

Understanding topic prevalence in oral debates on Artificial Intelligence

A comparative Natural Language Processing analysis of European Parliament and U.S. Congress debates held in 2021

Objective

- understand topic prevalence in EU and U.S. oral debates on AI in 2021
- fill a methodological gap: analyse previously unobserved data, using quantitative analysis for bilateral comparative analysis of AI-related policy debates

Building a new database

- web scraped 61 U.S. Congress and 21 EP debates
- processed into machine-readable text-as-data
- two samples:
- Debate Corpus = 19 debate documents (1,611,949 words)
- Speaker Corpus = 252 policymakers' contributions

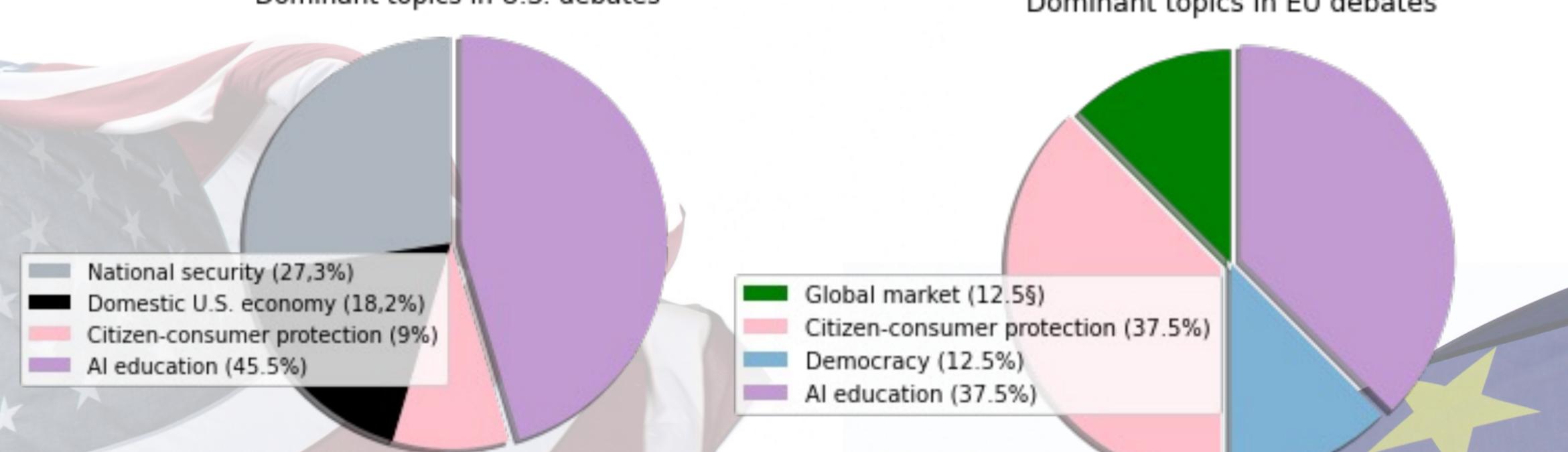
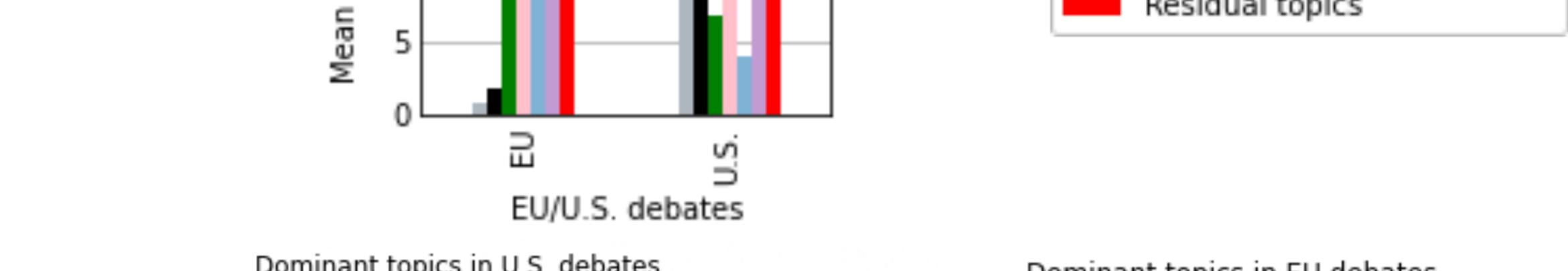
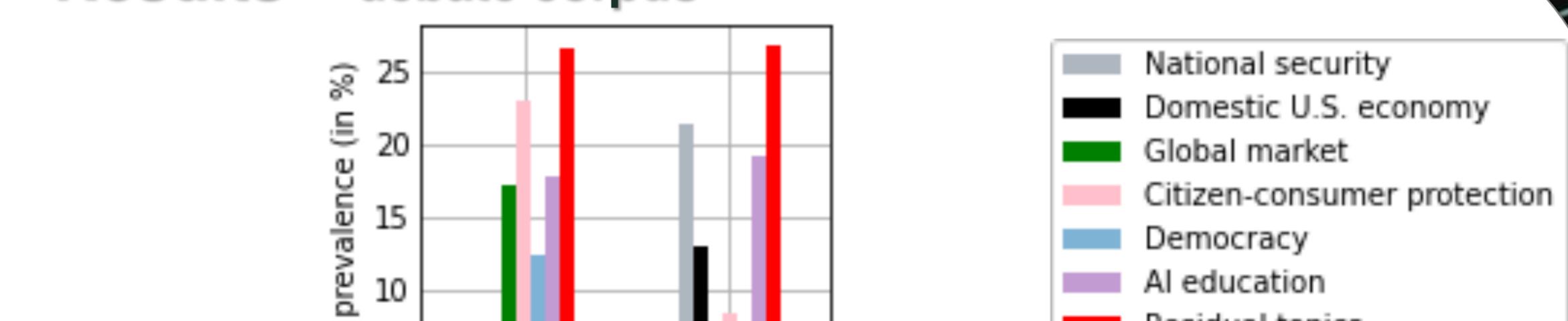
GitHub repo



Analyses

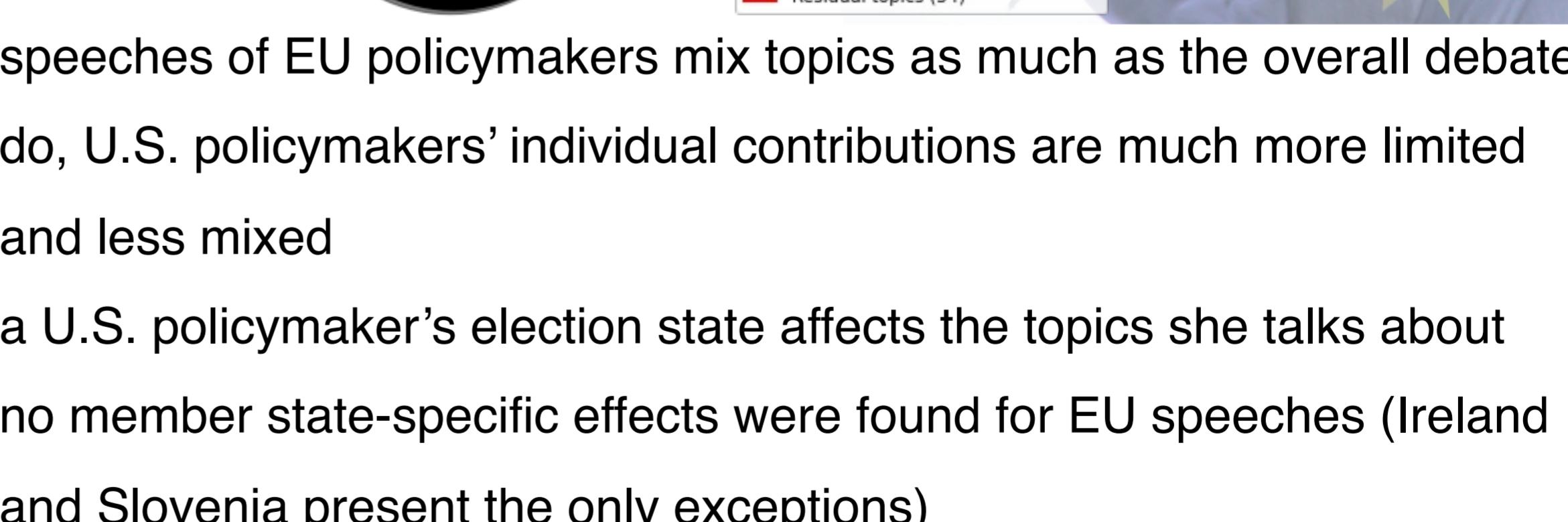
- Latent Dirichlet Allocation topic models on both corpora to uncover prevalent topics and their marginal contribution to debates/speeches
- independent-Samples test to validate the statistical significance of differences in topics between the U.S. and EU
- multivariate regression analyses to understand (member)state-specific effects on topic prevalence

Results – debate corpus



- differences between EU and U.S. topics are statistically significant
- on the debate level, many topics are shared across the EU and the U.S.; “AI education” is equally present

Results – speaker corpus



- speeches of EU policymakers mix topics as much as the overall debates do, U.S. policymakers' individual contributions are much more limited and less mixed
- a U.S. policymaker's election state affects the topics she talks about
- no member state-specific effects were found for EU speeches (Ireland and Slovenia present the only exceptions)

Implications

- replicated the perspective on transatlantic differences as established in qualitative research
- advanced the research in the field by identifying a specific policy field where the EU and U.S. may align on AI governance and discovered meaningful dynamics both in the U.S. and EU contexts
- showed that a quantitative research design that uses methods of NLP succeeds as a reliable, scalable, and productive alternative to qualitative text analysis for comparative policy analysis