SemEval 2026-Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, Özge Alaçam, Cengiz Acartürk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Simona Frenda, Alessandra Teresa Cignarella, Elena Tutubalina, Oleg Rogov, Aung Kyaw Htet, Xintong Wang, Surendrabikram Thapa, Kritesh Rauniyar, Tanmoy Chakraborty, Arfeen Zeeshan, Dheeraj Kodati, Satya Keerthi, Sahar Moradizeyveh, Firoj Alam, Arid Hasan, Syed Ishtiaque Ahmed, Ye Kyaw Thu, Shantipriya Parida, Ihsan Ayyub Qazi, Lilian Wanzare, Nelson Odhiambo Onyango, Clemencia Siro, Jane Wanjiru Kimani, Ibrahim Said Ahmad, Adem Chanie Ali, Martin Semmann, Chris Biemann, Shamsuddeen Hassan Muhammad, **Seid Muhie Yimam**

ACL 2026 SAN DIEGO JULY 2–7

https://polar-semeval.github.io/

- AI-MAP - The Story of POLAR
- POLAR
- POLAR@SemEval 2026

The Story of POLAR

AI-MAP – Social Media, Gender (Women) & Peacebuilding in Conflict Zones

Google Award for Inclusion Program 2023

How POLAR Started: The AI-MAP Project

- **AI-MAP** - AI-driven Monitoring of **Attitude Polarization** in Conflict-Affected Countries for **Inclusive Peace Process** and Women's Empowerment
- **Funded by the Google Award for Inclusion Research (~$60K)**
- Partners:

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Hub of Computing & Data Science (HCDS), Universität Hamburg, Germany

Bahir Dar University
"Wisdom at the Source of the Blue Nile"

Dept. of Journalism & Communication, Bahir Dar University, Ethiopia

- **Project goals**:
  - Study how digital platforms **intensify violence and spread hate**
  - Document the impact of the war on marginalized voices, especially **women, in conflict?**
  - Build AI tools that amplify, not suppress, underrepresented perspectives
  - Strengthen women's participation in peace processes
- **Case study: the Northern Ethiopia conflict (2020–2022)**

The Trigger: The Northern Ethiopia (Tigray) War

- Nov 2020 – Nov 2022: one of the deadliest conflicts in recent Ethiopia
- Estimated 311K–808K deaths; millions of people affected, especially women and children (although numbers differ among global reports)
- Social media (Facebook, X, YouTube, Telegram) are polarized and weaponized for propaganda, hate speech, and disinformation
- Government-imposed internet shutdowns; ethnic “us vs. them” narratives spilled from online into offline violence (and vice versa)
- Pretoria Peace Agreement (Nov 2022) ended active fighting, but digital harms persisted

Silenced Voices: The Motivation of Study

- Qualitative study: 10 key-informant interviews + 2 FGDs
- Finding: social media **weaponized** to deepen polarization; **women disproportionately** harmed (displacement, gender-based violence) and excluded from peace processes
- **Call to action:** use AI / NLP and LLMs to detect polarized content and support **inclusive digital peacebuilding**

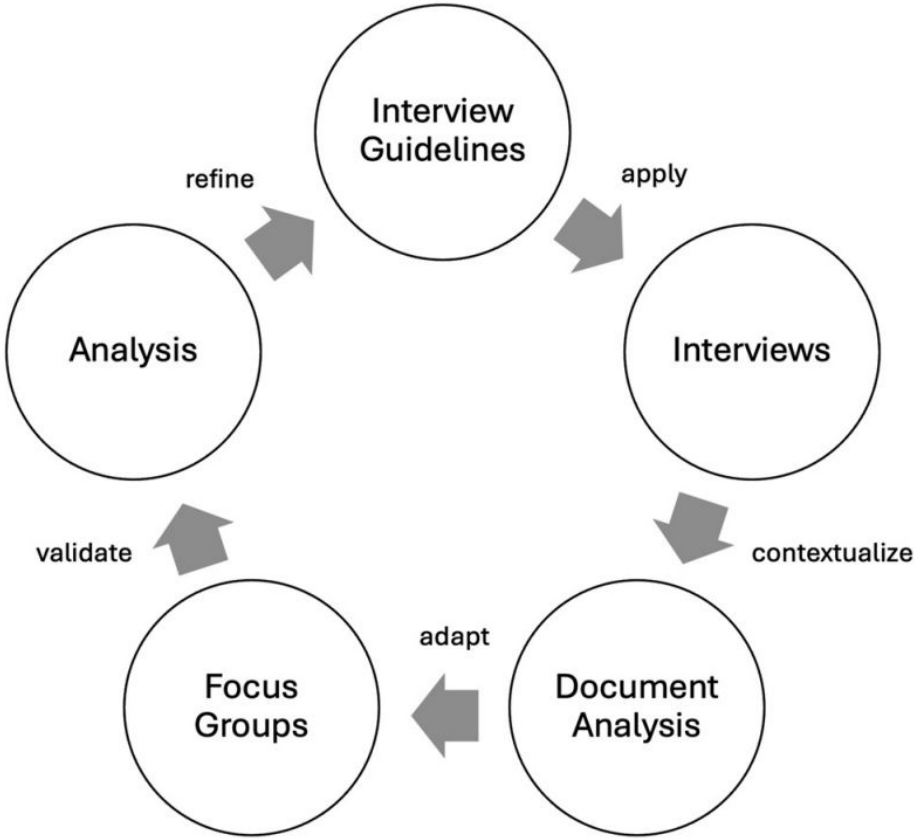
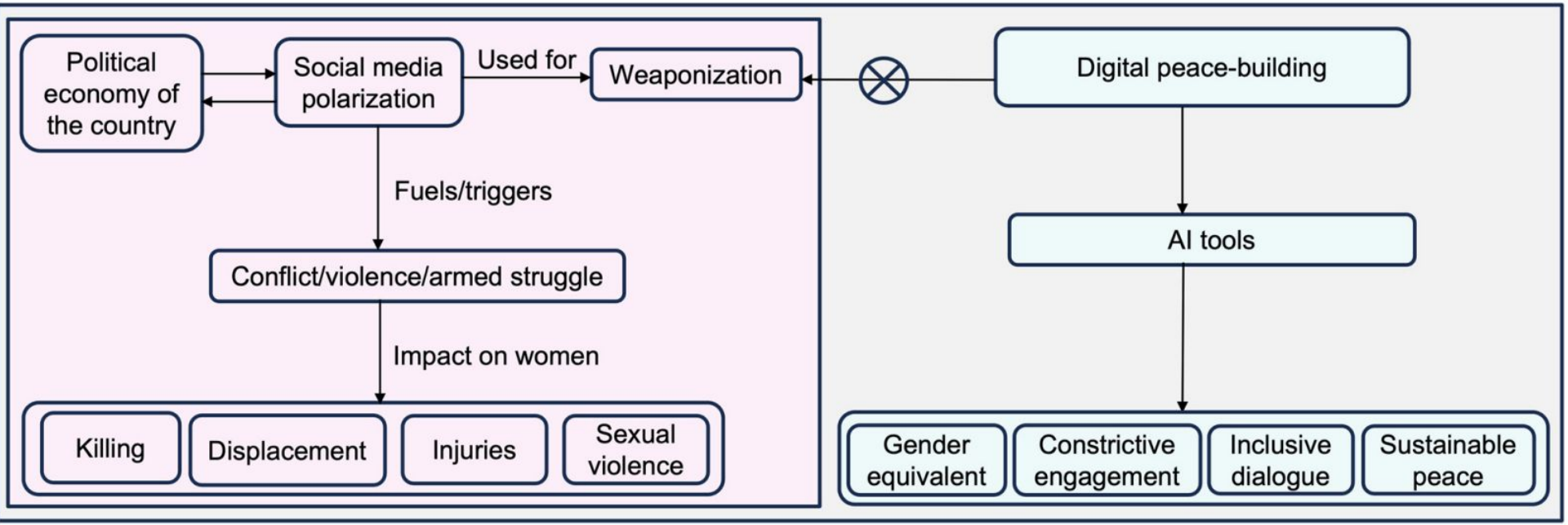


Figure 1: The iterative data collection and analysis process.

Social media polarization and the dearth of digital peacebuilding



A Digital Media Environment in Crisis

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

DE GRUYTER

i-com 2025; 24(2): 407–432

Research Article

Adem Chanie Ali*, Seid Muhie Yimam, Abinew Ali Ayele, Chris Biemann and Martin Semmann

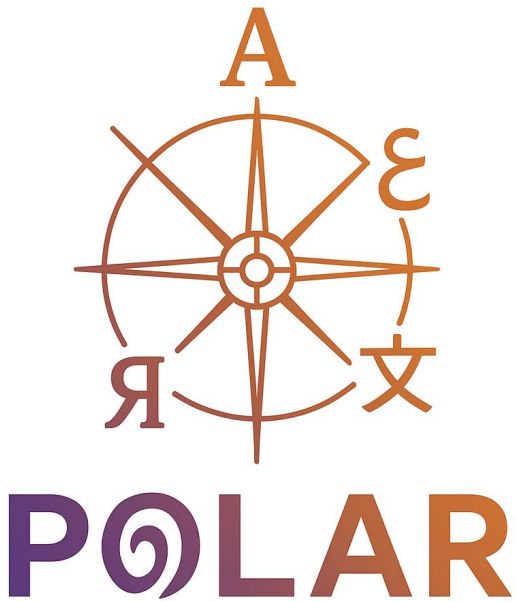
Silenced voices: social media polarization and women's marginalization in peacebuilding during the Northern Ethiopia War

https://doi.org/10.1515/icom-2025-0007
Received February 28, 2025; accepted August 4, 2025;
published online September 1, 2025

literacy. It recommends media literacy initiatives, inclusive peacebuilding frameworks, open and safe digital spaces, and gender-sensitive technological approaches. By center-

From AI-MAP to POLAR

- The gap in polarization NLP:
  - English-centric and high-resource only
  - Event-specific (e.g., U.S. elections)
  - Binary / topic-focused: ignores rhetorical tactics
- **Our response**: POLAR: a large-scale multilingual, multicultural, multi-event benchmark for fine-grained polarization
- From one conflict (Tigray) to 22 languages and many global events



Detecting Multilingual, Multicultural and MultiEvent Online Polarization

- Online polarization is dangerous for a democratic society
- POLAR is a multilingual, multicultural, and multi-event dataset with over **110K instances** in **22 languages** drawn from diverse online platforms and real-world events
- Polarization is annotated by **detection**, **type**, and **manifestation**, using various annotation platforms adapted to each cultural context
- Two main experiments: **fine-tuning six pretrained small language models** and evaluating a range of open and closed **large language models in few-shot and zero-shot** settings

POLAR - Pipeline

Data Processing (22 languages)

1.Data collection

- X(Twitter)
- Bluesky
- Local news

2. Data Preprocessing

- Keyword-based event selection
- Deduplication
- Anonymization

3. Annotation Tool Setup

- MTurk
- Prolific
- LabelStudio
- POTATO

POLAR Annotation

4. Is the text polarized?

YES NO

5. If polarized, select all applicable polarization types:

- [x] Political/Ideological
- [ ] Racial/Ethnic
- [ ] Religious
- [x] Gender/Sexual
- [ ] Others

6. If polarized, select all expressed manifestations of polarization:

- [x] Stereotyping
- [ ] Vilification
- [ ] Dehumanization
- [x] Absolutism
- [ ] Lack of Empathy
- [x] Invalidation

SLMs and LLMs Evaluation for 22 langauges

(amh, arb, ben, deu, eng, fas, hau, hin, ita, khm, mya, nep, ori, pan, pol, rus, spa, swa, tel, tur, urd, zho)

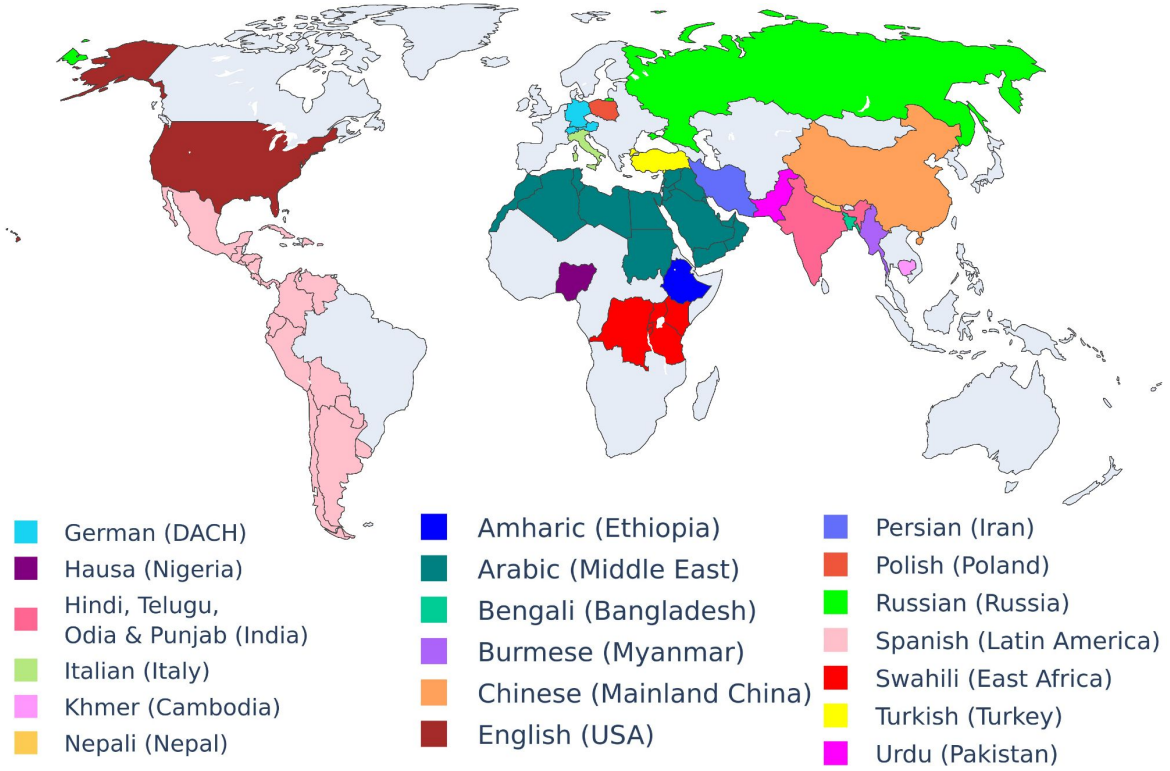
Multilingual, Multicultural

22
Languages

110K
Dataset

Multiple sources

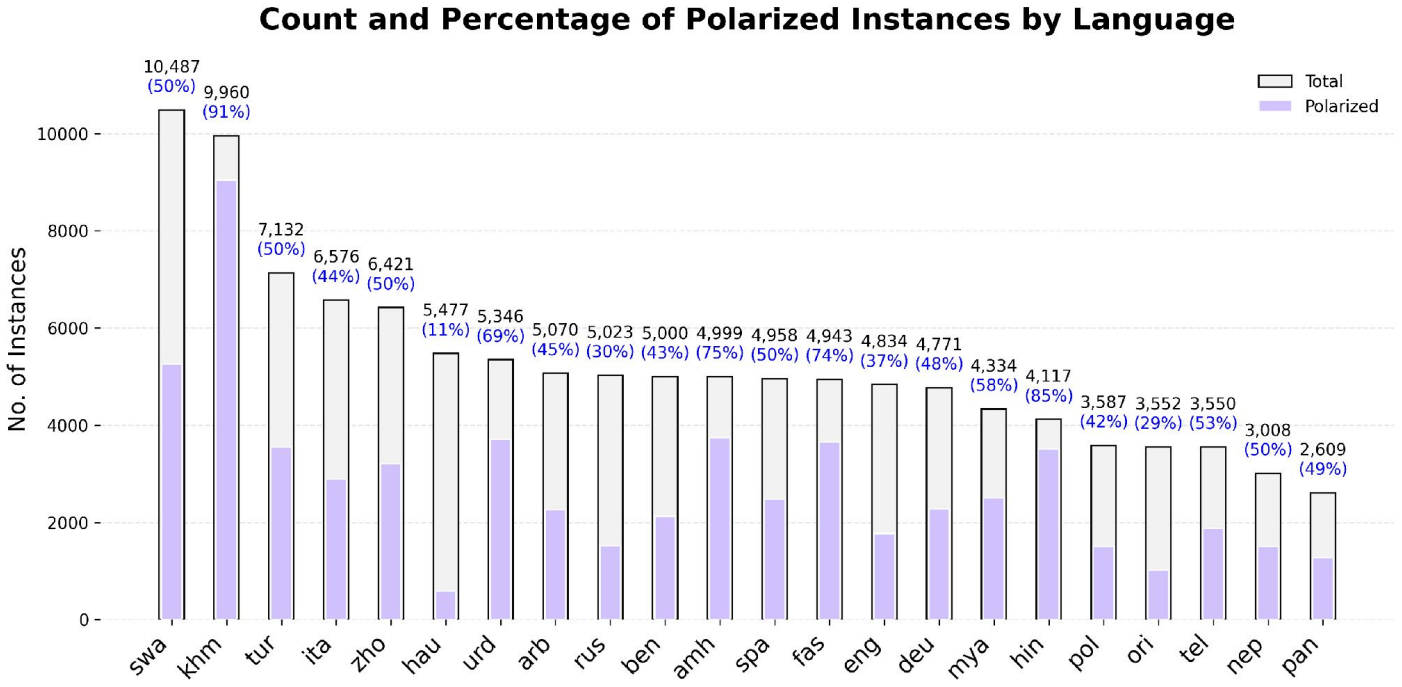
Source	No. of Instances	%
X (Twitter)	58,984	53.3%
News/websites	15,208	13.7%
Bluesky	8,099	7.3%
YouTube	7,159	6.5%
Reddit	5,705	5.2%
Existing Dataset	4,794	4.3%
Weibo	4,550	4.1%
Facebook	2,928	2.6%
Zhihu	1,088	1.0%
Threads	852	0.8%
Tieba	783	0.7%
Wikipedia	500	0.5%
Total	**110,650**	**100%**



Multi-Event


Justice Reform
Ethnicity
Popular Culture
Elections
Racism
Crime News
Social Issues
Pride Parade
Politics
Kenyan Elections
Tigray War
Sexuality
US Elections
Gender
Religion
Caste
Conflicts
Contemporary Topics
Immigration
Gender Inequality

Dataset Composition by Language



Count and share of polarized instances per language — high polarization in conflict-heavy, low-resource languages (e.g., Khmer 91%, Amharic 75%, Hindi 85%)

Dataset - Labels

Polarization
The increasing extremity of opinions, beliefs, or behaviors, resulting in heightened intergroup divisions and conflict.

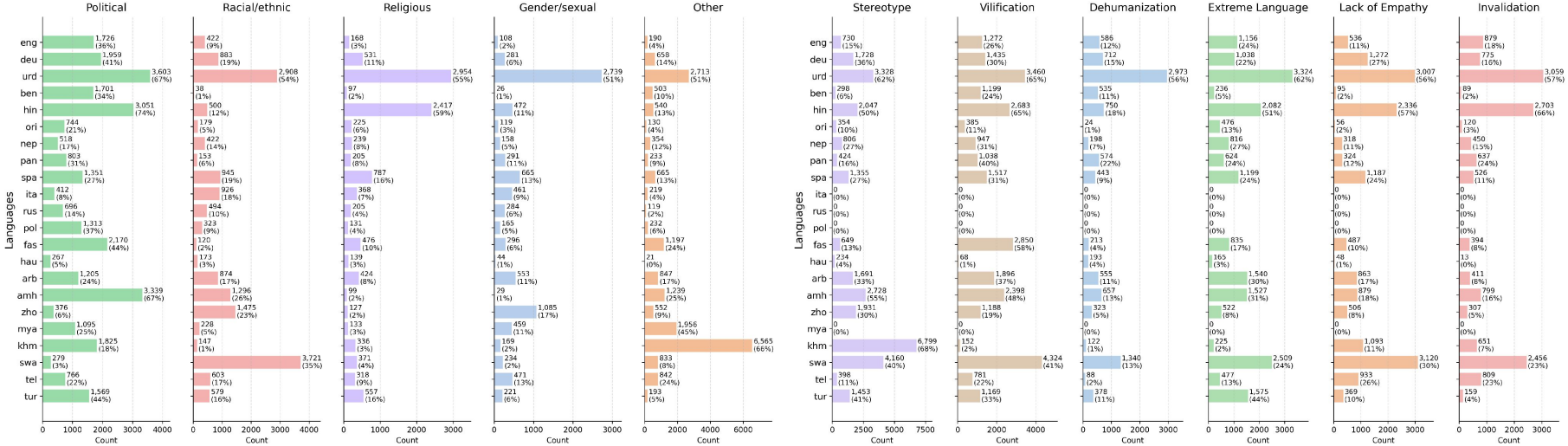
If yes

Types of polarization
- Political polarization
- Racial or ethnic polarization
- Religious polarization
- Gender/Sexual polarization
- Other

If yes

Manifestations of polarization
- Stereotype
- Vilification
- Dehumanization
- Extreme Language and Absolutism
- Lack of Empathy
- Invalidation

What Polarization Looks Like: Types & Manifestations



Left: polarization TYPES per language (political dominates). Right: rhetorical MANIFESTATIONS (stereotype & vilification most frequent).

Experimental Setup

- **PolarDetect**
  Classes: *polarizing, not polarizing*

- **PolarType**
  Classes: Political polarization,Racial or ethnic polarization,Religious polarization,Gender/Sexual polarization,Other

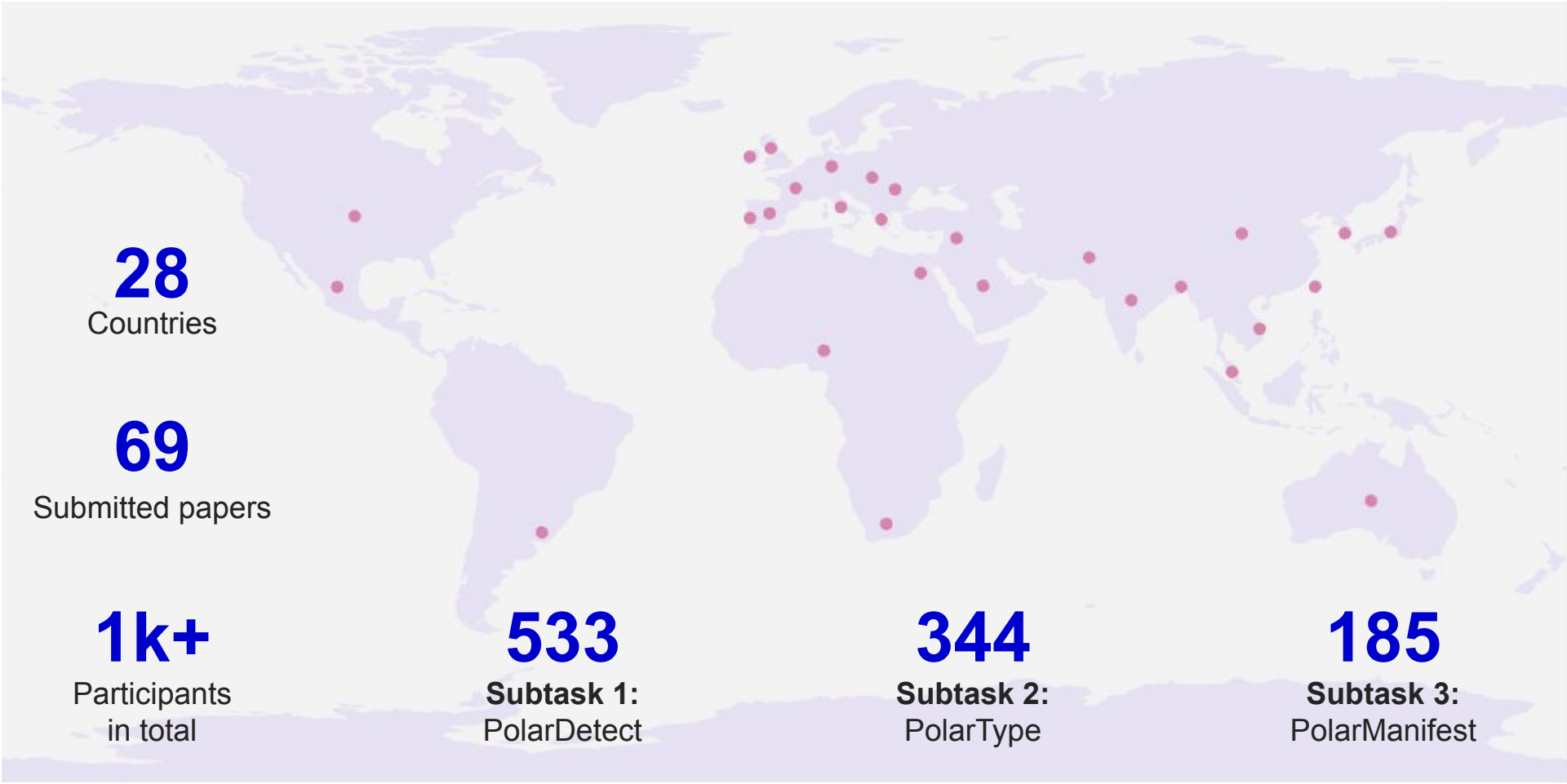
- **PolarManifest**
  Classes: Stereotype, Vilification, Dehumanization, Extreme Language and Absolutism, Lack of Empathy, Invalidation

Evaluation Metric	Baseline
Macro F-score	Fine-tuned LaBSE

POLAR @ SemEval-2026 Task 9

1000+ participants across 28 countries

Tasks Summary-Participation



28
Countries

69
Submitted papers

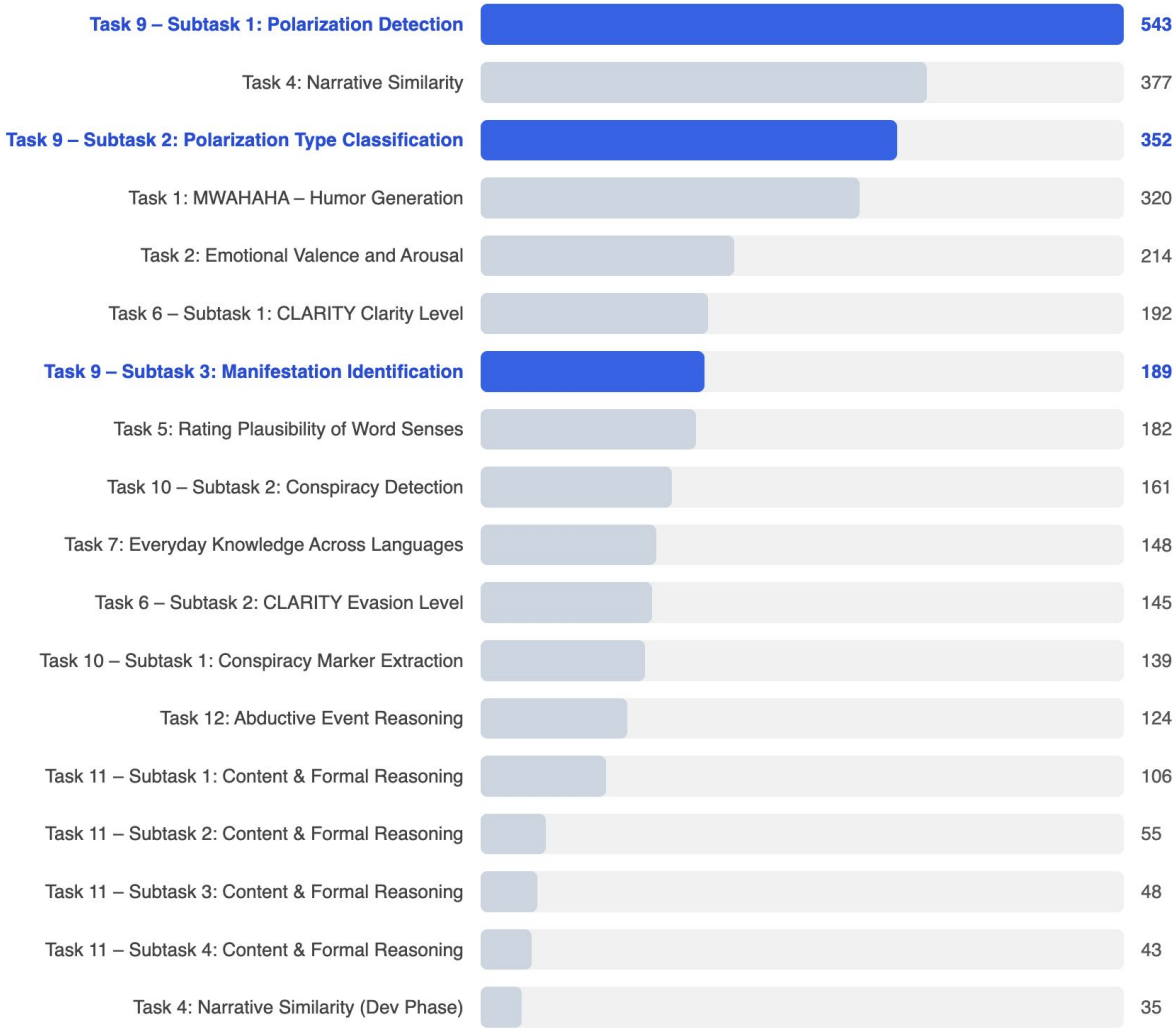
1k+
Participants in total

533
Subtask 1: PolarDetect

344
Subtask 2: PolarType

185
Subtask 3: PolarManifest

Most Populous SemEval Task in Decade



https://www.codabench.org/competitions/public/?page=1

Two More Upcoming Presentations of POLAR

POLAR - Benchmark - ACL Findings Poster

July 6, 9:00 Poster Session D

POLAR: A Benchmark for Multilingual, Multicultural, and Multi-Event Online Polarization

POLAR@SemEval 2026

July 4, 9:00

SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization

Thank you

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, Özge Alaçam, Cengiz Acartürk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Simona Frenda, Alessandra Teresa Cignarella, Elena Tutubalina, Oleg Rogov, Aung Kyaw Htet, Xintong Wang, Surendrabikram Thapa, Kritesh Rauniyar, Tanmoy Chakraborty, Arfeen Zeeshan, Dheeraj Kodati, Satya Keerthi, Sahar Moradizeyveh, Firoj Alam, Arid Hasan, Syed Ishtiaque Ahmed, Ye Kyaw Thu, Shantipriya Parida, Ihsan Ayyub Qazi, Lilian Wanzare, Nelson Odhiambo Onyango, Clemencia Siro, Jane Wanjiru Kimani, Ibrahim Said Ahmad, Adem Chanie Ali, Martin Semmann, Chris Biemann, Shamsuddeen Hassan Muhammad, Seid Muhie Yimam

https://polar-semeval.github.io/

Subtask 2: Polarization Type Classification

Looking at the given social media texts, the type or target polarization is classified as follows:

Political/ideological polarization: This type of extremism focuses on division, intolerance, and conflict between political parties and followers. Political polarization refers to political beliefs and affiliations becoming more extreme. People may identify more strongly with their political party, leading to deeper divides and a reduced willingness to compromise. It broadens ideological differences between political groups.

Racial or ethnic polarization: This type of polarization focuses on ethnic identity or racial origin and incites division, intolerance, and conflict between ethnic groups or races. This type of polarization arises when individuals identify more strongly with their own racial or ethnic group, leading to increased separation, mistrust, or conflict with individuals from other groups.

Religious polarization: This type of polarization focuses on religious identity and incites division, intolerance, and conflict between religious followers.

Gender polarization: This type of polarization refers to the exclusion, discrimination, and marginalization of individuals based on their gender. **Sexual orientation polarization:** This refers to the increasing division and distinction between different sexual orientations within society, often leading to heightened tensions, misunderstandings, conflicts, or marginalization among various groups.

Other: polarization texts targeting other groups/identities such as economy, technology, media, polarization, etc.

Subtask 3: Manifestation Identification

A message/ text on social media is considered to be polarizing if it exhibits one or more of the following characteristics:

Stereotype: This manifestation occurs when a message generalizes certain characteristics of individuals to all members of a group, ignoring individual differences. Stereotypes simplify complex personalities into one-size-fits-all representations.

Vilification: Vilification appears when a text defames or demonizes a particular group, person, or entity, often inciting fear through exaggeration, misrepresentation, or biased framing that portrays the subject in a harmful or negative light.

Dehumanization: This occurs when language strips a group or individual of human qualities or dignity, often by comparing them to animals, machines, or objects, or by otherwise denying their humanity and individuality.

Extreme Language and Absolutism: This manifestation involves the use of extreme or absolutist language that reflects polarized attitudes, such as words like “always,” “never,” “worst,” or “best.” It often presents issues in dichotomous terms such as “us vs. them” or “right vs. wrong.”

Lack of Empathy or Understanding: This occurs when the text shows no empathy or understanding for others’ perspectives or experiences. It may involve marginalizing alternative viewpoints or refusing to understand or relate to them.

Invalidation: Invalidation appears when a text denies or rejects the identity or existence of certain people or groups, dismissing their legitimacy or right to exist.