

# Detecting Hyperpartisanship and Rhetorical Bias in Climate Journalism: A Sentence-Level Italian Dataset



Funded by  
the European Union



UK Research  
and Innovation

## Usage/License



This work is licensed under CC BY 4.0; users may share and adapt it with attribution. The dataset includes annotations and links to articles that may (or may not) contain hyperpartisan content. Text data are not shared to preserve copyright. Data were collected per the website's Terms of Service and used only for research purposes.

## Download



GitHub Repo: [https://anonymous.4open.science/r/Climate\\_HP-RB-D5EF/README.md](https://anonymous.4open.science/r/Climate_HP-RB-D5EF/README.md)

## Researchers



- Michele Joshua Maggini
- Davide Bassi
- Pablo Gamallo

These researchers are affiliated with the CITIUS at the University of Santiago de Compostela.

## Summary

This dataset is the first Italian resource for joint hyperpartisan and rhetorical bias detection in climate change discourse. It contains 48 articles (1,010 sentences) from far-right Italian media, annotated at the sentence level for binary hyperpartisan classification and a multi-label taxonomy of 17 rhetorical bias categories. The dataset is designed to support research on hyperpartisanship and rhetorical bias in climate-related content.

## What We Offer

- A novel annotated dataset of 48 Italian climate change articles with over 1.5K rhetorical fallacy labels
- Detailed annotation guidelines in both English and Italian
- Full pipeline to recreate the dataset via article URLs (respecting copyright constraints)
- Baseline classification results across multiple model architectures and learning paradigms

## Key Features

- Sentence-level annotation for both binary hyperpartisan classification and 17 rhetorical bias categories
- Strong inter-annotator agreement (Cohen's kappa of 0.92 for hyperpartisan detection, 0.63 for rhetorical fallacies)
- Corpus analysis revealing significant correlations between hyperpartisan content and specific rhetorical techniques
- Baseline experiments using state-of-the-art models (GPT-4o, GPT-4o-mini, Italian BERTbase)

## Collaboration Objectives

- Advance research in hyperpartisan detection for underrepresented languages beyond English
- Provide empirical insights into how rhetorical techniques drive media manipulation in climate discourse
- Support the development of more robust, culturally-aware systems for detecting media bias across languages and contexts