AI4Democracy: Dynamic Dashboard for Analyzing Polarization and Extremism in Online Media

University of Hamburg, Bahir Dar University, FAKTENKONTOR, Zirkel Technologies GmbH, NORDAKADEMIE

Deep Learning Indaba 2024
September 1-7 2024, Senegal, Dakar

AI-MAP: AI-driven Monitoring of Attitude Polarization in Conflict-Affected Countries for Inclusive Peace Process and Women's Empowerment - Ethiopia

60%

- **Goal of the research**
  - Explore the impact of social media polarization in Sub-Saharan Africa (Ethiopia) on offline unrest, with a specific focus on women.
  - Focus Area: Ethiopia
  - Target Community: Women
  - Main Concern: Social media polarization **fuels armed conflict**
- **Objective**
  - Develop a **digital peacebuilding pipeline**, incorporating AI tools for automated classification.
  - Goal: Establish a digital infrastructure for peacebuilding
  - Key Tools: AI-driven **automated classification**
  - Outcomes: Streamlined digital processes for peace and reconciliation efforts
- **Issues**
  - Mitigate **marginalization of women's voices** in peacebuilding processes by providing viable solutions for addressing social media polarization.
  - Problem: Women's voices are often **marginalized in peacebuilding** efforts
  - Approach: Address social media polarization and **digital divide**
  - Solution: Provide actionable strategies to amplify women's participation and influence
- **Findings so far (from a survey - qualitative data):**

Threat — Social media fuels polarization and information disorder in Ethiopia

Weaponized — Social media used as weapon in Tigray war

Interconnected — Offline politics impact the online social media environment

Visible Conflict — Digital conflict is severe and evident in Ethiopia

Urgent Need — Absence of digital peace-building efforts demands immediate action

Alarming Demand — Ethiopia urgently needs digital peace-building initiatives

AI for Peace — High demand for AI, especially in peace-building

Online Impact — Few online affect millions without internet access

Unregulated — Social media in Ethiopia lacks proper regulation

Right-Wing Extremism Analysis in Germany

25%

Sentiment Trends:
- **Objective**: Track and understand sentiments of right-wing extremists.
- **Method**: NLP analysis of historical social media posts.
- **Focus**: Key events influencing sentiment changes and extremist sentiment patterns.

Network Analysis:
- **Objective**: Map communication networks.
- **Method**: Social Network Analysis to visualize influencers and echo chambers.
- **Focus**: Key actors and echo chamber effects.

Influence on Public Discourse:
- **Objective**: Assess the impact on public conversations.
- **Method**: Content analysis and sentiment comparison.
- **Focus**: Exploration of hate speech and public opinion shifts due to disinformation.

Polarization and Radicalization Patterns:
- **Objective**: Examine polarization driven by extremism.
- **Method**: AI analysis of social media interactions.
- **Focus**: Correlation between extremist messaging and increased polarization.

Policy and Intervention Strategies:
- **Objective**: Support with data-driven insights.
- **Approach**: Evidence-based recommendations, collaborations with NGOs and academics.
- **Impact**: Inform policymakers to reduce extremism and develop prevention strategies.

Goal:
Provide insights into **right-wing extremism** on social media, compare with other politicians, and understand the **impact on public discourse and algorithmic prioritization**. This work aims to generate deeper insights into social media's role in facilitating right-wing extremism.

Neo-fascist Classification - in the USA Context

15%

- **Neo-fascist Behavior Online:**
  - **Objective**: Detect neo-fascist discourse in online presence.
  - **Indicators**: Hate speech, politicization of minorities, social inequality justification.
  - **Method**: Machine learning algorithms.
  - **Focus**: Common themes and rhetoric identification
- **Hate Speech Identification:**
  - **Objective**: Detect hate speech prevalence and context.
  - **Method**: NLP and machine learning.
  - **Focus**: High frequency and targeted groups
- **Network and Influence Analysis:**
  - **Objective**: Map networks of neo-fascist influencers.
  - **Method**: Online media (forums) analysis.
  - **Focus**: Key influencers and network structures.
- **Influence on Public Discourse:**
  - **Objective**: Assess neo-fascist narratives' impact.
  - **Method**: Sentiment analysis and tracking.
  - **Outcome**: Shifts in discourse and normalization of extremist views.
- **Strategies for Mitigation:**
  - **Objective**: Provide insights to combat neo-fascist rhetoric.
  - **Mitigation**: Recommendations for policymakers, NGOs, and awareness programs.
  - **Outcome**: Inform policies and promote community resilience.
- **Goals:**
  - Normalizes referencing extremist views in public discourse.
  - Network and influence patterns are crucial for effective intervention.
  - Offer comprehensive analysis of neo-fascist behavior online.
  - Identify key themes.
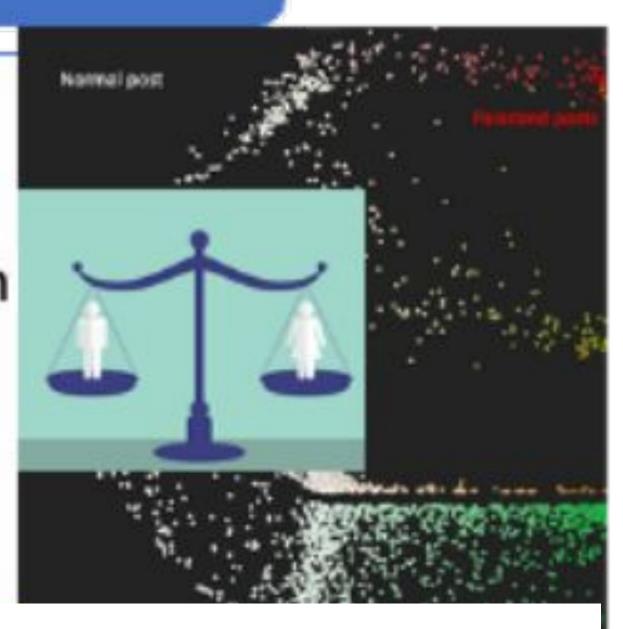 - Support strategies to combat their societal impact.

Data collection → **Data processing** → **Outcome**

- Facebook
- Twitter
- Telegram
- WhatsApp

- NLP
- ML approaches
- LLM finetuning
- Prompt Eng.

AI

- Comment classification
- Data visualization
- Collaborative encoding

Google Research