#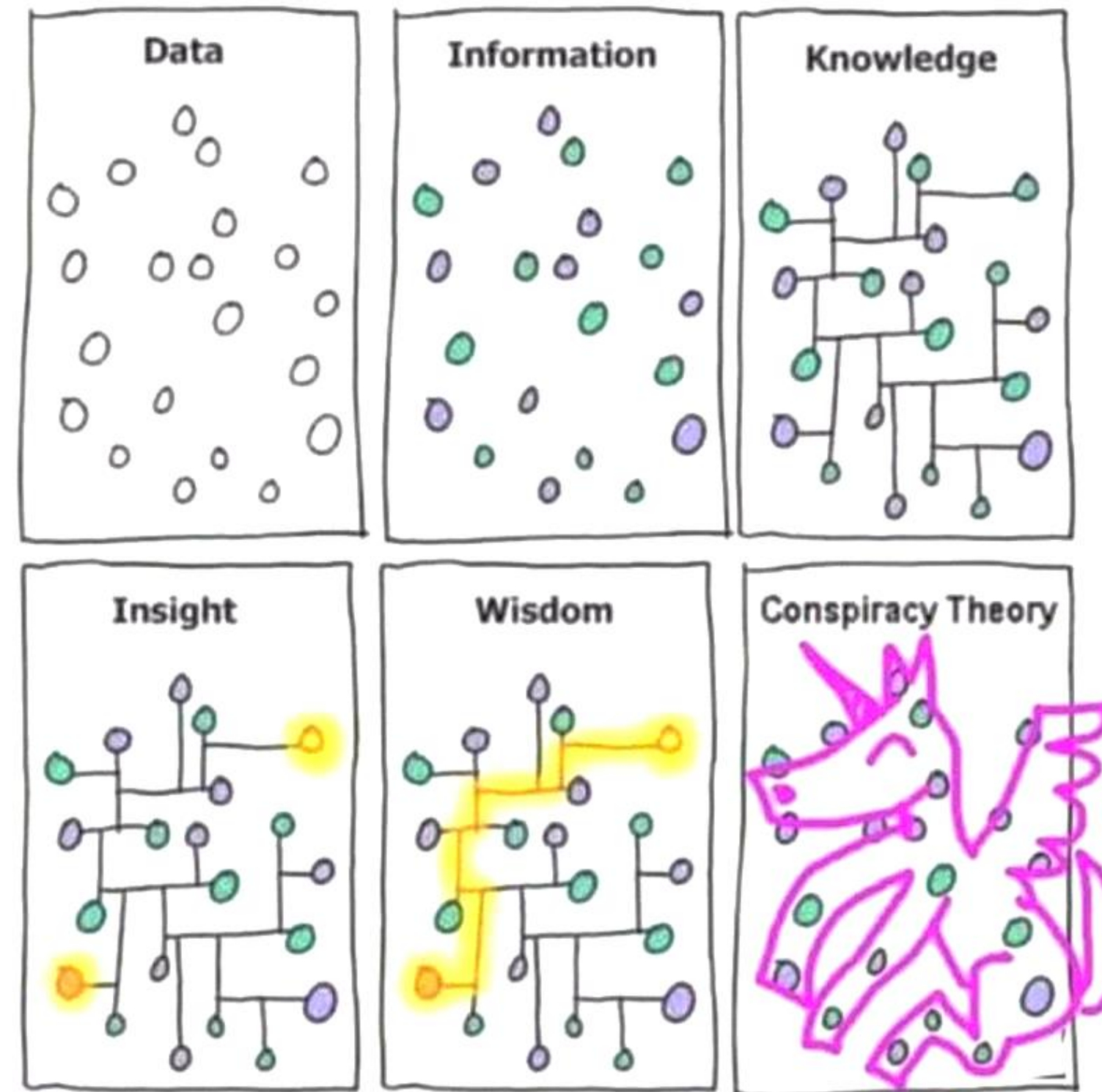 Towards a Computational Framework for Distinguishing Critical and Conspiratorial Texts by **Elaborating** on the Context and Argumentation with LLMs

Authors: Ariana Sahitaj, Premtim Sahitaj, Salar Mohtaj, Sebastian Möller, Vera Schmitt
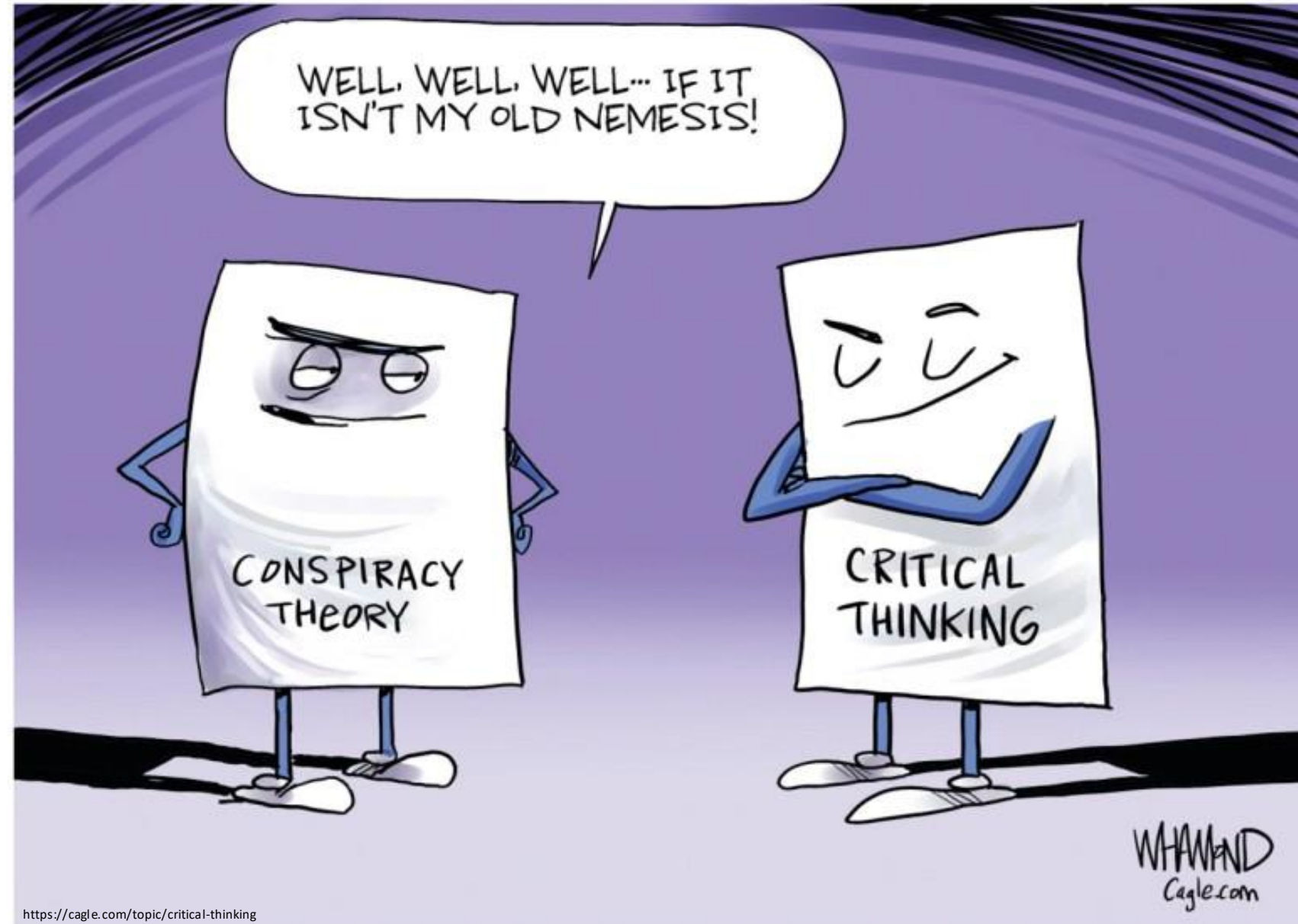
Problem Statement

- The COVID-19 pandemic not only led to a health crisis but also to an "infodemic" – an explosive increase in fake-news.

- **Problem / Task:**
  How do we distinguish between critical and conspiracy texts?

Critical Thinking vs Conspiracy Thinking

WELL, WELL, WELL... IF IT ISN'T MY OLD NEMESIS!

CONSPIRACY THEORY

CRITICAL THINKING

WHAMOND
Cagle.com

https://cagle.com/topic/critical-thinking

Critical Thinking:
- Based on evidence and rational argumentation.
- Inquisitive, evidence-based, supported by verifiable facts.

Conspiracy Thinking:
- Often rooted in unfounded assumptions and emotional persuasion.
- Emotionally charged, based on unverified assumptions, often oversimplified.

Our Goal:
- Analyze the context and argumentative structure of these texts to distinguish between critical and conspiratorial content.
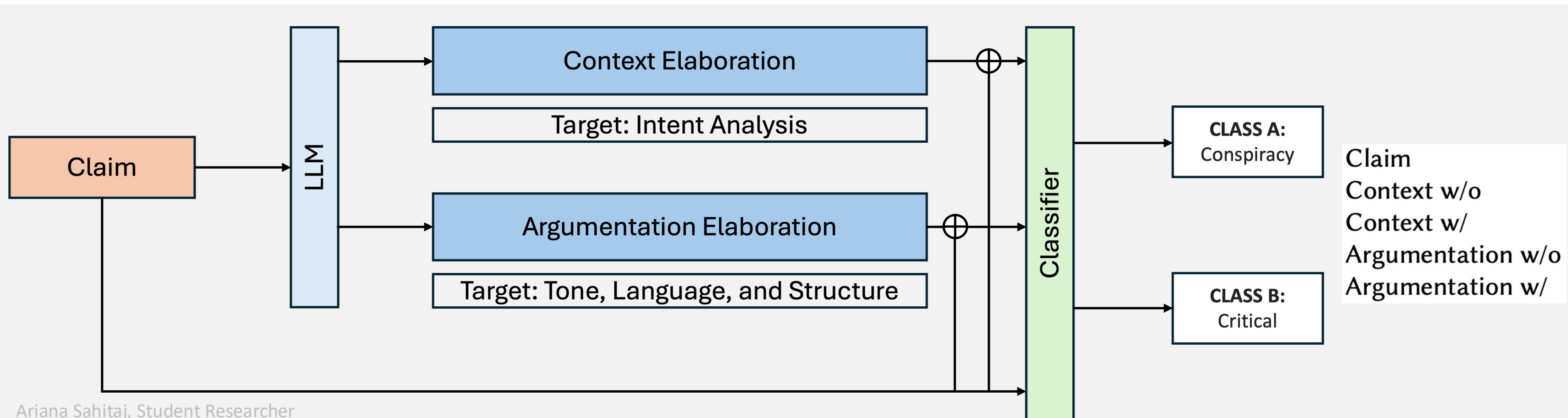
Ariana Sahitaj, Student Researcher

Dataset

- Collection of oppositional texts from Telegram
- Related to the COVID-19 pandemic
- Available in both English and Spanish

- Each entry contains a unique ID, text content, overall category (*CONSPIRACY* or *CRITICAL*), and a list of relevant annotations
- Only utilized textual claim (excluding annotations)

	English	Spanish
Text distribution		
Average token length	124.37	258.42
Minimum token length	15	31
Maximum token length	1307	1827
Median token length	87.0	189.0
Total token count	497,489	1,033,663
Label distribution		
CRITICAL	2621	2538
	65.53 %	63.45 %
CONSPIRACY	1379	1462
	34.47 %	36.55 %

Approach

■ Leveraging **Large Language Models** (LLMs)
for automatic feature generation

■ **Models Used:**
- **LLM:** Llama3 70B model for generating elaborations
- **Classifier:** DeBERTa as a base classifier, with and without LLM-generated elaborations

Ariana Sahitaj, Student Researcher

Context Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

LLM

Context Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

LLM

Context Elaboration

Target:
Intent Analysis

https://makeameme.org/meme/please-elaborate-5967cc

Context Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"
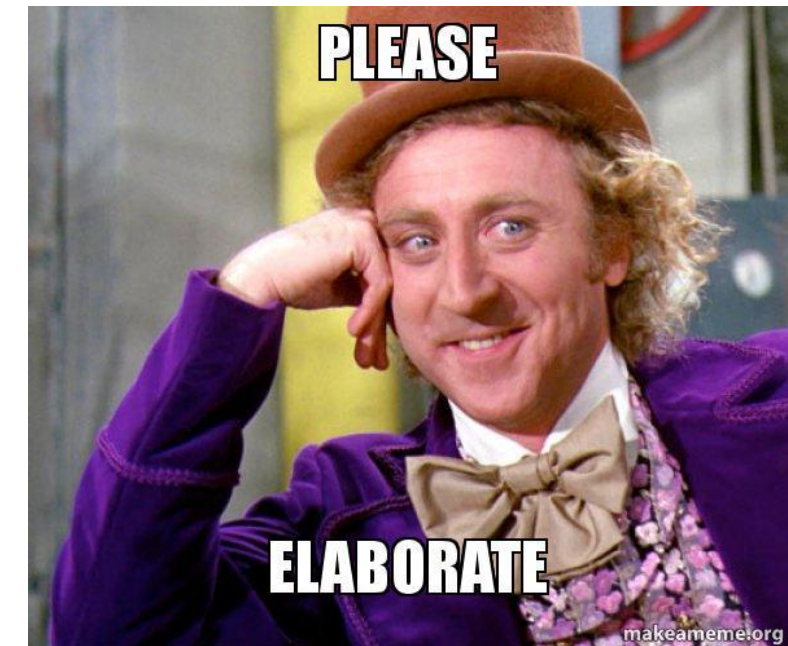
LLM

Context Elaboration

Intent: The intent appears to be to create a sense of injustice and anger among citizens by suggesting that illegal aliens are exempt from vaccine mandates, while also implying that pharmaceutical companies are avoiding accountability.

Argumentation Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

LLM

Argumentation Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

LLM

Argumentation Elaboration

Target:
Tone, Language, and Structure

https://makeameme.org/meme/please-elaborate-5967cc

Argumentation Elaboration

Claim: "Illegal aliens exempt from covid "vaccine" mandates because they can sue drug companies (but you can't) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

LLM

Argumentation Elaboration

Tone, Language, and Structure: The tone is provocative and accusatory, using phrases like 'Illegal aliens' and 'but you can't.' The language is inflammatory, implying unfair treatment and using scare quotes around the word 'vaccine.' The structure is geared towards sparking outrage and misinformation, as it presents a sensationalized claim without providing balanced context or credible sources. The inclusion of a link to a non-mainstream news website (Natural News) adds to the overall conspiratorial tone of the claim.

Experimental Setup

Parameter	Value
Random state	239
Batch size	20
Batch accumulation	1
Epochs	5
Dropout	0.05
Max length	512
Learning rate	3e-5
Warmup steps	100

Parameter	Value
Seed	239
Max tokens	512
Temperature	1.0
Top p	0.95
Min p	0.01

■ **Model Configurations:**
- **LLM Usage:** Llama3 70B model for generating elaborations.
- **vLLM:** Used for efficient memory management.

■ **Training Setup:** DeBERTa fine-tuning
- **Framework:** HuggingFace Transformers & PyTorch Lightning
- **Optimization:** AdamW optimizer with cross-entropy loss

■ **Evaluation Metrics:**
- MCC and F1-score

Results

Mode	Test			Train			Validation		
	Loss	F1	MCC	Loss	F1	MCC	Loss	F1	MCC
Claim	0.250	**0.917**	**0.838**	0.050	0.985	0.971	0.236	0.917	0.836
Context w/o	0.357	0.864	0.731	0.126	0.954	0.909	0.317	0.864	0.737
Context w/	0.253	0.913	0.827	0.074	0.972	0.945	0.196	0.918	0.836
Argumentation w/o	0.267	0.885	0.772	0.137	0.952	0.905	0.265	0.895	0.795
Argumentation w/	0.217	0.910	0.821	0.077	0.975	0.951	0.215	0.904	0.809

■ **Baseline vs. Elaborations:**
  baseline DeBERTa model outperformed models using additional context and argumentation elaborations

■ **Key Findings:**
  elaborations did not enhance the model's accuracy

Ariana Sahitaj, Student Researcher

Implications, Limitations and Future Work

Implications:
Theoretical benefits of elaborations need practical optimization.

Limitations:
Elaborations may overload the model with excessive information, reducing their effectiveness.

Future Directions:
Refining LLM-generated annotations to align better with the classification task.

Questions?

Questions?

Thank you!

CONTACT US
FOR FURTHER QUESTIONS

Ariana Sahitaj
Student Researcher
ariana.sahitaj@campus.tu-berlin.de

"Towards a Computational Framework for Distinguishing Critical and Conspiratorial Texts by Elaborating on the Context and Argumentation with LLMs"