INACH

Bringing the Online In Line with Human Rights

Dublin, Ireland

Technology Roundtable Report, 2026

HOPE & COURAGE COLLECTIVE

TikTok

Co-funded by the European Union

INACH Technology Roundtable 2026

On 3 June 2026, the International Network Against Cyber Hate (INACH), its member organisations, researchers, and representatives from civil society gathered at the TikTok offices in Dublin, Ireland, for the Technology Roundtable. The event was organised by INACH in collaboration with Hope and Courage Collective and TikTok, and aimed to create a dedicated space for dialogue between civil society organisations, policymakers, technology providers, and researchers.

The Roundtable focused on the growing role of AI-driven tools in strengthening the detection, analysis, and mitigation of online hate speech across digital platforms. Discussions explored how artificial intelligence can support civil society monitoring efforts while also examining broader issues at the intersection of cybersecurity, platform integrity, coordinated manipulation, bot-driven amplification, and hybrid threat environments.

Participants also addressed the implications of the EU Artificial Intelligence Act (AI Act) for civil society organisations. Particular attention was given to opportunities and risks associated with high-risk AI systems, transparency obligations, access to AI infrastructures, and the practical implications of emerging regulatory frameworks. The event further examined current limitations in platform data access and explored both policy and technical solutions to improve meaningful data-sharing mechanisms in line with the Digital Services Act (DSA).

A key objective of the Roundtable was to strengthen collaboration between stakeholders and contribute to enhanced compliance, accountability, and systemic risk mitigation across the digital ecosystem.

Panel Discussion on Governing AI Systems: Risks, Rights, and Responsibility

Following opening remarks from INACH, Hope and Courage Collective, and TikTok, participants moved directly into the first panel discussion, moderated by Emrah Gursel, Head of Fundraising at INACH.

The panel, titled *"Governing AI Systems: Risks, Rights, and Responsibility,"* examined the governance of AI systems through the lens of hate speech, digital rights, platform accountability, and democratic resilience.

The discussion featured the following speakers:

- **Carlos Saura**, Legal Officer at the European Commission and Team Leader on Regulatory Consistency within the DSA/DMA Coordination and Regulatory Compliance Unit (DG CNECT). Drawing on his experience negotiating key European digital legislation, including the AI Act, eIDAS2, and the Interoperable Europe Act, Carlos provided insights into the evolving European regulatory landscape and the importance of regulatory coherence across digital policy frameworks.

- **Jasmina Ploštajner**, co-founder and long-time collaborator of the Slovenian NGO *Danes je nov dan*. As a digital rights expert, Jasmina discussed the impact of artificial intelligence on human rights, democratic processes, and fundamental freedoms, highlighting the need for rights-based approaches to technology governance.

- **Lydia El-Khouri**, Project Manager at Textgain, who specialises in hate speech monitoring and AI applications. Lydia shared her extensive experience monitoring antisemitism and other forms of online hate across Europe and outlined how AI technologies can strengthen civil society monitoring efforts. She also presented perspectives from the European Observatory of Online Hate, an EU-funded initiative that supports civil society organisations and law enforcement agencies through AI-enabled monitoring tools.

- **Ellen Mok**, founder of Digitale Doetank and board member of Algorithm Audit. With a background spanning cyber strategy, national security, and technology policy, Ellen explores issues of digital sovereignty, responsible AI governance, and the relationship between emerging technologies, public accountability, and democratic resilience.

The panel highlighted both the opportunities and challenges presented by AI systems, emphasising the need for transparency, accountability, and meaningful stakeholder engagement in the governance of digital technologies.

Presentation

Regulatory Challenges: The DSA, the AI Act, and the Role of Civil Society

A central discussion point focused on the evolving European regulatory landscape and the relationship between the Digital Services Act (DSA) and the Artificial Intelligence Act (AI Act). Participants reflected on the significant progress made through recent digital legislation while also acknowledging the challenges of ensuring that regulatory frameworks remain effective in the face of rapidly evolving technologies.

The discussion highlighted the importance of consistency across legislative instruments. Participants noted that while the DSA provides important mechanisms to address online harms and platform accountability, emerging AI-enabled systems raise new questions that were not fully anticipated during the development of existing legislation. This has generated ongoing discussions about how AI governance can be effectively integrated with broader platform regulation and systemic risk mitigation efforts.

Questions were also raised regarding current efforts to simplify European digital legislation, including discussions surrounding the Omnibus process. Participants stressed the importance of ensuring that simplification initiatives do not undermine hard-won protections relating to transparency, accountability, and fundamental rights.

Jasmina Ploštajner reflected on civil society engagement during the negotiations of the AI Act. While acknowledging the significance of establishing the world's first comprehensive AI regulatory framework, she expressed disappointment that the final legislation moved away from a stronger human rights-based approach and instead adopted a more product-oriented regulatory model. This shift, participants noted, has implications for how harms and societal impacts are assessed within the regulatory framework.

The panel also discussed transparency obligations under the AI Act, including requirements for users to be informed when interacting with AI systems and provisions relating to synthetic or AI-generated content such as deepfakes. However, concerns were raised about implementation timelines and the phased application of obligations. Participants noted that delays in the enforcement of certain provisions, particularly those relating to high-risk systems, may create incentives for companies to accelerate deployment before regulatory requirements come fully into force.

Another recurring theme concerned the role of civil society organisations in shaping and evaluating digital regulation. Participants expressed concern that opportunities for meaningful civil society participation during legislative and implementation processes remain limited. The discussion emphasised that civil society organisations play an essential role not only in identifying harms and monitoring compliance, but also in advocating for stronger protections and educating the public about their digital rights.

The discussion concluded with a recognition that effective digital governance requires ongoing dialogue between regulators, platforms, researchers, and civil society actors. Participants stressed that ensuring accountability, transparency, and protection of fundamental rights will depend not only on legislative frameworks themselves, but also on the extent to which diverse stakeholders are able to contribute to their implementation and oversight.

Presentation

Can AI Fix the Business Model Behind Online Harm?

In a presentation titled "Can AI Fix the Business Model Behind Online Harm?", Dr. Lex Zard explored the relationship between platform business models, digital advertising, and the emergence of online harms. Drawing on work related to the Digital Omnibus, the Digital Fairness Act, and the enforcement of the Digital Services Act (DSA), Online Safety Act (OSA), and Digital Markets Act (DMA), the presentation examined whether technological solutions alone can address problems that are rooted in platform incentives.

A key argument presented was that the common understanding of platforms as simple "user-to-user" services is increasingly outdated. While platforms continue to facilitate interactions between users, their operations are shaped by a much broader ecosystem involving content creators, advertisers, recommendation systems, advertising technologies, and large-scale data collection practices.

Participants were encouraged to consider platforms not merely as communication tools but as complex commercial systems in which user attention and behavioural data function as valuable economic resources. Within this model, content creators generate engagement, advertisers provide revenue, and platforms use recommendation and optimisation algorithms to maximise attention, interaction, and advertising performance.

The presentation highlighted the growing role of artificial intelligence throughout the digital advertising ecosystem. AI systems are increasingly being deployed for media planning, media buying, creative content generation, advertising optimisation, and emerging forms of

agentic commerce. These technologies are becoming central to how advertising campaigns are designed, targeted, and delivered.

Particular attention was given to the increasing automation of advertising processes. AI-powered advertising products, such as Google's Performance Max, were discussed as examples of systems that make decisions about audience targeting and ad delivery with limited transparency. As AI assumes greater responsibility for determining how advertisements are distributed, concerns arise regarding accountability, oversight, and the ability of regulators, advertisers, and users to understand how decisions are being made.

The discussion raised broader questions about whether AI can meaningfully reduce online harms if the underlying business incentives that reward engagement and data extraction remain unchanged. Participants reflected on whether existing legal frameworks are sufficient to address these challenges and considered what additional measures may be necessary to align platform incentives with public interest objectives.

The presentation concluded by encouraging participants to look beyond individual technologies and consider the structural drivers of online harms. While AI may offer tools for detection, moderation, and risk management, meaningful change may also require addressing the economic and advertising models that shape how digital platforms operate.

Presentation

MaxPlain & AI-Supported Hate Speech Detection: Opportunities and Limitations

Marten Risius and Nils Riekers from Neu-Ulm University of Applied Sciences presented their research on AI-supported hate speech detection, focusing on the challenges associated with identifying harmful content in increasingly complex digital environments. Their work explores multimodal hate speech detection, explainable AI, and the responsible preparation of AI training data.

The discussion challenged the assumption that artificial intelligence alone can solve the problem of online hate speech. While AI systems have demonstrated significant improvements in identifying harmful content at scale, participants acknowledged that many forms of hate speech remain highly contextual, ambiguous, and increasingly multimodal, combining text, images, video, symbols, and cultural references that may be difficult for automated systems to interpret accurately.

As part of the session, the researchers introduced MaxPlain, a research prototype designed to move beyond traditional statistical classification models. Rather than simply categorising content as hateful or non-hateful, the system seeks to provide argument-based and explainable assessments of potentially harmful content. This approach aims to make AI-generated decisions more transparent and understandable for human users.

A key discussion point concerned the practical value of being able to explain operational settings. Participants explored whether providing moderators and analysts with clear reasoning behind AI-generated assessments could improve decision-making processes, particularly in environments where content reviewers are required to make rapid judgments

under significant time pressure. The discussion highlighted that explainability may increase trust in automated systems, support consistency in moderation decisions, and assist reviewers in understanding why particular content has been flagged.

At the same time, participants recognised that explainability alone does not eliminate the challenges associated with hate speech detection. AI systems continue to face limitations when interpreting nuanced language, sarcasm, coded expressions, and evolving forms of online communication. The discussion therefore reinforced the importance of maintaining human oversight and expertise alongside automated detection tools.

Presentation

The Cyberhate Neutralization Hub: Understanding Bot-Driven Amplification and Online Hate

Participants were introduced to the Cyberhate Neutralization Hub, a collaborative initiative between INACH and Textgain operating within the framework of the European Digital Media Observatory (EDMO) network. The project focuses on identifying and analysing bot-driven activity on TikTok and examining how coordinated amplification influences online discourse and the spread of harmful narratives.

The project was established in response to growing concerns about the role of automated accounts in shaping public debate. While much attention is often directed toward individual pieces of illegal or highly toxic content, the research highlighted that online harms can also emerge through the sustained amplification of lower-toxicity content that gradually normalises exclusionary narratives, discriminatory language, and extremist viewpoints.

Researchers demonstrated how automated and coordinated networks contribute to the spread of narratives relating to immigration, religion, nationalism, and conspiracy theories. In the Dutch context, the project identified recurring themes including antisemitic, anti-Muslim, anti-immigrant, and xenophobic rhetoric, as well as narratives associated with concepts such as the "Great Replacement" theory and slogans such as "Eigen Volk Eerst" ("Our Own People First").

A key finding presented during the session was that harmful narratives are not always disseminated through overt hate speech. Instead, coordinated activity frequently promotes language and concepts that appear less toxic on the surface but can contribute to the gradual normalisation of exclusionary or discriminatory ideas. Participants discussed examples such as the increasing use of terms like "remigration," which may circulate in mainstream discussions while simultaneously serving as entry points for more extreme narratives.

The project's analysis focused on TikTok discussions surrounding the 2026 Dutch local elections. Researchers collected data using election-related keywords, public figures, and political topics across multiple municipalities, while also developing control datasets based on general TikTok activity. This enabled comparisons between political and non-political discussions and provided insights into the role of automated behavior within election-related discourse.

The methodology combined several behavioral indicators associated with bot activity, including posting frequency, timing patterns, responsiveness, and coordination between accounts. Researchers explained that higher scores reflected increasingly bot-like behaviour and that network analysis techniques were used to identify coordinated amplification patterns.

One of the most significant findings presented was that approximately one-third of the observed online conversation appeared to be driven by accounts exhibiting bot-like characteristics. Network visualisations demonstrated clusters of coordinated accounts that amplified content relating to political figures and controversial social issues, suggesting organised efforts to increase the visibility of particular narratives.

The discussion emphasised that coordinated amplification should be understood as a systemic risk rather than solely a content moderation issue. Participants noted that bot-driven campaigns can distort public discourse even when individual posts do not violate platform policies or legal thresholds. By artificially increasing the visibility and perceived popularity of certain viewpoints, coordinated networks can influence public perceptions of what constitutes mainstream opinion.