



INACH

Bringing the Online In Line with Human Rights

**Compiled by
Cosima A. Hofacker &
Julie Heezius
2024**

**INACH's 22nd Anti-Cyber Hate
Summit -
Artificial Intelligence and the
Changing Legal Climate**

TABLE OF CONTENTS

International Network Against Cyber Hate - INACH	2
INTRODUCTION	3
CONFERENCE DAY – 27TH OF SEPTEMBER	4
Welcome Address	4
Keynote Speech - How Extremism Goes Viral by Dr. Julia Ebner	5
Panel Opening – Navigating Digital Harms: A Collaborative Approach by Agustina Callegari	6
Panel: Preventing and Addressing the Use of AI Technologies in Spreading Hate	7
Ronald and Suzette Award Ceremony by Magenta	9
INACH’s Cyber Hate Neutralization Hub by Tamás Berecz	10
Presentation - Impacts of the DSA	10
Breakout room Sessions	11
Room 1 - The DSA and Trusted Flaggers from a National DSC Perspective	11
Room 2 - Algorithmic transparency and the wider impact of AI on society	12
Room 3 - The industry’s response to online hate	12
Closing Remarks	13

International Network Against Cyber Hate – INACH

INACH was founded in 2002 to use intervention and other preventive strategies against cyber hate. The member organisations are united in a systematic fight against cyber hate, for example as complaints offices, monitoring offices or online help desks. In their respective countries, they provide important contacts for politicians, internet providers, educational institutions, and users.

Funding for INACH is provided by its members, the European Commission, the BPB and other donors. The International Network Against Cyber Hate (INACH) unites multiple organizations from the EU, Israel, Russia, South America, and the United States. While starting as a network of online complaints offices, INACH today pursues a multi-dimensional approach of educational and preventive strategies.

This publication has been produced with the financial support of the The Citizens, Equality, Rights and Values (CERV) Programme of the European Union. The contents of this publication are the sole responsibility of the International Network Against Cyber Hate (INACH) and can in no way be taken to reflect the views of the European Commission.



Supported by the Citizenship, Equality, Rights and Values (CERV) Programme of the European Union

*In cooperation with Subjective Values and Háttér Society. Sponsored by **Google**, TikTok, Meta and Microsoft*



Introduction

The INACH Annual Conference 2024, titled “Artificial Intelligence and the Changing Legal Climate,” was held on September 27th at the Rubin Wellness & Conference Hotel in Budapest, Hungary – hosted by the two Hungarian member organisations Háttér Society and the Subjective Values Foundation.

The event brought together over 75 participants, attending both in person and virtually, to discuss the challenges, dangers but also opportunities connected to Artificial Intelligence and Online Hate Speech, and the new regulatory climate to hate speech, disinformation and AI. The issues were viewed and discussed from different angles due to the diverse backgrounds of participants from academia, representatives of NGOs, civil societies, political institutions, other stakeholders and social media platforms, among others.

The conference featured a diverse agenda, including two keynote speakers, a panel, a presentation, three breakout rooms and two networking fairs to connect as many organisations, companies and individuals as possible. The Keynote speeches addressed the growing mainstream acceptance of extremist ideas and the role of AI in amplifying digital harms like misinformation and online abuse. The panel discussion centred on AI's impact on various forms of hate, including gender-based violence, racism, and its use in spreading false narratives. Presentations and breakout sessions provided deeper insights into the Digital Services Act (DSA), trusted flagger mechanisms, algorithmic transparency, and the industry's response to combating online hate.

The event emphasised the importance of collaborative global efforts in creating safer digital environments while navigating the complexities of AI and maintaining a balance between human rights and innovation. The Ronald & Suzette Award was presented for the fourth time, this year to two organisations: Fighting Online Antisemitism (FOA) and the NEVER AGAIN Association.

Conference Day – 27th of September

Welcome Address



The conference was opened by moderator, **Tamás Berecz** and **Dr. Steffen Eisentraut** of the INACH Board.

Dr. Eisentraut welcomed and thanked everyone on behalf of the INACH board and members for attending. He addressed the many crises and wars in our

society and the increase of the spread of hate speech and disinformation to new levels. Hate speech and disinformation are not new phenomena as INACH had to deal with masses of hateful content already in 2002 when it was founded. However, the dynamics and the manifestations have changed fundamentally. Artificial intelligence is a key driver in this development. Technology has made a huge leap in the last two years and is accessible to almost anyone, facilitating the spread of fake news and discrediting individuals or social groups.

Dr. Eisentraut emphasised the importance of having a diverse group of participants at the conference from industry, research, state actors and NGOs who may come with different perspectives but are united by the goal of finding solutions to these problems and enabling a better, peaceful coexistence in our society. He closed by thanking the Hungarian members, Háttér Society and the Subjective Values Foundation on behalf of the INACH board and thanking the sponsors of the conference: the European Union, Google and TikTok, Microsoft and Meta for their support and participation. He wished everyone a great conference with stimulating discussions, new insights, and sustainable networking in the spirit of the INACH slogan, to bring the online in line with human rights.

Keynote Speech - How Extremism Goes Viral by Dr. Julia Ebner



The first Keynote speech was given virtually by **Dr. Julia Ebner**, a Postdoctoral Researcher at the Calleva Centre for Evolution and Human Science at the University of Oxford (Magdalen College). She is the Leader of the Violent Extremism Lab at Oxford's Centre for the Study of

Social Cohesion and a Senior Research Fellow at the Institute for Strategic Dialogue.

Dr. Ebner shared her experiences infiltrating different extremist movements, including those centred on hate speech, misogyny, Muslim extremism, and other conspiracy-driven groups such as QAnon. Drawing on insights from her three books, she highlighted the importance of the socialisation process for the recruitment and radicalization of people and therefore becoming familiar not merely with the ideology and ideological indoctrination of an extremist group but with their language, subculture references and jokes, and thus becoming part of an exclusive counterculture. The start is often the creation of a small subculture that begins to cultivate networks on alternative media platforms to amplify its presence, aiming to provoke backlashes from political opponents and then mobilise the masses, who then have to choose sides in a hyper-polar conflict. The entry gates, to how average people get absorbed into extremist ideologies or groups, are manifold but mostly through antifeminism, racism, LGBTQ+ rights and topics around science, such as climate change, or the Corona pandemic.

Dr. Ebner also emphasized how extremist groups use the three strategies of *gamification*, meaning to transform radicalization into something entertaining to engage with and to recruit younger audiences, *global networking and international collaboration*, and the building of an *alternative information ecosystem*, with equivalents to platforms such as YouTube or Facebook, to spread their ideas and come together.

A central point of her speech was the concept of *Identity Fusion*, where individuals blend their personal identity with that of a group identity, motivating them to take extreme actions to defend and promote the group's cause. Social Media can accelerate this as this type of identity fusion is usually created by shared negative and transformative experiences and dramatic images which are often shared online. Besides AI, social media, and other new technologies changing how extremist movements can go viral, mobilise or spread their ideas and conspiracy myths, she emphasised the yet overlooked phenomena of blockchain-based technology and decentralized Autonomous Organizations (DAOs). These can play a role in potentially giving rise to a new form of decentralized mobilization that relies entirely on crypto-based transactions and on self-governance spaces that are not regulated by any laws and as it sits outside of any traditional institutions, making it very interesting to explore in the future.

Panel Opening – Navigating Digital Harms: A Collaborative Approach by Agustina Callegari

The second speech – the panel opening – was given virtually by **Agustina Callegari** who currently leads the Digital Safety Initiative at the World Economic Forum. In her speech, she focused on the growing complexity and prevalence of digital harms resulting from advancing technologies and the work of the World Economic Forum's Global Coalition for Digital Safety.



The second speech – the panel opening – was given virtually by **Agustina Callegari** who currently leads the Digital Safety Initiative at the World Economic Forum. In her speech, she focused on the growing complexity and prevalence of digital harms resulting from advancing technologies and the work of the World Economic Forum's Global Coalition for Digital Safety.

Ms Callegari addressed the evolving digital safety landscape and showcased the current trends in online harms. These include child sexual abuse material and sextortion, hate speech, human exploitation and online sex trafficking, the spread of mis-, and disinformation in connection with deepfakes – particularly related to elections – as well as online terrorism that makes use of decentralised networks, cryptocurrency and social

media. She pointed out the challenges posed by AI-enhanced technologies, which can be created and used to spread disinformation and intensify extremist ideologies and highlighted how AI has the *dual role* and capacity to amplify both positive uses, such as assisting with complex challenges and negative uses of digital platforms and enabling violent ideologies. However, she underlined that it is important to also recognise generative AI as part of solutions to these challenges as there is a growing body of initiatives, tools and technologies that tackle harmful content online effectively with the help of AI. Another challenge she addressed is the lack of shared understanding and fragmented definitions for terms such as cyberbullying, doxing, hate speech and explicit content, causing an inability to comprehensively address the issue. The coalition has created a *typology of online harms* designed to serve as a comprehensive framework for understanding the many varieties of online harm and developing a foundational language. **Ms Callegari** emphasised the need for global collaboration among tech companies, governments, international organizations, and civil society to effectively tackle these challenges and foster safer online environments while upholding human rights and innovation.

Panel: Preventing and Addressing the Use of AI Technologies in Spreading Hate



The third programme point of the conference was the panel on the title “Preventing and Addressing the Use of AI Technologies in Spreading Hate”. It featured **Omny Miranda Martone**, the Founder and CEO of the Sexual Violence Prevention Association (SVPA), **Elena Haig**,

a Legal and Policy officer at the European Commission, **Sigurdur Ragnarsson**, the CEO of Videntifier and **Marcus Scheiber** from the research project Decoding Antisemitism, and was moderated by **Dr. Selma Muhič Dizdarevič** of INACH.

The panel discussed the impact and contribution of AI innovations to gender-based violence, anti-Semitism, racism and other forms of systemic oppression, the limits of AI and its role in the spread of online hate, as well as the issue of a race between AI developments and detection tools and strategies to mitigate them, including the role of governments, laws, policies and courts in addressing the harms of AI.

Mr Ragnarsson emphasized that the rise of deepfake technology has blurred the line between what is real and what is fake, making it difficult to identify genuine content. This trend poses a significant challenge, especially as the quality of deepfakes improves and becomes almost undetectable. He shared how his company has been working on AI-based tools to counter this threat by improving detection and enforcement against AI-generated fake content.

Mr Scheiber focused on the danger of antisemitic content being spread online via AI-generated media and text-based algorithms. He elaborated on how AI technologies can be programmed to spread hate and manipulate public perception. He underscored the need for AI-driven solutions that detect hate speech and counter the automated spread of antisemitism.

Miranda Martone highlighted the rapid increase in digital violence, especially the rise of non-consensual pornography and realistic AI-generated images that depict exploitation. They argued that AI, while useful, is also being used in predatory ways, especially in cyber harassment and abuse of women. They called for social media platforms to take preventive measures by integrating better screening tools that detect and prevent digital sexual violence before it spreads.

Ms Haig discussed the real-world impacts of online hate, stressing that the borderless nature of the internet allows for hate speech and extremism to spread unchecked across countries. Elena spoke about EU efforts, such as the Code of Conduct on Hate Speech, which seeks to establish guidelines for social media platforms to regulate harmful content, highlighting the challenges in enforcing these regulations across different jurisdictions.

Eventually, The panellists agreed that addressing the growing challenge of AI-enhanced hate speech will require collaborative efforts between governments, private companies, and civil society. They acknowledged that while AI can be a powerful tool to tackle hate speech, it is equally essential to constantly evolve these tools to stay ahead of those who exploit the technology for harmful purposes.

Ronald and Suzette Award Ceremony by Magenta

The Ronald and Suzette Award by Magenta honoured two of INACH’s founders, Ronald Eissens and Suzette Bronkhorst. The award was established to recognize individuals and organizations who have made significant contributions to combating online hate speech and advancing civil rights in the digital space.

The recipients included **Fighting Online Antisemitism (FOA)**, an organization founded in 2022 to combat antisemitism with the help of over 3,000 volunteers, monitoring 60,000 social media posts. Another recipient was the **“Never Again” Association**, an anti-Nazi and anti-racism group that successfully pushed for the conviction of several Neo-Nazis and reduced the sale of Nazi paraphernalia. Both organizations have made a significant impact in removing hate speech from the internet. **FOA** continues to expand awareness through educational workshops, while **“Never Again”** has remained dedicated to advocacy and research even under politically challenging circumstances.



Anna Tatar from the “Never Again” Association Nir Kaplan from Fighting Online Antisemitism

INACH's Cyber Hate Neutralization Hub by Tamás Berecz



Tamás Berecz (INACH) and **Laurent Janse** from the Dutch member organisation and reporting centre for online discrimination **Meld.Online Discriminatie (MOD)** introduced the collaborative *Cyber Hate Neutralization Hub* in further cooperation with Textgain, who has developed the algorithm behind it. Two videos describing the Hub were shown, explaining the efforts to report online insults and threats and its importance in combating online hate was emphasised. It was focused on how bots are used to divide society and influence democratic processes. **Mr Janse** thanked INACH for the collaboration and indicated further plans to shift their focus from online group discrimination to also reporting online insults and threats, with more information about it in the future.

Presentation - Impacts of the DSA



The Presentation on the SafeNet project was held by **Julie Heezius** a dedicated Project and Advocacy Officer who specialises in monitoring and reporting online hate speech, utilising legal expertise to support human rights advocacy. It focused on the SafeNet project which seeks to apply a comprehensive and intersectional

approach to prevention and fight against intolerance, racism and xenophobia. **Ms Heezius**, the monitoring officer for INACH on the Safenet project, presented an overview of this EU-funded initiative aimed at combating intolerance, racism, and xenophobia online. The project spans 24 months, with 21 partners from 19 EU countries, many of whom are members of INACH. Safenet's key objectives include monitoring illegal hate

speech, raising awareness, and sharing progress through bimonthly factsheets from each partner country.

Ms Heezius highlighted the importance of the Digital Services Act (DSA), which mandates faster platform responses, transparency in content moderation, and regular audits of social media companies. She shared data on the removal of hate speech, response times within 24 hours, and cases with no feedback, noting the DSA's final implementation halfway through the project. Though **Ms Heezius** represented the project as a whole, she emphasized that different data applies to different countries, reflecting diverse regions when monitoring hate. She concluded by stressing the need for ongoing collaboration to ensure social media platforms' accountability in combating online hate speech.

Breakout room Sessions

Room 1 - The DSA and Trusted Flaggers from a National DSC Perspective



Breakout room 1 was held by **Andrea Cox** from the Slovak Media Services Council and Director of DigiQ; and **Dana Paraschiv**, the Director of Regulatory Operations at Coimisiún na Meán in Ireland, and was moderated by **Natalia**

Babicova from DigiQ.

The session focused on the crucial role of trusted flaggers under the Digital Services Act (DSA) and created a space for discussion on obstacles and challenges potential trusted flaggers face. These entities are essential for identifying and reporting illegal online content, acting as a bridge between platforms and regulators to ensure swift and accurate content moderation. The breakout room highlighted once more the importance of their work in upholding digital safety and the challenges encountered in this complex landscape. The role of Digital Service Coordinators (DSCs) in implementing the DSA, with a focus on the approach to the assessment of trusted flagger applications and awarding

the trusted flagger status was also touched upon and an overview of the guidance and application form adopted by the Irish DSC and the challenges they encountered to date was provided.

Room 2 - Algorithmic transparency and the



**wider
impact of AI
on society**



This breakout room was presented by **Jens Linge** from the Joint Research Centre, **Gijs van Beek** from Textgain, and **Jordy Nijenhuis** from Dare to be Grey and was moderated by **Kelly Grossthal** from the Estonian Human Rights Centre. It was debated how AI is used to analyse disinformation narratives, particularly those targeting minority groups and global conflicts like Ukraine/Russia, climate change, and migration and insights showcasing AI's role in dissecting these narratives were provided. Finally, it was explored how AI and algorithms are quietly but powerfully influencing everybody's lives, the challenges posed by the lack of transparency of these systems, particularly concerning the spread of disinformation and amplification of hate speech and how these systems can be gamed to generate hate and manipulative content, emphasising the need for transparency in algorithmic decision-making.

Room 3 - The industry's response to online hate



Breakout room 3 was held by **Amelia Connor-Afflick** from Meta, **Anna Zizola** from X (formally Twitter), and **Fergal Browne** from TikTok, and moderated by **Daniel Heller** from CEJI.

The session looked at how social media platforms are addressing online hate, with a particular emphasis on antisemitism and anti-

Muslim hate. Furthermore, it also examined the role of the Digital Services Act (DSA) and artificial intelligence in moderating online content. The speakers shared their professional insights and experiences and aimed to offer an understanding of effective responses and policy frameworks. The breakout room was based significantly on a comprehensive, very interactive and dynamic Q&A session including questions that had been collected by the participants in advance.

Closing Remarks

Tamás Berecz thanked everyone for attending the conference and announced that next year's summit will be held in Sofia, Bulgaria.

All photographs included in the report were taken by © Claudia Kamergorodski.