
# Protecting Electoral Integrity in the Age of Disinformation: AI-Powered Social Media Monitoring

Başak Bozkurt
Oxford Internet Institute, University of Oxford
basak.bozkurt@oii.ox.ac.uk

**EDMO Scientific Conference on Disinformation 27 February 2024, University of Amsterdam**

# Safeguarding Electoral Integrity in the Digital Age

- **Challenge:** The rise of disinformation on social media threatens the integrity of elections.
- **Global Response:** Governments have begun to monitor social media to counteract disinformation.
- **Research Focus:** Investigating how governments utilise social media monitoring and the role of AI-powered tools.

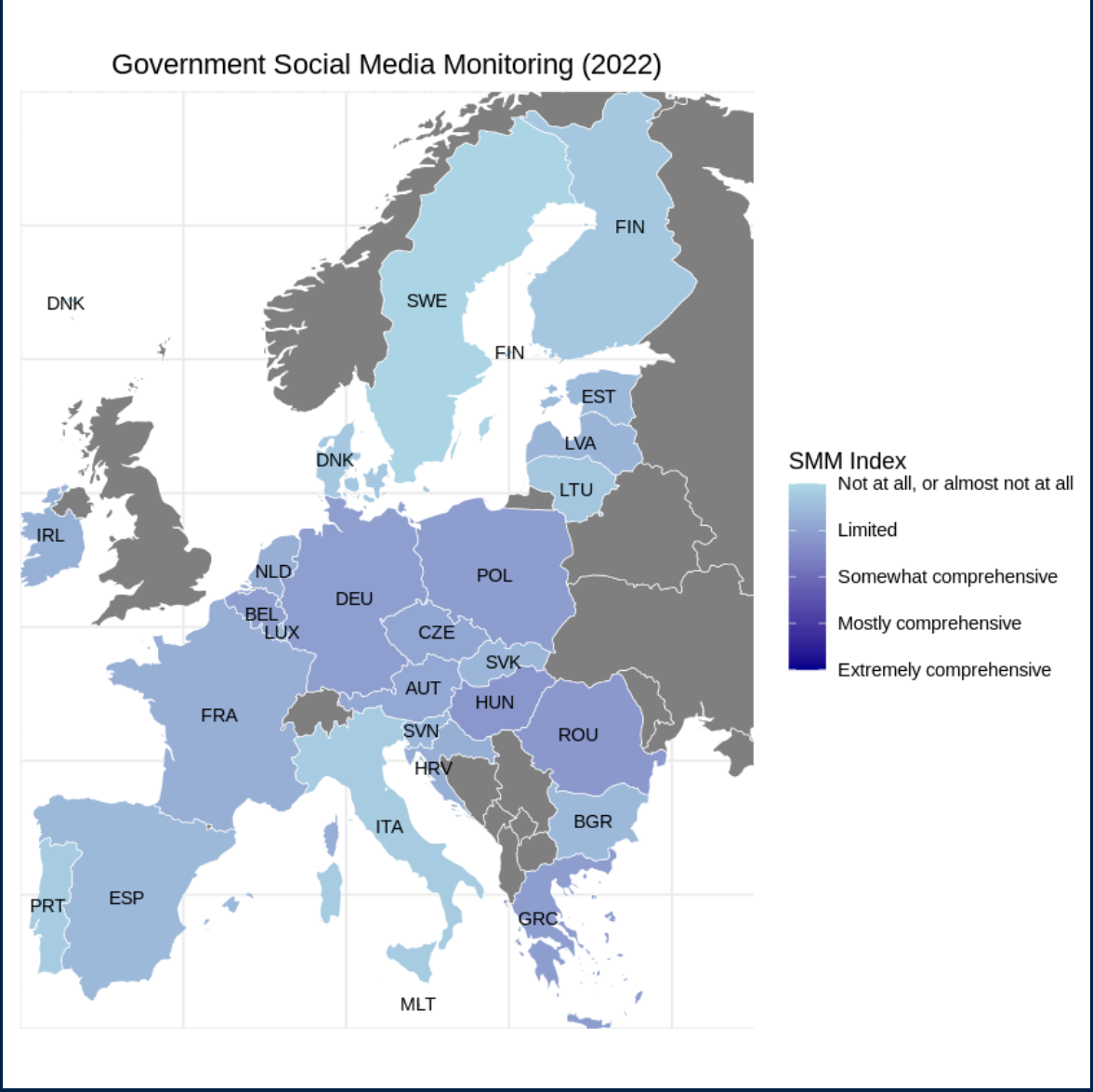
- **Research Questions:**
  - (RQ1): How does the use of social media monitoring by government influence the integrity of elections?
  - (RQ2): How do governments use AI in SMM during elections?
  - (RQ3): What are the opportunities and challenges created by Incorporating AI-powered social media monitoring tools?

# Methodology

- Scope - Time frame: EU members, 2016-2022.
- Literature on SMM*, reports by election observation missions (OSCE ODIHR), government publications, policy documents and election-related news articles.
- Electoral integrity data (V-Dem Dataset), Press Freedom Index (Reporters Without Borders), economic development data (World Bank)
- Government disinformation responses dataset's "technical and algorithmic responses" category Cipers et al. (2023), Global Surveillance Index (Feldstein, 2022), list compiled by Schmuziger Goldzweig et al. (2019) on government initiatives for social media monitoring

- * Using Boolean operators, I searched for the terms "social media monitoring" "social media intelligence", "social media surveillance" and "social media listening" along with the following three-word stems: "government*", "state*", "artificial intelligence*," on Scopus, Web of Science and Google Scholar.

“How comprehensive is the surveillance of political content in social
the government or its agents?” (Coppedge et al., 2022)



Regression Models Summary

| | Dependent variable: | |
|---|---|---|
| | Electoral integrity | |
| | Model 1 (1) | Model 2 (2) |
| Electoral management body autonomy | 0.484*** (0.046) | 0.472*** (0.051) |
| Electoral management body capacity | 0.045 (0.070) | 0.036 (0.069) |
| CSO | 0.263*** (0.047) | 0.209*** (0.051) |
| Press freedom | −0.007*** (0.003) | −0.006** (0.003) |
| Judiciary high court independence | 0.217*** (0.033) | 0.206*** (0.033) |
| Judiciary lower court independence | −0.147*** (0.054) | −0.107* (0.054) |
| GDP per capita (log) | −0.100*** (0.034) | −0.100*** (0.034) |
| Government social media monitoring | | 0.901*** (0.261) |
| Party dissemination of false information | | 0.287** (0.143) |
| Foreign governments dissemination of false information | | 0.855*** (0.237) |
| Interaction: Party * Govt SMM | | −0.086** (0.043) |
| Interaction: Foreign governments * Govt SMM | | −0.238*** (0.070) |
| Constant | 2.630*** (0.279) | −0.490 (0.921) |
| Observations | 189 | 189 |
| R2 | 0.715 | 0.744 |
| Note: | *p<0.1; **p<0.05; ***p<0.01 | |

- Social media monitoring is a significant predictor in Model 2.
- Less social media monitoring is associated with more robust elections.
- The presence of disinformation negatively affects this relationship.


Czechia
Denmark
France
Germany
Hungary
Italy
Lithuania
Slovakia
Sweden
EU

https://electionmonitorai.com/dataset/ (Bozkurt, 2024)

# The surveillance AI framework

*The seven axes of surveillance (Caines, 2022, with additional explanations)*

| Axis | Description |
|---|---|
| **The surveyor** | Government entities, special task forces, departmental networks |
| **The surveyed** | Both domestic and foreign entities across social media platforms |
| **Methods and instrumentalities** | NLP, machine learning algorithms, social media analytics |
| **Passive/Real time monitoring** | Real time: Dynamic tracking of election-misinformation on social media<br>Passive: Retrospective analysis |
| **Subsequent effect** | Legal sanctions<br>Removal of accounts or content<br>Diplomatic interventions |
| **Accountability** | Challenging to Evaluate |
| **Transparency** | Limited |

## Opportunities

- Automation (Feldstein, 2019; Shahbaz & Funk, 2019; Patel et al., 2019 )
- Disinformation and hate speech detection e.g. HaterNet (Pereira-Kohatsu et al., 2019)
- Real-time crisis management (Morović, 2020)

## Challenges

- Misclassification
- Linguistic barriers (Vytautas Magnus University, 2021).
- Shift towards less regulated platforms, e.g. Germany 2021 election (ODIHR Election Expert Team, 2022).
- Privacy violations, infringements on freedom of speech

# Conclusion

- Potentials:
  - Monitoring media coverage
  - Enforcing campaign silence periods
  - Regulating political ads

- Need to ensuring a balanced approach to maintain electoral integrity without infringing on rights

# Thank you!

basak.bozkurt@oii.ox.ac.uk