# Fallacious Argument Classification In Political Debates

**Pierpaolo Goffredo,**

**Serena Villata, Elena Cabrio,**

**Shohreh Haddadan, Vorakit Vorakitphan**

pierpaolo.goffredo@etu.univ-cotedazur.fr
3IA, Université Côte d'Azur, Inria, CNRS, I3S
3IA Chair**: Artificial Argumentation For Humans**
Supervisors: **Serena Villata, Elena Cabrio**

## Introduction

The main aim of this work is to **identify** different categories of **fallacious arguments** in political debates after defining and annotating them on an existing dataset of U.S. presidential debates:

- Ad Hominem
- Appeal To Authority
- Appeal To Emotion
- False Cause
- Slogan
- Slippery Slope

Our core contributions are twofold:

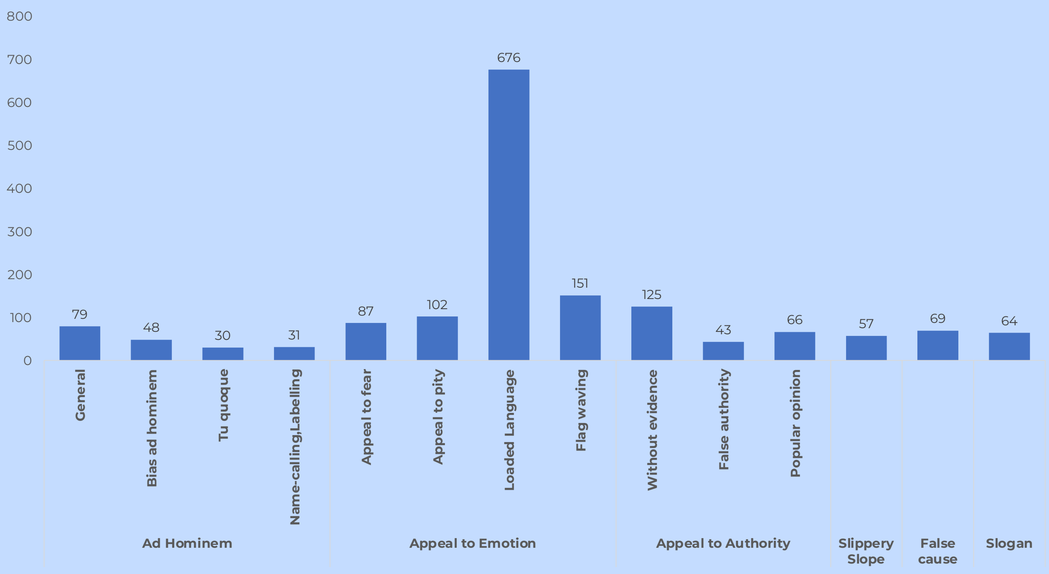
1. Novel large **linguistic resource** of political debates
2. New **transformer-based model architecture**, fine-tuned on argumentation features.

## The ElecDeb60To16-fallacy Dataset

To investigate fallacious arguments in political debates, we extend the annotations of the **ElecDeb60To16** dataset, that collects televised debates of the presidential election campaigns in the U.S. from **1960** to **2016**.

| FALLACY CATEGORY | SUB-CATEGORY | SAMPLE |
|---|---|---|
| Ad Hominem | General | you were totally out of control |
| | Name-calling,Labeling | Such a nasty woman! |
| Appeal to Emotion | Appeal to fear | These terrorists are serious, they're deadly, and they know nothing except trying to kill. |
| | Flag waving | Communism is the enemy of all religions; and we who do believe in God must join together. |
| Appeal to Authority | False authority | I don't think General Douglas MacArthur would like that too much. |
| Slippery Slope | | Now what do the Chinese Communists want? They don't want just Quemoy and Matsu; they don't want just Formosa; they want the world |
| False Cause | | Make America great again! |

## Statistics And Data Analysis




## Fallacies Classification In Political Debates

The approach applies is based on:

- **multi-class sequence classification task** to classify the fallacies observed in the debates.
- enhancing a classifier with **argumentation-based features** (i.e., argument components and relations) within each fallacious argument.

**BERT** and **RoBERTa** are considered as a baseline.

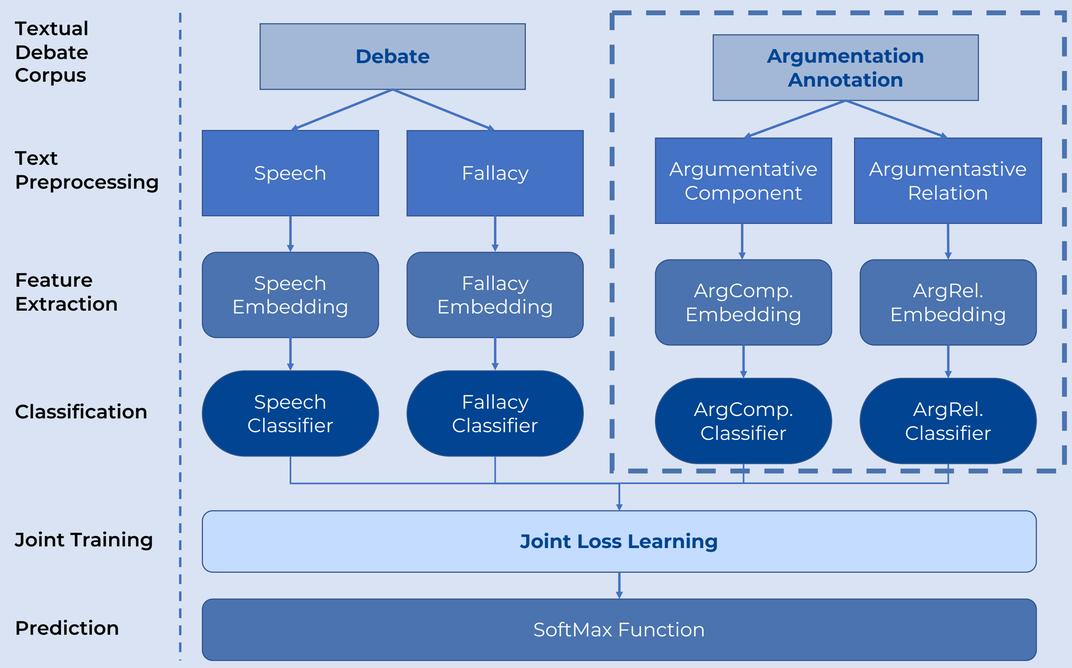
It was necessary to use more advanced **PLMs** to tackle such lengthy speeches, i.e., **Longformer** and **TransformerXL** which can capture long-input texts to perform the classification.

## Proposed Architecture

The approach is based on the Longformer model empowered with the argumentation features, and the context of the fallacious argument in the debate. Each debate is processed into four components:

- **Dialogue context**
- **Fallacious argument snippet**
- **Argument component**
- **Argument relation**

Each component is then extracted in the embedded vectors using the PLM of interest. Each embedding has its own transformer-based classifier to finally obtain a logit (L).




Pipeline for the task of fallacious argument classification.

## Experimental Setup

| MODEL | CATEGORY | $LOSS_{JOINT\_LOSS}$ | ARG. FEAT. | PREC | REC | M F1 |
|---|---|---|---|---|---|---|
| BERT | Main | No | - | 0,62 | 0,55 | 0,55 |
| RoBERTa | Main | No | - | 0,58 | 0,56 | 0,53 |
| Longformer | Main | No | - | 0,64 | 0,60 | 0,57 |
| Longformer | Main | Yes | - | 0,66 | 0,61 | **0,61** |
| TransformerXL | Main | No | - | 0,61 | 0,45 | 0,47 |
| TransformerXL | Main | No | - | 0,61 | 0,51 | 0,53 |
| Longformer | Sub | Yes | - | 0,44 | 0,45 | 0,42 |
| Longformer | Main | Yes | Component | 0,88 | 0,81 | **0,83** |
| Longformer | Main | Yes | Relation | 0,87 | 0,81 | **0,83** |
| Longformer | Main | Yes | Comp. + Rel. | 0,84 | 0,85 | **0,84** |

## Conclusions & Future Work

- Novel approach to the **task of fallacious argument classification**.
- Transformer with an **attention mechanism that** can take advantage of the **context** of the fallacious argument.
- **Generation** of sound arguments out of the identified fallacies and their context
- Investigation of how to counter the **formal invalidity** of these fallacious arguments through newly generated counter-arguments