# Leveraging Context for Multimodal Fallacy Classification in Political Debates

Alessio Pittiglio
DISI, University of Bologna, Italy

Scan for the full paper!

## Motivation

- Not all persuasive techniques rely on false facts
- Fallacies rely on misleading reasoning
- Text + audio = **deeper understanding**

## Task setup

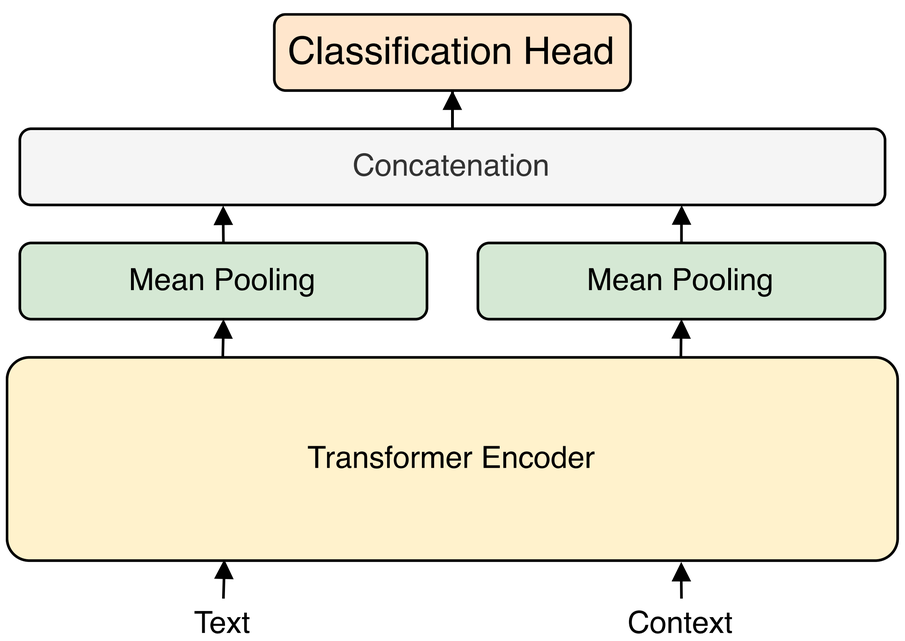
- **Goal**: Classify fallacies in political debates
- **Input types**: Text, Audio, or Multimodal
- **Dataset**: MM-USED-fallacy
- **My focus**: Incorporating **context information**

## Text setting

**Backbones**: RoBERTa-large, ModernBERT, DeBERTaV3
**Context strategies**

- *Concat*: `[text] [SEP] [context]`
- *ContextPool*: Shared encoder + mean pooling
- *CrossAttn*: Text queries context via encoder with gated fusion and attentive pooling

**Best**: ContextPool + RoBERTa-large




**ContextPool**

## Audio setting

**Backbones**: HuBERT-Base, WavLM, Wav2Vec2
**Approaches**:

- Fine-tuned HuBERT
- TemporalaAvg (with context, less effective)

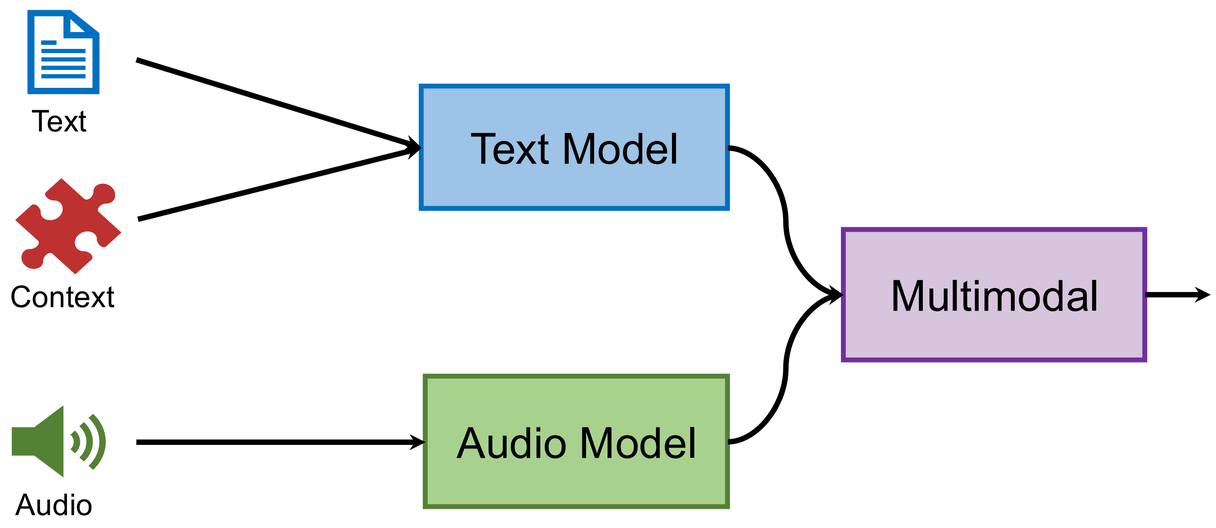
Audio truncated to 15s due to OOM limitations.

## Fusion strategy

**Late Fusion**:

- **Weighted average of logits** via Bayesian optimization
- **Majority Voting** (Text, Audio, Fusion) used in final submission

**Result**: *"faded" version* of text-only
**Hypothesis**: Need **better modality interaction** (cross-modal attention, etc.)




## Result

| Input | Team | F1 |
|---|---|---|
| Text-Only | Team NUST | 0.4856 |
| | Baseline BiLSTM | 0.4721 |
| | **My team** | **0.4444** |
| Audio-Only | **My team** | **0.3559** |
| | Team EvaAdriana | 0.1858 |
| | Team Nust | 0.1588 |
| Text-Audio | Team Nust | 0.4611 |
| | **My team** | **0.4403** |
| | Baseline RoBERTa + WavLM | 0.3816 |

# Key takeaways

1. **Text context** is critical; best results at $N = 5$ window.
2. **Audio-only model (HuBERT)** demonstrates the value of audio for fallacy classification.
3. **Multimodal late fusion** underperforms, needs more **cross-modal learning**.
4. **Dataset challenges**: class imbalance, audio truncation.