



# REAL UP EUROPEAN PROJECT

*Hate speech, racism and xenophobia: Alert Mechanisms and Response, analysis of the Upstander speech*

## ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

**CONCEPT NOTE - 12TH MARCH FROM 10:00 A.M. TO 12:00 P.M (CET). (1/2)**

The **REAL-UP** project “**Hate Speech, Racism and Xenophobia: Alert Mechanisms and Response, analysis of the Upstander speech**” is cofounded through the Citizens, Equality, Rights and Values (CERV) program. The project has been structured in five work packages (WP), distributed between the project leader, the Spanish Observatory on Racism and Xenophobia (OBERAXE), and the four partners: the National Office for Combating Hate Crimes of the Ministry of Interior of Spain (ONDOD), the Community of Research on Excellence of the University of Barcelona (UB-CREA), the University of Valencia (UV) and a civil consultant agency Cidalia.

The general objective of the Real UP project is to improve the capacities of state authorities to identify, analyze, monitor and evaluate online hate speech. The project aims at developing and strengthening counter-narrative strategies (upstander speech) against hate speech motivated by racism, xenophobia, Islamophobia, anti-Semitism and anti-Gypsyism. These will be done by discussing and generating a map of good practices and effective didactic tools for counter-narrative hate speech online; as well as, describing the profiles of perpetrators and victims, through which the intensity, severity, distribution and potential impact of hate speech can be assessed; and to establish action-oriented recommendations based on the upstander speech.

Among the partners, the National Office for Combating Hate Crimes (ONDOD), is assigned with the development of WP3. This work package focuses on the design and implementation of an artificial intelligence tool aimed at monitoring online hate speech, which will facilitate the detection and investigation of hate crimes by law enforcement agencies.

According to that workpackage the ONDOD, will organize an international online seminar (Spanish/English) scheduled for the 12th March of 2024. This training event will be structured in three main sections. The first section will include a welcome speech and the presentation of the main milestones and the most innovative tools developed under the Real UP project”.



This Project has received funding from the EU



## ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

**CONCEPT NOTE - 12TH MARCH FROM 10:00 A.M. TO 12:00 P.M (CET). (2/2)**

The second module will focus on the presentation of the academic research conducted by the four specialists who have participated in this work package. They will present topics such as: the use of the “Haterbert” algorithm and the implementation of a tracker specialized in hate speech; the characteristics of the profiles of hate-mongering users and the victims of this type of speech; the impact of hate speech on social networks; the most effective strategies for the automatic generation of counter-narratives and the effectiveness of the Upstanders’ speech.

Finally, the third section will be dedicated to exploring the application and impact of these investigations, in line with Digital Services Act (DSA) guidelines, focusing on two key areas: cyber-policing best practices and the role of trusted flaggers combating illegal online hate speech.

<https://real-up.eu/>



# REAL UP EUROPEAN PROJECT

*Hate speech, racism and xenophobia: Alert Mechanisms and Response, analysis of the Upstander speech*

## ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

**PROGRAMME - 12TH MARCH FROM 10:00 A.M. TO 12:00 P.M.**

### INSTITUTIONAL MODULE: Total 20 minutes

<b>10:00</b>	<p><b>Welcome:</b></p> <ul style="list-style-type: none"> <li>Ms. Pilar Cancela Rodríguez Secretary of State for Migration. Ministry of Inclusion, Social Security and Migration.</li> <li>Mr. Rafael Pérez Ruiz, Secretary of State for Security, Ministry of Home Affairs.</li> </ul>
<b>10:10</b>	<p><b>REAL UP Project:</b></p> <ul style="list-style-type: none"> <li>Positioning of the REAL UP project in the framework of the European Commission, Ms. Karoline Fernández de la Hoz, Director of OBERAXE.</li> <li>Main milestones of the European REAL UP project, Ms. Carmen Girón Tomás, Technical Advisor of the OBERAXE.</li> </ul>
<b>10:15</b>	<p>Approximation to the main innovative tools analysed in REAL UP, with a specialised multidisciplinary approach, Lieutenant Colonel, Ms. Pilar Muniesa Tomás, Head of the Statistical System and Victim Services, ONDOD.</p>

### ANALYSIS AND ACADEMIC RESEARCH MODULE: Total 50 minutes

<b>10:25</b>	<p>Automatic monitoring of hate speech in social network "X" and other parts of the web: use of "Haterbert" and other artificial intelligence algorithms, D<sup>a</sup> Lara Quijano, professor, UAM.</p>
<b>10:35</b>	<p>Characteristics of the hater user profile and victims of hate speech. The adaptation of the counter-narrative to each user in X: Rosa Elvira Lillo, Professor, Universidad Carlos III.</p>
<b>10:45</b>	<p>The impact of hate speech on social networks, José Camacho Collados, Professor, Cardiff University.</p>
<b>10:55</b>	<p>The best strategies for automatic generation of counter-narratives, Ms. Maite Martín Valdivia, Professor, University of Jaén.</p>
<b>11:05</b>	<p>Effective cyberpatrolling resulting in criminal conviction, Mr. Pedro Antonio Martínez Herranz, EVO Valencia</p>



## ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

**APPLICATION AND IMPACT MODULE: 1) EXAMPLE OF GOOD PRACTICE IN CYBERPATROLLING AND 2) CONTRIBUTION TO POSITIVE ONLINE CYBERACTIVISM ALIGNED WITH THE DSA. Total 20 minutes**

<b>11:15</b>	Trusted flaggers refresher course to prevent and combat illegal online hate speech: Commander, Mr. Joaquín González Salas, Information Service of the Guardia Civil and the University Centre of the Guardia Civil, in the framework of the Inter-institutional Agreement to combat racism, xenophobia, LGTBphobia and other related forms of intolerance.
<b>11:55</b>	Closing session, ONDOD.
<b>12:00</b>	<b>END OF WORKSHOP</b>

**QUESTIONS FROM THE AUDIENCE: Total 30 minutes**

<https://real-up.eu/>



# ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

## SPEAKERS

1/5



### **Ms. Pilar Cancela Rodríguez**

Secretary of State for Migration since December 2023. She holds a degree in Law from the University of Santiago de Compostela and a Higher Degree in Occupational Risk Prevention. She is a civil servant in the Senior Corps of the General Administration of the Xunta de Galicia (1993). Throughout her career she has held various positions such as: President of the Equality Commission of the Congress (2016-2021); President of the Commission on the State Pact on Gender Violence; Secretary of State for International Cooperation - Ministry of Foreign Affairs, European Union and Cooperation - (2021-2023); President of the Spanish Agency for International Development Cooperation (2021- 2023); and President of the Advisory Group of the UN Secretary General for the localisation of the SDGs of the 2020 Agenda.

### **Mr. Rafael Pérez Ruiz**

Secretary of State for Security. Born in 1981. He is a magistrate of the General Council of the Judiciary, where he was Head of the Judicial Office Section. Until his appointment, he held the position of Director of the Cabinet of the Minister of the Interior. Previously, he worked in various courts, such as those of First Instance and Instruction in Almadén (Ciudad Real) and Montoro (Córdoba) and the Criminal Court No. 1 in Córdoba. He has been a lawyer for the General Council of the Judiciary, acting as Head of the Judicial Office Section.

### **Ms. Karoline Fernández de la Hoz Zeitler**



Doctor in Medicine from the University of Barcelona, Master in epidemiology at the London School of Hygiene & Tropical Medicine, diploma in public health and pharmacoconomics. In 2022 I received the Order of Merit of the Civil Guard. Since 2015 I am Director of the Spanish Observatory on Racism and Xenophobia. My commitment is to public service, collaboration between institutions and civil society and the encouragement of human teams. I started my professional career in the regional and national public administration with responsibilities in epidemiological surveillance and health alerts. In 2005 I joined the European Centre for Disease Prevention and Control (ECDC) participating in its establishment. Back in Spain, in the Ministry of Health I dealt with international affairs and participated in the management of crises such as the influenza A pandemic (2009) as a member of the European Health Security Committee. In 2015 I received the ECDC medal.





## ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

### SPEAKERS

2/5

#### **Ms. Carmen Girón Tomás**

Technical Advisor of the Spanish Observatory of Racism and Xenophobia (OBERAXE), in the Secretariat of State for Migration of the Ministry of Inclusion, Social Security and Migration since 2019, civil servant of the Higher Scale of Traffic Technicians since 2003, Head of International Relations at the Directorate General of Traffic from 2008 to 2018, PhD student in Law and Social Sciences at the UNED, author of several publications, national and international institutional speaker. She participates in the coordination of European projects: CISDO Inter-Police and Social Cooperation against Hate Crimes <https://cisdoproject.eu/> , REAL UP Hate speech, racism, and xenophobia: Alert and Response Mechanisms <https://real-up.eu/> , Upstander Discourse Analysis, IMMERSE, Integration Mapping of Refugee and Migrant Children in Schools and Other Experiential Environment in Europe, <https://www.immerse-h2020.eu/es/> HELCI <https://helci.usal.es/> Higher Education Learning Community for Inclusion, ALRECO, Hate Speech, Racism and Xenophobia: Alert and Coordinated Response Mechanisms, <https://www.inclusion.gob.es/oberaxe/es/ejes/delitosodio/alreco/index.htm>

#### **Ms. Pilar Muniesa Tomás**

She is the Head of the Statistical System and Victim Support, and Lieutenant Colonel in the Guardia Civil, where the National Office for the Fight against Hate Crimes (ONDOD) is located, within the General Directorate for Coordination and Studies of the Secretary of State for Security in the Ministry of the Interior.

<https://real-up.eu/>



# ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

## SPEAKERS

3/5



### Ms. Lara Quijano-Sánchez

Born in Madrid, Spain. She is a professor at the Universidad Autónoma de Madrid and an active member of the UC3MSantander Big Data Institute of the Universidad Carlos III de Madrid. She is an expert in data science techniques, recommender systems, user modelling, information search and retrieval, social network analysis and the design of Big Data models. Currently, her research focuses on machine learning models for police, security and personal safety. He actively collaborates with law enforcement on projects such as predictive policing, smart patrolling and femicide prevention. Her achievements include the Spanish Police Foundation Research Awards in 2016-2017 and 2017-2018.

### Ms. Rosa E. Lillo



Rosa Elvira Lillo. She was born in Mérida (Badajoz). She graduated in Mathematics with Extraordinary Prize from the Complutense University of Madrid in 1992 and obtained her PhD in Mathematical Sciences from the same university in 1996. Since 2010 she is Professor of Statistics and Operations Research in the Department of Statistics at the University Carlos III of Madrid. She has been the director of the Degree in Statistics and Business, director of the Department of Statistics and is currently the director of the uc3m-Santander Big Data Institute ([www.ibidat.es](http://www.ibidat.es)). She has published more than 90 articles in international journals. She has supervised 18 doctoral theses and has collaborated in a large number of projects both in research and applications to the business world and public institutions through transfer contracts. In 2012, she obtained the Excellence Award for young researchers granted by Banco Santander and in the past she was a visiting professor at the University of California (Berkeley) and the University of Arizona (Tucson). Since 2005 she has won several awards for her teaching and research work and since 2014 she has been a member of several evaluation commissions at ANECA and the AEI. She is also very committed to outreach work and is the creator of the “STAT WARS” events that have reached national and international dimensions.



# ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

## SPEAKERS

4/5



### Mr. José Camacho Collados

Professor at Cardiff University in the UK where he heads the Natural Language Processing group. Before joining Cardiff University, he completed his PhD at Sapienza University in Rome and was a Google Doctoral Fellow. His research has focused on a variety of natural processing topics, particularly in the multilingual and language understanding domains. He wrote the book “Embeddings in Natural Language Processing” and is the general chair of the \*SEM 2024 conference dedicated to these topics. More related to the topic of the talk, in recent years José has been working on social network analysis and applications, developing computational tools for this domain, such as methods for hate speech detection.

### Ms. M. Teresa Martín-Valdivia



Professor in the Department of Computer Science at the University of Jaén. She belongs to the SINAI (Sistemas INteligentes de Acceso a la Información) research group, a national and international reference in Natural Language Processing (NLP). Currently, his main lines of research are focused on the study and development of PLN systems applied to biomedicine and, on the other hand, in systems based on sentiment and emotion analysis for social purposes such as hate speech detection systems, detection of mental health problems (anorexia, depression, suicide...) or detection of offensive language as well as the generation and integration of linguistic resources that allow adaptation to Spanish to improve this type of systems.

### Mr. Pedro Antonio Martínez Herranz

National Police Inspector, Head of the Far Right and Latin Gangs Group of the Provincial Information Brigade of Valencia, Head of the Violent Extremism and Hate Group (EVO), COPRA Valencia Coordinator of the European project CISDO, Professor of the subject “Hate Crimes” of the Valencian Institute of Public Safety and Emergencies in the training of Local Police of the Valencian Community, Trainer of Trainers of the European project TAHCLE.





# ADVANCES IN AI FOR DETECTING AND COMBATING HATE SPEECH IN SOCIAL NETWORKS

## SPEAKERS

5/5

### Mr. Commander Joaquín González Salas

He is stationed at the Information Headquarters, currently Central Social Interlocutor and head of the National Hate Crimes Response Team (REDO Nacional). Joined the Guardia Civil in 2009. During his tenure as Lieutenant and Captain, he was stationed at the Madrid Command, where he served for 7 years as Head of the Information Group of the Madrid Command, a unit responsible for investigating hate crimes, especially those linked to violent extremism. After a period of two years in the Traffic Group of the Guardia Civil, in 2021 he was assigned to the Information Headquarters as Central Social Interlocutor, where he is dedicated to the promotion and supervision of the prevention and investigation tasks included in the Action Plan of the Guardia Civil against hate crimes, coordinating the actions carried out by the 54 REDO Teams distributed throughout Spain. Throughout his career, he has collaborated in multiple activities related to the training of judicial and prosecutorial authorities, Security Forces and other public and private entities, in the fight against hate crimes and hate speech, having participated in different forums, events and activities at national and international level.

<https://real-up.eu/>