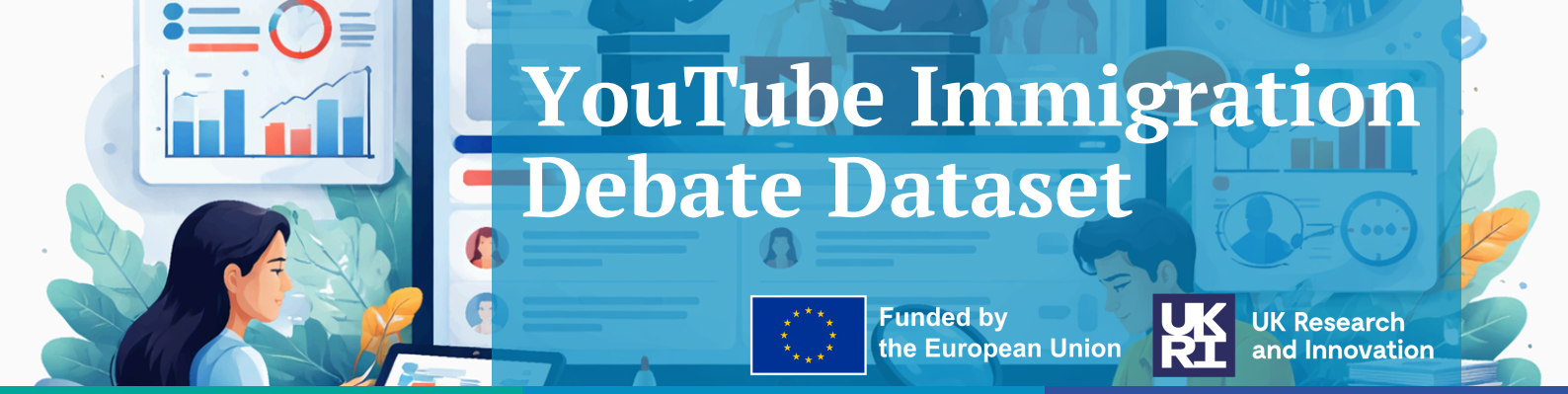

YouTube Immigration Debate Dataset

Funded by the European Union

UK Research and Innovation

Usage/License

The dataset includes social media comments that may contain offensive content.

Data were collected in line with YouTube's Terms of Service and processed in anonymized, aggregated form.

Download

Complete resources related to the project are available at: https://zenodo.org/records/18925414

Researchers

- Davide Bassi — CiTIUS, University of Santiago de Compostela
- Renata Vieira — CIDEHUS, University of Évora
- Martín Pereira-Fariña — iHUS, University of Santiago de Compostela

Summary

This dataset contains YouTube comments on U.S. immigration debates (2020–2024), focusing on recurring users and their interactions across ideological communities. It includes stance annotations using an NLI-based pipeline and supports analysis of polarization, user behavior, and communication dynamics in online discussions.

Data Source: YouTube comments collected from 25 high-visibility U.S. news and political channels categorized as left-leaning or right-leaning.

Dataset Composition: ~3,500 YouTube videos with comments and longitudinal user tracking.

Key Features

- Longitudinal dataset (2020–2024) covering a full U.S. electoral cycle
- Large-scale corpus from thousands of YouTube videos
- NLI-based stance detection (pro, contra, neutral)
- Reconstructed conversation threads and interaction networks
- Tracking of persistent users across time
- Linguistic and behavioral features (e.g., toxicity, identity attacks)

What We Offer

- Large-scale dataset of immigration debates on YouTube
- Computational pipeline (data collection, reconstruction, stance detection, interaction analysis)
- Support for research on polarization and online political dynamics

Collaboration Objectives

- Advance the computational analysis of political discourse on social media.
- Provide resources for studying polarization and cross-ideological interaction.
- Support research on online disagreement dynamics and communicative strategies.
- Facilitate interdisciplinary collaboration between computational social science, linguistics, and political communication

hybrids

HYBRIDS Website
www.hybridsproject.eu

Get in Touch
davide.bassi@usc.es