

Expression of Interest to Join a Horizon Europe Consortium

Call: Advisory support and network for countering and preventing radicalisation, extremism, hate speech and polarisation

* HORIZON-CL2-2027-01-DEMOCRACY-01

Deadline: 23 September 2027

Research Partner Expression of Interest

dr. Ilia Markov

Department of Language, Literature and Communication (LLC), Vrije Universiteit Amsterdam (VU)

Expertise: Computational Linguistics, Natural Language Processing, Artificial Intelligence

i.markov@vu.nl

Expertise and Contribution to the Call

I investigate how hate speech is expressed in online environments, developing computational methods for its automatic detection, analysis, and explanation. My work focuses on the automatic identification of hate speech in text and multimodal content (e.g., harmful memes) across multiple languages. I also work on the automatic generation of counterspeech as a response strategy to mitigate the spread and impact of online hate speech.

This expertise aligns directly with the call's priorities by:

- Providing robust AI-based tools for detecting and analysing online hate speech and extremist narratives, supporting a better understanding of the mechanisms underlying the circulation of such content across online environments and communities.
- Contributing to the development and evaluation of scalable countermeasures, including counterspeech generation and intervention strategies, that can be adapted by public authorities, NGOs, local authorities, and practitioners.
- Enabling cross-linguistic and multimodal analysis of hate ecosystems, addressing the cross-border nature of online hate speech, radicalisation, extremism, and polarisation, and their circulation across digital environments.
- Strengthening interdisciplinary collaboration between computer science, social sciences, and policy stakeholders, ensuring that technical solutions are grounded in real-world societal needs and end-user contexts.

My methodological approach combines machine learning, deep learning techniques, and generative large language models, with a focus on explainability, robustness, and real-world applicability in socially sensitive domains.

Possible Roles within the Consortium

I can contribute to or lead work packages related to:

- AI-based detection and analysis of online hate speech, extremist narratives, and polarising content across languages and modalities (text and memes).
- Development of generative AI and NLP-based counterspeech and intervention strategies to mitigate online hate speech and support prevention efforts.
- Evaluation of AI-based tools and interventions in real-world contexts, in collaboration with stakeholders such as NGOs, public authorities, and practitioners.

Specific activities include:

- Designing and implementing machine learning and deep learning models for automatic hate speech detection and classification
- Developing and evaluating generative AI methods for counterspeech production in online environments
- Conducting model evaluation, error analysis, and interpretability studies to improve robustness and trustworthiness
- Supporting evidence-based insights for stakeholders, including NGOs, public authorities, and practitioners
- Contributing to real-world evaluation of AI tools and their applicability in societal contexts

Institutional Profile: Vrije Universiteit Amsterdam (VU)

VU Amsterdam is a research-intensive university in the Netherlands with strong interdisciplinary expertise across Social Sciences, Humanities, Law, Sciences, and Medicine. Research within the Faculty of Social Sciences and Humanities focuses on digital society, communication, language, and social cohesion, with growing strengths in computational social science and AI-driven analysis of online communication.

The Department of Language, Literature and Communication (LLC) has expertise in computational linguistics, multilingual language processing, and the analysis of online discourse, including hate speech, misinformation, and polarisation. Research combines AI-based methods with social science approaches to study complex digital societal phenomena.

VU Amsterdam is based in Amsterdam and actively participates in Horizon Europe, ERC, and NWO-funded projects, with strong international collaborations across academia, public institutions, and industry in the field of digital society and inclusive communication.

Contact

dr. Ilia Markov

Department of Language, Literature and Communication (LLC), Vrije Universiteit Amsterdam (VU)

i.markov@vu.nl