A Computational Social Science Approach to Online Violence Against Female Political Candidates in the 2023 Turkish National Elections

Işık Topçu*, Amanda Zaner**, Melih Can Yardı*, Fırat Durusan*, Erdem Yörük*, Müge Finkel**, Steven E. Finkel**, Jungmin Han**, Ali Hürriyetoğlu***

*Koç University, **University of Pittsburgh, ***Wageningen University

Background

Full and equal participation of women in politics is a necessary component of democratic governance and sustainable development. However, as women's participation and representation in politics increases, so does their risk of becoming victims of political violence. Women in politics are 27 times more likely to experience online abuse than men in equivalent positions (Amnesty International, 2023).

Online violence against women in politics (OVAW-P) refers to a spectrum of digitally perpetrated acts, including cyberstalking, sexual harassment, doxing, intimate image abuse, hateful or offensive speech, and disinformation that attack and undermine women as political actors on the basis of gender or other identity factors (Krook, 2020).

Abstract

A focus on Türkiye extends the scope of analysis beyond US-centered content, enriching our understanding of OVAW-P as a globally digital phenomenon (Fontanella, Chulvi, Ignazzi et al. 2024). Further, as over 88% of world languages lack access to LLM resources (Sajid, 2024), training the deep-learning algorithms with a corpus of Turkish language content helps build linguistically and regionally-specific speech classification systems.

This research examines online violence in X (Twitter) posts directed against politicians in the 2023 parliamentary and 2024 municipal elections in Türkiye using a Large Language Model (LLM). We refined deep-learning models capable of detecting and classifying instances of threats, insults, abuse, and hate speech directed at female and male candidates. Our work expands upon existing literature in both Gender-Based Violence (GBV) and LLM research.

Research Questions

1. Are there differences in the extent and nature of online violence directed towards female and male candidates in Turkish elections?
2. Are there differences in the extent and nature of online violence directed towards candidates at the national versus municipal election levels?
3. Is gender-based online violence more prevalent than violence related to other characteristics of candidates, such as ethnicity or religion?
4. Do female candidates from different political parties, ideologies, or with different levels of prominence experience more targeted online violence?
5. Can we build accurate LLMs to detect hate speech in Türkiye based on gender and other identity-related candidate characteristics?

Live Q&A with our Research Team

Want to Learn More? Ask Our Researchers!

Other members from our Computational and Policy research teams are available to answer specific questions via Zoom chat during the presentation.

Q&A Link

Methodology

Data:

Tweets were acquired from ERC-funded Politius Analytics database[1].

1. Local elections data[2]:
Dates 6 months before 31 March 2024:

- Total tweets: 89,670 (64.5% female and 35.5% male)
- Offensive speech tweet count: 10,702 (11.9% of total tweets)
- Hate speech tweet count: 3,411 (3.8% of total tweets)

2. General Elections Data[3]:
Dates 6 months before 14 May 2023:

- Total tweets: 6,025,597 (87.5% male and 12.5% female)
- Offensive speech tweet count: 813,713 (13.5% of total tweets)
- Hate speech tweet count: 123,782 (2.1% of total tweets)

Active Learning Strategies in Annotation:

1. LLM Approach:

Hate speech binary annotation with ChatGPT-4o[4] was conducted on a sample of ten thousand tweets to create a balanced sample (equal number of tweets from each gender-party combination) to help increase the identification of hate speech instances for human annotation.

2. Hate Speech Model[5]:

A Hugging Face hate speech model was run on a large sample of tweets. After processing, a balanced sample was obtained, ensuring an equal number of tweets from each gender-party combination for human annotation.

Classification Categories:

1. Offensive Speech or Not: Classifies tweets into offensive language or not.
2. Offensive Speech Intensity: Classifies offensive speech tweets into the following categories: Wishing or Threatening Harm, Insult (Profanity), Insult (Demeaning and Ridicule.
3. Hate Speech or Not: Classifies tweets into hate speech or not.
4. Hate Speech Types: Classifies hate speech tweets into the following categories: Gender, Religion or Ethnicity.

BERT-Based Language Models

Pretrained transformers model used:
dbmdz/bert-base-turkish-128k-cased[6]

Performance Metrics:

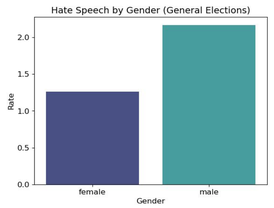
Hate Speech Types:
F1 Macro: 0.77
Precision Macro: 0.88
Recall Macro: 0.71

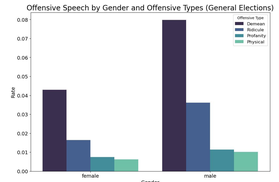
Offensive Speech Types:
F1 Macro: 0.70
Precision Macro: 0.73
Recall Macro: 0.69

[1] https://politusanalytics.com/
[2] Parties analyzed for the local elections are AKP, CHP, MHP, IYI, DEM/YSP.
[3] All parties have been analyzed for the general elections.
[4] https://platform.openai.com/docs/models
[5] https://huggingface.co/ctoraman/hate-speech-berturk
[6] https://huggingface.co/dbmdz/bert-base-turkish-128k-cased

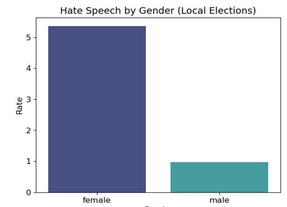
Results

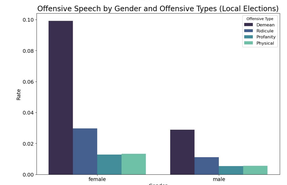
General Elections:



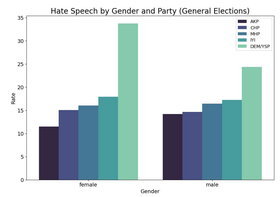


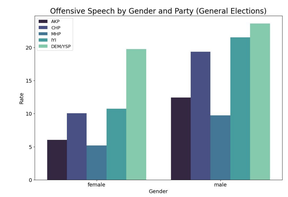
Local Elections:

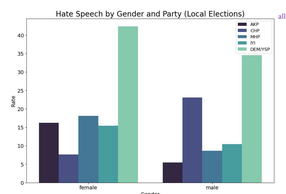


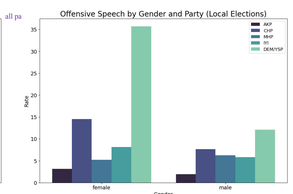


Male candidates received higher rates of hate and offensive tweets than female candidates in the 2023 Turkish national elections while the opposite was true during the municipal elections in 2024.

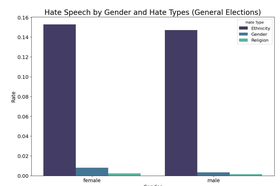


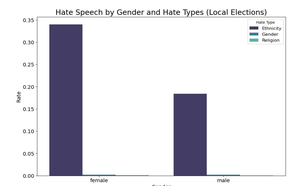


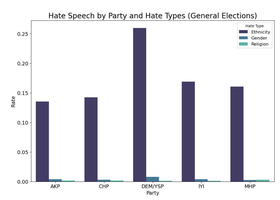


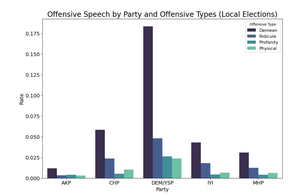


Both female and male candidates from the DEM/YSP (Kurdish) party received the highest rates of hate and offensive tweets than any other political party both at the national and local elections.









The type of hate speech directed against both female and male candidates from all parties is overwhelmingly ethnicity-based and less so based on gender or religion.

Scan the QR code to view all figures.

Conclusion

1. **Gender-Based Targeting:** Female candidates face disproportionately high hate speech in local elections, while male candidates are more targeted in national elections.
2. **Intersectionality:** Female candidates from DEM/YSP (Kurdish) parties in local elections receive the highest levels of hate and offensive tweets.
3. **Ethnicity-Based Hate and Offensive Tweets:** For both genders in both elections, hate and offensive speech predominantly targets ethnicity, rather than gender or religion.
4. **Methodology:** Synthesizing expert annotations, language models, and large language models was crucial for achieving high performance in hate speech detection, and for creating benchmark models specific to Turkish Twitter's hate speech landscape.

Further Research

- The team will investigate candidate characteristics, such as ideology, incumbency, and party list placement, that influence the prevalence of hate and offensive tweets.
- Future research will aim to predict the profiles of individuals who engage in hate and offensive speech on social media.
- Methodologically, the focus will be on developing advanced models to detect hate and offensive speech specifically targeting candidates.
- Expanding the annotation dataset to include more ethnically and culturally diverse samples will enhance the robustness and applicability of the models.

References

Key References:

Alkomah, Fatimah, and Xiaogang Ma. 2022. "A Literature Review of Textual Hate Speech Detection Methods and Datasets" Information 13, no. 6: 273.

Bjarnegård, Elin, and Zetterberg, Pär, eds. 2023. Gender and Violence Against Political Actors. Philadelphia: Temple University Press. Accessed May 27, 2024.

Bardall, Gabrielle, et al. "How Is Political Violence Gendered? Disentangling Motives, Forms, and Impacts." Political Studies, vol. 68, no. 4, 3 Dec. 2019.

Krook, Mona Lena. Violence against Women in Politics. New York, Ny, Oxford University Press, 2020.

Nicholas, G. and Bhatia, A. (2023) Lost in Translation: Large Language Models in Non-English Content Analysis. Center for Democracy & Technology.

Complete Reference List:

Acknowledgements

This research was conducted by the "Online Violence Against Women in Politics" working group at the Ford Institute for Human Security, and supported by the Kabak Grant through the Global Studies Center at the University of Pittsburgh.

For more information visit - FORD INSTITUTE FOR HUMAN SECURITY Fordinstitute.pitt.edu

KOÇ UNIVERSITY CENTER FOR COMPUTATIONAL SOCIAL SCIENCES Politus Analytics