sup> <https://huggingface.co/datasets/krishnagarg09/SemEval2016Task6>

<sup>14</sup> [https://huggingface.co/datasets/ibm-research/argument\\_quality\\_ranking\\_30k](https://huggingface.co/datasets/ibm-research/argument_quality_ranking_30k)

estratégia de controlo adequada comprometem a estabilidade do modelo, fazendo com que este se "esqueça" de aprendizagens anteriores e gere resultados inconsistentes.

Numa última tentativa de *fine-tuning* foi testado o modelo "microsoft/deberta-v3-large"<sup>15</sup>, utilizando um *dataset* mais refinado (com cerca de 17 mil linhas), sendo uma combinação de dados dos dois utilizados anteriormente. Devido à dimensão do *dataset*, o *finetuning* demorou ainda mais do que os antecedentes e a *performance* do modelo ficou novamente aquém do esperado.

Ainda numa perspetiva de exploração local, foi considerada a utilização de modelos locais através de ferramentas como o Ollama. No entanto, os tempos de treino e reposta encontrados superaram os benefícios esperados.

#### **Adoção de Modelos via API (OpenAI)**

Após estas tentativas, foi adotada uma abordagem baseada em modelos disponibilizados via API pela OpenAI. Esta escolha trouxe inúmeras vantagens: tempos de resposta rápidos, elevada qualidade linguística, capacidade de generalização e flexibilidade na resposta, sobretudo quando combinada com técnicas de *prompt engineering*. Esta solução permitiu concentrar esforços no desenho das *prompts* e na lógica do sistema, ao invés de investir em infraestrutura de treino local.

#### **4.1.2 Tentativas de Construção Automática do Grafo via LLM**

No que toca à criação dos grafos de conhecimento dos debates, inicialmente, ponderou-se também recorrer aos modelos generativos para gerar automaticamente as *queries* Cypher responsáveis pela criação do grafo no Neo4j. No entanto, esta abordagem revelou-se instável: os grafos criados tinham erros de estrutura, relações ausentes ou mal definidas, e não respeitavam a lógica hierárquica necessária à organização dos dados. Mesmo com múltiplos ajustes de *prompt*, o modelo não conseguia manter consistência na criação das estruturas. Por essa razão, a extração, classificação e agrupamento de argumentos foram mantidas a cargo do LLM, enquanto a criação do grafo passou a ser feita diretamente, utilizando a driver oficial do Neo4j e *queries* explícitas geradas programaticamente.

---

<sup>15</sup> <https://huggingface.co/microsoft/deberta-v3-large>

### 4.1.3 Classificação de Comentários e Respostas ao Invés de Argumentos

Antes da adoção da classificação específica dos argumentos extraídos, foi inicialmente implementada uma abordagem mais direta, na qual a classificação de posicionamento era atribuída apenas a nível dos comentários e respostas. Nesta versão preliminar, cada comentário ou resposta era analisado como um todo, sendo classificado como "FOR", "AGAINST" ou "NEUTRAL" em relação ao tópico da discussão, tal como ainda acontece para a análise inicial. No entanto, depois de reflexão, chegou-se à conclusão que faria sentido classificar especificamente e individualmente cada argumento extraído.

Embora a estratégia anterior permitisse um fluxo e integração mais rápidos dos dados no grafo, revelaram-se várias limitações. Primeiramente, muitos comentários apresentam múltiplos argumentos com orientações distintas, o que tornava a classificação binária ou ternária de todo o texto imprecisa e enganadora. Finalmente, como o sistema tratava blocos textuais inteiros como unidades de análise, a granularidade necessária para tarefas como o agrupamento de argumentos ou análise semântica fina era mais difícil de alcançar. Estas limitações motivaram a transição para uma abordagem centrada numa segunda classificação, agora explícita e individualizada dos argumentos, o que trouxe ganhos significativos na precisão e valor analítico da solução.

### 4.1.4 Agrupamento de Argumentos – Primeiras Abordagens

No que diz respeito ao agrupamento de argumentos semelhantes, foram testadas diferentes abordagens, tentando chegar a uma que contribuísse para o sucesso esperado desta tarefa. Numa primeira abordagem, considerou-se a possibilidade de utilizar métricas de similaridade semântica como *cosine similarity* ou outras baseadas em *embeddings*. No entanto, já que o objetivo do projeto passa pela utilização de LLMs, optou-se por manter uma abordagem unificada e continuar a utilizar um modelo generativo. Assim, desenvolveu-se uma *prompt* que, numa única chamada, apresentava ao modelo todos os argumentos extraídos, independentemente da sua classificação de posicionamento, solicitando a criação de grupos semânticos coerentes.

Contudo, verificou-se que esta abordagem levava frequentemente a agrupamentos conceptuais que ignoravam a orientação dos argumentos, independentemente do modelo

utilizado, unindo, por exemplo, argumentos "a favor" e "neutros" apenas por partilharem um tema semelhante. Apesar de tentativas sucessivas de reforçar essa distinção via *prompt engineering*, o modelo continuava a valorizar a afinidade temática em detrimento da polaridade argumentativa, o que comprometia a integridade analítica dos resultados.

#### **Abordagem Final para *Clustering***

A solução passou por uma alteração simples, mas eficaz: os argumentos passaram a ser agrupados separadamente por classe de posicionamento. Cada conjunto (*FOR*, *AGAINST*, *NEUTRAL*) era processado individualmente, garantindo que o modelo agrupava apenas ideias dentro da mesma orientação. Esta abordagem permitiu preservar a coerência argumentativa de cada grupo e produzir agrupamentos mais consistentes e interpretáveis.

#### **4.1.5 Reflexão Final**

Estas tentativas demonstram os limites práticos de algumas ferramentas de inteligência artificial generativa, bem como a importância do processo iterativo de design. Embora os LLMs apresentem capacidades impressionantes, o seu desempenho é fortemente dependente do contexto, da formulação das *prompts* e das estratégias de integração. A solução final adotada representa o ponto de equilíbrio entre desempenho, robustez e controlo, e resultou diretamente das aprendizagens obtidas nas fases exploratórias.

## **4.2 Metodologia de Desenvolvimento**

Partindo de uma abordagem metodológica baseada em *Action Research*, procurou-se articular a investigação teórica com a intervenção prática, através de ciclos iterativos de planeamento, ação e reflexão.

A fase inicial consistiu na aquisição de competências técnicas (*hard skills*) essenciais ao desenvolvimento da solução proposta. Visto que o autor não possuía experiência prévia no desenvolvimento de sistemas baseados em LLMs ou Grafos de Conhecimento, foi necessário investir numa etapa de capacitação prática. Para esse fim, foram explorados recursos formativos disponibilizados por plataformas como a Databricks Academy e a Neo4j Graph

Academy, os quais forneceram os fundamentos técnicos necessários para a compreensão e aplicação das tecnologias envolvidas.

Com essa base adquirida, seguiu-se o desenvolvimento de pequenas aplicações experimentais, utilizando LLMs e Streamlit para realizar tarefas elementares de NLP, tais como tradução automática e sumarização de textos curtos, contando já com a presença de UI. Esta fase prática teve como objetivo consolidar os conhecimentos adquiridos e validar a viabilidade técnica da abordagem. Superada esta etapa preparatória, procedeu-se ao planeamento e desenho da arquitetura do que seria o sistema final a desenvolver, tendo como base os requisitos identificados e os objetivos delineados para a solução.

#### **4.2.1 Considerações Éticas e Conformidade Durante o Desenvolvimento**

Tendo em conta a natureza sensível dos dados analisados, nomeadamente textos opinativos produzidos por utilizadores em redes sociais, foram respeitados princípios fundamentais de ética e privacidade ao longo de todo o desenvolvimento do projeto. Todos os dados recolhidos provêm de discussões públicas no Reddit, uma plataforma onde os conteúdos são disponibilizados sob licenças abertas e acessíveis via API pública, e onde os utilizadores usam nomes de utilizador que, por padrão, não correspondem à sua identidade real. Para reforçar ainda mais o compromisso com a privacidade, o sistema foi desenhado de forma a não armazenar e omitir qualquer referência a nomes de utilizadores, mesmo os gerados automaticamente pela plataforma. Desta forma, garante-se que os comentários e respostas analisados são apresentados de forma totalmente anónima, não permitindo a identificação dos seus autores. Foram, então, seguidas as recomendações estabelecidas pelo Regulamento Geral sobre a Proteção de Dados, assegurando que nenhuma informação pessoal identificável é armazenada ou exposta durante o desenvolvimento e teste do sistema.

Adicionalmente, todo o processamento realizado teve como foco o conteúdo argumentativo e não o perfil dos utilizadores, sendo descrito e detalhado nos capítulos de desenvolvimento seguintes. Não foram feitas análises de traços pessoais, nem construídos perfis ou inferências que pudessem violar a privacidade dos indivíduos. Este cuidado reforça o compromisso do projeto com a minimização de risco ético e respeito pela autonomia dos utilizadores cujos textos foram analisados.

Por fim, todo o processamento dos dados utilizados foi documentado e as tecnologias utilizadas, incluindo os LLMs, foram sempre aplicadas em conformidade com os termos de utilização das APIs e licenças dos serviços envolvidos (como OpenAI, LangChain e Neo4j).

## 4.3 Desenvolvimento dos Diferentes Módulos

Como referido anteriormente, o desenvolvimento foi conduzido de forma modular, permitindo uma maior flexibilidade, reutilização de componentes e facilitação na deteção e resolução de problemas. Cada módulo foi projetado com responsabilidades bem definidas, alinhadas com os objetivos funcionais da solução: desde a recolha de dados até à sua análise, representação e visualização. Este subcapítulo descreve em detalhe o processo de implementação de cada um dos módulos que compõem a solução.

### 4.3.1 Arquitetura Base com FastAPI

Todos os módulos desenvolvidos no âmbito deste trabalho foram construídos em Python com base na *framework* FastAPI, ferramenta já apresentada. Esta *framework* facilita a comunicação entre módulos através de *endpoints* REST acessíveis por qualquer cliente HTTP, incluindo outros serviços internos da aplicação. Esta característica foi essencial para permitir a orquestração de tarefas entre os diversos módulos.

A estrutura de cada *endpoint* segue um padrão consistente: definição de modelos de entrada e saída com utilização do BaseModel do Pydantic, validação automática dos dados, tratamento de erros com códigos HTTP apropriados, e retorno de respostas JSON bem estruturadas. Esta base técnica comum é aplicada a todos os submódulos apresentados nas secções seguintes.

O ficheiro “main.py” (disponível no Apêndice A1 – Definição dos *Endpoints* do Sistema) constitui o ponto de entrada da aplicação FastAPI, centralizando a definição da instância da aplicação e dos seus diversos *endpoints*. Cada *endpoint* está associado a um componente específico da *pipeline*. Esta estrutura modular permite separar claramente as responsabilidades de cada componente, facilitando a manutenção e escalabilidade do sistema.

### 4.3.2 Extração e Recolha de Dados - *Reddit Scraper*

A recolha de dados constitui a base fundamental do sistema de análise. No que toca ao acesso facilitado dos dados, o Reddit disponibiliza uma funcionalidade muito útil para desenvolvedores que pretendam extrair dados de *threads*, ao permitir que qualquer URL de uma *thread* pública (Figura 9) seja convertido diretamente num ficheiro JSON apenas adicionando o sufixo “.json” ao final do URL (Figura 10). Esta funcionalidade expõe a estrutura completa da publicação e dos comentários num formato hierárquico e legível por máquinas, facilitando a extração automatizada de dados sem necessidade de recorrer à API autenticada do Reddit. Ao aceder ao URL modificado, o servidor retorna a informação da *thread*, incluindo meta dados do *post* principal e os comentários correspondentes, o que permitiu construir um *scraper* simples, mas eficiente.

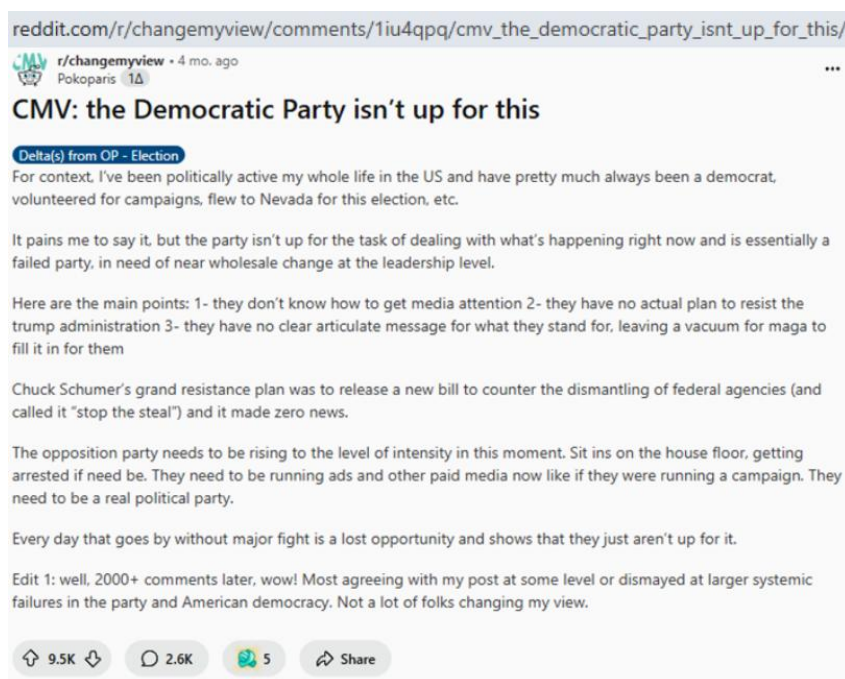


Figura 9 - Exemplo de URL e Thread do Reddit

```

reddit.com/r/changemyview/comments/1iu4qpq/cm_v_the_democratic_party_isnt_up_for_this/.json
{
  "kind": "listing",
  "data": {
    "after": null,
    "dist": 1,
    "modhash": "",
    "geo_filter": "",
    "children": [
      {
        "kind": "t3",
        "data": {
          "approved_at_utc": null,
          "subreddit": "changemyview",
          "selftext": "For context, I've been politically active my whole life in the US and have pretty much always been a democrat, volunteered for campaigns, flew to Nevada for this election, etc. I don't paine me to say it, but the party isn't up for the task of dealing with what's happening right now and is essentially a failed party, in need of near wholesale change at the leadership level. Where are the main points? they don't know how to get media attention, they have no actual plan to resist the trump administration (sh- they have no clear articulate message for what they stand for, leaving a vacuum for maga to fill it in for them (chuck schumer's grand resistance plan was to release a new bill to counter the dismantling of federal agencies (and called it "stop the steal") and it made zero news. the opposition party needs to be rising to the level of intensity in this moment. Sit on the house floor, getting arrested if need be. They need to be running ads and other paid media now like if they were running a campaign. they need to be a real political party. every day that goes by without major fight is a lost opportunity and shows that they just aren't up for it. edit: i well, 2000+ comments later, wow! Most agreeing with my post at some level or dismayed at larger systemic failures in the party and american democracy, but a lot of folks changing my view. ",
          "user_reports": [],
          "saved": false,
          "mod_reason_title": null,
          "gilded": 0,
          "clicked": false,
          "title": "Oh: the Democratic Party isn't up for this",
          "link_flair_richtext": [],
          "subreddit_name_prefixed": "r/changemyview",
          "hidden": false,
          "pwls": 6,
          "link_flair_css_class": "Deltaelection",
          "downs": 0,
          "top_awarded_type": null,
          "hide_score": false,
          "name": "t3_limgqg",
          "quarantine": false,
          "link_flair_text_color": "light",
          "upvote_ratio": 0.96,
          "author_flair_background_color": "",
          "subreddit_type": "public",
          "ups": 848,
          "total_awards_received": 0,
          "media_embed": {}
        }
      }
    ],
    "author_flair_template_id": null,
    "is_original_content": false,
    "author_fullname": "t3_6k8as",
    "secure_media": null,
    "is_reddit_media_domain": false,
    "is_meta": false,
    "category": null,
    "secure_media_embed": {}
  }
}

```

Figura 10 - Exemplo de URL e Thread do Reddit em Formato JSON

O módulo de *scraping* do Reddit foi implementado nesta base. O *endpoint* criado (*/reddit\_scraper*) aceita pedidos do tipo POST, recebendo um objeto JSON que inclui o URL de uma *thread* do Reddit. Ao receber o pedido, o módulo executa o processo de extração e devolve uma estrutura JSON contendo os dados da publicação e os comentários mais relevantes. A função principal do módulo (disponível no Apêndice A2 – Função Principal do Reddit *Scraper*) é responsável por construir o URL de acesso à API pública do Reddit, adicionando o sufixo (.json) ao URL fornecido. A extração é realizada através de uma requisição HTTP com requests, utilizando um cabeçalho *User-Agent* customizado para evitar bloqueios por parte do Reddit.

A resposta JSON obtida é dividida em duas partes: a primeira contém os metadados da publicação (título, autor, texto, data de criação, número de votos, etc.), e a segunda os comentários. Os comentários são filtrados para excluir entradas automáticas (como o AutoModerator) e são ordenados por pontuação (*upvotes*), retendo apenas os 10 mais votados. Ainda, são extraídas também as 5 respostas com mais *upvotes* feitas a cada um destes comentários. Esta medida constituiu um passo importante no desenvolvimento da solução e foi tomada já que as *threads* dos subreddits mencionados chegam a ter centenas de comentários e respostas, muitas vezes, com um número elevado destes sem qualquer *upvote* ou relevância para o debate. Extraindo os 10 comentários mais votados, e ainda, as 5 respostas mais votadas de cada comentário, permite ter uma ideia bastante detalhada e granular da discussão e dos argumentos utilizados, contribuindo também para um desempenho e tempos de resposta aceitáveis. Com esta medida, é possível extrair até 50 comentários e repostas, o que será mais que suficiente para identificar um número elevado e considerável de argumentos. Cada comentário é processado com uma *helper function* que os estrutura forma recursiva com base

na hierarquia de encadeamento. Este processamento recursivo é feito tanto a nível de comentários como de respostas. Ambas mantêm o id original do Reddit e registam o corpo do comentário, o autor, a data, o score e o corpo do comentário pai (`parent_body`), permitindo manter o contexto argumentativo de cada resposta. O corpo textual de cada entrada é também limpo com a função `process_text` (disponível no Apêndice A3 – *Helper Function* `process_text`) que identifica e assinala blocos de citação (identificados pelo símbolo `>` convertido para `&gt;`) com marcadores como `“Citing:”` e `“End of Citation”`, o que facilita a análise posterior por modelos de linguagem ao distinguir claramente entre texto original e texto citado, bem como a leitura na interface. A resposta final da API é um objeto estruturado que inclui o conteúdo do *post* e uma lista de comentários e respostas, pronta para ser enviada para os módulos de análise seguintes (disponível no Apêndice A4 - Objeto `"thread_data"`).

### 4.3.3 Identificação do Tópico de Discussão

Após a definição e contextualização do processo de extração e recolha dos dados, segue-se a componente central da *pipeline* de processamento — a fase em que entram em ação os *Large Language Models*, responsáveis por interpretar, classificar e sintetizar a informação textual recolhida.

A correta identificação do tópico principal de uma *thread* é uma etapa crítica para garantir a coerência e precisão das análises subsequentes, nomeadamente a *stance classification* e a extração de argumentos. No presente sistema, essa responsabilidade recai sobre um LLM, mais concretamente o GPT-4o-mini, acedido por meio da *framework* LangChain e da API da OpenAI. Esta escolha não se prende apenas com a elevada capacidade de inferência semântica do modelo, mas também com a facilidade de integração e modularidade oferecida pela LangChain, permitindo encapsular com clareza o processo de construção de *prompts* e a chamada ao modelo. O funcionamento do módulo *Topic Identifier* baseia-se num conceito essencial de *prompt engineering*: guiar o modelo com instruções explícitas, orientadas para uma tarefa bem definida, mas sem restringir excessivamente a sua capacidade de raciocínio. A *prompt* utilizada (Figura 11) apresenta ao modelo o texto de entrada (o título e o corpo do *post* inicial da *thread*) e solicita que este identifique o tópico central da discussão. Mais ainda, é-lhe pedido que descreva sucintamente quais seriam, nesse contexto, as posições "a favor" e "contra", explicitando o que significa estar de um lado ou de outro do debate. Esta dupla tarefa — identificação do tópico e contextualização das posições — visa preparar o terreno para que as

análises seguintes interpretem corretamente os argumentos expressos nos comentários, mapeando-os para as posturas corretas e compreendendo o enquadramento da discussão, bem como providenciar ao eventual utilizador toda a informação necessária para a compreensão da discussão.

```
# Function that identifies the discussion topic and stances
prompt = PromptTemplate(
    input_variables=["text"],
    template="""
    Text: "{text}"

    Instruction: This is a social media discussion thread header.
    Based on it, identify the main topic being discussed in the thread, so that it is possible to know, reading the topic,
    what the two stances (for and against) would mean in that context.
    Then, really briefly explain what the "for" and "against" stances refer to in the context of this topic.
    Specifically, really briefly describe what these stances represent regarding the actions or policies being discussed,
    and clarify what "for" and "against" are supporting or opposing."
    Response:
    """
)
```

Figura 11 - *Prompt* Utilizada para a Identificação do Tópico da Discussão

Este tipo de tarefa requer uma capacidade de raciocínio contextual que os LLMs, treinados em vastos volumes de dados, conseguem executar com relativa eficácia. Apesar da entrada poder ser relativamente curta ou ambígua o modelo é capaz de recorrer ao conhecimento implícito nos seus parâmetros para inferir o tema implícito prever com razoável precisão quais seriam os argumentos típicos de cada posição. Esta capacidade é extremamente útil em contextos reais, onde os utilizadores muitas vezes não expressam de forma explícita o tópico da discussão. A temperatura do modelo foi ajustada para 0.1, de forma a promover respostas consistentes e com baixa variabilidade. Este valor reduz o grau de criatividade do modelo, tornando-o mais determinístico — uma escolha apropriada para tarefas em que se pretende extrair informação objetiva e coerente, em vez de fomentar diversidade de resposta. O parâmetro “max\_tokens” foi limitado a 200, suficiente para permitir uma explicação clara, mas concisa das posições sem correr o risco de gerar saídas excessivamente longas ou desviadas do objetivo (Figura 12).

```
# Discussion topic identifier model (via Langchain)
topicIdentifier_model = ChatOpenAI(
    model = "gpt-4o-mini",
    max_tokens = 200,
    temperature = 0.1,
    openai_api_key = openai_api_key
)
```

Figura 12 - Definição do Modelo para Identificação do Tópico de Discussão

A resposta gerada pelo modelo é um texto estruturado que inclui o tópico principal e as descrições das posturas opostas, sendo posteriormente armazenada e utilizada pelos módulos

seguintes. Este encapsulamento claro da função de identificação temática permite não só uma separação de responsabilidades na arquitetura da aplicação, mas também facilita futuras atualizações do modelo, ajustes ao *prompt* ou integração com outros LLMs.

#### 4.3.4 Classificação do Posicionamento dos Comentários e Respostas – *Stance Classification*

Após a identificação do tópico de discussão, a *pipeline* prossegue com uma tarefa essencial para a compreensão da dinâmica argumentativa dos debates: a classificação do posicionamento de cada comentário relativamente ao tópico em questão. Esta etapa recorre novamente ao uso LLMs, explorando a sua capacidade de interpretar linguagem natural em contextos argumentativos.

Neste módulo, o objetivo é atribuir a cada comentário uma classificação que represente a sua orientação argumentativa face ao tema identificado, distinguindo entre três possibilidades: *FOR* (a favor), *AGAINST* (contra), ou *NEUTRAL* (neutro). A complexidade desta tarefa reside não apenas na natureza ambígua da linguagem informal usada nas redes sociais, mas também no facto de muitos comentários conterem ironia, sarcasmo, ou posicionamentos implícitos e indiscriminados.

O modelo selecionado para esta tarefa foi o GPT-4o, superior ao *4o-mini* usado anteriormente, já que esta tarefa necessita de um desempenho mais robusto e fiável para lidar com as nuances discursivas presentes. A temperatura foi configurada a 0 para garantir consistência e objetividade nas respostas, dado que a tarefa exige classificações determinísticas e sem ambiguidade (Figura 13).

```
# Initialize LangChain stance detection model
stance_model = ChatOpenAI(
    model="gpt-4o",
    openai_api_key=openai_api_key,
    temperature=0,
    max_tokens=10
)
```

Figura 13 - Definição do Modelo para Stance Classification

Tal como no módulo anterior, recorreu-se à biblioteca LangChain para estruturar o processo de *prompt engineering*, utilizando um "*PromptTemplate*" desenhado especificamente para a

tarefa de deteção de posicionamento. Esta *prompt* (Figura 14) fornece ao modelo o contexto completo, incluindo o título da *thread*, o conteúdo textual do *post* original, o tópico identificado e, por fim, o corpo do comentário a ser classificado. Esta riqueza de contexto é fundamental para garantir que o modelo compreende a totalidade do discurso antes de atribuir uma classificação.

```
# Define stance classification prompt using LangChain's PromptTemplate
stance_prompt = PromptTemplate(
    input_variables=["thread_title", "thread_selftext", "identified_topic", "comment_body"],
    template="""
You are an AI trained in stance detection. Your task is to classify a Reddit comment's stance toward the discussion topic based on the full thread context.
Analyze the given Reddit post and comment carefully, and return ONLY ONE of the following labels:
- AGAINST
- FOR
- NEUTRAL

Keep in mind that the comments can contain sarcasm and irony. Do NOT provide any explanation, analysis, or additional text.

Reddit Thread:
Title: "{thread_title}"
Post Content: "{thread_selftext}"
Identified Discussion Topic: "{identified_topic}"

Reddit Comment:
"{comment_body}"

Label:
""")
```

Figura 14 - *Prompt* Utilizada para *Stance Classification*

Uma das características da *prompt* utilizada é a sua restrição explícita à devolução de apenas uma das três *labels* definidas. Esta limitação tem como objetivo evitar divagações, explicações ou classificações adicionais, forçando o modelo a focar-se apenas na decisão categórica. Outra das características é a presença da frase “*Keep in mind that the comments can contain sarcasm and irony.*”, que é muito importante para que o modelo perceba que não está na presença de um ambiente argumentativo formal e direto, mas sim num ambiente informal e ambíguo em que os utilizadores são muitas vezes irónicos, sarcásticos e, por vezes, insultuosos. Este processo de classificação do posicionamento constitui um passo crítico na estruturação da informação recolhida, permitindo não só distinguir os vários lados de um debate, mas também preparar os dados para etapas seguintes como a sumarização e extração de argumentos.

#### 4.3.5 Sumarização das Posições – *Stance Summarization*

Com os comentários classificados por posicionamento, a etapa seguinte visa condensar os principais argumentos expressos por cada grupo (a favor, contra e neutro) de forma clara, concisa e estruturada. Esta fase é especialmente relevante para transformar discussões fragmentadas e comentários individuais em representações da informação acessíveis e informativas. Neste módulo, e assim como em todos os módulos da *pipeline* (exceto o de

extração dos dados), recorre-se novamente a um LLM — neste caso, voltando ao GPT-4o — cuja tarefa é gerar resumos estruturados sob a forma de pontos-chave. Desta vez, a temperatura foi definida como 0.2, de forma a permitir uma ligeira variabilidade na formulação dos resumos, sem comprometer a consistência ou a objetividade (Figura 15). Este valor equilibra a criatividade do modelo com a necessidade de manter uma estrutura coerente e informativa nos pontos sintetizados. Não foi definido um limite explícito de *max tokens*, permitindo ao modelo gerar respostas tão longas quanto necessário dentro dos limites padrão da API. Esta decisão visa garantir que o modelo consiga incluir todos os pontos relevantes na síntese, sem ser truncado prematuramente.

```
# Initialize summarizer model
summarizer = ChatOpenAI(
    model="gpt-4o",
    openai_api_key=openai_api_key,
    temperature=0.2
)
```

Figura 15 - Definição do Modelo para Sumarização

Tal como nas etapas anteriores, a construção da *prompt* é fundamental. Através desta, o modelo é instruído a atuar como um especialista na síntese de discussões online. São fornecidas orientações claras que moldam a resposta pretendida: os pontos devem ser apresentados como tópicos curtos seguidos de uma descrição, não devem existir repetições e o resumo deve ser autoexplicativo (Figura 16).

```
# Define summarization prompt (back to the original approach)
summarizer_prompt = PromptTemplate(
    input_variables=["comments"],
    template="""
You are an expert at synthesizing online discussions.
Given the following user comments that share a similar stance in an online discussion, summarize the key points into a few concise bullet points.

### Guidelines:
- Begin each bullet point with a short title, followed by a colon and the full idea.
- The summary should be self-contained and reflect the main reasoning shared across the comments.
- Do not repeat similar points. Focus on distinct insights.

### Example Output:

- Economic Barriers: Many people want children but face high costs of childcare, healthcare, and housing.
- Cultural Priorities: Modern values emphasize personal freedom and career growth, delaying or reducing interest in parenthood.
- Policy Skepticism: Financial incentives like $5,000 are seen as insufficient or ineffective in changing birth rates.

### Comments:
{comments}

### Summary (bullet points):
""")
```

Figura 16 - *Prompt* Utilizada para Sumarização Argumentativa

Um aspeto distintivo desta abordagem é o facto de os resumos serem gerados separadamente para cada grupo de posicionamento (FOR, AGAINST, NEUTRAL). A função que orquestra esta

lógica percorre os comentários agrupados por classe e submete cada grupo ao modelo, garantindo que os resumos respeitem o ponto de vista correspondente. Esta medida foi tomada para assegurar que o modelo não mistura comentários de *stances* diferentes no mesmo resumo.

Esta estratégia de sumarização não só contribui para a inteligibilidade da discussão, como também prepara os dados para visualizações posteriores ou para o enriquecimento de grafos de conhecimento. Além disso, permite que utilizadores finais — humanos ou outros sistemas — possam aceder rapidamente ao essencial de um debate sem terem de percorrer centenas de comentários.

#### 4.3.6 Extração, Classificação e Agrupamento de Argumentos

Depois de finalizada a primeira fase de análise, o sistema prossegue com a identificação, extração e classificação dos argumentos expressos em cada comentário e resposta, bem como o agrupamento destes argumentos. Esta fase é talvez a mais importante de toda a *pipeline*, sendo central na construção da representação argumentativa do debate. Estas duas tarefas assentam novamente na utilização de um LLM, (GPT-4o) acessível via LangChain (Figura 17). Embora, do ponto de vista arquitetural, a extração de argumentos, o agrupamento semântico e a criação do grafo estejam integrados no mesmo módulo, opta-se nesta secção por focar apenas nas duas primeiras tarefas, dado o seu carácter puramente linguístico e dependente de modelos.

```
# Initialize OpenAI model
llm = ChatOpenAI(
    model="gpt-4o",
    temperature=0,
    openai_api_key=os.getenv("OPENAI_API_KEY")
)
```

Figura 17 - Definição do Modelo para Extração, Classificação e Agrupamento de Argumentos

#### Identificação e Extração de Argumentos

Para orientar o modelo nesta tarefa, foi construída uma *prompt* com abordagem *few-shot learning*, na qual se fornecem exemplos explícitos do output que se pretende pelo modelo (Figura 18). A estrutura elaborada guia o modelo a ignorar declarações vagas, ofensivas ou redundantes, focando-se apenas nas frases que exprimam conteúdo argumentativo relevante.

Um dos exemplos fornecidos ao modelo, no contexto desta abordagem, contém as seguintes afirmações: *“Parents that give kids smarthphones are stupid. Smartphones distract students from learning.”* Neste caso, o modelo é instruído a apenas extrair e classificar como argumento a segunda proposição - *“Smartphones distract students from learning.”* - já que apenas esta contém conteúdo argumentativo relevante. A primeira frase, por ser ofensiva, subjetiva e desprovida de justificção, deve ser ignorada, de acordo com as instruções da *prompt*.

Esta técnica permite melhorar a performance do modelo sem necessidade de treino adicional, ajustando o seu comportamento ao que se pretende, com base em exemplos concretos inseridos diretamente na *prompt*.

```
# Prompt to extract arguments
argument_extraction_prompt = PromptTemplate(
    input_variables=["text", "topic"],
    template="""
    You are an AI trained to extract formal, self-contained, non-redundant arguments from Reddit discussions.

    Your task is to extract only arguments that can be used in scientific, academic, or logical contexts.
    Arguments should clearly support or oppose a specific claim related to the discussion topic.
    Avoid insults, and vague statements.

    ### INSTRUCTIONS:
    - Extract arguments that include reasoning, causal relationships, or factual claims.
    - Each argument must be self-contained: don't use pronouns like "this" or "it" without defining them.
    - Avoid repeating the same idea in different words.
    - Do not extract vague, general statements or insults.
    - If no clear arguments exist, return: "No clear arguments found."

    ### EXAMPLES:

    Topic: "Should schools ban smartphones?"
    Text:
    Parents that give kids smarthphones are stupid. Smartphones distract students from learning.
    Kids use them during class to cheat on tests.
    Smartphones allow students to stay connected with parents in emergencies.

    Output:
    1. Smartphones distract students from learning.
    2. Students use smartphones to cheat during exams.
    3. Smartphones can help students stay in touch with parents during emergencies.

    Topic: "Is Trump a threat to democracy?"
    Text:
    Trump is very stupid.
    Trump is cunning and wants revenge.
    He pressures officials to do what he wants.
    Trump uses his power to discredit investigations.

    Output:
    1. Trump has pressured government officials to influence investigations.
    2. Trump has used his power to discredit investigations and investigators.
    3. Trump seeks to consolidate power for personal gain, undermining democratic norms.

    ### Now extract arguments for the following post:

    Reddit Topic: "{topic}"

    Content:
    {text}

    Only output a numbered list of arguments, or "No clear arguments found."
    """)
    )
```

Figura 18 - *Prompt* Utilizada para Extração de Argumentos

Cada comentário ou resposta é submetido ao modelo individualmente, juntamente com o tópico da discussão, de forma a orientar o raciocínio do modelo para um domínio específico. O resultado é uma lista numerada de argumentos extraídos, os quais são posteriormente ligados, via grafo de conhecimento, aos comentários ou respostas de origem, mantendo assim a rastreabilidade entre discurso original e argumento identificado.

### Classificação do Posicionamento dos Argumentos Extraídos

Uma vez extraídos os argumentos individuais a partir dos comentários e respostas, cada um deles é submetido a uma segunda chamada ao mesmo modelo (Figura 17) com o objetivo de determinar a sua posição relativa ao tópico da discussão.

A classificação do posicionamento de cada argumento é feita com base numa *prompt* que apresenta ao modelo o tópico da discussão e o texto do argumento, solicitando que este identifique se a proposição expressa está “a favor”, “contra” ou é “neutra” relativamente ao tópico. O modelo é instruído a considerar apenas o conteúdo lógico e factual do argumento, evitando enviesamentos que possam surgir da linguagem emocional ou do autor.

A Figura 19 ilustra a estrutura da *prompt* utilizada para esta tarefa de classificação individual de argumentos, onde o modelo recebe o tópico e o argumento isolado, e devolve uma das três etiquetas possíveis: *FOR*, *AGAINST* ou *NEUTRAL*.

```
# Prompt to classify arguments
stance_classification_prompt = PromptTemplate(
    input_variables=["argument", "topic"],
    template="""
You are given the Reddit discussion topic:
"{topic}"

Classify the stance expressed in the following argument with one of these labels: FOR, AGAINST, or NEUTRAL.

Argument:
"{argument}"

Instructions:
- FOR: The argument expresses clear or implicit support for the topic.
- AGAINST: The argument expresses clear or implicit opposition to the topic.
- NEUTRAL: The argument neither supports nor opposes the topic, or is ambiguous.

Consider the overall position conveyed, including implicit meanings or nuances, not only explicit statements.

Return exactly one word: FOR, AGAINST, or NEUTRAL.
"""
)
```

Figura 19 - *Prompt* Utilizada para Classificação de Argumentos

## Clustering dos Argumentos Extraídos

Concluída a etapa de extração dos argumentos, segue-se o *clustering*, com recurso ao mesmo modelo definido anteriormente (Figura 17). Esta fase visa detetar conjuntos de argumentos que exprimem a mesma ideia fundamental, mesmo que com formulações distintas.

Na *prompt* criada (Figura 20), solicita-se ao modelo a criação de grupos coesos, cada um acompanhado por um resumo detalhado e autoexplicativo da ideia comum defendida nesse conjunto. Este resumo deve deixar claro o que está a ser defendido ou criticado, e porquê, sem depender da leitura das frases originais. O modelo é ainda orientado a não fundir ideias distintas, evitando generalizações como “críticas à esquerda” ou “preocupações com corrupção”, tentando que os resumos criados não sejam vagos e com falta de detalhe.

```
##Prompt to create argument clusters
argument_grouping_prompt = PromptTemplate(
    input_variables=["stance", "arguments"],
    template="""
    You are an AI system trained to group semantically similar arguments from Reddit discussions.

    ### Objective:
    Given a list of arguments that all share the same stance ("stance") toward a topic, group those that express the same specific idea – even if they use different wording.
    For each group, write a "detailed and self-contained summary" that explains the "shared claim" and the "main reasoning" behind it.

    ### Instructions:
    - The summary must make it "completely clear what the core argument is" without reading the individual comments.
    - Be specific. Clearly express what the grouped users believe or argue, and "why" they argue it.
    - Do NOT use vague generalizations like "Criticism of the left" or "Concerns about corruption."
    - Avoid summarizing with abstract categories – instead, explain the "actual position" being taken.
    - Only group arguments that express the same idea. Do not merge arguments that are different, even if related.
    - Use the original wording for the bullet points. Do not rewrite or paraphrase them.
    - If an argument doesn't belong in a group, place it alone with its own summary.

    ### Format:

    Group: <detailed explanation of the shared argument and reasoning>
    - <original argument 1>
    - <original argument 2>

    Group: <another detailed explanation>
    - <original argument 3>

    ### Arguments:
    {arguments}
    """)
```

Figura 20 - Prompt Utilizada para o *Clustering* de Argumentos

A combinação destas duas tarefas — extração e *clustering* — permite transformar um grupo de comentários e respostas desorganizado e informal num conjunto estruturado de proposições argumentativas organizadas, que captam os principais núcleos discursivos de uma discussão. A sua eficácia resulta diretamente do poder inferencial dos LLMs e da clareza dos *prompts* que os guiam, demonstrando o papel crucial do *prompt engineering* na mediação entre linguagem natural e representação estruturada do conhecimento.

#### 4.3.7 Criação dos Grafos de Conhecimento

A fase final do processamento culmina na organização estruturada dos elementos argumentativos extraídos, através da criação de um grafo de conhecimento que represente, de forma explícita, a lógica e as relações presentes na discussão analisada. Esta estrutura assume um papel central no sistema, funcionando como a base para a navegação semântica dos argumentos, a geração de visualizações, e a realização de consultas complexas que não seriam facilmente possíveis com dados não estruturados.

O grafo representa diversos tipos de entidades — nomeadamente o tópico da discussão, os comentários e respostas, os argumentos extraídos e os clusters criados. Cada um destes elementos corresponde a um nó no grafo, enquanto as relações entre eles são modeladas como arestas direcionadas, indicando por exemplo que um comentário contém um argumento, ou que determinado argumento pertence a um grupo temático.

Além das relações estruturais, o grafo inclui também a orientação argumentativa de cada comentário e dos seus argumentos associados, marcando explicitamente se pertencem à *stance* "a favor", "contra" ou "neutra". Esta informação é essencial para permitir que o grafo não seja apenas uma representação topológica, mas sim um mapa semântico da posição dos participantes no debate.

A construção do grafo é feita com base nos dados previamente processados pelos modelos, mas não se recorre a LLMs nesta fase. A estruturação e inserção dos dados no grafo é feita de forma programática, recorrendo a uma instância da base de dados Neo4j e um conjunto de *helper functions* em Cypher que interagem com esta (Figura 21). Esta integração permite explorar, por exemplo, todas as ideias defendidas por uma determinada posição, encontrar comentários que partilham argumentos semelhantes, ou identificar os principais agrupamentos conceptuais dentro de uma discussão.

```

# Cypher helpers
def check_existing_discussion_by_url(tx, url: str):
    result = tx.run("""
        MATCH (t:Topic {url: $url})
        RETURN t.discussion_id AS discussion_id
    """, url=url)
    record = result.single()
    return record["discussion_id"] if record else None

def check_comment_exists(tx, comment_id, discussion_id):
    result = tx.run("""
        MATCH (c:Comment {id: $comment_id, discussion_id: $discussion_id})
        RETURN count(c) > 0 AS exists
    """, comment_id=comment_id, discussion_id=discussion_id)
    record = result.single()
    return record["exists"] if record else False

def check_reply_exists(tx, reply_id, discussion_id):
    result = tx.run("""
        MATCH (r:Reply {id: $reply_id, discussion_id: $discussion_id})
        RETURN count(r) > 0 AS exists
    """, reply_id=reply_id, discussion_id=discussion_id)
    record = result.single()
    return record["exists"] if record else False

def merge_topic(tx, title, discussion_id, url):
    tx.run("""
        MERGE (t:Topic {url: $url})
        SET t.title = $title,
            t.discussion_id = $discussion_id,
            t.updated_at = datetime()
    """, title=title, discussion_id=discussion_id, url=url)

```

Figura 21 - Exemplos das funções Cypher criadas

#### 4.3.8 Exploração dos Grafos de Conhecimento

Para além da sua função enquanto estrutura de representação semântica, o grafo de conhecimento criado pelo sistema é também explorável de forma dinâmica através de um conjunto de *queries* pré-definidas. Estas consultas foram implementadas com o objetivo de permitir ao utilizador aceder a diferentes dimensões analíticas da discussão, extraíndo rapidamente informação relevante sobre os argumentos, comentários e relações envolvidas.

A interface criada permite selecionar uma das *queries* disponíveis, sendo o resultado da consulta apresentado automaticamente. A execução destas *queries* é feita diretamente sobre a base de dados Neo4j, utilizando a linguagem Cypher para navegação eficiente pelos nós e relações que compõem o grafo.

A Figura 22 apresenta a estrutura destas *queries* no sistema, que estão diretamente associadas a funcionalidades de exploração específicas, descritas em detalhe na secção de demonstração a seguir.

Este mecanismo de exploração orientada por *queries* oferece uma alternativa complementar à visualização gráfica do grafo, funcionando como uma ferramenta de análise interativa e dirigida. Permite ao utilizador obter rapidamente respostas a questões específicas, sem necessidade de percorrer a estrutura visual ou consultar manualmente todos os nós e relações.

```
# Define queries that can be applied to the selected topic
query_options = {
  "List of Arguments by Stance": {
    "query": """
      MATCH (t:Topic {title: $title})
      MATCH (a:Argument)-[:EXTRACTED_FROM]->(n)
      WHERE (n:Comment)-[:SUPPORTS|OPPOSES|NEUTRAL]->(t)
      OR (n:Reply)-[:REPLY_TO]->(:Comment)-[:SUPPORTS|OPPOSES|NEUTRAL]->(t)
      RETURN a.text AS Argument, a.stance AS Stance
    """
  },
  "Argument Groups by Popularity": {
    "query": """
      MATCH (t:Topic {title: $title})
      MATCH (a:Argument)-[:EXTRACTED_FROM]->(n)
      WHERE (n:Comment)-[:SUPPORTS|OPPOSES|NEUTRAL]->(t)
      OR (n:Reply)-[:REPLY_TO]->(:Comment)-[:SUPPORTS|OPPOSES|NEUTRAL]->(t)
      MATCH (a)-[:HAS_GROUP]->(g:ArgumentGroup)
      RETURN
        g.summary AS GroupSummary,
        a.text AS ArgumentText,
        a.stance AS Stance
    """
  },
  "Replies to Supporting Comments": {
    "query": """
      MATCH (c:Comment)-[:SUPPORTS]->(t:Topic {title: $title})
      MATCH (r:Reply)-[:REPLY_TO]->(c)
      RETURN c.body AS ParentComment,
        collect(r.body) AS Replies,
        size(collect(r.body)) AS ReplyCount
      ORDER BY ReplyCount DESC
      LIMIT 10
    """
  },
  "Replies to Opposing Comments": {
    "query": """
      MATCH (c:Comment)-[:OPPOSES]->(t:Topic {title: $title})
      MATCH (r:Reply)-[:REPLY_TO]->(c)
      RETURN c.body AS ParentComment,
        collect(r.body) AS Replies,
        size(collect(r.body)) AS ReplyCount
      ORDER BY ReplyCount DESC
      LIMIT 10
    """
  },
  "Popular Comments (by score)": {
    "query": """
      MATCH (c:Comment)-[:SUPPORTS|OPPOSES|NEUTRAL]->(t:Topic {title: $title})
      RETURN c.body AS Comment,
        c.score AS Score, type(r) AS Stance
      ORDER BY c.score DESC
      LIMIT 10
    """
  },
  "Discussion url": {
    "query": """
      MATCH (t:Topic {title: $title})
      RETURN t.title AS Title, t.url AS URL
    """
  }
}
```

Figura 22 - Estrutura das *Queries* Pré-definidas para Exploração do Grafo de Conhecimento

### 4.3.9 *Framework* de Avaliação

A *framework* desenvolvida permite analisar, de forma sistemática, a eficácia de três componentes críticos da *pipeline*: (1) a extração de argumentos, (2) a classificação das *stances* de cada argumento e (3) o *clustering* dos argumentos. Todas as avaliações são realizadas por um LLM, GPT-4o, instruído com *prompts* específicas desenhadas para cada tarefa, que se enquadram no estilo *zero-shot*. As instruções fornecidas ao modelo procuram simular a análise de um avaliador humano, com critérios claros e escalas de pontuação que variam entre 1 e 5.

As *prompts* de avaliação desenvolvidas são aplicadas em dois contextos distintos, mas complementares: na análise individual de cada elemento processado — seja um argumento, uma classificação de *stance* ou um *cluster* — e na avaliação geral de cada uma destas tarefas. Em ambos os casos, são utilizados critérios consistentes e bem definidos para aferir a qualidade das respostas geradas pelo sistema. As avaliações individuais permitem identificar com precisão os elementos que apresentam erros ou inconsistências, enquanto as avaliações gerais sintetizam o desempenho global de cada componente, destacando padrões de erro, pontos fortes e oportunidades de melhoria. Esta abordagem combinada reforça o carácter sistemático e analítico da *framework* de avaliação.

#### **Avaliação da Extração de Argumentos**

A qualidade da extração é avaliada com base no conteúdo original (comentário ou resposta) e na lista de argumentos identificados pelo sistema. A *prompt* (Figura 23) instrui o modelo a classificar a extração de acordo com quatro dimensões: completude, precisão, qualidade e relevância dos argumentos extraídos.

O modelo responde com uma pontuação global de 1 a 5, acompanhada de uma justificação textual e, caso aplicável, uma lista de problemas identificados (por exemplo, omissões, argumentos mal formulados ou irrelevantes).

```

ARGUMENT_EXTRACTION_PROMPT = PromptTemplate(
    input_variables=["content_text", "extracted_arguments", "content_type"],
    template="""
You are evaluating the quality of argument extraction from a {content_type}.

Original {content_type}:
"{content_text}"

Arguments Extracted by the System:
{extracted_arguments}

Evaluate the extraction based on:
1. COMPLETENESS: Did the system capture all main arguments present in the text?
2. ACCURACY: Are the extracted arguments actually present in the original text?
3. QUALITY: Are the arguments well-formulated and self-contained?
4. RELEVANCE: Are the extracted arguments actually arguments (not just vague statements)?

Evaluation scale:
1 = Very poor (missed important arguments or extracted irrelevant content)
2 = Poor (captured some arguments but with many problems)
3 = Average (captured main arguments but with some flaws)
4 = Good (good extraction with minor problems)
5 = Excellent (complete and accurate extraction)

Respond in the format:
Score: X
Justification: [detailed explanation in 2-3 sentences]
Identified problems: [list main problems, if any]
""")

```

Figura 23 - *Prompt* Utilizada para Avaliação da Extração de Argumentos

### Avaliação da Classificação de Posicionamento

Para cada argumento extraído, são avaliadas as classificações das *stances* atribuídas pelo sistema a cada argumento, em função do tópico analisado. A *prompt* (Figura 24) apresenta ao modelo o tópico da discussão, o texto do argumento e a classificação original atribuída. Com base nestes elementos, o modelo avalia se a classificação atribuída é precisa, ambígua ou errada, atribuindo uma pontuação de 1 (completamente errada) a 5 (totalmente correta e inequívoca). Esta avaliação é ainda essencial para validar se a transição da classificação ao nível dos comentários para uma classificação a nível argumentativo resultou num aumento da precisão semântica.

```

STANCE_EVALUATION_PROMPT = PromptTemplate(
    input_variables=["topic", "argument_text", "detected_stance"],
    template="""
You are evaluating the accuracy of stance classification for an argument used in the context of the following topic:
"{topic}"

Extracted argument:
"{argument_text}"

Stance assigned by the system:
{detected_stance}

Evaluate the correctness of the classification based on the argument content and the topic in question. Consider:

- FOR: the argument supports or agrees with the main statement of the topic.
- AGAINST: the argument disagrees with or contradicts the main statement of the topic.
- NEUTRAL: the argument is informative, tangential, or does not take a clear position.

Evaluation scale:
1 = Incorrect (completely wrong classification)
2 = Inaccurate (Indicates wrong direction or is ambiguous)
3 = Acceptable (can be interpreted as correct, but with doubts)
4 = Correct (classification appropriately matches the content)
5 = Very accurate (clearly correct and unambiguous classification)

Respond in the format:
Score: X
Justification: [brief explanation in 1-2 sentences]
"""
)

```

Figura 24 - *Prompt* Utilizada para Avaliação das *Stances* Atribuídas

### Avaliação do *Clustering* de Argumentos

Por fim, cada grupo de argumentos gerado na fase de *clustering* é submetido a avaliação automática. A *prompt* utilizada nesta fase (Figura 25) apresenta ao modelo o resumo explicativo do grupo e a lista de argumentos incluídos. A avaliação baseia-se em quatro critérios: coerência semântica, qualidade do resumo, granularidade do grupo e completude. O modelo retorna uma pontuação de 1 a 5, uma explicação detalhada e sugestões de melhoria, caso existam. Esta avaliação visa identificar *clusters* demasiado genéricos, incoerentes ou resumos que não representem adequadamente as ideias agrupadas.

```

CLUSTERING_EVALUATION_PROMPT = PromptTemplate(
    input_variables=["cluster_summary", "cluster_arguments"],
    template="""
You are evaluating the quality of an argument cluster.

Cluster Summary:
"{cluster_summary}"

Arguments in Cluster:
{cluster_arguments}

Evaluate the clustering based on:
1. SEMANTIC COHERENCE: Do the grouped arguments really share the same central idea?
2. SUMMARY QUALITY: Does the summary adequately capture the essence of the grouped arguments?
3. GRANULARITY: Is the cluster neither too specific nor too generic?
4. COMPLETENESS: Does the cluster include all relevant arguments for this idea?

Evaluation scale:
1 = Very poor (unrelated arguments grouped together)
2 = Poor (some related arguments but confusing clustering)
3 = Average (reasonable clustering but with some inconsistencies)
4 = Good (coherent clustering with minor problems)
5 = Excellent (semantically perfect clustering)

Respond in the format:
Score: X
Justification: [detailed explanation in 2-3 sentences]
Suggestions: [how to improve clustering, if applicable]
""")

```

Figura 25 - Prompt Utilizada para Avaliação dos Clusters Criados

## Avaliações Gerais

Além das avaliações individuais, a *framework* inclui uma componente de avaliação geral que sintetiza o desempenho global de cada um destes módulos. Esta análise é também gerada por um LLM, utilizando prompts específicas que compilam estatísticas agregadas (como número total de elementos analisados, média de pontuação e distribuição dos scores atribuídos) bem como um conjunto representativo das justificações geradas durante as avaliações individuais. Estas justificações, que explicam as razões subjacentes às pontuações atribuídas, são recolhidas e agregadas de forma a refletir padrões recorrentes de desempenho, permitindo identificar pontos fortes, limitações e inconsistências. O conteúdo agregado serve de base à geração automática de uma explicação geral, que destaca tendências observadas e apresenta recomendações, pontos fortes e limitações concretas dos módulos do sistema.

As *prompts* de avaliação geral (Figura 26, Figura 27 e Figura 28) foram desenhadas para produzir um parágrafo de síntese com 3 a 4 frases, focado na qualidade global da tarefa em causa, na consistência dos resultados e no impacto esperado nos módulos seguintes da *pipeline*.

```

GENERAL_EXTRACTION_PROMPT = PromptTemplate(
    input_variables=["topic", "total_content", "total_args", "avg_score", "score_distribution", "justifications"],
    template="""
You are providing a general evaluation of argument extraction performance for the topic: "{topic}"

Dataset Statistics:
- Total pieces of content (comments + replies): {total_content}
- Total arguments extracted: {total_args}
- Average extraction score: {avg_score}/5
- Score distribution: {score_distribution}

Evaluation Justifications (sample of system-level feedback from individual evaluations):
{justifications}

Based on these justifications and scores, summarize the overall quality of argument extraction. Focus on:
1. Overall performance and consistency
2. Common strengths and weaknesses mentioned
3. Recommendations for improving extraction quality

Respond in 3-4 sentences providing an insightful performance summary.
"""
)

```

Figura 26 - *Prompt* de Avaliação Geral da Extração de Argumentos

```

GENERAL_CLUSTERING_PROMPT = PromptTemplate(
    input_variables=["topic", "total_clusters", "total_args", "avg_score", "score_distribution", "justifications"],
    template="""
You are providing a general evaluation of argument clustering performance for the topic: "{topic}"

Dataset Statistics:
- Total clusters created: {total_clusters}
- Total arguments clustered: {total_args}
- Average clustering score: {avg_score}/5
- Score distribution: {score_distribution}

Evaluation Justifications (sample of system-level feedback from individual evaluations):
{justifications}

Based on these justifications and scores, summarize the clustering performance. Focus on:
1. Overall semantic coherence of clusters
2. Issues in argument grouping or summary quality
3. Recommendations for improving clustering effectiveness

Respond in 3-4 sentences providing actionable insights.
"""
)

```

Figura 27 - *Prompt* de Avaliação Geral do *Clustering*

```

GENERAL_STANCE_PROMPT = PromptTemplate(
    input_variables=["topic", "total_args", "avg_score", "score_distribution", "stance_distribution", "justifications"],
    template="""
You are providing a general evaluation of stance classification performance for the topic: "{topic}"

Dataset Statistics:
- Total arguments classified: {total_args}
- Average stance classification score: {avg_score}/5
- Score distribution: {score_distribution}
- Stance distribution: {stance_distribution}

Evaluation Justifications (sample of system-level feedback from individual evaluations):
{justifications}

Based on these justifications and scores, provide a global assessment of the system's stance classification. Focus on:
1. Accuracy and consistency of the classifications
2. Presence of any notable patterns or systematic issues
3. Suggestions for improving classification reliability

Respond in 3-4 sentences providing a clear and critical performance summary.
"""
)

```

Figura 28 - *Prompt* de Avaliação Geral da Classificação de *Stances*

## **Considerações Finais**

A adoção desta *framework* automatizada de avaliação, embora dependa novamente de um LLM, permite validar iterativamente os resultados da *pipeline* criada e orientar melhorias baseadas em *feedback* objetivo. Ao mesmo tempo, garante uma avaliação rápida e alinhada com os critérios pretendidos, constituindo uma peça essencial para a robustez e fiabilidade do sistema implementado.

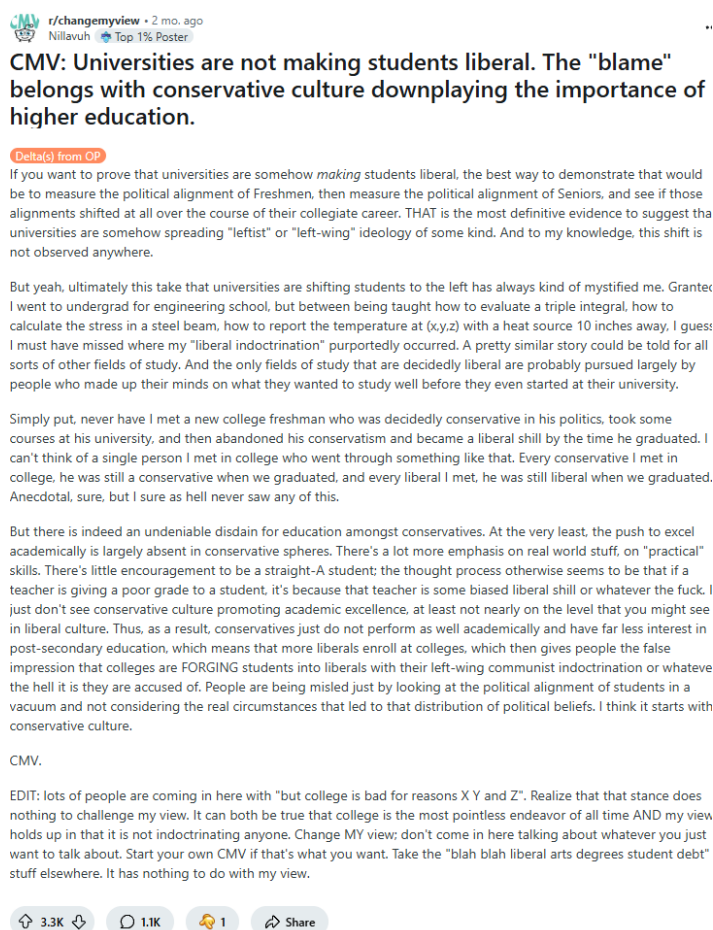
## 5 Resultados e Discussão

Este capítulo apresenta e analisa os principais resultados obtidos com a implementação da solução desenvolvida. O objetivo central é demonstrar, de forma prática e fundamentada, o funcionamento da *pipeline* de análise argumentativa, bem como refletir criticamente sobre o desempenho dos seus diferentes módulos e sobre as escolhas técnicas e metodológicas adotadas ao longo do projeto. A primeira parte do capítulo é dedicada a uma demonstração funcional detalhada, baseada na análise de uma discussão real retirada da plataforma Reddit. Esta demonstração percorre todas as fases do sistema, desde a recolha inicial dos dados até à construção final do grafo de conhecimento, ilustrando como os módulos interagem entre si para estruturar e representar o discurso argumentativo de forma automatizada. Na segunda parte, é realizada uma avaliação crítica da solução, onde se analisam quantitativamente e qualitativamente os pontos fortes e limitações do sistema desenvolvido. Por fim, esta análise é confrontada com os objetivos delineados inicialmente, avaliando o grau de concretização da solução proposta e o seu potencial de aplicação em contextos reais de análise discursiva e síntese argumentativa em ambientes digitais.

**Nota:** Os tópicos abordados nas *threads* analisadas neste capítulo foram selecionados exclusivamente com o objetivo de demonstrar as capacidades técnicas da solução desenvolvida. As discussões analisadas foram retiradas de uma plataforma pública e não refletem, em caso algum, as opiniões ou posicionamentos do autor desta dissertação. Qualquer posicionamento político, ideológico ou social eventualmente identificado nos comentários analisados é da exclusiva responsabilidade dos seus autores originais.

## 5.1 Demonstração Funcional

Com o objetivo de ilustrar, de forma concreta, o funcionamento da solução desenvolvida em contexto real, esta secção apresenta a execução completa da *pipeline* de análise sobre uma discussão autêntica extraída da plataforma Reddit. O propósito desta demonstração é evidenciar, passo a passo, a forma como os diferentes módulos interagem entre si para processar, estruturar e representar o conteúdo discursivo de uma *thread*. A discussão analisada foi retirada do subreddit r/ChangeMyView, previamente identificado como especialmente adequado ao modelo conceptual do sistema. Para efeitos de demonstração, efetua-se a análise da *thread* de título: “*CMV: Universities are not making students liberal. The 'blame' belongs with conservative culture downplaying the importance of higher education.*” Nesta publicação (Figura 29), o autor inicia a discussão com uma opinião clara — contrariando a perceção comum de que as universidades têm uma influência liberalizante nos estudantes — e atribui, em alternativa, essa transformação a fatores culturais internos ao conservadorismo.



r/changetomyview • 2 mo. ago  
Nillavuh • Top 1% Poster

**CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.**

**Delta(s) from OP**

If you want to prove that universities are somehow *making* students liberal, the best way to demonstrate that would be to measure the political alignment of Freshmen, then measure the political alignment of Seniors, and see if those alignments shifted at all over the course of their collegiate career. THAT is the most definitive evidence to suggest that the universities are somehow spreading "leftist" or "left-wing" ideology of some kind. And to my knowledge, this shift is not observed anywhere.

But yeah, ultimately this take that universities are shifting students to the left has always kind of mystified me. Granted I went to undergrad for engineering school, but between being taught how to evaluate a triple integral, how to calculate the stress in a steel beam, how to report the temperature at (x,y,z) with a heat source 10 inches away, I guess I must have missed where my "liberal indoctrination" purportedly occurred. A pretty similar story could be told for all sorts of other fields of study. And the only fields of study that are decidedly liberal are probably pursued largely by people who made up their minds on what they wanted to study well before they even started at their university.

Simply put, never have I met a new college freshman who was decidedly conservative in his politics, took some courses at his university, and then abandoned his conservatism and became a liberal shill by the time he graduated. I can't think of a single person I met in college who went through something like that. Every conservative I met in college, he was still a conservative when we graduated, and every liberal I met, he was still liberal when we graduated. Anecdotal, sure, but I sure as hell never saw any of this.

But there is indeed an undeniable disdain for education amongst conservatives. At the very least, the push to excel academically is largely absent in conservative spheres. There's a lot more emphasis on real world stuff, on "practical" skills. There's little encouragement to be a straight-A student; the thought process otherwise seems to be that if a teacher is giving a poor grade to a student, it's because that teacher is some biased liberal shill or whatever the fuck. I just don't see conservative culture promoting academic excellence, at least not nearly on the level that you might see in liberal culture. Thus, as a result, conservatives just do not perform as well academically and have far less interest in post-secondary education, which means that more liberals enroll at colleges, which then gives people the false impression that colleges are FORGING students into liberals with their left-wing communist indoctrination or whatever the hell it is they are accused of. People are being misled just by looking at the political alignment of students in a vacuum and not considering the real circumstances that led to that distribution of political beliefs. I think it starts with conservative culture.

CMV.

EDIT: lots of people are coming in here with "but college is bad for reasons X Y and Z". Realize that that stance does nothing to challenge my view. It can both be true that college is the most pointless endeavor of all time AND my view holds up in that it is not indoctrinating anyone. Change MY view; don't come in here talking about whatever you just want to talk about. Start your own CMV if that's what you want. Take the "blah blah liberal arts degrees student debt" stuff elsewhere. It has nothing to do with my view.

3.3K ↑ 1.1K ↓ 1 🏆 Share

Figura 29 – Demonstração: *Thread* a Analisar

Esta *thread* foi selecionada por constituir um exemplo de debate politicamente carregado, com um ponto de partida bem definido e suscetível de gerar reações fortes tanto a favor como contra. Assim, serve como base ideal para ilustrar o funcionamento completo da *pipeline* de processamento.

### 5.1.1 Header da Discussão – Tópico, Descrição, *Stances* e Distribuições

Depois da introdução do URL da *thread* na interface do sistema, é iniciada a análise automática da discussão. Na Figura 30 é possível observar a primeira parte da página principal do sistema, que contém o sumário da discussão analisada.

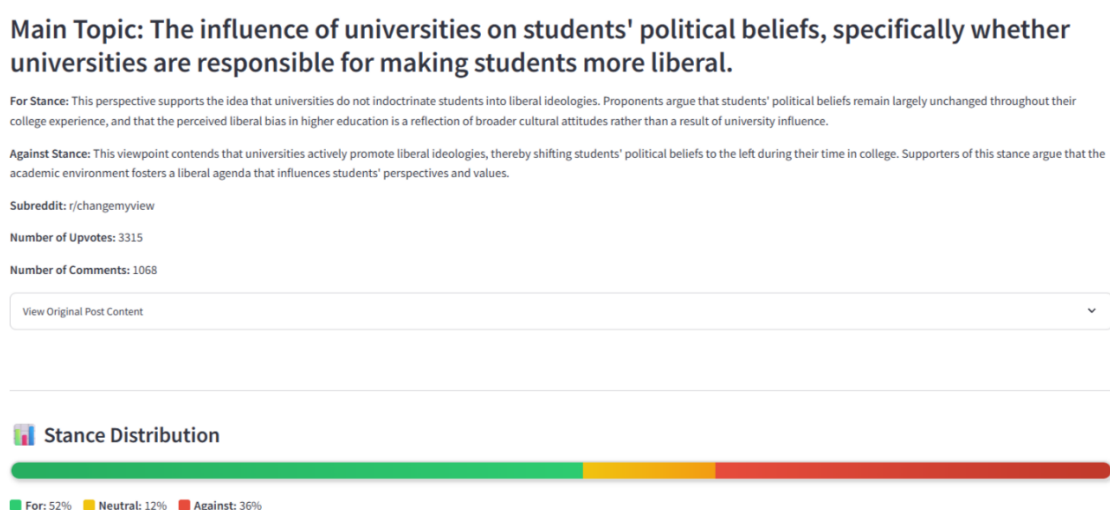


Figura 30 – Demonstração: *Header* da Discussão

A página inicial de resumo da discussão apresenta uma visão geral da análise efetuada sobre a *thread* introduzida, disponibilizando informação essencial relativa ao tópico em debate, distribuição de posicionamentos e estatísticas de participação. O tópico principal identificado foi "***A influência das universidades nas crenças políticas dos estudantes, especificamente no que diz respeito à questão de saber se as universidades são responsáveis por tornar os estudantes mais liberais.***"

A interface apresenta de seguida, de forma estruturada, explicações sobre os dois posicionamentos "a favor" e "contra" no contexto desta discussão (o posicionamento "neutro" mantém-se para todas as discussões):

- **A favor:** Defende que as universidades não têm um papel ativo na transformação ideológica dos estudantes, e que a percepção de enviesamento liberal é consequência de fatores culturais mais amplos;
- **Contra:** Argumenta que as universidades promovem ativamente ideologias liberais, influenciando diretamente a orientação política dos alunos;
- **Neutro:** Inclui comentários que não expressam uma posição clara ou que contribuem de forma informativa sem apoiar diretamente nenhum dos lados.

Além desta informação, são apresentados indicadores quantitativos relevantes, nomeadamente o número de *upvotes* da publicação original (3315) e o total de comentários presentes (1068). A distribuição dos comentários e respostas por posicionamento pode ser observada através do *meter* de “*Stance Distribution*” e foi estimada da seguinte forma: 52% a favor, 36% contra e 12% neutros, revelando uma leve predominância de concordância com a posição inicial do autor do tópico.

### 5.1.2 Resumo dos Argumentos por *Stance*

A segunda secção da página (Figura 31) apresenta um resumo estruturado dos principais argumentos utilizados por cada grupo de posicionamento identificado na discussão — favorável, contra e neutro (neste caso não há comentários neutros).

#### Arguments by Stance - Summary

##### ■ Favorable Arguments

- **Exposure to Diversity:** Many individuals experience a shift from conservative to liberal views due to exposure to diverse people and ideas, often encountered for the first time in college settings.
- **Critical Thinking Development:** Colleges teach critical thinking skills, enabling students to evaluate and question their long-held beliefs, which can lead to a shift in political ideology.
- **Cultural and Economic Influences:** Conservatives often come from rural, blue-collar backgrounds where higher education is less emphasized, contributing to fewer conservatives in academia.
- **Non-Causal Correlation:** While there is a correlation between higher education and liberal views, it is not necessarily causal; exposure to diverse experiences plays a significant role.
- **Misconceptions of Indoctrination:** The belief that colleges indoctrinate students with liberal ideas is challenged by the argument that exposure to new experiences and critical thinking naturally leads to ideological shifts.
- **Systemic Factors:** Societal pressures and incentives for traditionally underrepresented demographics to attend college contribute to the liberal lean in academia, rather than direct indoctrination.

##### ■ Against Arguments

- **Anecdotal Evidence Concerns:** Many arguments rely on personal experiences rather than empirical data, questioning the validity of claims about education's impact on political beliefs.
- **Liberal Academia Influence:** There is a perception that academia, particularly in liberal arts, leans left, potentially influencing students' political views through course material and professor biases.
- **Self-Censorship Among Conservatives:** Conservative students often feel pressured to self-censor in academic settings due to fear of grade penalties or social backlash, leading to a lack of diverse political discourse.
- **Exposure to One-Sided Views:** Students, especially those without strong pre-existing political beliefs, may adopt left-leaning views due to continuous exposure to predominantly liberal perspectives in their coursework.
- **Field-Specific Bias:** While STEM fields are perceived as more objective, liberal biases are still present, though less pronounced compared to liberal arts, affecting students' political development.
- **Empirical Studies on Education and Politics:** Some studies suggest higher education correlates with shifts towards less authoritarian and more left-leaning attitudes, though causality remains debated.
- **Political Realignment and Education:** Research indicates a growing political divide based on educational attainment, with college graduates trending more Democratic and non-graduates more Republican.

Original Favorable Comments

Original Opposing Comments

Figura 31 - Demonstração: Resumos dos Argumentos por *Stance*

Estes sumários foram gerados automaticamente com base no agrupamento dos argumentos utilizados nos comentários classificados previamente, sendo sintetizados de forma a refletir os pontos de vista mais recorrentes e distintivos dentro de cada classe.

Estes resumos argumentativos, divididos por posicionamento, permitem ao utilizador ter uma visão clara dos principais pontos de debate sem necessidade de percorrer manualmente todos os comentários individuais. Estes, estão presentes na interface através dos *dropdowns* “*Original Favorable Comments*” e “*Original Opposing Comments*”, onde é possível consultar os comentários na íntegra, bem como as *stances* atribuídas a cada *reply*. A Figura 32 apresenta um exemplo desta representação. Outros exemplos adicionais podem ser consultados no Apêndice B – Demonstração da Solução.

The image shows a side-by-side comparison of two comment threads. The left thread, 'Original Favorable Comments (660 points)', contains a main comment and four replies with point values of 159, 30, 28, and 18. The right thread, 'Original Opposing Comments (72 points)', contains a main comment and two replies with point values of 20 and 0. The interface uses color-coded icons (green, orange, red) to indicate the status of each reply.

Figura 32 – Demonstração: Comentários Originais e Respetivas *Replies* Classificadas

### 5.1.3 Criação do Grafo de Conhecimento

Após o processamento da discussão, o sistema procede à criação automática do grafo de conhecimento, que constitui uma das componentes centrais da solução desenvolvida. Este grafo é construído a partir da informação recolhida e processada ao longo das etapas anteriores da *pipeline*.

No caso desta discussão em particular, o sistema criou e adicionou à base de dados um grafo com um total de 9 comentários, 16 respostas e 48 argumentos distintos, cada um associado à sua origem textual e à respetiva posição no debate. A geração do grafo é concluída automaticamente após a execução dos módulos de extração e agrupamento de argumentos, garantindo que a totalidade da informação relevante se encontra representada sob a forma de *nodes* e ligações semânticas. A Figura 33 mostra a mensagem exibida ao utilizador após a conclusão bem-sucedida da criação do grafo de conhecimento. O grafo criado fica agora disponível para análise, permitindo a exploração detalhada e minuciosa da discussão.

## Knowledge Graph

Nodes Added: Comments: 9, Replies: 16, Arguments: 48

✓ Knowledge graph created from this discussion!

Figura 33 - Demonstração: Criação do Grafo de Conhecimento

### 5.1.4 Grafo de Conhecimento Criado

A Figura 34 ilustra o Grafo de Conhecimento que representa visualmente a discussão analisada, presente na base de dados Neo4j.

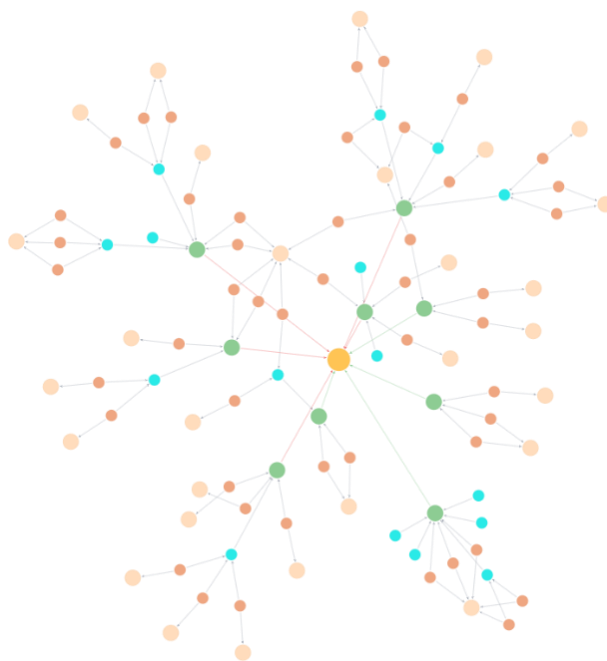


Figura 34 - Demonstração: Grafo de Conhecimento Criado

O grafo apresenta a seguinte estrutura:

- O tópico principal é representado como um nó amarelo no centro do grafo;
- Os comentários ao *post* original são representados por nós verdes; estes são ligados diretamente ao nó principal através de setas de cores diferentes (verde, vermelho ou cinzento), representando as suas *stances*;
- As respostas a esses comentários surgem como nós azul-claro, ligados aos respectivos comentários de origem.
- Os argumentos extraídos, tanto de comentários como de respostas, são representados por nós laranja, cada um ligado ao comentário ou resposta de onde foi extraído.
- Os *clusters* de argumentos, representados por nós bege, ligam-se a todos os argumentos identificados como semanticamente semelhantes. Nos casos em que um argumento não apresenta qualquer correspondência semântica com outros, é ainda assim criado um agrupamento, contendo apenas esse argumento.

Utilizando a interface do Neo4j (Figura 35), é possível perceber que existem 103 nós no total e 29 *clusters* diferentes. Quatro dos comentários extraídos apoiam a ideia do criador do *post* original, e 5 opõem-se à mesma. É também visível que todos os 48 argumentos extraídos foram conectados a um *cluster* através da relação “*HAS\_GROUP*”.



Figura 35 - Demonstração: *Overview* do Grafo

No Apêndice B – Demonstração da Solução, estão presentes representações visuais mais detalhadas do grafo, os seus nós e relações.

### 5.1.5 Exploração Inicial do Grafo de Conhecimento

Após a construção do grafo de conhecimento, é permitido ao utilizador a exploração da discussão previamente analisada e armazenada na base de dados, através da interface de exploração, proporcionando uma navegação intuitiva e orientada por tópicos. Selecionando a discussão em causa, é possível visualizar imediatamente uma visão geral dos argumentos identificados, organizados por categoria de posicionamento (Figura 36): 24 argumentos a favor, 21 contra e 3 neutros.

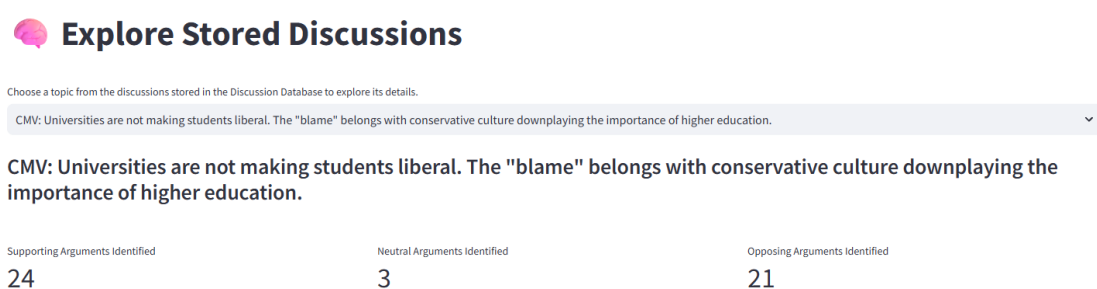


Figura 36 - Demonstração: Contagem de Argumentos Identificados por Posição

### 5.1.6 Exploração de Cada Post e Respetivos Argumentos Extraídos

A interface de exploração permite também a análise detalhada de cada comentário ou resposta incluídos na discussão. Selecionando um *post* de cada *stance* para efeitos de demonstração, como por exemplo, o *Post #32* (Figura 37), classificado como “*Against*”, e o *Post #22* (Figura 38), classificado como “*For*”, obtêm-se os seguintes resultados demonstrados nas figuras:

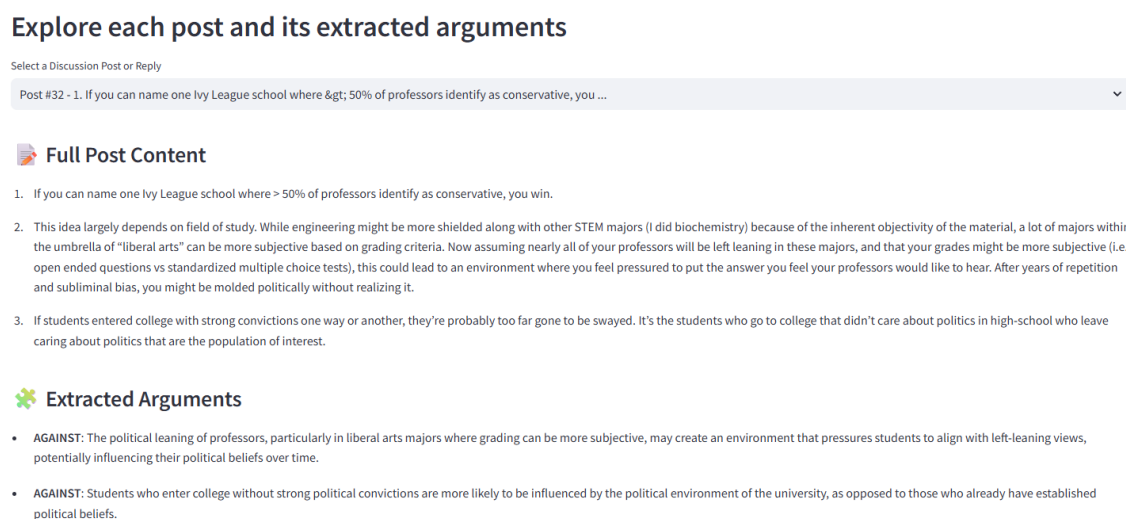


Figura 37 - Demonstração: Análise dos Argumentos Extraídos de Um Post (1)

## Explore each post and its extracted arguments

Select a Discussion Post or Reply

Post #22 - Agreed with your first point. Studies have found a correlation with political ideology and higher ed...

### Full Post Content

Agreed with your first point. Studies have found a correlation with political ideology and higher education but haven't found it to be causal.

Half agree with your second point. Conservatives are less interested in higher education, but I think this has more to do with other demographics aspects that correlate with conservatism but not conservatism itself. For example, conservatives are more likely to be from rural areas with economies that are driven by blue collar work where a college degree is not as advantageous. Additionally, society has been pushing and incentivizing more traditionally underrepresented demographics, which tend to lean liberal, to attend college and incentivize colleges to recruit more students from those demographics. This same pressure doesn't exist for conservative demographics.

I don't necessarily disagree with your conclusion that it might be advantageous for conservative to embrace higher education more, but I find it interesting that people understand systemic factors when analyzing a problem with group they empathize with but lose the ability when it is one they don't.

### Extracted Arguments

- **FOR:** Studies have found a correlation between political ideology and higher education but have not established a causal relationship.
- **FOR:** Conservatives are less interested in higher education due to demographic factors, such as being from rural areas where blue-collar work is more prevalent and a college degree is less advantageous.
- **FOR:** Society has been incentivizing traditionally underrepresented demographics, which tend to lean liberal, to attend college, and colleges are encouraged to recruit more students from these demographics, a pressure not present for conservative demographics.

Figura 38 - Demonstração: Análise dos Argumentos Extraídos de Um Post (2)

### 5.1.7 Exploração do Grafo através de Queries Pré-Definidas

Na última secção de exploração do grafo, o sistema disponibiliza uma interface dedicada à exploração estruturada e minuciosa dos dados processados, através da execução de consultas pré-definidas. Esta funcionalidade permite aceder rapidamente a diferentes perspetivas sobre a discussão analisada, sem a necessidade de interagir diretamente com a visualização em grafo.

A interface apresenta várias opções de exploração, onde o utilizador pode selecionar a consulta pretendida (Figura 39). As queries disponíveis são as seguintes:

- **List of Arguments by Stance:** devolve todos os argumentos extraídos da discussão, agrupados pela sua respetiva stance. Esta consulta permite obter uma visão global e estruturada sobre os principais pontos defendidos em cada lado do debate.
- **Argument Groups by Popularity:** permite visualizar os grupos de argumentos semanticamente semelhantes, incluindo o texto de cada argumento, o resumo gerado para o grupo e a respetiva stance. Esta consulta é particularmente útil para identificar os núcleos argumentativos mais recorrentes.
- **Replies to Supporting Comments** e **Replies to Opposing Comments:** mostram as respostas dirigidas aos comentários com posição “a favor” ou “contra”, respetivamente. Para cada comentário original, são listadas as respostas recebidas e o número total de interações, permitindo identificar focos de debate mais ativos dentro de cada posicionamento.

- **Popular Comments (by score):** lista os comentários mais votados da *thread*, independentemente da *stance*. Esta métrica de popularidade pode indicar os contributos considerados mais relevantes ou impactantes pela comunidade.
- **Discussion URL:** fornece o link direto para a *thread* original no Reddit, permitindo validar o conteúdo original e consultar o contexto completo da discussão.

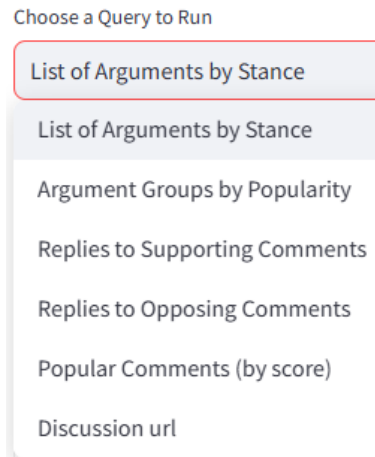


Figura 39 - Demonstração: Consultas disponíveis

Para ilustrar a utilidade prática da funcionalidade de exploração por *queries*, são apresentadas nesta secção os resultados de duas consultas representativas. A Figura 40 apresenta um exemplo da execução da consulta “*List of Arguments by Stance*”. Os argumentos são listados numa tabela, acompanhados da sua classificação de posicionamento. Esta consulta é particularmente útil para uma leitura sintética dos principais pontos defendidos por cada lado, sem necessidade de navegar pela totalidade dos comentários da discussão.

**Results for 'CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.':**

Argument	Stance
Exposure to diverse life experiences in college can lead to a gradual shift in political views.	FOR
College education provides students with skills to evaluate arguments and conduct research, which can lead to a critical reassessment of previously held cons	FOR
Colleges serve as an antidote to indoctrination by equipping students with tools for critical thinking, which is a concern for conservatives.	FOR
Colleges expose students to diverse perspectives and options, which can challenge long-held beliefs and potentially shift individuals from conservative to libe	FOR
Higher education institutions teach critical thinking skills, which can lead individuals to reevaluate and possibly change their political beliefs.	FOR
Humanities and arts majors tend to lean left, whereas majors like business, engineering, and general STEM fields tend to be more moderate or right, suggestin	FOR
Courses in humanities electives are often more politically or socially charged and tend to be left-leaning, which could influence students' views if their entire d	FOR
Professors in STEM fields are predominantly left-leaning, which may influence the political orientation of students.	AGAINST
The political affiliation ratio in engineering fields is 1.6:1 in favor of left-leaning professors, indicating a significant imbalance.	AGAINST
Studies by Professor Robert Anthony Altemeyer indicate that trending away from conservative views occurs primarily due to exposure to diverse individuals, €	FOR

Figura 40 - Demonstração: Execução da Consulta “*List of Arguments by Stance*”

Na Figura 41 está presente a execução da consulta “Argument Groups by Popularity”. As restantes consultas disponíveis são apresentadas no Apêndice B – Demonstração da Solução, a fim de não comprometer a fluidez da leitura deste capítulo.

Results for 'CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.':

	Group Summary	Stance	Argument Count
0	The predominance of left-leaning professors in academia, particularly in subjective fields, influences students' political	AGAINST	7
1	College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift from conservat	FOR	5
2	Demographic and cultural factors contribute to conservatives being less interested in higher education. This includes th	FOR	3
3	Political education should begin at home, with individuals taking responsibility for seeking diverse political ideas and m	FOR	3
4	Certain academic disciplines, particularly in the humanities and arts, are perceived as more left-leaning, which may influ	FOR	2
5	Exposure to diverse environments and experiences, rather than university education itself, is a significant factor in shifti	FOR	2
6	The scientific consensus on issues like climate change and the perception of conservative policies as anti-science may d	FOR	2
7	Cultural preferences influence career paths, with more liberals pursuing academic careers and conservatives opting for	FOR	2
8	The integration of progressive ideologies into course material can lead students to equate these ideologies with factual	AGAINST	2
9	Homogeneous educational environments, such as Bible Colleges, do not lead to a liberal shift, suggesting that diversity	FOR	1

Figura 41 - Demonstração: Execução da Consulta “Argument Groups by Popularity”.

Esta abordagem orientada a *queries* reforça o papel do grafo de conhecimento não apenas como estrutura de armazenamento, mas como base de exploração analítica, permitindo extrair toda a informação relevante de forma flexível, intuitiva e adaptada às necessidades do utilizador.

## 5.2 Avaliação Crítica da Solução

Concluída a demonstração funcional da *pipeline*, este subcapítulo tem como objetivo refletir criticamente sobre o desempenho global da solução desenvolvida. A avaliação não se limita à análise de resultados pontuais, mas procura considerar aspetos fundamentais como a eficácia dos LLMs nas tarefas atribuídas, a coerência dos resultados gerados e a robustez da estrutura de armazenamento e visualização.

Como complemento à análise qualitativa dos resultados, a *framework* de avaliação automática desenvolvida, baseada num modelo GPT-4o, permite analisar criticamente a performance do sistema nas tarefas de classificação de posicionamento, extração de argumentos e agrupamento semântico (*clustering*). Esta abordagem recorre ao mesmo tipo de modelo utilizado na geração inicial dos resultados, tirando partido da sua capacidade de realizar avaliações metalinguísticas, como a verificação da coerência, relevância e consistência argumentativa dos outputs produzidos.

Esta secção encontra-se organizada em torno dos principais módulos da *pipeline*, sendo cada um avaliado de forma individual e crítica, destacando os pontos fortes, as limitações observadas e possíveis caminhos para melhoria futura.

### **5.2.1 Identificação do Tópico de Discussão – Avaliação**

Com o objetivo de avaliar de forma mais rigorosa e independente o desempenho do módulo de identificação do tópico de discussão, foi selecionada uma nova *thread* do mesmo subreddit, distinta da utilizada na demonstração funcional. A análise desta nova discussão permite observar o comportamento do sistema num contexto real, validando a sua capacidade de generalização para tópicos variados.

A *thread* escolhida tem o seguinte título: “*CMV: Trump Will Mark the Beginning of the End of the U.S. as a Global Superpower.*” A escolha desta *thread* (Figura 42) justifica-se pelo facto de abordar um tema sociopolítico atual com forte carga ideológica e moral que permite posições bastante polarizadas — um cenário ideal para testar a *performance* do modelo na identificação do tópico central e das respetivas posições argumentativas.

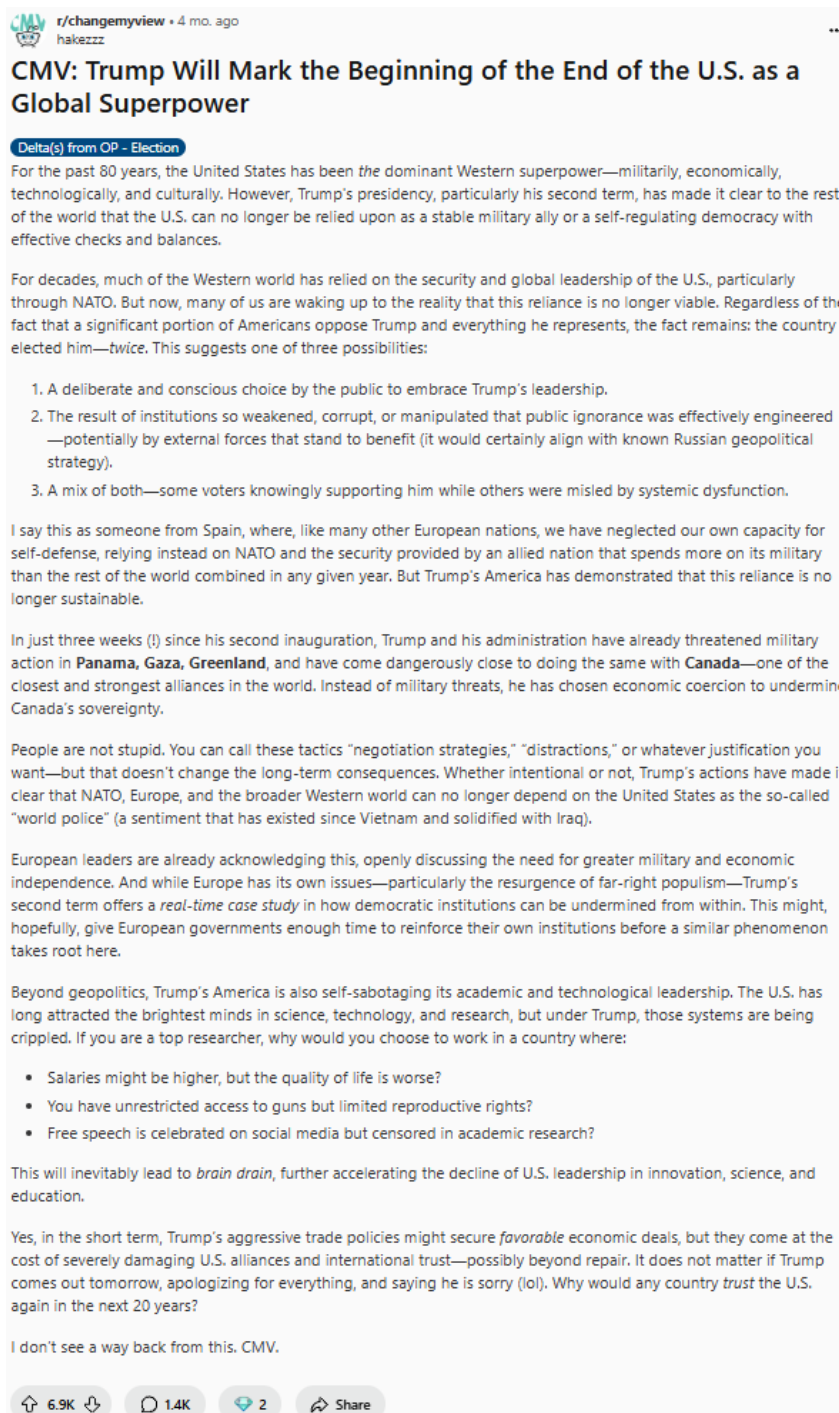


Figura 42 - Avaliação: Thread Utilizada

O tópico identificado pelo sistema foi: ***“The impact of Trump's presidency on the United States' status as a global superpower.”*** A Figura 43 apresenta os resultados obtidos.

## Main Topic: The impact of Trump's presidency on the United States' status as a global superpower.

**For Stance:** This perspective supports the idea that Trump's actions and policies are leading to a decline in U.S. global influence and reliability as a superpower, emphasizing the need for Europe and other nations to seek greater independence in military and economic matters.

**Against Stance:** This viewpoint argues that Trump's presidency does not necessarily signify the end of U.S. global dominance, suggesting that his policies could lead to beneficial outcomes or that the U.S. can recover its standing through future leadership and reforms.

Figura 43 - Avaliação: Tópico Identificado e Explicações das Respetivas *Stances*

Tendo em conta o título da publicação — *"CMV: Trump will mark the beginning of the end of the United States as a global superpower"* — o tópico identificado pelo modelo mostra-se adequado, claro e suficientemente abrangente. A reformulação do título original num enunciado mais neutro e genérico permite enquadrar a discussão num registo mais analítico, sem perder o foco na essência do debate.

No que diz respeito à descrição das *stances*, o sistema demonstrou capacidade interpretativa satisfatória. A posição *"For"* foi corretamente associada à ideia de que as ações e políticas de Trump contribuíram para a erosão da influência global dos EUA, destacando o enfraquecimento da sua fiabilidade e o impulso para que outras potências adotem uma postura mais independente. Já a posição *"Against"* apresenta um contra-argumento válido, ao defender que a presidência de Trump não representa necessariamente o fim da hegemonia norte-americana, e que o país poderá recuperar a sua posição através de reformas e liderança futura.

Ambas as perspetivas foram bem formuladas, refletindo as principais linhas de argumentação que a discussão tende a seguir. A resposta do modelo evidencia clareza, distinção entre os lados do debate e contexto suficiente para suportar os módulos seguintes da *pipeline*. Esta tarefa de identificação do tópico e caracterização das *stances* demonstra-se como a mais fiável do sistema devido à simplicidade da mesma e beneficiando da natureza explícita do *post* original.

### 5.2.2 Extração de Argumentos – Avaliação

Para avaliar o desempenho do módulo de extração de argumentos, foi utilizada a mesma *thread* analisada na secção anterior, garantindo continuidade na avaliação e permitindo observar como os argumentos são extraídos a partir dos comentários e respostas identificados na discussão.

Após o processamento do conteúdo, o sistema identificou um total de **10 comentários**, **31 respostas** e destes, foram extraídos **87 argumentos** e criados **34 clusters**. A Figura 44 apresenta

uma visualização do grafo gerado, onde é possível observar a estrutura hierárquica entre os nós, bem como as ligações entre estes e o tópico da discussão.

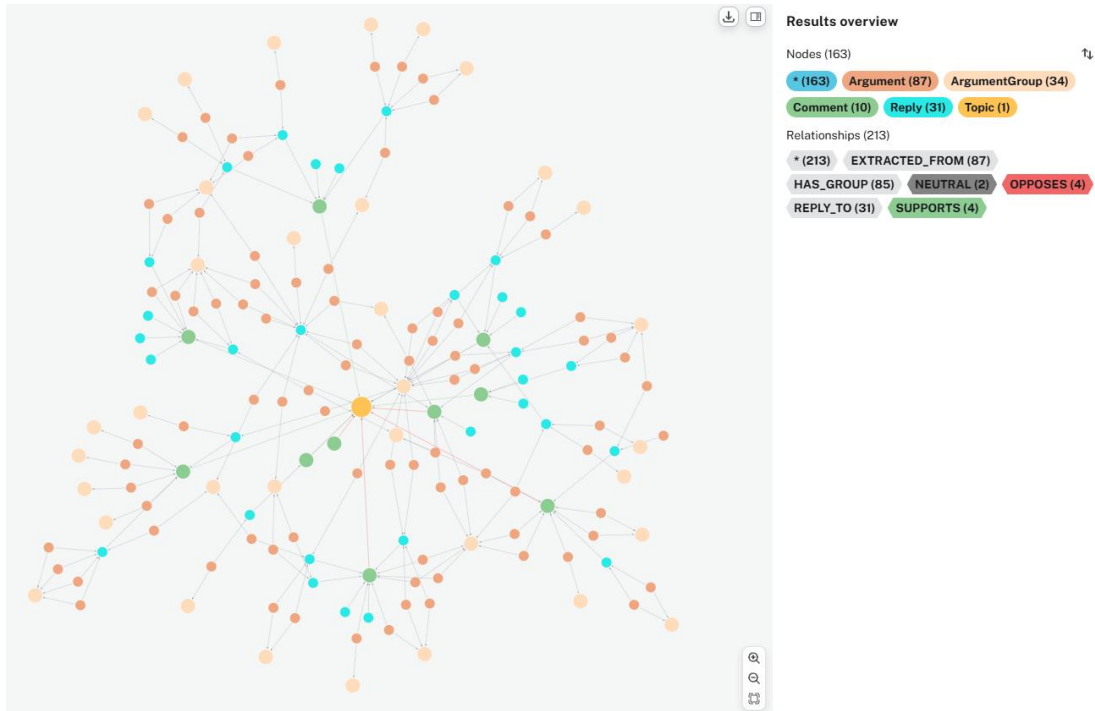


Figura 44 - Avaliação: Grafo de Conhecimento Criado

Dos 87 argumentos gerados, **39 foram classificados como *supporting*, 31 como *neutral* e 17 como *opposing*** (Figura 45). Esta distribuição sugere uma predominância de argumentos que apoiam a premissa apresentada no título original da *thread*, ainda que exista uma presença relevante de posições neutras e contrárias, o que evidencia diversidade argumentativa.

#### CMV: Trump Will Mark the Beginning of the End of the U.S. as a Global Superpower

Supporting Arguments Identified

39

Neutral Arguments Identified

31

Opposing Arguments Identified

17

Figura 45 - Avaliação: Argumentos Extraídos por Classificação

Utilizando a *framework* de avaliação descrita anteriormente, verificou-se que 5 dos 10 comentários e 19 das 31 respostas continham argumentos válidos, o que perfaz um total de 24 conteúdos argumentativos num universo de 41 unidades de texto.

A média de 2.1 argumentos por comentário/resposta com conteúdo argumentativo indica uma densidade argumentativa relevante, reforçando o potencial da discussão para uma análise crítica estruturada.

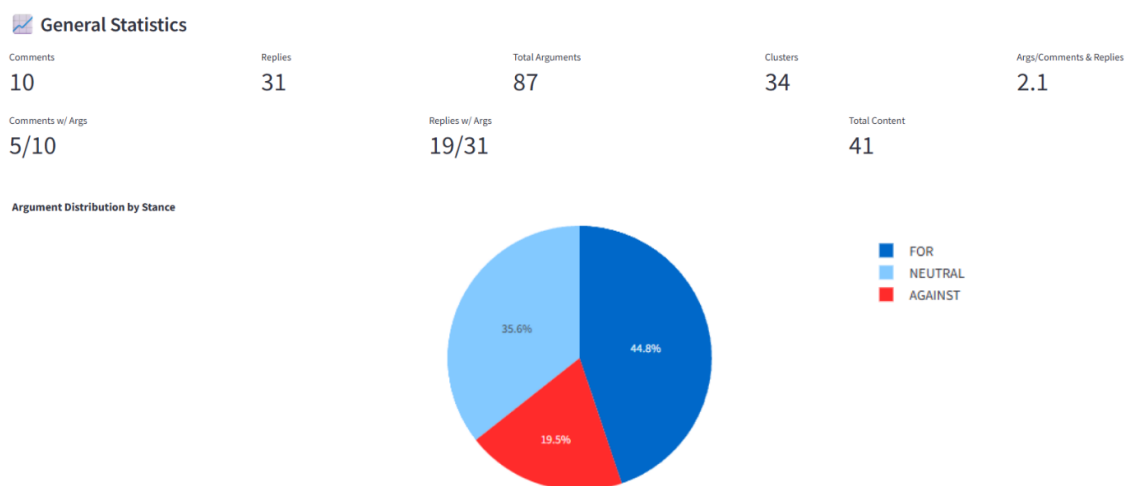


Figura 46 - Avaliação: Estatísticas Gerais da Discussão

### Avaliação Geral da Extração de Argumentos

Com a mesma ferramenta, cada extração foi analisada individualmente com base nos critérios de completude, precisão, qualidade e relevância, e posteriormente foi gerada uma avaliação geral com base nos resultados atribuídos e nas justificações associadas. Os resultados são apresentados na Figura 47.

Argument Extraction

3.96/5

General Analysis:

The overall performance of the argument extraction system is commendable, with an average score of 3.96 out of 5, indicating a generally high level of accuracy and relevance in capturing the main arguments. The system consistently excels in formulating well-structured and self-contained arguments, effectively summarizing key points related to U.S. economic, military, and geopolitical dynamics. However, common weaknesses include minor omissions of context and specificity, such as missing nuanced details or specific mentions that could enhance completeness. To improve extraction quality, the system should focus on capturing more contextual details and ensuring that all relevant nuances and specific references from the original text are included.

Figura 47 - Avaliação: Classificação Geral da Extração de Argumentos

A pontuação média obtida de 3.96 em 5, indica um desempenho globalmente satisfatório na identificação e formulação dos argumentos. De acordo com a explicação geral gerada pelo modelo, o sistema demonstrou forte capacidade em estruturar argumentos de forma clara e autocontida, especificando que foram sintetizados eficazmente os pontos centrais relacionados com as dinâmicas económicas, militares e geopolíticas dos Estados Unidos. Apesar do desempenho positivo, foram identificadas algumas fragilidades recorrentes, sobretudo ao nível da omissão de detalhes contextuais e referências específicas presentes no texto original. Estas lacunas afetam ligeiramente a completude da extração, ainda que não comprometam de forma significativa a validade dos argumentos formulados. Como recomendação, sugere-se que futuras versões do sistema reforcem a atenção à contextualização e à preservação de elementos específicos relevantes, de forma a garantir uma representação ainda mais fidedigna do conteúdo original.

### **Avaliação Individual da Extração de Argumentos**

Para demonstrar a forma como a avaliação é realizada ao nível individual, apresentam-se dois exemplos concretos de comentários analisados. O primeiro exemplo (Figura 48) corresponde a um comentário que explora “a ideia de *brain drain* como consequência indireta da presidência de Trump”. O sistema foi capaz de identificar dois argumentos centrais: “a perda de liderança dos EUA nas áreas de ciência, inovação e educação”, e “a possibilidade de emigração por parte de cidadãos qualificados, agravando essa tendência.”

Original Comment:

\*\*Citing:\*\*

This will inevitably lead to brain drain, further accelerating the decline of U.S. leadership in innovation, science, and education.

\*\*End of Citation\*\*


My only argument here is that you seem to have narrowed your focus to attracting foreign minds.

Consider how many Americans now want to leave the US. Trump supporters won't be the people moving away, for many obvious reasons — it'll be people who have the education and money to do so. The brain drain could happen faster than "inevitably," if you have the double whammy of qualified Americans leaving and others not coming.

Extracted Arguments:

1. The potential brain drain from the U.S. could accelerate the decline of U.S. leadership in innovation, science, and education.
2. Qualified Americans may leave the U.S. due to dissatisfaction, compounding the effects of a brain drain alongside reduced foreign talent attraction.

Evaluate Extraction

 Score: 4/5

Justification:

The system successfully captured the main arguments present in the original comment, focusing on the potential brain drain and its implications for U.S. leadership in various fields. The extracted arguments are accurate, relevant, and well-formulated, providing a clear and self-contained summary of the concerns expressed in the comment. However, the extraction could be slightly improved by explicitly mentioning the context of dissatisfaction among Americans as a driving factor for the brain drain.

Figura 48 - Avaliação: Classificação Individual de Extração (1)

A avaliação automática deste exemplo resultou numa pontuação de 4 em 5, com uma justificação que reconhece a precisão e relevância dos argumentos extraídos, bem como a sua formulação clara e autocontida. No entanto, foi também assinalado que o sistema poderia ter explorado de forma mais explícita o fator de insatisfação interna como motivador da saída dos cidadãos qualificados, o que teria aumentado a completude da extração.

O segundo exemplo presente na Figura 49 diz respeito a uma resposta que aprofunda a tese de que Trump estaria a ser usado para interesses russos, numa estratégia de longo prazo para beneficiar a Rússia. O sistema conseguiu identificar corretamente dois argumentos principais: “a existência de vários agentes russos próximos de Trump”, e “a ideia de que os planos de Putin visam a diminuição da influência dos EUA, alinhando-se com o comportamento do ex-presidente.”

A avaliação resultou numa pontuação de 3 em 5, refletindo uma extração parcialmente correta, mas incompleta. O modelo de avaliação apontou que, apesar de os argumentos extraídos serem geralmente relevantes e bem formulados, estes não captaram totalmente a nuance da intenção consciente ou inconsciente abordada no comentário original. Além disso, foi salientado que o argumento extraído introduz uma conclusão explícita sobre o enfraquecimento dos EUA como superpotência, que não estava diretamente presente no texto original, evidenciando uma leve extrapolação.

Original Reply:

I think you're right and I also think he is fully aware of that and it's intentional. Or if he's not aware then at least there are many around him that are very aware of it and using him as a means to that end. There are way too many Russian assets in his circle to be a coincidence and Putin has been playing the long game. This is all very much according to plan.

Context (Parent Comment):

Parent comment:

I think Trump is merely accelerating what was already slowly taking place. If I were European, I'd be concluding that the people in charge here and the people supporting our leaders are either really...

Extracted Arguments:

1. The presence of numerous Russian assets in Trump's circle suggests a deliberate strategy to weaken the U.S. as a global superpower.
2. Putin has been strategically planning to diminish U.S. influence, and Trump's actions align with this long-term goal.

Evaluate Extraction

Score: 3/5

Justification:

The system captured the main arguments related to the presence of Russian assets and Putin's long-term strategy, which are central to the original reply. However, the extraction lacks completeness as it misses the nuance of intentionality and awareness mentioned in the original text. Additionally, while the extracted arguments are generally accurate, they introduce a specific claim about weakening the U.S. as a global superpower, which is not explicitly stated in the original reply.

Figura 49 - Avaliação: Classificação Individual de Extração (2)

## Considerações Finais

A avaliação qualitativa e quantitativa da extração de argumentos revela um desempenho globalmente satisfatório, com uma pontuação média de 3.96/5, refletindo uma boa capacidade do sistema em identificar e formular argumentos relevantes a partir da linguagem utilizada em discussões informais em ambiente digital. Os exemplos analisados demonstram que, em grande parte dos casos, os argumentos extraídos são claros, autocontidos e semanticamente alinhados com o conteúdo original. Contudo, foram também identificadas limitações recorrentes, nomeadamente a omissão de detalhes contextuais importantes e, por vezes, a extrapolação de ideias não explicitamente mencionadas nos comentários. Estes aspetos afetam a completude e a precisão de algumas extrações. Ainda assim, a consistência geral do módulo e a sua utilidade para os módulos subsequentes da *pipeline* — como a classificação de *stances* ou o *clustering* — confirmam a viabilidade e robustez da abordagem baseada em LLMs para esta tarefa.

### 5.2.3 Classificação de *Stances* dos Argumentos – Avaliação

Após a extração dos argumentos, o sistema procede à atribuição de uma *stance* a cada argumento. Seguindo a mesma metodologia, estas classificações foram avaliadas de maneira geral e individual pela mesma *framework* de avaliação.

## Avaliação Geral da Classificação de *Stances*

A média obtida neste módulo foi de 3.94/5, um valor que indica um desempenho consistente e maioritariamente preciso por parte do sistema de classificação. A análise geral (Figura 50) destaca que o modelo demonstrou uma interpretação robusta das posições argumentativas, com explicações claras tanto para classificações "FOR" como "AGAINST". A distribuição de *stances* revelou um ligeiro predomínio de classificações "FOR", embora isso pareça refletir o conteúdo da discussão original mais do que um viés sistemático do classificador.

Apesar dos bons resultados, a análise também evidenciou alguns aspetos a melhorar. A ausência de classificações com pontuação mínima (1/5) pode indicar uma falta de rigor na identificação de erros graves, ou uma subavaliação de classificações claramente incorretas. Este comportamento pode limitar a capacidade do sistema de identificar falhas mais evidentes e, conseqüentemente, restringir o potencial de melhoria contínua do modelo.

Analisando individualmente este referente à ausência de classificações mínimas, é evidente que o mais provável será a falta de existência de classificações completamente erradas já que a capacidade interpretativa do modelo utilizado é bastante alta.

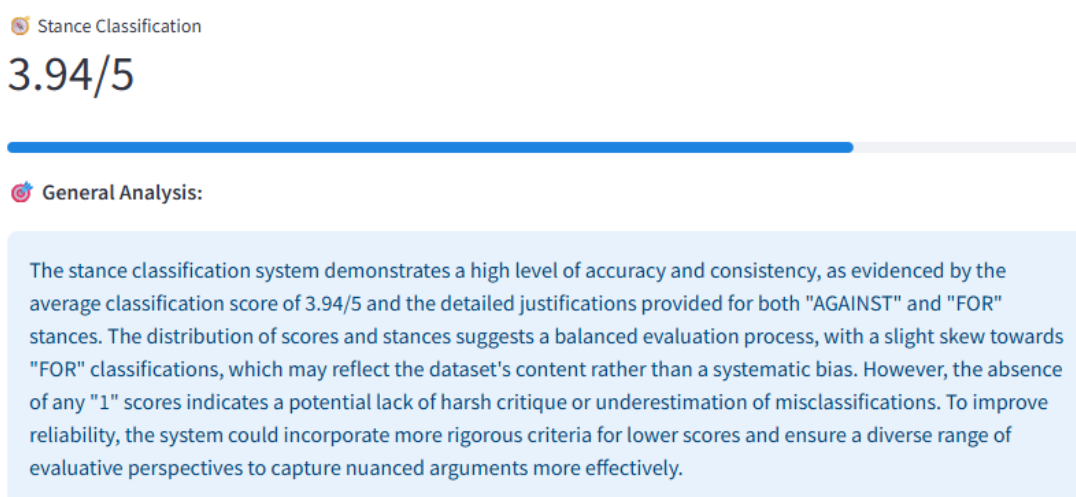


Figura 50 - Avaliação: Média Geral da Classificação de *Stances*

## Avaliação Individual da Classificação de *Stances*

A seguir (Figura 51) apresenta-se um exemplo representativo de uma avaliação individual que recebeu a pontuação máxima (5/5), evidenciando um caso de classificação altamente precisa.


Argument text

America's power is based on its military and world-leading economy, which have not been changed by Trump's actions.

Assigned stance:

AGAINST

Evaluate Stance

 **Score: 5/5**

---

Justification:

The argument states that America's power, based on its military and economy, remains unchanged by Trump's actions, which directly contradicts the topic's claim that Trump will mark the beginning of the end of the U.S. as a global superpower. Therefore, the classification of AGAINST is clearly correct and unambiguous.

Figura 51 - Avaliação: Classificação Individual da Atribuição de *Stances* (1)

Neste exemplo, o argumento apresentado nega diretamente a ideia central do tópico, ao afirmar que o poder dos Estados Unidos, sustentado pela sua força militar e pela economia de topo, não sofreu alterações com as ações de Trump. Esta declaração contrária de forma explícita a premissa de que Trump terá marcado o início do declínio da hegemonia americana, justificando claramente a classificação como “*AGAINST*”. A *stance* atribuída demonstra uma correspondência semântica exata entre o conteúdo do argumento e o posicionamento assumido. A ausência de ambiguidade ou de múltiplas interpretações plausíveis reforça a validade da classificação. O sistema evidenciou aqui um entendimento preciso do sentido do argumento e da sua relação com o tópico em análise, alcançando uma avaliação exemplar.

Outro exemplo (Figura 52) ilustra um caso de maior dificuldade interpretativa, com uma pontuação média de 3 em 5. O argumento apresentado descreve o comportamento dos eleitores norte-americanos como sendo motivado por fatores circunstanciais, como a pandemia ou a inflação, e não como uma rejeição explícita às políticas ou figura de Trump. Esta observação sugere uma abordagem descritiva que evita tomar uma posição clara face ao impacto de Trump enquanto agente no declínio dos EUA. Contudo, o argumento também pode ser interpretado de forma indireta como minimizando o papel de Trump nesse processo, o que pode inclinar ligeiramente a classificação para “*AGAINST*”. Esta ambiguidade justifica a pontuação intermédia, uma vez que, apesar da escolha de “*NEUTRAL*” ser defensável, o modelo

poderia ter aprofundado a análise para clarificar se o argumento realmente não assume qualquer posicionamento relevante no debate. O caso ilustra bem os desafios na distinção entre argumentos verdadeiramente neutros e aqueles que apenas aparentam neutralidade devido à sua linguagem mais descritiva.


Argument text

U.S. voters' actions, such as Biden's 2020 victory, are often reactions to immediate issues like the pandemic or inflation, rather than a repudiation of Trump or his policies.

Assigned stance:

NEUTRAL

Evaluate Stance

 Score: 3/5

Justification:

The argument suggests that U.S. voters' decisions are driven by immediate issues rather than a direct response to Trump or his policies, which could imply a neutral stance. However, it could also be interpreted as indirectly supporting the idea that Trump's influence is not the primary factor in the U.S.'s potential decline as a superpower, making the classification somewhat ambiguous.

Figura 52 - Avaliação: Classificação Individual da Atribuição de *Stances* (2)

### Considerações Finais

A avaliação da componente de classificação de *stances* demonstra um desempenho consistente e maioritariamente fiável, com uma média de pontuação de 3.94/5 atribuída pelo modelo avaliador. A análise qualitativa das classificações revela uma boa correspondência entre os argumentos extraídos e os posicionamentos atribuídos, com especial destaque para a precisão em argumentos que expressam posições claras e diretas face ao tópico.

Contudo, em casos mais ambíguos, onde os argumentos assumem um tom descritivo ou abordam causas e consequências sem ligação explícita ao tópico central, a atribuição de *stance* torna-se mais desafiante, levando a interpretações divergentes.

Em suma, o desempenho observado nesta fase da *pipeline* confirma a eficácia da abordagem proposta, sendo esta componente uma mais-valia para o entendimento estrutural dos debates analisados.

## 5.2.4 Clustering dos Argumentos – Avaliação

A avaliação do desempenho do módulo de *clustering* seguiu a mesma abordagem sistemática aplicada nos anteriores, utilizando a *framework* desenvolvida para avaliar cada grupo de argumentos com base na coerência semântica do *cluster*, na qualidade do resumo explicativo, na completude do agrupamento e na granularidade adequada. No Apêndice C – Avaliação: *Clusters Criados*, encontra-se uma tabela representativa de alguns *clusters* criados e respetivos argumentos para efeitos de demonstração de resultados.

### Avaliação Geral do Clustering

O módulo de *clustering* demonstrou um desempenho sólido, tendo obtido uma avaliação geral média de 4 em 5 (Figura 53). Segundo o modelo avaliador, a maioria dos agrupamentos apresentam uma forte coerência semântica e uma organização temática clara. A elevada proporção de *clusters* avaliados com pontuações de 4 ou 5 indica que os argumentos foram, em geral, corretamente agrupados com base em semelhanças conceptuais e discursivas.

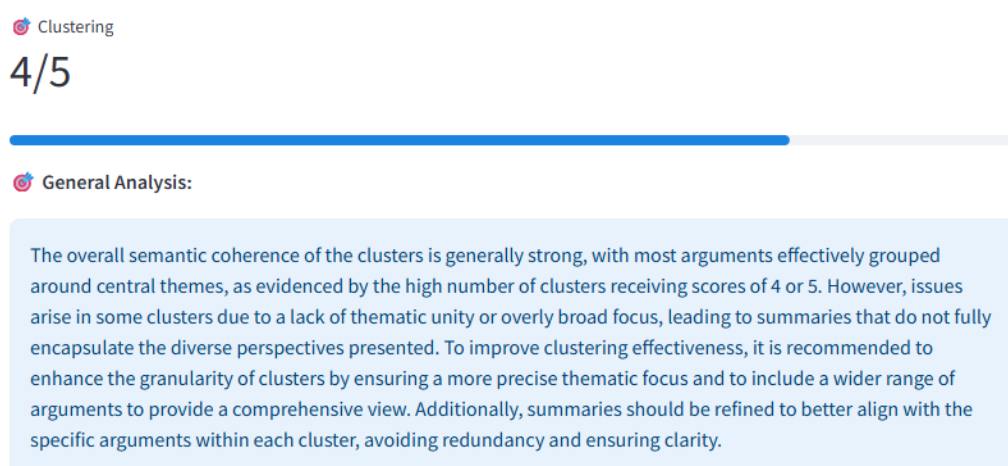


Figura 53 - Avaliação: Classificação Geral do Clustering

Entre os pontos fortes, é destacada a capacidade do sistema para identificar e organizar argumentos em torno de ideias centrais bem definidas, facilitando a leitura e compreensão dos principais eixos argumentativos da discussão. No entanto, algumas limitações foram também observadas, nomeadamente na existência de *clusters* com temas demasiado abrangentes ou heterogéneos, o que compromete a especificidade dos resumos gerados.

Para melhorar o desempenho deste módulo, recomenda-se uma maior granularidade na criação dos grupos, assegurando que cada *cluster* aborde uma ideia distinta e bem delimitada.

Adicionalmente, os resumos poderiam ser ajustados para refletir com mais precisão os argumentos incluídos, evitando redundâncias e melhorando a clareza global da estrutura argumentativa.

### Avaliação Individual do *Clustering*

A Figura 54 apresenta um exemplo de avaliação a um *cluster* individual. O agrupamento escolhido é o segundo da discussão com maior número de argumentos, incluindo 7 argumentos extraídos de comentários e 1 de respostas.

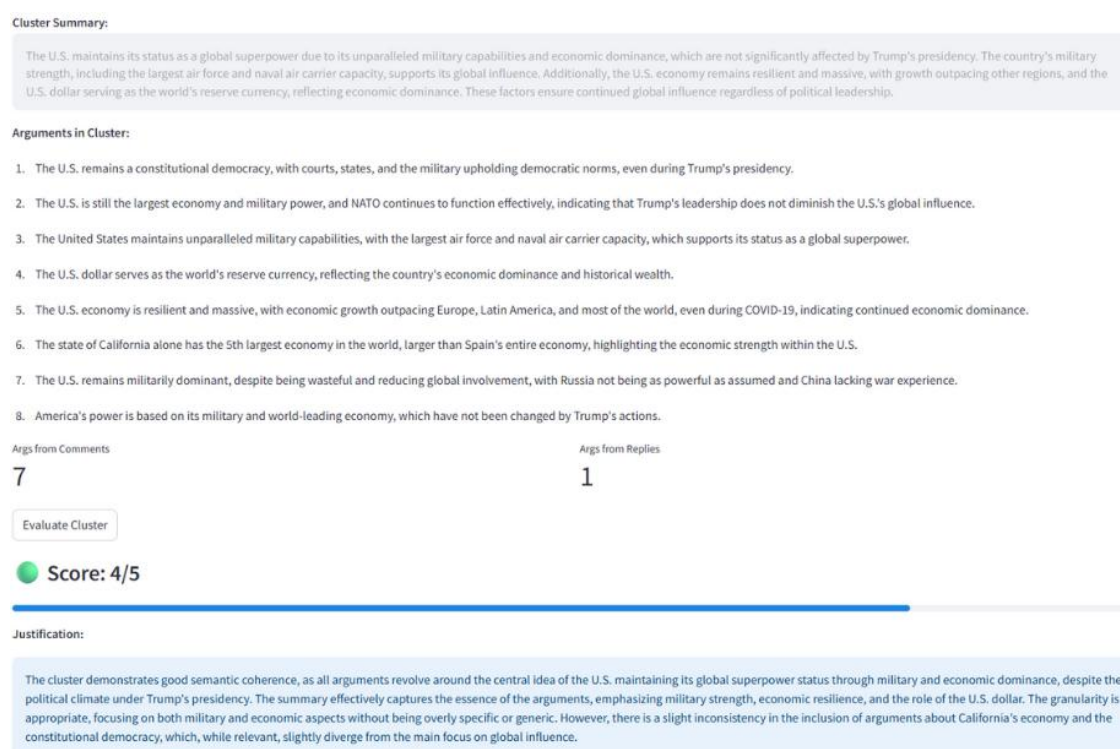


Figura 54 - Avaliação: Classificação Individual do *Clustering* (1)

Este cluster apresenta uma organização tematicamente sólida em torno da ideia central de que os Estados Unidos mantêm o seu estatuto de superpotência global, independentemente da presidência de Donald Trump. Os argumentos incluídos abordam, de forma consistente, os pilares económicos e militares que sustentam essa posição, desde “a dimensão da economia norte-americana” e “o papel do dólar como moeda de reserva global”, até à “supremacia militar expressa em números e capacidades”.

O resumo gerado reflete adequadamente esta diversidade argumentativa, conseguindo condensar as principais ideias num parágrafo coeso e informativo. A granularidade do

agrupamento também é apropriada, permitindo capturar diferentes dimensões da mesma tese (económica e militar) sem dispersar o foco principal.

Ainda assim, a inclusão de argumentos que referenciam coisas como “a estabilidade institucional” (como a referência à democracia constitucional) ou a economia de um estado específico (Califórnia) introduzem uma ligeira quebra na coesão temática. Embora relevantes, estes pontos desviam parcialmente da narrativa central sobre a projeção de poder global.

A avaliação de 4/5 revela, assim, um desempenho bastante positivo, com pequenas oportunidades de melhoria ao nível da seleção e delimitação temática dos argumentos no *cluster*.

O segundo exemplo de avaliação individual (Figura 55) apresenta uma pontuação de 2 em 5, refletindo limitações significativas na coerência semântica e na precisão do resumo gerado.

The screenshot displays a cluster evaluation interface. At the top, the 'Cluster Summary' is 'The historical and strategic implications of Trump's "America First" ideology'. Below this, 'Arguments in Cluster' lists three points: 1. Trump's disruption of the neoliberal trade order places significant strain on a global system that will soon face unprecedented challenges due to climate change. 2. The "America First" ideology, as noted by Mitch McConnell, contributed to the acceleration of World War II, suggesting it could again increase global uncertainty and fear. 3. Historical examples like Vietnam, Iraq, and Afghanistan show that financial investment alone is insufficient without public support for military actions, and currently, public support for Trump's military actions is waning. The interface shows 'Args from Comments' at 0 and 'Args from Replies' at 3. A red 'Evaluate Cluster' button is visible. Below the button, the score is 'Score: 2/5' with a red circle next to the number 2. A blue bar indicates the score level. The 'Justification' section states: 'The arguments in the cluster lack semantic coherence as they address different aspects of Trump's "America First" ideology without a clear central theme. The first argument focuses on trade and climate change, the second on historical parallels with World War II, and the third on military actions and public support. The summary does not adequately capture the essence of these diverse arguments, as it suggests a focus on historical and strategic implications, which is not consistently reflected in the arguments. The cluster is too broad, attempting to cover trade, military, and historical aspects without a unifying thread.'

Figura 55 - Avaliação: Classificação Individual do *Clustering* (2)

Apesar de todos os argumentos se relacionarem com a presidência de Donald Trump e possíveis consequências das suas políticas, cada um aborda uma dimensão distinta: um foca-se no “impacto da política comercial no contexto das alterações climáticas”, outro nas “implicações históricas da ideologia «*America First*»”, e o terceiro na “perda de apoio público a ações militares”. A falta de um fio condutor evidente entre estes subtemas compromete a unidade temática do grupo.

Além disso, o resumo atribuído ao cluster é vago e genérico, ao referir “implicações históricas e estratégicas” sem capturar com clareza a diversidade dos conteúdos presentes. Esta avaliação demonstra que, embora os argumentos individualmente sejam válidos, o agrupamento carece de foco temático e o resumo falha em representar adequadamente a complexidade do conjunto. O exemplo evidencia a importância de uma maior granularidade e de critérios mais rigorosos na formação dos *clusters*.

### **Considerações Finais**

A avaliação do módulo de *clustering* evidencia um desempenho sólido, com uma média geral de 4/5, refletindo uma capacidade consistente do sistema em agrupar argumentos com base em semelhanças temáticas. A maioria dos *clusters* gerados demonstrou boa coerência semântica, com resumos representativos e agrupamentos centrados em eixos argumentativos bem definidos.

Contudo, também foram identificadas limitações importantes em casos mais complexos. Em particular, alguns *clusters* apresentaram excesso de heterogeneidade temática, reunindo argumentos que, embora relacionados de forma ampla ao tópico geral, não partilhavam um foco argumentativo claro. Esta fragilidade afetou a qualidade dos resumos, que por vezes não captavam adequadamente as nuances ou diversidade dos argumentos incluídos. Adicionalmente, foram também identificados *clusters* com sumários relativamente vagos, não excluindo a necessidade de leitura dos argumentos constituintes.

Os exemplos analisados ilustram bem esta dualidade: por um lado, *clusters* bem organizados e sumários precisos; por outro, agrupamentos demasiado abrangentes com resumos vagos ou genéricos. A performance atual do módulo é, ainda assim, bastante promissora, podendo ser reforçada através da melhoria da granularidade dos agrupamentos e de uma maior sensibilidade semântica no processo de agrupamento. O refinamento dos critérios de agrupamento e a revisão automática dos resumos são áreas recomendadas para desenvolvimento futuro.

#### **5.2.5 Considerações Finais da Avaliação Global da Solução**

A avaliação dos três módulos centrais da *pipeline* (Figura 56) — extração de argumentos, classificação de posicionamento e *clustering* — revela um desempenho globalmente positivo,

com médias de pontuação elevadas e respostas coerentes do sistema ao longo de todas as fases. Entre os módulos avaliados, o de extração de argumentos demonstrou bastante robustez, com uma média de 3.96/5, evidenciando capacidade para captar ideias centrais de forma clara e estruturada, ainda que com algumas omissões contextuais pontuais.

A classificação de *stances* apresentou resultados igualmente satisfatórios, com uma média de 3.94/5, destacando-se pela consistência na atribuição de posicionamentos argumentativos, embora tenha revelado ligeira ambiguidade em argumentos mais neutros ou interpretáveis.

Já o *clustering* obteve a média mais alta de 4/5, mostrando boa organização temática na maioria dos grupos formados, mas revelando também fragilidades quando os argumentos são semanticamente dispersos.

Em termos gerais, observou-se uma tendência do sistema para favorecer classificações moderadas (pontuações entre 3 e 4), com escassez de avaliações extremamente negativas (1 ou 2), o que poderá indicar uma ligeira suavização do critério avaliativo adotado pelo modelo. A integração de justificações detalhadas em cada avaliação permitiu enriquecer a análise e identificar padrões de erro que não seriam detetáveis apenas com base na pontuação.

Estas avaliações automáticas, embora dependentes de um LLM, mostram-se eficazes na detecção de falhas específicas e na orientação de melhorias. A principal recomendação global incide na necessidade de aperfeiçoar a granularidade e precisão semântica ao longo de toda a *pipeline*, bem como na eventual inclusão de mecanismos de validação cruzada para reforçar a fiabilidade dos resultados.



Figura 56 - Avaliação: Classificação Global dos Módulos



## 6 Conclusões

O presente capítulo visa apresentar uma reflexão final sobre o trabalho desenvolvido ao longo desta dissertação. Para tal, procede-se a uma análise sistemática do grau de concretização dos objetivos inicialmente propostos, bem como à apresentação dos principais resultados obtidos durante a conceção, implementação e avaliação da solução. São igualmente discutidas as limitações identificadas e possíveis direções para trabalho futuro, que poderão contribuir para a evolução da abordagem aqui proposta. Por fim, é discutido o impacto científico do projeto, nomeadamente no que diz respeito à sua relevância para a comunidade académica e ao contributo gerado através da produção e submissão de um artigo científico baseado no trabalho realizado.

### 6.1 Objetivos Cumpridos

A presente dissertação cumpriu integralmente os objetivos definidos na fase inicial do projeto, tendo resultado numa solução funcional, robusta e alinhada com as metas conceptuais e técnicas previamente traçadas. Esta secção visa retomar os objetivos previamente definidos no Subcapítulo 1.3, avaliando o seu grau de concretização ao longo da dissertação. Para cada objetivo, é indicada a sua realização e apresentada uma explicação sucinta da forma como foi cumprido, incluindo referências às secções onde a respetiva implementação se encontra descrita.

- Objetivo 1: Realizar uma revisão abrangente da literatura e estado da arte sobre argumentos e negociação baseada em argumentação, *Large Language Models*, Grafos de Conhecimento e análise de argumentos em redes sociais.

Este objetivo foi alcançado através da análise crítica e estruturada da literatura, apresentada no Capítulo 2. As Secções 2.1 a 2.4 cobrem os principais pilares do projeto, abordando a teoria da argumentação, os fundamentos e aplicações de *Large Language Models*, a utilização de grafos de conhecimento, bem como o estado da arte da análise e classificação de argumentos em redes sociais.

- Objetivo 2: Conceber e implementar uma *pipeline* de processamento baseada em LLMs, capaz de sumarizar uma discussão realizada numa rede social, identificando o tópico central de uma discussão, e extraíndo, classificando e agrupando os argumentos centrais utilizados.

A *pipeline* foi concebida de forma modular e descrita detalhadamente na Secção 3.4, onde são apresentados os seus componentes e o fluxo de informação. No que toca ao seu desenvolvimento, o Capítulo 4, descreve todo o processo. Quanto ao desenvolvimento do módulo de identificação do tópico, este está descrito em 4.3.3, enquanto a extração, classificação e agrupamento de argumentos estão desenvolvidos no subcapítulo 4.3.6.

- Objetivo 3: Integrar um sistema de dupla classificação de posicionamento, atribuindo uma primeira classificação aos comentários e uma segunda, mais granular, aos argumentos extraídos individualmente, de forma a atingir a maior precisão de classificação possível e utilidade analítica da representação discursiva.

Esta funcionalidade encontra-se descrita nas Secções 4.3.4 e 4.3.6, onde se implementa a classificação de *stance* tanto para os comentários como para os argumentos extraídos. Esta abordagem permite uma análise mais rica e detalhada do posicionamento discursivo dos participantes da discussão.

- Objetivo 4: Agrupar argumentos semelhantes através de agrupamento semântico com LLMs, garantindo a coerência temática e a preservação da orientação argumentativa (a favor, contra, neutro).

O processo de agrupamento semântico foi implementado com recurso a modelos generativos e está descrito na Secção 4.3.6, mais especificamente na subsecção de título “*Clustering dos Argumentos Extraídos*”, sendo avaliado na Secção 5.2.4, onde se analisam a coerência e relevância dos grupos formados.

- Objetivo 5: Construir e armazenar um grafo de conhecimento estruturado, onde se representem explicitamente as entidades da discussão (tópico, comentários, respostas, argumentos e grupos de argumentos), bem como as relações entre estas.

O processo de desenvolvimento do módulo de construção dos grafos é descrito na Secção 4.3.7, com a sua visualização demonstrada no Subcapítulo 5.1.4. O grafo inclui todos os elementos fundamentais da discussão e apresenta as relações semânticas e argumentativas de forma clara e estruturada.

- Objetivo 6: Permitir a análise detalhada da discussão analisada, através de consultas ao grafo de conhecimento criado.

Foram desenvolvidas várias *queries* de exploração dos grafos criados, descritas em 4.3.8 e ilustradas com exemplos práticos nas Secções 5.1.5 a 5.1.7, permitindo ao utilizador obter *insights* e padrões discursivos com base nas relações entre entidades.

- Objetivo 7: Avaliar a qualidade da extração, classificação e agrupamento de argumentos, utilizando *prompts* específicas para análise automática da performance de cada componente do sistema com base em modelos generativos e critérios qualitativos.

A avaliação automática foi realizada com base numa *framework* construída especificamente para este fim, apresentada na Secção 3.4.1, com o seu desenvolvimento descrito em 4.3.9. Os resultados detalhados da avaliação encontram-se no Capítulo 5.2, divididos por componente (extração, classificação e *clustering*), com destaque para a utilização de LLMs como juízes automatizados da qualidade.

- Objetivo 8: Analisar as limitações, falhas e enviesamentos do sistema, identificando pontos de melhoria futura e discutindo o equilíbrio entre desempenho, controlo e interpretabilidade na utilização de modelos generativos em contextos argumentativos.

A análise crítica dos pontos fracos da solução encontra-se na Secção 6.3.1, sendo complementada por sugestões de melhoria e expansão futura do sistema na Secção 6.3.2. São discutidas limitações relacionadas com a generalização, a dependência de APIs externas, e os desafios associados à mitigação de enviesamentos algorítmicos.

Paralelamente ao cumprimento dos objetivos conceptuais, foram também validados todos os requisitos funcionais e não funcionais definidos para o sistema. A ferramenta desenvolvida cumpre com sucesso os requisitos funcionais, incluindo a recolha de dados (RF1), a identificação do tópico (RF2), a classificação de comentários (RF3) e argumentos (RF6), o agrupamento com base na posição (RF4) e na semântica (RF7), a geração de resumos (RF5, RF8), a criação do grafo de conhecimento (RF9) e a apresentação clara e acessível dos resultados (RF10). A avaliação do sistema por LLMs responde ainda ao requisito funcional de validação automática da performance (RF11).

Relativamente aos requisitos não funcionais, a interface criada assegura a usabilidade de todo o tipo de utilizadores (RNF1). Devido à estratégia aplicada de extração máxima de 50 elementos discursivos (comentários e respostas), discussões de grande dimensão não afetam negativamente o desempenho do sistema (RNF2). Toda a arquitetura segue princípios de modularidade (RNF3), integração entre componentes (RNF4) e escalabilidade (RNF5). A modularidade e integração são provadas no *design* e arquitetura da *pipeline* desenvolvida. Já a escalabilidade do sistema é provada através das várias abordagens e tentativas experimentais descritas, em que o sistema não sofreu qualquer mudança estrutural drástica (por exemplo, a adição de classificações individuais aos argumentos, que passou apenas pela adição de uma nova *prompt* e lógica correspondente ao módulo de criação de grafos existente). Adicionalmente, o sistema recorre a LLMs (RNF6) e a uma base de dados de grafos (RNF7) conforme definido. Foram ainda asseguradas boas práticas de ética e conformidade (RNF8), através da anonimização total dos utilizadores e foco exclusivo nos conteúdos argumentativos das discussões analisadas.

## 6.2 Resultados Alcançados

Os resultados obtidos demonstram a viabilidade e o valor prático da abordagem proposta. A solução desenvolvida foi testada com múltiplas discussões públicas retiradas do Reddit, com temas política e socialmente sensíveis. As principais conquistas alcançadas incluem:

- **Clareza e Organização da Informação:** A transformação de debates caóticos em estruturas argumentativas claras permitiu evidenciar os principais pontos de vista presentes nas discussões. A separação dos argumentos por posicionamento e a criação de *clusters*, acompanhados de um sumário descritivo, contribuíram para a desconstrução e simplificação dos debates.
- **Extração de Argumentos Relevantes e Distintos:** A *pipeline* é capaz de extrair proposições argumentativas auto-contidas e distintas, ignorando declarações vagas, ofensivas ou redundantes. O modelo demonstrou boa capacidade de interpretar o contexto dos comentários e identificar apenas as proposições com valor argumentativo real.
- **Classificação Precisa de Posicionamentos:** A utilização de um modelo dedicado à tarefa de *stance detection* após a extração dos argumentos permitiu garantir que cada argumento, e não o comentário como um todo, fosse corretamente classificado segundo a sua orientação argumentativa.
- **Criação de *Clusters* Argumentativos Coerentes:** O sistema conseguiu agrupar argumentos que expressam a mesma ideia central, mesmo quando formulados de maneiras diferentes. Os sumários atribuídos a cada grupo mostraram-se úteis para captar rapidamente o conteúdo comum de cada *cluster*, promovendo uma compreensão mais rápida e eficiente do debate.
- **Visualização e Capacidade de Análise:** A apresentação visual e interativa dos resultados, possibilita a navegação e exploração dos dados, tornando possíveis análises detalhadas das discussões e elementos argumentativos que as compõem.
- **Avaliação automatizada:** Foram definidos três módulos de avaliação baseados em *prompts* que classificam a qualidade da extração, da classificação de posicionamento e

do *clustering*. Esta avaliação qualitativa, baseada em modelos, permite identificar falhas, inconsistências e oportunidades de melhoria.

Estes resultados evidenciam não só o cumprimento dos objetivos propostos como também o potencial da solução para aplicação em contextos reais, tanto no domínio acadêmico e científico como em plataformas de moderação de conteúdos e redes sociais.

## **6.3 Limitações e Trabalho Futuro**

Com o sistema implementado e avaliado, importa agora refletir sobre as suas principais limitações e identificar caminhos viáveis para o seu aperfeiçoamento e evolução futura. Para tal, este subcapítulo encontra-se dividido em duas partes: uma dedicada à análise das limitações observadas e outra às propostas de trabalho futuro.

### **6.3.1 Limitações Identificadas**

Apesar dos resultados alcançados e da eficácia geral da solução proposta, é importante reconhecer que a solução apresenta algumas limitações que se tornaram evidentes ao longo do desenvolvimento e da fase de avaliação. Estas limitações não comprometem a validade do sistema, mas constituem oportunidades relevantes para o seu aperfeiçoamento futuro.

#### **Inconsistência nas decisões do modelo**

Não obstante de serem utilizados LLMs de elevado desempenho, verificou-se que, por vezes, o mesmo modelo pode gerar classificações ou agrupamentos diferentes para argumentos semelhantes, ou até dar respostas incoerentes com as suas próprias instruções. Esta inconsistência torna mais difícil validar automaticamente os resultados e compromete a reprodutibilidade exata do processo em múltiplas execuções. Ainda, em discussões com vários comentários e respostas, é comum que o sistema não identifique argumentos em vários destes, mesmo que pareçam conter elementos argumentativos relevantes, contribuindo para a imprevisibilidade do modelo.

## **Deteção de Ironia, Sarcasmo e Mensagens Subliminares**

Na mesma linha de raciocínio, os modelos utilizados ainda demonstram alguma dificuldade na deteção de ironia, sarcasmo ou em situações que o utilizador deixa mensagens subliminares, criando problemas nas tarefas de classificação de *stances* ou extração de argumentos.

### **Limitação na cobertura das discussões**

Apesar de se ter adotado a abordagem de selecionar os 10 comentários mais relevantes (por *score*) permitindo obter uma ideia bastante profunda das discussões, e permitindo tempos de resposta e desempenhos decentes, esta seleção pode excluir algumas intervenções argumentativamente ricas que não receberam tantos votos positivos. Isto afeta a representatividade da amostra e pode enviesar a análise para perspetivas mais populares ou consensuais, deixando de fora argumentos minoritários, mas possivelmente relevantes.

### **6.3.2 Trabalho Futuro**

Várias direções promissoras podem ser seguidas para a evolução da solução. Em primeiro lugar, destaca-se a possibilidade de melhorar e diversificar os *prompts* utilizados, afinando instruções, adicionando exemplos (através de *few-shot learning*) e testando variações para tarefas mais desafiantes, como a distinção entre argumentos neutros e pouco explícitos. Em segundo lugar, poderá ser benéfico explorar modelos alternativos aos utilizados atualmente, testando outras variantes de LLMs mais especializadas (como os modelos de *reasoning*) ou modelos *fine-tuned* para as tarefas específicas realizadas. Outra direção pertinente seria a implementação de uma abordagem Human-in-the-Loop, permitindo ao utilizador validar, corrigir ou complementar automaticamente os resultados obtidos em cada módulo. Esta integração de *feedback* humano no processo de extração e análise de argumentos seria útil para reforçar a fiabilidade da solução e permitir a sua utilização em contextos reais e mais críticos.

Adicionalmente, poderia ser adotada uma abordagem contínua em que depois do processamento e representação inicial dos dados de uma discussão, esta continuaria a ser monitorizada para serem adicionados e incluídos todos os novos comentários que fossem feitos. Para isto, seria também necessário extrair e processar discussões inteiras, o que dependeria muito da capacidade do sistema em que corre a solução e dos modelos utilizados.

Finalmente, outro possível melhoramento passa pela expansão da solução para outras plataformas sociais além do Reddit, incluindo dados provenientes do X (Twitter), fóruns públicos, ou plataformas de debate estruturado, o que implicaria ajustes nos mecanismos de recolha e pré-processamento, mas aumentaria substancialmente o alcance e aplicabilidade da solução.

## 6.4 Impacto Científico

O desenvolvimento deste projeto culminou na produção e submissão de um artigo científico, em coautoria com o colega Hugo Jorge, intitulado *“Extracting Arguments and Motivations in Large-Scale Group Decision-Making: An approach using Large Language Models and Knowledge Graphs”*, o qual foi submetido à conferência *“AIDEM 2025 - International Tutorial and Workshop on Artificial Intelligence, Data Analytics and Democracy”*<sup>16</sup>, evento paralelo na *“European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases - ECML PKDD”*<sup>17</sup>. Este artigo representa, possivelmente, um contributo relevante para a investigação aplicada na área da análise argumentativa em ambientes deliberativos online, com foco no apoio à tomada de decisão coletiva em larga escala (LSGDM).

O trabalho submetido apresenta uma abordagem dual: por um lado, centra-se na extração e organização de argumentos em debates informais provenientes do Reddit, e por outro, complementa esta estrutura com a análise das motivações subjacentes a cada argumento, com base no modelo de Desenvolvimento em Escala Humana de Max-Neef. A dissertação aqui apresentada corresponde à primeira dessas abordagens — de extração, classificação, *clustering* e representação argumentativa com recurso a LLMs e Grafos de Conhecimento — estando plenamente representada na primeira metade do artigo. A inclusão desta investigação em formato de publicação científica evidencia não apenas o rigor metodológico e a relevância prática da solução proposta, mas também o seu potencial de aplicação em contextos mais formais de deliberação digital, como plataformas de *e-democracy* e participação cívica promovidas por instituições públicas. Este artigo estabelece assim uma base possivelmente útil e interessante para futuros trabalhos na área de deliberação coletiva assistida por inteligência artificial, servindo como um exemplo prático concreto.

---

<sup>16</sup> <https://aidem2025.isti.cnr.it/>

<sup>17</sup> <https://ecmlpkdd.org/2025/>

# Referências

- Abbas, H. (2024). How to Develop Modular LLM Pipelines? - Hakeem Abbas - Medium. Medium. <https://medium.com/@hakeemmysd/how-to-develop-modular-llm-pipelines-31faa8fae136>
- Abu-Salih, B., Al-Tawil, M., Aljarah, I., Faris, H., & Wongthongtham, P. (2020). Relational Learning Analysis of Social Politics using Knowledge Graph Embedding. <http://www.foaf-project.org/>
- AI Act - EU Artificial Intelligence Act. (2024). EU Artificial Intelligence Act. <https://artificialintelligenceact.eu/ai-act-explorer/>
- ALDayel, A., & Magdy, W. (2021). Stance detection on social media: State of the art and trends. *Information Processing and Management*, 58(4). <https://doi.org/10.1016/j.ipm.2021.102597>
- Alpert, L. (2016). Major Issues in Philosophy: I. What is an Argument?
- Amaratunga, T. (2023). Understanding Large Language Models. In *Understanding Large Language Models*. Apress. <https://doi.org/10.1007/979-8-8688-0017-7>
- Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R., & Minor, M. (2011). Cats Rule and Dogs Drool!: Classifying Stance in Online Debate.
- Antila, A. (2024). X as a Platform for Political Engagement: Analyzing Information Encountering and Participatory Democracy Among Candidates in the 2023 Finnish Parliamentary Election. In *Anniina Antila*.
- Antypas, D., Ushio, A., Camacho-Collados, J., Neves, L., Silva, V., & Barbieri, F. (2022). *Twitter Topic Classification*. <https://huggingface.co/datasets/>
- AWS. (2024). What Are Transformers? - Transformers in Artificial Intelligence Explained - AWS. Amazon Web Services, Inc. <https://aws.amazon.com/what-is/transformers-in-artificial-intelligence/>
- Broadhead, G. (2023). A Brief Guide To LLM Numbers: Parameter Count vs. Training Size. Medium. <https://gregbroadhead.medium.com/a-brief-guide-to-llm-numbers-parameter-count-vs-training-size-894a81c9258>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901
- Cambridge Dictionary. (2020). ARGUMENT | meaning in the Cambridge English Dictionary. Cambridge.org.
- Carneiro, J. M. R. (2017). Negotiation in Group Decision Support Systems - An Approach based in Argumentation and Satisfaction.
- Christensen, J. (2024). *What is generative AI? Definition, applications, examples*. Neo4j. <https://neo4j.com/blog/genai/what-is-generative-ai/>
- Clancy, L. (2022). Americans differ by party, ideology over the impact of social media on U.S. democracy. Pew Research Center. <https://www.pewresearch.org/short-reads/2022/12/14/americans-differ-by-party-ideology-over-the-impact-of-social-media-on-u-s-democracy/>

- Copestake Ann. (2004). Natural Language Processing Lecture Synopsis. <http://www.cl.cam.ac.uk/users/aac/>
- Darwish, K., Stefanov, P., Aupetit, M., & Nakov, P. (2020). Unsupervised user stance detection on Twitter. In Proceedings of the international AAAI conference on web and social media, vol. 14 (pp. 141–152).
- Databricks (2024). What Is LLMops? Databricks. <https://www.databricks.com/glossary/llmops>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <http://arxiv.org/abs/1810.04805>
- Diehl, T., Weeks, B. E., & Gil de Zúñiga, H. (2015). Political persuasion on social media: Tracing direct and indirect effects of news use and social interaction. *New Media and Society*, 18(9), 1875–1895. <https://doi.org/10.1177/1461444815616224>
- Drost, D. (2023). Different ways of training LLMs. Medium. <https://towardsdatascience.com/different-ways-of-training-llms-c57885f388ed>
- Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. <http://www.semantic-web-journal.net/content/>
- Ekin, S. (2023). *PROMPT ENGINEERING FOR CHATGPT: A Quick Guide To Techniques, Tips, And Best Practices*. TechRxiv.
- El-Sisi, A., & Mousa, H. (2014). Argumentation Based Negotiation in Multi-agent System. In *International Arab Journal of e-Technology* (Vol. 3, Issue 3).
- Erdl, A. (2023). *Unifying LLMs & Knowledge Graphs for GenAI: Use Cases & Best Practices*. Neo4j. <https://neo4j.com/blog/unifying-llm-knowledge-graph/>
- Feger, M. (2021). Online argumentation and social media - what they can learn from each other.
- Ghosh, S., Singhanian, P., Singh, S., Rudra, K., & Ghosh, S. (2019). Stance detection in web and social media: A comparative study. In *Experimental IR meets multilinguality, multimodality, and interaction* (pp. 75–87). Cham: Springer International Publishing.
- Gül, I., Lebet, R., & Aberer, K. (2024). Stance Detection on Social Media with Fine-Tuned Large Language Models. <https://huggingface.co/blog/llama2>
- Habermas, J. (1984). *The theory of communicative action* vol. 1 - Reason and the rationalization of society.
- Hasan, K. S., & Ng, V. (2013). Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. <http://www.createdebate.com/>
- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 328–339
- Hugging Face. (2024). Hugging Face – On a mission to solve NLP, one commit at a time. Huggingface.co. <https://huggingface.co/models>
- Hugging Face (2024). Hugging Face – Prompt Engineering. Huggingface.co. <https://huggingface.co/docs/transformers/main/tasks/prompting>
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>

- Ji, Z., Yu, T., Xu, Y., Lee, N., Ishii, E., & Fung, P. (2023). Towards Mitigating Hallucination in Large Language Models via Self-Reflection. [https://en.wikipedia.org/wiki/Noonan\\_syndrome](https://en.wikipedia.org/wiki/Noonan_syndrome).
- Jing, Z., Su, Y., & Han, Y. (2024). *When Large Language Models Meet Vector Databases: A Survey*. <http://arxiv.org/abs/2402.01763>
- Kau, A., He, X., Nambissan, A., Astudillo, A., Yin, H., & Aryani, A. (2024). Combining Knowledge Graphs and Large Language Models. <http://arxiv.org/abs/2407.06564>
- Krauthoff, T., Baurmann, M., Betz, G., & Mauve, M. (2016). Dialog-Based Online Argumentation. *Frontiers in Artificial Intelligence and Applications*, 287, 33–40. <https://doi.org/10.3233/978-1-61499-686-6-33>
- Küçük, D., & Fazli, C. A. N. (2020). Stance detection: A survey. In *ACM Computing Surveys* (Vol. 53, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3369026>
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22(4), 545–552. <https://doi.org/10.1177/0956797611402512>
- Kunz, J. & Kuhlmann, M. (2024). *Properties and Challenges of LLM-Generated Explanations*. Proceedings of the Third Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCINLP), 13–27.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*.
- Li, Y., & Caragea, C. (2019). Multi-task stance detection with sentiment and stance lexicons. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (pp. 6299–6305). Hong Kong, China: Association for Computational Linguistics, <http://dx.doi.org/10.18653/v1/D19-1657>.
- Markowitz, D. (2021) Transformers, Explained: Understand the Model Behind GPT-3, BERT, and T5. Daleonai.com. <https://daleonai.com/transformers-explained>
- Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2024). *Prompt Engineering in Large Language Models* (pp. 387–402). [https://doi.org/10.1007/978-981-99-7962-2\\_30](https://doi.org/10.1007/978-981-99-7962-2_30)
- Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19(1), 1–22. <https://doi.org/10.1007/s10506-010-9104-x>
- Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. (2016). SemEval-2016 Task 6: Detecting Stance in Tweets. <http://alt.qcri.org/semeval2016/task6/>
- Murel, J., & Noble, J. (2024). What is LLM Temperature? IBM. <https://www.ibm.com/think/topics/llm-temperature>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A Comprehensive Overview of Large Language Models. <http://arxiv.org/abs/2307.06435>
- OpenAI. (2024). What Are Tokens and How to Count them? [Help.openai.com. https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them](https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them)
- Paulheim (2016). Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semantic Web Journal*, (Preprint):1–20.
- Peeperkorn, M., Kouwenhoven, T., Brown, D., & Jordanous, A. (2024). *Is Temperature the Creativity Parameter of Large Language Models?*

- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1–67
- Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S., & Sonenberg, L. (2003). Argumentation-based negotiation. *Knowledge Engineering Review*, 18(4), 343–375. <https://doi.org/10.1017/S0269888904000098>
- Rapp, Christof, "Aristotle's Rhetoric", *The Stanford Encyclopedia of Philosophy* (2023), Edward N. Zalta & Uri Nodelman (eds.). <https://plato.stanford.edu/archives/win2023/entries/aristotle-rhetoric/>.
- Rauniar, R., Rawski, G., Johnson, B., & Yang, J. (2013). Social Media User Satisfaction-Theory Development and Research Findings. *Journal of Internet Commerce*, 12(2), 195–224. <https://doi.org/10.1080/15332861.2013.817864>
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications*. <http://arxiv.org/abs/2402.07927>
- Shanahan, M. (2024). Talking about Large Language Models. *Communications of the ACM*, 67(2), 68–79. <https://doi.org/10.1145/3624724>
- Smith, A. (2014). Politics and advocacy in the social media era. Pew Research Center. <https://www.pewresearch.org/internet/2014/07/29/politics-and-advocacy-in-the-social-media-era/>
- Somekh, B. (2006). *Action Research: A Methodology For Change And Development: a methodology for change and development*. McGraw-Hill Education (UK).
- Statista. (2025). *Digital population worldwide as of February 2025*. Retirado em junho 2025, de <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- Stegeman, J. (2024). What Is a Knowledge Graph? - Neo4j Graph Database & Analytics. <https://neo4j.com/blog/what-is-knowledge-graph/>
- Stryker C., Scapicchio, M. (2024). Generative AI. [Ibm.com. https://www.ibm.com/think/topics/generative-ai](https://www.ibm.com/think/topics/generative-ai)
- Sycara, K. P. (1992). *Argumentation: Planning Other Agents' Plans*.
- Toulmin, S. (2003). *The Uses of Argument*, Updated Edition.
- União Europeia. (2016). Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados (RGPD). *Jornal Oficial da União Europeia*, L 119, 1–88. <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32016R0679>
- União Europeia. (2024). Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de março de 2024 relativo à inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) 2018/858, (UE) 2019/2144 e (UE) 2020/1056, e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento Inteligência Artificial – AI Act). *Jornal Oficial da União Europeia*, L 168, 1–254. <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689>
- van Eemeren, F., Grootendorst, R., & Snoeck Henkemans, F. (1996). *FUNDAMENTALS OF ARGUMENTATION THEORY: A Handbook of Historical Backgrounds and Contemporary Developments*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *arXiv (Cornell University)*, 30, 5998–6008. <https://arxiv.org/pdf/1706.03762v5>

Walton, D. N. (2007). *Media argumentation: dialectic, persuasion, and rhetoric*. Cambridge University Press.

Wydanski, W. (2022). Self-attention vs attention in transformers - Medium. <https://medium.com/@wwydanski/whats-the-difference-between-self-attention-and-attention-in-transformer-architecture-3780404382f3>

Zhang, B., Dai, G., Niu, F., Yin, N., Fan, X., Wang, S., Cao, X., & Huang, H. (2024). *A Survey of Stance Detection on Social Media: New Directions and Perspectives*. <http://arxiv.org/abs/2409.15690>

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2024). A Survey of Large Language Models. <http://arxiv.org/abs/2303.18223>



# Apêndices

## Apêndice A – Excertos de Código

### Apêndice A1 – Definição dos *Endpoints* do Sistema

```
# Define the API app
app = FastAPI()

# API Setup
# Summarization endpoint
@app.post("/summarizer")
async def summarize_comments_endpoint(request: Dict[str, Dict[str, List[str]]]):
    grouped_comments = request.get("grouped_comments", {})
    summaries = summarize_grouped_comments(grouped_comments)
    return {"summaries": summaries}

# Reddit scraper endpoint
@app.post("/reddit_scraper", response_model=RedditResponse)
async def scrape_reddit_thread(request: RedditRequest):
    return fetch_reddit_data(request.url)

# Topic identifier endpoint
@app.post("/topicIdentifier")
async def identify_topic(request: TopicIdentifierRequest):
    return topicIdentifier(request)

# Stance classifier endpoint
@app.post("/stanceClassifier")
async def classify_stance(request: StanceClassificationRequest):
    return stance_classifier(request)

# Knowledge graph creator endpoint
@app.post("/kgCreator")
def build_kg(request: KGRequest):
    return create_knowledge_graph(request.thread_data)
```

Figura 57 - Estrutura do ficheiro "main.py"

### Apêndice A2 – Função Principal do *Reddit Scraper*

```
# Function to extract thread data
def fetch_reddit_data(url: str):
    """Fetch Reddit thread data and comments, keeping top 10 comments and their replies."""
    try:
        # Ensure the URL ends with a slash before appending '.json' for API access
        if not url.endswith('/'):
            url += '/'
        api_url = f"{url}.json"

        # Standard user-agent header to avoid being blocked by Reddit's servers
        headers = {
            'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36'
        }

        # Send GET request to Reddit's API
        response = requests.get(api_url, headers=headers)
```

Figura 58 - Função Principal do *Reddit Scraper*

## Apêndice A3 – Helper Function “process\_text”

```
# Helper function to clean up post text
def process_text(text):

    lines = text.split("\n") # Split the text into individual lines
    inside_citation = False # Flag to track if we're inside a quoted block
    processed_lines = []

    for line in lines:
        # Detect if the line starts with a Reddit quote ('&gt;')
        if re.match(r"^\s*&gt;", line): # If the line starts with '>'
            if not inside_citation:
                processed_lines.append("***Citing:**\n") # Add citation start marker
                inside_citation = True
            # Remove '&gt;' and any extra spaces from the start of the line
            processed_lines.append(re.sub(r"^\s*&gt;\s*", "", line))
        else:
            if inside_citation:
                processed_lines.append("\n**End of Citation**") # Add citation end marker
                inside_citation = False
            processed_lines.append(line)

    # If the last lines were quotes, close the citation
    if inside_citation:
        processed_lines.append("\n**End of Citation**")

    return "\n".join(processed_lines) # Reassemble processed text
```

Figura 59 - Helper Function "process\_text"

## Apêndice A4 - Objeto "thread\_data"

```
# Build structured thread data
thread_data = {
    'post': {
        'title': post_data.get('title', ''),
        'author': post_data.get('author', ''),
        'created_utc': datetime.fromtimestamp(post_data.get('created_utc', 0)).strftime('%Y-%m-%d %H:%M:%S'),
        'score': post_data.get('score', 0),
        'upvote_ratio': post_data.get('upvote_ratio', 0),
        'url': post_data.get('url', ''),
        'selftext': process_text(post_data.get('selftext', '')),
        'num_comments': post_data.get('num_comments', 0),
        'subreddit': post_data.get('subreddit', '')
    },
    'comments': processed_comments
}

return {"thread_data": thread_data}
```

Figura 60 - Objeto "thread\_data"

## Apêndice B – Demonstração da Solução

**(15 points)**

I think your experience may have been different because you took a major that isn't known for doing that.

It's pretty well documented that humanities and arts majors tend to lean left whereas majors like business, engineering, and general STEM fields tend to be more moderate or right. So one could infer it's only certain departments that would push said "liberal indoctrination".

A bit anecdotal here, but I'm an engineering major as well. Of all of the courses I took throughout university, the only ones that were more politically/socially charged were any humanities electives I took, and of course they were quite left leaning.

I can only assume if your entire degree consisted of that and your grades were at least partially based on it, I wouldn't quite call it "indoctrination", but it would probably shift most peoples' views a bit more liberal.

**L, (3 points) ■**

That is not true. Even in STEM fields, professors are overwhelmingly left-leaning:  
[https://www.nas.org/academic-questions/31/2/homogenous\\_the\\_political\\_affiliations\\_of\\_elite\\_liberal\\_arts\\_college\\_faculty](https://www.nas.org/academic-questions/31/2/homogenous_the_political_affiliations_of_elite_liberal_arts_college_faculty)

The only major that's even close is "Engineering" (which is like 12 different majors in one) and it's still a 1.6:1 ratio, which is significant

**(13 points)**

Hey there, studies executed and referenced by Professor Robert Anthony Altemeyer have been done on this topic. Those studies have been repeated several times in many formulations, in many parts of North America. The finding is reasonably simple and quite complete: Trending away from Conservative views tends to occur mostly strongly with exposure to individuals, experiences, and viewpoints beyond the curated norms of childhood (parents and early influences).

University education has very little impact on it - it is the fact that for many people, this is the first time in their life they are forced to interact with people who are different than the homogeneity of their neighbourhoods. This is also estimated to be a factor why urban areas are more liberal - you simply meet and experience more different people and more different things, because those different things are close by and can't be avoided.

**End of Citation**

Reality has a well known liberal bias to quote Colbert. If 99% of climate scientists agree man made climate change is real but the republicans continue to pretend it's a made up plot of course no climate scientist is going to identify as republican. Now look at other anti science policies like vaccine skepticism, creationism, etc.

You can claim this is some left wing conspiracy to keep out conservative scientists but you'd have to prove that the universities which receive billions from wealthy donors and corporations are actually secretly run by communists which is kind of silly, why would businesses pour billions of dollars into funding institutions that are trying to destroy them

**L, (10 points) ■**

**Citing:**

If you can name one Ivy League school where > 50% of professors identify as conservative, you win.

**End of Citation**

- "Win"? This isn't a contest. It's a discussion.
- Conservative culture downplaying education is just as likely to lead to fewer professionals working in the field of academia as it is to lead to fewer liberal students. It's still the same end, with the same root cause, so what have you proven here?

**Citing:**

assuming nearly all of your professors will be left leaning in these majors, and that your grades might be more subjective (i.e. open ended questions vs standardized multiple choice tests), this could lead to an environment where you feel pressured to put the answer you feel your professors would like to hear. After years of repetition and subliminal bias, you might be molded politically without realizing it.

**End of Citation**

Frankly this just seems like a HUGE stretch and some major spitballing to say that giving an answer that a person wants to hear leads to you actually changing the way you think, to such a major degree that your entire political affiliation shifts. I have an incredibly hard time buying that one.

**Citing:**

It's the students who go to college that didn't care about politics in high school who leave college

Figura 61 - Comentários Originais e Respetivas *Replies* Classificadas

**Neutral Arguments**

- Short-Term Memory in Politics:** Political events and controversies, such as those during Trump's presidency, are often quickly forgotten, especially when overshadowed by new global events like the Ukraine crisis.
- Enduring Superpower Status:** Despite fluctuations in leadership and policy, the US continues to maintain its superpower status, partly due to the lack of viable alternatives and its nuclear capabilities.
- Economic Influence on Politics:** The American economic system, which allows significant wealth accumulation, often influences political leanings, with wealthier individuals tending to support Republican policies for tax benefits.
- Historical Policy Impact:** Past US policies, such as those during Reagan's era, have had long-term global impacts, contributing to issues like the rise of Islamic terrorism and affecting the US's global standing.
- Transatlantic Relationship Dynamics:** While transatlantic relationships may evolve, the US remains a crucial partner for Europe, as alternatives like China do not offer the same level of support, particularly against threats like Russia.

Figura 62 - Resumo dos Argumentos Neutros

**Results for 'CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.':**

Group Summary	Argument Text	Stance
College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift	Exposure to diverse life experiences in college can lead to a gradual shift in political views.	FOR
College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift	College education provides students with skills to evaluate arguments and conduct research, which can	FOR
College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift	Colleges serve as an antidote to indoctrination by equipping students with tools for critical thinking, whi	FOR
College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift	Colleges expose students to diverse perspectives and options, which can challenge long-held beliefs and	FOR
College education fosters critical thinking and exposure to diverse perspectives, which can lead to a shift	Higher education institutions teach critical thinking skills, which can lead individuals to reevaluate and f	FOR
Certain academic disciplines, particularly in the humanities and arts, are perceived as more left-leaning.	Humanities and arts majors tend to lean left, whereas majors like business, engineering, and general STE	FOR
Certain academic disciplines, particularly in the humanities and arts, are perceived as more left-leaning.	Courses in humanities electives are often more politically or socially charged and tend to be left-leaning.	FOR
Exposure to diverse environments and experiences, rather than university education itself, is a significan	Studies by Professor Robert Anthony Altemeyer indicate that trending away from conservative views occ	FOR
Exposure to diverse environments and experiences, rather than university education itself, is a significan	Urban areas tend to be more liberal because they offer more opportunities to meet and experience diver	FOR
Homogeneous educational environments, such as Bible Colleges, do not lead to a liberal shift, suggestin	Institutions of higher education that are ethnically, religiously, and geographically homogeneous, such a	FOR

Figura 63 - Demonstração: Execução da Consulta "Argument Groups by Popularity – Full Detail"

**Results for 'CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.':**

Comment 1 ^

**Parent Comment:**

I am not sure your test captures the longterm impact of the college experience. I started college very conservative and now am quite liberal. I didn't really change my views until about 10 years after graduation. It was a gradual process, but the seeds of my liberalism were sown in undergraduate classes.

1. I met people who had different life experiences from me.
2. I learned how to evaluate arguments and do research.
3. As I continued reading and learning, I started seeing flaws in the conservative views I held and had the tools to evaluate those arguments more critically.

I don't think that colleges are indoctrinating students, but I do think they serve as an antidote for indoctrination, which is what conservatives fear.

**Replies:**

**Reply #1:** This is exactly what happened to me as well, down to the exact steps you enumerated.

Fun twist though, I went to an extremely conservative Christian university. In fact, I think the most interesting observation of all in my personal life is that every single person (and I know multiple!) who shifted from conservative to liberal/leftist went to a conservative Christian college. In contrast, every single conservative that I know who went to one of those "big evil public Marxist" colleges is still conservative.

Very interesting observation I've noticed.

---

**Reply #2:** Fair enough; you could still evaluate a person's shift in political beliefs at any point after college and compare them to the person who never went to college. Do you have any data on that?

---

**Reply #3:** Odd I'm the reverse. Started out liberal, liberal friends, then I opened my mind and did research and became more conservative.

Figura 64 - Demonstração: Execução da Consulta "Replies to Supporting Comments"

**Results for 'CMV: Universities are not making students liberal. The "blame" belongs with conservative culture downplaying the importance of higher education.':**

Comment	Score	Stance
I am not sure your test captures the longterm impact of the college experience. I started college very conservative and now am qui	654	SUPPORTS
Your entire point of view is an anecdotal fallacy, which means it relies on your personal experience. You point out that a measurem	70	OPPOSES
1. If you can name one Ivy League school where &gt; 50% of professors identify as conservative, you win. 2. This idea largely depe	61	OPPOSES
I actually can speak to this. I studied Sociology in undergrad despite being a conservative and even did my senior capstone studyi	43	OPPOSES
I am currently a university student, so my argument may be a bit anecdotal. It has been my experience that all of my teachers, at le	31	OPPOSES
First two studies are longitudinal, which is a specific word for the study design you're asking for in your first paragraph. Later ones	19	OPPOSES
I think your experience may have been different because you took a major that isn't known for doing that. It's pretty well docum	16	SUPPORTS
Hey there, studies executed and referenced by Professor Robert Anthony Altemeyer have been done on this topic. Those studies h	16	SUPPORTS
Agreed with your first point. Studies have found a correlation with political ideology and higher education but haven't found it to	15	SUPPORTS

Figura 65 - Demonstração: Execução da Consulta "Popular Comments by Score"



Node details		Node details	
<b>Argument</b>		<b>ArgumentGroup</b>	
Key	Value	Key	Value
<id>	4:94ec970e-8fc4-4d51-83d6-59b1b78f1b2b:26	<id>	4:94ec970e-8fc4-4d51-83d6-59b1b78f1b2b:1518
updated_at	2025-06-24T15:51:23.15800000Z	summary	"The predominance of left-leaning professors and course materials in universities influences students' political views towards liberal ideologies. Thi... <a href="#">Show all</a>
discussion_id	"53a3e382-77dd-4dae-b0d6-5f546c5789d2"	updated_at	2025-06-24T15:52:50.03600000Z
text	"The assertion that professors and academia are generally left-leaning suggests that the educational environment may influence students' political vie... <a href="#">Show all</a>	discussion_id	"53a3e382-77dd-4dae-b0d6-5f546c5789d2"
stance	"AGAINST"	stance	"AGAINST"

Figura 69 - Detalhes de Nós de Argumento e *Cluster* de Argumentos

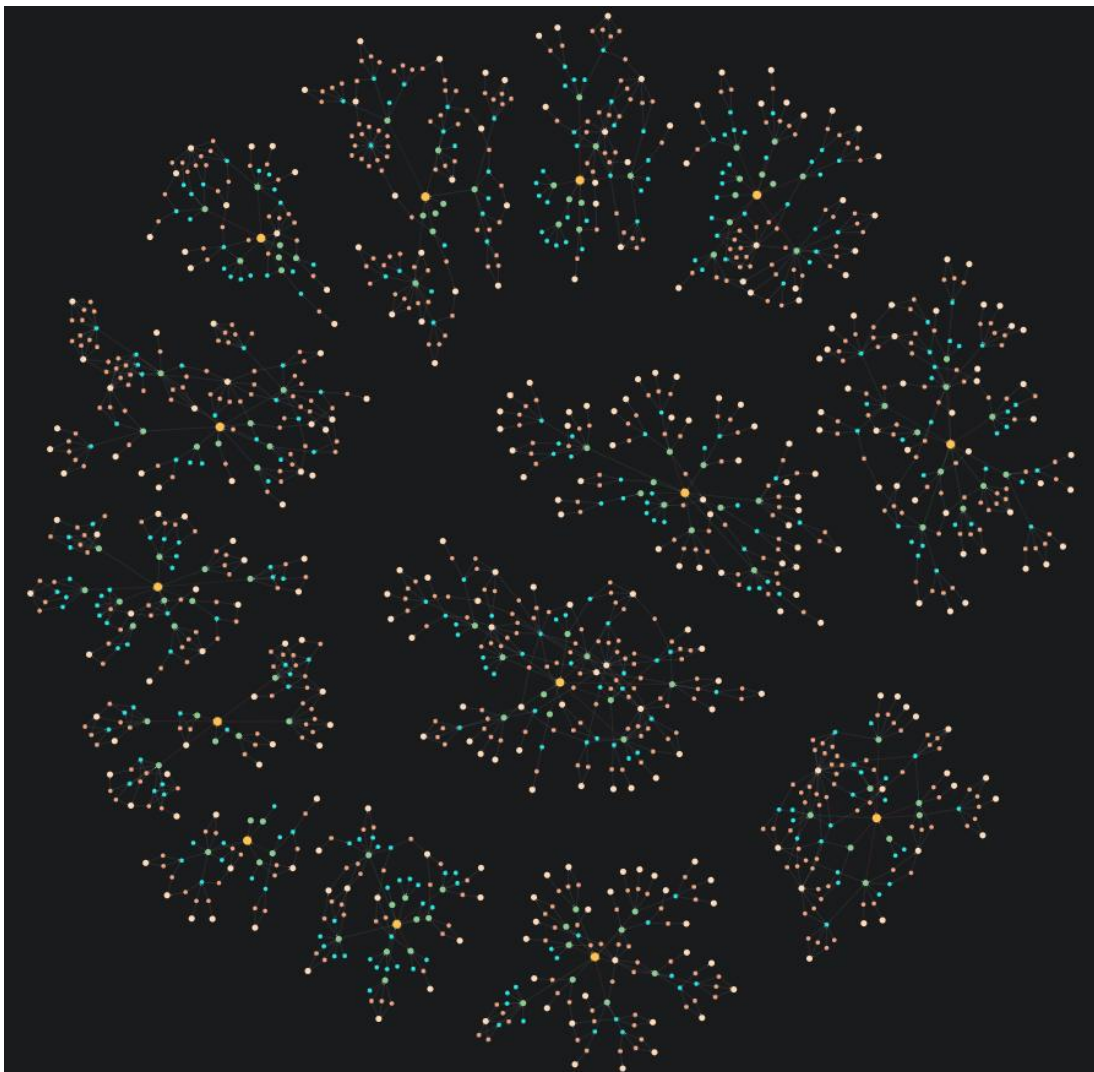


Figura 70 - Base de Dados de Grafos com Várias Discussões Analisadas

## Apêndice C – Avaliação: Clusters Criados

Sumário do Cluster	Argumentos
<p><i>Trump's policies and rhetoric are causing a shift away from U.S. reliance, prompting countries to diversify alliances and reduce dependency on the U.S.</i></p>	<p><i>The long-term shift towards diversification of alliances is already occurring, indicating a move away from reliance on the United States as the primary global power.</i></p>
<p><i>(same as above)</i></p>	<p><i>The second Trump administration has highlighted the strategic vulnerability of relying on the United States or any single power, prompting a reevaluation of global alliances.</i></p>
<p><i>(same as above)</i></p>	<p><i>European and global strategic shifts, such as Germany and France's push for European defense autonomy and Japan and South Korea increasing defense spending, indicate a move away from reliance on the U.S.</i></p>
<p><i>(same as above)</i></p>	<p><i>Countries like Canada and India are actively diversifying their economic and military strategies, reflecting a reduced dependence on the U.S. and a response to perceived instability in U.S. politics.</i></p>
<p><i>The U.S. is likely to maintain international engagement and avoid permanent isolation, as countries will continue to engage with it due to its economic and military viability, despite any temporary strains in trust.</i></p>	<p><i>Countries are likely to continue engaging with the U.S. as long as it remains economically and militarily viable, despite any strained trust.</i></p>
<p><i>(same as above)</i></p>	<p><i>Trump's erratic leadership is unlikely to result in permanent international isolation, as nations have historically adapted to different U.S. administrations.</i></p>

---

*Trump's policies are harming rural states and his own supporters, leading to potential internal displacement and economic decline.*

*(same as above)*

*Trump's policies disproportionately harm highly rural states, potentially displacing residents as internal refugees due to economic decline.*

*Trump's actions dismantle mechanisms that transfer wealth from affluent regions to poorer areas, negatively impacting his own supporters who benefit from such transfers.*

---

# Anexos

## Anexo A – Constituição de um Argumento segundo Alpert (2016)

*Uma alegação isolada não constitui um argumento. Por exemplo:*

- A. *Barack Obama foi o melhor presidente que os EUA já tiveram.*
- B. *Cada um de nós possui uma alma que sobrevive mais anos do que o corpo.*
- C. *Cães são melhores animais de estimação do que gatos.*

*Estas são apenas alegações, pois não estão acompanhadas de premissas que ofereçam razões para acreditar que a alegação seja verdadeira. Se adicionarmos premissas que sustentem essas alegações, estas tornam-se argumentos:*

- A. *Barack Obama foi o melhor presidente que os EUA já tiveram, dado que tornou o seguro de saúde acessível a todos os americanos.*
- B. *Cada um de nós possui uma alma que sobrevive mais anos do que o corpo, porque a Bíblia assim o afirma.*
- C. *Cães são melhores animais de estimação do que gatos, uma vez que todos os cães que conheci foram amigáveis.*

*A inclusão de premissas fornece justificativas racionais que permitem sustentar a veracidade das conclusões, transformando as simples alegações em argumentos estruturados.*

(Alpert, 2016)