Secure Societies 2024: Horizon Europe Cluster 3 Brokerage Event in Istanbul Pitch Presentation

Serena Bressan

Fondazione Bruno Kessler (Italy)

DE-ESCALHATE

Digital EmpowErment for combating hate Speech on soCial mediA and in the reaL world through advanced and trustwortHy AI-based Tecnologies and Education

Contact Details

- Name: Serena Bressan
- Organisation: Fondazione Bruno Kessler – FBK
- E-mail: bressan@fbk.eu
- Telephone: +39 0461314054
- Country: Italy

Proposal activity

HORIZON-CL3-2024-FCT-01-05 | Combating hate speech online and offline

Description of the Organization

- **FBK** *is a prestigious Italian research institution that excels in* **technology, innovation, humanities, and social sciences**. *The consortium will comprise its* **Digital Society Center and Advanced Intelligence Center**.
- *FBK boasts* **extensive expertise in countering hate speech and disinformation**, *developed through participation in European projects like* **Hatemeter, Hatedemics, AI4TRUST, and AI-CODE**.

.

- **Expertise** *related to the topic*
  - ○ *A) Natural Language Processing, Hate Speech and Disinformation Detection*
  - ○ *B) Online Social Networks Mining and Analysis*
  - ○ *C) Counter-Speech Generation, Generative AI*

Proposal idea/content #1

- *FBK proposes itself as **Coordinator of the DE-ESCALHATE proposal***

- *Description of the proposed **project idea:***

  - **General objective**
  - ➢ To develop innovative **AI-based methods** such as EU-GDPR compliant **social media tracking, social network analysis (SNA) and counter-narrative generation** as well as interactive **educational toolkits** for combating online and offline **hate speech**, while preserving **user privacy** and promoting **social cohesion** in European societies.

Proposal idea/content #2

- **Specific objectives**
  - To mine and analyse **messages shared through social media** to identify and understand by means of **social network analysis** the spread of hate speech within online communities, enabling targeted interventions **to limit the dissemination of harmful content and promote positive discourse**.
  - To develop and **implement generative AI-based counter-narrative techniques** to mitigate the spread and impact of hate speech on online platforms, with a focus on identifying and **addressing language patterns indicative of hate speech** across multiple languages and cultures **in full dialogical settings**.
  - To design and deploy **interactive educational tools and resources** aimed at promoting tolerance and inclusivity **among online communities and in the real world**. This objective involves the creation of educational modules tailored to different age groups and societal demographics based on the innovative digital tools developed during the project.

- **Expected results**
  - To develop a **scalable and adaptable AI model capable of generating counter-narratives** in dialogical settings characterised by multiple languages and cultural contexts, **reducing the virality and impact of hate speech** across diverse online platforms.
  - To create an integrated **suite of educational tools and resources** aimed at **raising awareness about the dangers of hate speech and fostering constructive dialogue** among individuals from various backgrounds, ultimately contributing to a more inclusive and tolerant society.

Project participants

Existing Consortium - Profile of Known Partners

No	Partner Name	Type	Country	Role in the Project
01	FINCONS Group	Company	CH	System Integrator
02	Sabanci University	Academia	TK	Social Network Analysis
03	Amnesty International	NGO	IT	Education, Training & Activism
04	StopHate UK	NGO	UK	Education, Training & Activism
05	SAHER EU	SME	EE	Law, Ethics & Training

Project participants

Existing Consortium – Required Partners

No	Partner Name	Type	Country	Role in the Project
01	Relevant LEAs	LEA	Eligible Country	Pilot participants
02	Relevant NGOs	NGO	Eligible Country	Pilot participants
03	Experts in pilot management	TBA	Eligible Country	Pilot leaders
04	Sociology	RTO/A	Eligible Country	Literature review, collection of socio-technical requirements & evaluation
05	Criminology	RTO/A	Eligible Country	Literature review, collection of socio-technical requirements & evaluation