# GATE Disinformation Research:
# Expanding the Boundaries of Disinformation Research on Bulgarian Social Media Content

**Milena Dobreva**,
Irina Temnikova, Silvia Gargova,
Hristiana Krasteva, Ivo Dzhumerov

GATE
big data for smart society

## Why this topic?

- Bulgaria is ranked on the 30th position in the European Media Literacy index for 2021
  - The top: Finland (1st), Denmark (2nd), Estonia (3rd), Sweden (4th) and Ireland (5th)
  - The bottom: North Macedonia (35th), Bosnia and Herzegovina (34th), Albania (33rd), Montenegro (32nd) and Turkey (31st).
- Bulgaria is also among the countries with higher vulnerability and lower resilience towards disinformation in the region
- Undersupply of local fact-checkers (and subsequently low visibility in aggregated fact checks, e.g. https://www.poynter.org/ifcn-covid-19-misinformation/ -
  only 6 from 9,000+ fact checks)

# Cyberbalkanization and the Balkans

- The regional vs the global trends

  "Cyberbalkanization refers to the idea of segregation of the Internet into small political groups with similar perspectives to a degree that they show a narrow-minded approach to those with contradictory views."

Engin Bozdag, Jeroen V D Hoven (2015) Breaking the filter bubble: democracy and design. Ethics and Information Technology, 17 (4), pp. 249–265.

- Local example


# Why this topic?

- Background
  - Disinformation research (NLP, ML, AI applications) is very prominent, but not on Bulgarian content
  - Paradox: Bulgarian companies develop tools for analysis of disinformation in English
  - There are useful first steps done – but there is also a gap to bridge
- After some preparatory work in 2020, GATE institute started building a team to address this topic in the summer of 2021

# Ambition

To develop tools which will allow to support disinformation detection for Bulgarian.

Societal impact: support for fact-checkers, media agencies, academic/educational institutions, the citizens.

What do we need?

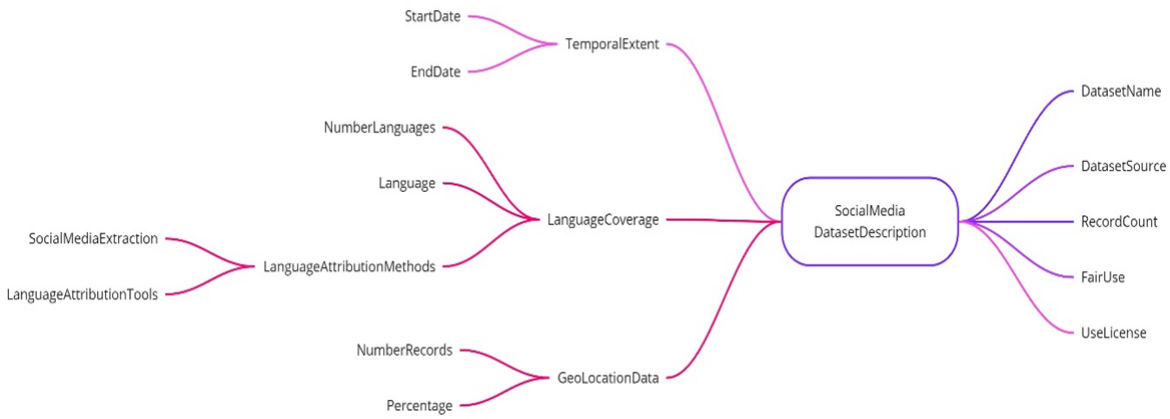
- Datasets, instruments, models

# What have we done so far?

- Extraction of a set of 52,000 tweets, analysis of 3,200 of them
  - 20% in languages other than Cyrillic
  - 4% clickbaits
  - Tools for language identification
  - Language analysis (BoW) almost done

  More details in the video
- Extraction and annotation of facebook messages in Bulgarian
  (aligned to the international study of FirstDraft on English, French, Spanish)

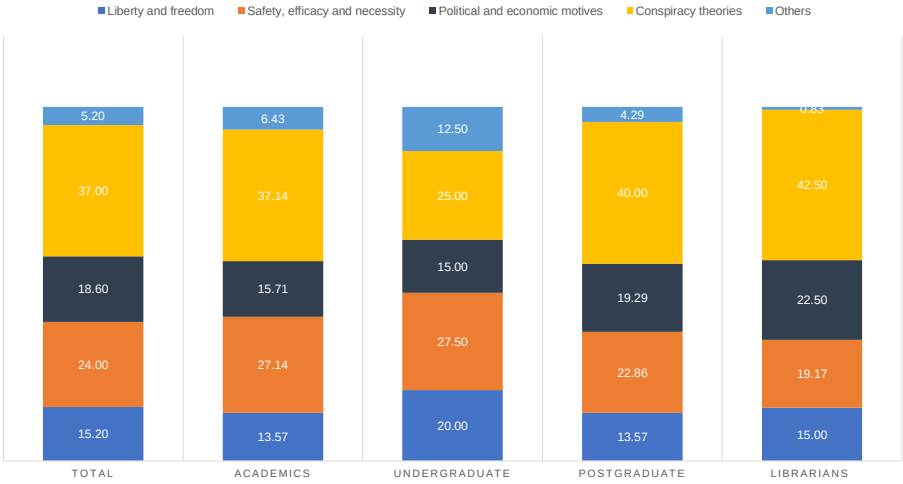
# Sideline: Characterisation of datasets




# Classification experiment

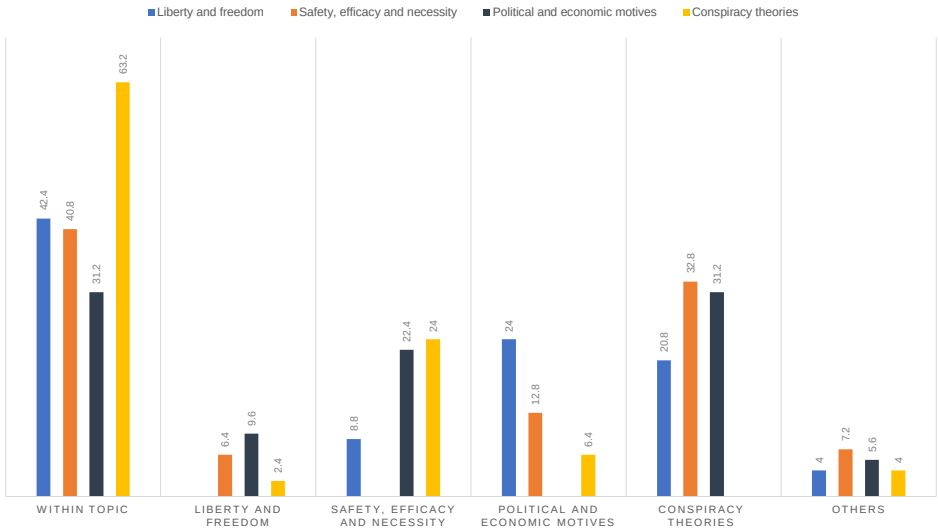
- Two sets of 20 real social media messages (Facebook)
- One with long and one with short messages (long paragraph / sentence)
- Each set has 5 examples across 4 categories
- Development; Religion excluded – no examples
- Annotators can choose the categories and enter a value for OTHER when they have an opinion the statement does not fit any of the four categories

# Results: annotation exercise




*Distribution of the annotations by four groups of users (7 academics, 5 undergraduate, 7 postgraduate students and 6 librarians).*

# Results: annotation exercise

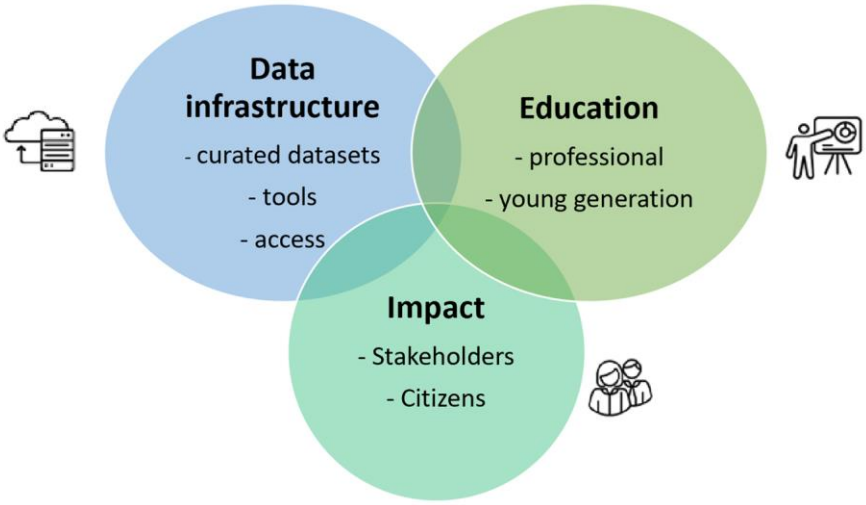



Distribution of the annotations which match the original annotation and mismatched annotations per category

# We also offer educational programmes!

Summer school, 27-28 July 2021, Plovdiv

# Summary



# Thank you!

**Projects**

- Disinformation pilot project
- Social Explainable AI

13