DELIVERABLE 8 – Focus groups report

Project Title:	**Right-wing extremist ecosystems driving hate speech: dissemination and recruitment strategies (RECO-DAR)**
Project Number:	**101084729**
Call/Topic:	**CERV-2022-EQUAL**
Author(s):	**Elena Jung (modus|zad), Erik Hacker (SCENOR)**
Date:	**16 December 2024**

Table of Contents

Section 1: Abstract

A crucial part of the RECO_DAR project is to involve key stakeholders and incorporate their expertise into project outputs. To do so, two focus groups were implemented to validate the research findings and generate recommendations on their basis. The following report elaborates on the objectives of these focus groups, their planning, implementation, and results.

17 experts from various disciplines participated in the two focus groups, covering the fields of social sciences, legal research, and youth work. In structured breakout sessions, participants discussed the key issues identified in the course of the RECO_DAR projects and developed tailored recommendations for countering them. The questions raised were addressed to specific groups of attendants in the focus groups (practitioners, social science or legal researchers), to ensure in-depth reflection and relevant feedback. This report outlines all collected recommendations for each of the two focus groups. The collected recommendations are actionable, feasible, and realistic, addressed specifically to platforms, and/or policymakers, and/or frontline practitioners.

The focus groups fulfilled their purpose and generated numerous solutions to pressing issues related to far-right extremist hate speech and recruitment online. The two focus groups made a significant contribution to the RECO_DAR project by delivering tailored recommendations for specific target groups, informed by the multidisciplinary expertise of participants. These outputs will be integrated into a forthcoming policy brief.

Participants highlighted the relevance and timeliness of the findings, particularly in the context of recent far-right electoral successes across Europe. Insights into ecosystem linkages were considered crucial for understanding the factors behind this trend. Additionally, the project's methodology was recognised as adaptable to other ideological and geographical contexts.

Section 2: Basic Project Data

Project Number	101084729
Project Title	Right-wing extremist ecosystems driving hate speech: dissemination and recruitment strategies (RECO-DAR)
Consortium Partners	• SCENOR – Verein zur Erforschung aktueller gesellschaftlicher Herausforderungen • Modus – Zentrum für angewandte Deradikalisierungsforschung (modus|zad)
Duration	27 months (1 January 2023 – 31 March 2025)
Budget	€ 269,310.65
Project Coordinator	Dr Daniela Pisoiu (SCENOR)

Section 3: Technical Report

1. Introduction

The project's workplan foresaw the implementation of two focus groups with the purpose of (1) feeding into the final recommendations for action to policymakers, researchers, and practitioners; (2) evaluating the project outcomes' relevance, and (3) discussing the potential for transferring the methodology and insights to other contexts. These focus groups were scheduled to take place after finalising all research outputs but before producing the recommendations based on the findings, i.e. in December 2024.

2. Preparatory activities

The recruitment for focus groups started with the compilation of a list of researchers and practitioners that have experience and expertise on online far-right extremism and related matters. This list also included individuals that had already contributed to the RECO_DAR project by participating in expert interviews earlier. Once the list was completed, these individuals were contacted using a standardised invitation, that was then slightly tailored to each invitee. For the first focus group organised by modus|zad, 17 experts were contacted. Of these, 12 confirmed their intention to participate, and 5 declined to participate or did not reply. For the second focus group organised by SCENOR, 17 experts were contacted. Of these, 11 confirmed their intention to participate, 3 declined to participate, 1 declined to participate but referred a colleague as an alternative who agreed to participate, and 2 did not reply. A TikTok representative was also contacted, but after expressing initial interest, communication abruptly stopped.

Simultaneously to the invitations, an executive summary of the key project findings was drafted (see annex 3), elaborating on the most important issues found that should be addressed by recommendations. This 4-page summary also included plots and screenshots.

An agenda and script were also drafted for the focus groups. The focus groups were planned to last 90 minutes, of which 30 minutes were foreseen for breakout sessions, where experts with similar backgrounds can discuss the questions in smaller groups. For these sessions, the experts were assigned to three groups based on their expertise: social science researchers, legal researchers, and practitioners. A key element of the script were the criteria for recommendations (actionable, specific, feasible, scalable; see Annex) and nine specific questions (see chapter 3 below), with three questions tailored to each expert group. These questions were designed to help participants develop specific, feasible, and actionable recommendations to a particular problem identified in RECO_DAR research reports.

To facilitate the work in breakout sessions, a Miro board was prepared. Miro is an openly available online service that allows users to interact in real time on a graphic surface. This board (see Annex) included the agenda for the focus groups, the key findings of the project, and the specific questions for

each expert group, including the criteria for recommendations. The board was shared with participants a few days before each focus group as an additional element in helping experts prepare for the session.

The focus groups were designed to take place under the Chatham house rules, with no recording and no media presence. Participants were nonetheless offered to be named in public reports of the project in the context of their output in the focus groups, yet no participant made use of this offer.

3. First focus group on 3 December 2024

The first focus group led by modus|zad took place on 3 December 2024, between 14.00 and 15.45 hrs online on Microsoft Teams. After two cancellations on short notice due to sickness and one unexpected no-show, nine experts participated, alongside four project staff. The session began with an introduction of the team and self-introductions of the participants, followed by a summary of the RECO_DAR project, its objectives, and key findings. These findings were discussed with participants, including the methodology used. Participants' questions about the results mainly concerned the definitions used, e.g. how the project defines implicit hate speech and pro-Russian propaganda, but also some of the key findings on the evolution of hate speech across platforms.

Participants were also asked to share their impressions about the results, highlighting aspects that were surprising and innovative. Most answers emphasised the linking dynamics across platforms as an unexpected yet fascinating finding, as well as the concentration of young users in one 'community' in the TikTok sample.

Participants were then assigned to their respective breakout sessions, with five people in the social science researchers' group, two people in the practitioners' group, and two people in the legal researchers' group. Using the previously prepared Miro board (see Annex), experts discussed the project's findings for 35 minutes and developed recommendations, answering three tailor-made questions in each group. After this discussion, participants presented their recommendations in the plenum and discussed benefits, risks, and challenges associated with them.

The session concluded with summing up the results of the focus group and informing the experts about upcoming project publications, webinars, and the final conference. The following sections elaborate on the questions asked and the corresponding recommendations of each group.

3.1. Social science researchers

Question 1: The TikTok community guidelines already prohibit implicit hate against protected groups, yet the enforcement seems to have gaps. How could platforms improve the detection and removal of implicit hate/borderline content? (Target group of recommendations: platforms)

> *Recommendations*: There are already lots of technological solutions, but the public (or even researchers) does not know what these solutions are capable of, and what exactly platforms are doing to moderate content. There needs to be more transparency, so it can be evaluated what is happening. Without that, no recommendation can be given.

Question 2: Some far-right figures' accounts get frequently removed, yet they maintain a sustained presence on TikTok. How can platforms detect and react to account sustainment strategies (back-up accounts, content mirroring, cross-platform ecosystem) quicker or even proactively? (Target group of recommendations: platforms)

> *Recommendations*: This issue is more specific to certain platforms (Instagram, TikTok). Some platforms already have solutions for this, e.g. YouTube, thus there are strategies already established to counter this problem. The solution is thus to share these technologies, strategies and knowledge within the industry.

Question 3: Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. How can policymakers, and/or platforms, and/or practitioners limit the impact of disinformation by linkages? (Target group of recommendations: policymakers, platforms)

> *Recommendations*: Links to different platforms are already hard to moderate, and it is unclear if such moderation is even desirable. Large social media platforms already have a problematic monopoly, and by moderating links, this issue would only get worse as it would create even more centralisation. Also, the issue is not that platforms cannot do anything about this, but that they do not have enough incentives. The DSA could change this.

3.2.   Legal researchers

Question 1: Current domestic and EU regulations appear to have gaps regarding borderline content and implicit hate speech, that in turn also lead to gaps in content moderation. How could policies be adapted to counter implicit hate while maintaining freedom of expression? (Target group of recommendations: policymakers)

> *Recommendations*: There are lots of criminal law regulations already on (hate) speech both on the domestic and EU levels (e.g. DSA systemic risk assessment, coordinators). It would be challenging to create new laws without creating higher risks for freedom of speech. Thus, it wouldn't make sense to create new regulations and obligations for platforms and service providers. It would instead make more sense to have more specific policies on how to enforce existing regulations.

Question 2: While TikTok's community guidelines prohibit implicit hate against protected groups, this appears to be insufficiently enforced or too vaguely defined. How could platforms' community guidelines be adapted to counter implicit hate and borderline content? (Target group of recommendations: platforms)

> *Recommendations*: Do not focus on what kind of speech is not desired, as trying to delete every problematic account and post is not possible. Instead, focus on what kind of speech is desired on social media platforms, and make it more visible by amplifying its engagement. It is commonly known that the platform design and algorithm feed into distorted civic discourse which favours emotional, conflict-loaded (in-group vs. out-group) content that provokes outrage creates a lot of engagement on digital media. The key to resolve this issue is to tackle the algorithms responsible for this and level the playing field by changing the logic of content recommendations towards increasing the visibility of "positive" content. This simultaneously

would limit the visibility of borderline content. However, this raises lots of ethical questions in this context, e.g. who gets to define democratic discourse?

Question 3: Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. How can policymakers and/or platforms limit the impact of disinformation by linkages while maintaining freedom of expression? (Target group of recommendations: policymakers, platforms)

> *Recommendations*: This is already handled by the EU's Digital Services Act (DSA). The question is rather how to enforce it.

3.3. Practitioners

Question 1: Identity politics-based framing polarizes societies and can incite hatred yet is hard to counter legally. What (counter- and alternative) narratives can limit the impact of polarizing identity politics-based frames claiming existential threats by scapegoats (e.g. migrants, elites)? (Target group of recommendations: practitioners)

> *Recommendations*: Change the norms of the social media landscape by not only educating people (especially young people) with 'edutainment' (educational entertainment), but giving them fact-based posts, which is a fundamental basis for building opinions. Do not only present opinions, but present facts. Educational narratives should impact the psychological literacy of the young audience. It is also important to allow those affected to have their say and share their experiences with discrimination, or coming to terms with stereotypes.

Question 2: The ecosystem approach not only helps the far-right to sustain presence by linking to back-up accounts but can also be used for campaigns to boost selected content and for luring users to ecosystems of 'alternative news'. What can platforms do to limit the impact of the cross-platform ecosystem approach for boosting content, luring users to alternative ecosystems and rebuilding back-up accounts? (Target group of recommendations: platforms)

> *Recommendations*: AI could help to identify, scan, locate, and check cross-platform content as well as users that share content. Platforms should also lower the hurdle for blocking accounts after they had already been banned on another platform, although this requires consultation between the service providers.

Question 3: Implicit hate and identity politics-based narratives are less obvious and more palatable to individuals. Besides counter- and alternative narratives, how to make vulnerable individuals and/or societies more resilient against identity politics-based frames and implicit hate? (Target group of recommendations: policymakers, practitioners)

> *Recommendations*: Platforms should share data on people who are sanctioned by moderation, so that other platforms know who to look for. AI could also be used to detect 'shitstorms' that are temporarily targeting one person, so that this person can automatically receive links to services that help in such scenarios. It is also recommended to try to embed psychosocial education into counter-narratives and awareness raising. Practitioners should also embed self-awareness into the messaging that interests this target audience, and use the same strategies used by violent extremists, e.g. boosting certain content.

4. Second Focus group on 10 December 2024

The second focus group led by SCENOR took place on 10 December 2024, also between 14.00 and 15.45 hrs online on Microsoft Teams. Eight experts attended the session, after one participant cancelled due to sickness, two others cancelled due to unexpected emergencies on the day of the focus group, and one expert did not show up despite confirming their participation earlier. The session began with an introduction of the project team, an elaboration on the focus group setting, and an overview of the RECO_DAR project's context, objectives, methodology and main results. A flowchart summing up the key project findings was added to the Miro board for this session to provide participants with a quick overview in addition to the summary sent earlier. This segment was followed by a Q&A with participants. The questions were related to content moderation practices on TikTok (e.g. what methods are used to detect harmful content, what share of the reported content gets removed), the group of 'Reichsbürger' (sovereignists) in the project's analysis, and the presumable age of the TikTok users in the sample. Participants also talked about their impressions of the results. They emphasised their surprise that the linkage between TikTok and Telegram showed a ratio of 3:1. These reactions are aligned with the reactions of the first focus groups' participants.

After explaining the guidelines for recommendations, including an example, the experts were assigned into three breakout sessions: three participants in the group of social science researchers, three in the group of legal researchers, and two in the group of practitioners. These groups drafted recommendations for the specific issues and questions defined on the Miro board (see Annex). After 40 minutes of discussion in the breakout rooms, participants returned to the plenum and elaborated on their recommendations and the specific downsides of these.

The session concluded with the project team thanking the participants, summarising the recommendations, and informing about upcoming project events and publications. The following sections elaborate on the questions asked and the corresponding recommendations of each group.

4.1. Social science researchers

Question 1: The TikTok community guidelines already prohibit implicit hate against protected groups, yet the enforcement seems to have gaps. How could platforms improve the detection and removal of implicit hate/borderline content? (Target group of recommendations: platforms)

Recommendations:

- Experts should work together with platforms to find and flag borderline content. They however require API access for this.
    - Downside: borderline content can be useful for platform's business model, as it is provocative and generates traffic. Resources are also likely to be insufficient to find and flag everything.
- To streamline this process, train AI-models on phenomena such as dog whistles and borderline content.
    - Downside: dog whistles are constantly adapting, so the dataset needs to be continuously reviewed.

Question 2: Some far-right figures' accounts get frequently removed, yet they maintain a sustained presence on TikTok. How can platforms detect and react to account sustainment strategies (back-up accounts, content mirroring, cross-platform ecosystem) quicker or even proactively? (Target group of recommendations: platforms)

> *Recommendations*:
>
> - Implement network theory-based detection models to find mirror accounts quicker, e.g. via the pictures and texts they use.
>   - Downside: users could be misclassified; false negatives.
> - Analyse the metadata of accounts to identify the origins of trends, e.g. whether they come from a specific foreign country.
>   - Downside: easy to bypass using fake IDs and IPs (e.g. with a VPN), especially by state actors.
> - Identify and sanction/prosecute "professional haters" (i.e. viral accounts, influencers) in order to limit the extent of dissemination of hate. Platforms can share the data of accounts that belong to the same person with authorities.
>   - Downside: costly in terms of human resources.

Question 3: Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. How can policymakers, and/or platforms, and/or practitioners limit the impact of disinformation by linkages? (Target group of recommendations: policymakers, platforms)

> *Recommendations*: Policymakers should work together on the European/supranational level to sanction TikTok and to facilitate cooperation between social media platforms.

4.2. Legal researchers

Question 1: Current domestic and EU regulations appear to have gaps regarding borderline content and implicit hate speech, that in turn also lead to gaps in content moderation. How could policies be adapted to counter implicit hate while maintaining freedom of expression? (Target group of recommendations: policymakers)

> *Recommendations*:
>
> - The legal framework is different in each country. Article 16 of the DSA is too complicated, considering that most users are young people that don't want to go through the entire process just to submit a report. The European Commission should lay down guidelines to simplify mechanisms for reporting hate speech.
>   - Downside: 'overblocking'; false accusations.
> - Mandate transparency reporting requirements for platforms, including measures against implicit hate speech.
>   - Downside: Significant financial and operational burden for platforms.

Question 2: While TikTok's community guidelines prohibit implicit hate against protected groups, this appears to be insufficiently enforced or too vaguely defined. How could platforms' community

guidelines be adapted to counter implicit hate and borderline content? (Target group of recommendations: platforms)

Recommendations:

- Refine community guidelines: draft a more detailed definition of implicit hate speech that is comprehensive and understandable for users, with specific examples of implicit hate speech. Clarify what the consequences are for posting ineligible content. This could also be done in partnership with social scientists and affected communities.
- Conduct public consultations and incorporate this user feedback into regular updates of the guidelines.
- Focus less on content-based moderation and more on system-based moderation.
  - o Downside: High implementation costs for moderation improvements.
- Make it easier to flag content
- Improve algorithms to flag potentially (implicit) hate speech content in a visible way, especially taking into consideration that more and more content is AI generated. Combine AI-efforts with culturally trained human moderators to identify and address subtle forms of hate.
  - o Downside: Risk of false positives/negatives with AI moderation, Possible perception of over-censorship by users.
- Counterspeech and recommendation systems: seeing a specific kind of content should not result in the algorithm pushing more radical content.

Question 3: Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. How can policymakers and/or platforms limit the impact of disinformation by linkages while maintaining freedom of expression? (Target group of recommendations: policymakers, platforms)

Recommendations:

- Establish inter-platform collaboration to share threat intelligence and verification mechanisms. If needed, policymakers should mandate cross-platform data-sharing protocols.
- Implement tools to flag or warn users about links to platforms known for spreading disinformation, explicit hate speech or implicit hate speech. For this, link-verification systems are needed that provide reputation scoring for external links.
  - o Downside: Privacy concerns with link tracking and analysis.
  - o Challenge: Technical challenges in improving and developing algorithms and tools to implement link-verification systems, demote mechanisms and reputation scoring for external links.
- Develop and/or improve algorithms to demote users/content linking to fringe platforms known for spreading disinformation, explicit hate speech or implicit hate speech.
  - o Downside: Risk of over-regulation leading to unintended censorship (problem of lawful versus harmful).
- Launch public media literacy campaigns to educate users about disinformation and implicit hate speech tactics. This could be done in partnership with fact-checking organizations and civil society to design educational materials.

4.3.   Practitioners

Question 1: Identity politics-based framing polarizes societies and can incite hatred yet is hard to counter legally. What (counter- and alternative) narratives can limit the impact of polarizing identity politics-based frames claiming existential threats by scapegoats (e.g. migrants, elites)? (Target group of recommendations: practitioners)

Recommendations:

- Spread counter narratives via relatable individuals (from the perspective of youth), e.g. influencers. Such narratives are less likely to be effective when disseminated by old, white man.
- Counter narratives: critical media literacy and proper usage of social media should be taught in school. This should be accompanied by educating practitioners (youth workers and teachers), too, e.g. on how dog whistles work.
- 'Prebunk' disinformation and hateful narratives instead of just using counter narratives.
- Establish monitoring projects that provide frequent (e.g. weekly) briefings on content and trends to relevant professions (e.g. professionals in education, youth work); eventually even with a weekly option of 'office hours' for Q&A.
- Cooperate with different professions on counter narratives (multi-perspective approach).
- Counter narratives should emphasise common values, such as respect and solidarity, as a counter pole to divisive aspects. They should highlight positive, constructive anecdotes about successful cooperation and integration, showing how diversity can lead to positive changes, thereby countering negative stereotypes.
- Counter narratives should also involve local voices from communities to counter stereotypes. Empowering multiplicators can help spread such counter and alternative narratives.

Question 2: The ecosystem approach not only helps the far-right to sustain presence by linking to back-up accounts but can also be used for campaigns to boost selected content and for luring users to ecosystems of 'alternative news'. What can platforms do to limit the impact of the cross-platform ecosystem approach for boosting content, luring users to alternative ecosystems and rebuilding back-up accounts? (Target group of recommendations: platforms)

Recommendations:

- Establish a "council of experts" that makes decisions about guidelines for the specific platform.
  - Downside: boosts the narrative of 'censoring the truth'.
- Develop technological solutions that automatically recognise networks of back-up accounts.
- Adjust algorithms to disrupt the spread of content from networks associated with hate speech and disinformation.
- Facilitate cooperation between platforms to disrupt cross-platform ecosystems.
- Ensure transparency regarding the algorithm.

- Educate users about manipulative ecosystems and their risks.
- Blocking IP addresses instead of accounts.
  - Downside: '1984', accusations of censorship.

Question 3: Implicit hate and identity politics-based narratives are less obvious and more palatable to individuals. Besides counter- and alternative narratives, how to make vulnerable individuals and/or societies more resilient against identity politics-based frames and implicit hate? (Target group of recommendations: policymakers, practitioners)

> ***Recommendations***:

- Improve media education and critical thinking skills in schools, including by educating practitioners on how hate speech works, what dog whistles are and how the far-right is strategically breaking taboos.
- Support programs that increase social ties and belonging to prevent isolation and vulnerability to hate speech.
- Develop monitoring and early warning systems to identify trends in the spread of implicit hate speech and identity politics, so that counter measures can be taken in a timely manner.

5. Reflection

A total of 17 international experts from various disciplines participated in the two focus groups and delivered numerous recommendations to the issues identified in RECO_DAR research reports. The summary of findings received positive feedback in general, although the explanation of the methodology appears to have fallen short for a few participants. These uncertainties were resolved in the Q&A session in the focus groups.

The Miro board proved to be a highly successful tool to facilitate online collaboration among participants. Its effectiveness was seemingly increased by providing the participants access to the board prior to the focus group, so they can prepare for the questions, as well as by guiding the process of developing recommendations via structuring the 'sticky notes'. The board also helped project staff in accurately documenting the results.

Small breakout sessions were a key element for ensuring that all participants' voices are heard, and a lively discussion can take place, which would not necessarily be possible in the plenum. Some participants in the second focus group mentioned that more time would have been preferred, as 40 minutes for answering three questions proved to be too short.

The output of the two focus groups is assessed to add significant value to the project results. Thanks to the multidisciplinary background of participants, the recommendations that were drafted are tailored to specific, narrow target groups by subject matter experts with extensive experience. The output of the focus groups will be incorporated into the upcoming recommendations in the form of a policy brief.

The participants assessed the findings to be relevant and timely, given the recent electoral successes of far-right parties across Europe. Some of the findings, particularly about the direction of linkages across the ecosystem, were said to be a crucial element for efforts trying to pinpoint the reasons behind the European far-right's surge. The project's methodology was deemed to be transferable for other contexts, both in terms of ideology and geography. Overall, the two focus groups fulfilled their purpose and advanced the RECO_DAR project.

6. Annexes

Annex 1: Screenshot: Miro board – first focus group (3 December 2024)

Annex 2: Screenshot Miro board – second focus group (10 December 2024)

Annex 3: Executive summary provided to participants

Annex 1: Screenshot: Miro board – first focus group (3 December 2024)

RECO_DAR
RESEARCH PROJECT

Welcome to the online focus groups
as part of EU CERV "RECO_DAR" Project

Today's Agenda

- Introduction
- RECO_DAR: context, methodology, findings
- Q&A in Plenum
- Breakout Sessions in working groups
- Result Discussion
- Summary and Closing

Session background

- The session is held under Chatham House Rules.
- There will be no recording of the session.
- We will provide a written summary of the results.

Basic reading material

- Conceptual approach to the far-right ecosystem online
- A new framework for analysing far-right hate speech on TikTok
- Disentangling the Web of Hate on TikTok
- Online Far-Right Framing Exposed

Introduction

Methodology, Key Findings and Q&Q

Hate Speech

is a form of communication, verbal, written, or visual, that intentionally or unintentionally degrades, discriminates against, devalues or threatens individuals or groups based on their inherent characteristics such as race, religion, ethnicity, sexual orientation, or any other societal categorisation. Even without explicit derogatory language, hate speech can exclude and impact individuals by constructing an alien 'other.'

GROUP 1

1. The TikTok community guidelines already prohibit implicit hate against protected groups, yet the enforcement seems to have gaps. **How could platforms improve the detection and removal of implicit hate/borderline content?** (Target group of recommendations: platforms)

2. Some far-right figures' accounts get frequently removed, yet they maintain a sustained presence on TikTok. **How can platforms detect and react to account sustainment strategies (back-up accounts, content mirroring, cross-platform ecosystem) quicker or even proactively?** (Target group of recommendations: platforms)

3. Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. **How can policymakers, and/or platforms, and/or practitioners limit the impact of disinformation by linkages?** (Target group of recommendations: policymakers, platforms)

Mainstreaming far-right ideas

Reach a vast amount of people → Convince people
Maintain presence on major social media platforms → Make extremism more palatable
Rebuild accounts | Evade moderation | Create urgency | Provide sources
Cross platform ecosystem | Conceal hate: borderline content | Conceal hate: framing strategy | Links to 'alternative' news

Focus Groups
30 min - 10 / Question

Guidelines for recommendations
- **which specific problem** that needs to be addressed?
- **who** should **act**?
- **what action** should be taken?
- **how** to implement this? **How** does it solve the issue?
- **why**? (what is the objective and expected outcome?)
- **what indicates success?**
- what are potential **challenges/downsides** of this recommendation?

GROUP 2

1. Current domestic and EU regulations appear to have gaps regarding borderline content and implicit hate speech, that in turn also lead to gaps in content moderation. **How could policies be adapted to counter implicit hate while maintaining freedom of expression?** (Target group of recommendations: policymakers)

2. While TikTok's community guidelines prohibit implicit hate against protected groups, this appears to be insufficiently enforced or too vaguely defined. **How could platforms' community guidelines be adapted to counter implicit hate and borderline content?** (Target group of recommendations: platforms)

3. Many narratives, including implicit hate and incitement as well as disinformation, are made plausible by links to 'alternative' platforms from TikTok. **How can policymakers and/or platforms limit the impact of disinformation by linkages while maintaining freedom of expression?** (Target group of recommendations: policymakers, platforms)

Guidelines for recommendations
- **which specific problem** that needs to be addressed?
- **who** should **act**?
- **what action** should be taken?
- **how** to implement this? **How** does it solve the issue?
- **why**? (what is the objective and expected outcome?)
- **what indicates success?**
- what are potential **challenges/downsides** of this recommendation?

GROUP 3

1. Identity politics-based framing polarizes societies and can incite hatred yet is hard to counter legally. **What (counter- and alternative) narratives can limit the impact of polarizing identity politics-based frames claiming existential threats by scapegoats (e.g. migrants, elites)?** (Target group of recommendations: practitioners)

2. The ecosystem approach not only helps the far-right to sustain presence by linking to back-up accounts but also be used for campaigns to boost selected content and for luring users to ecosystems of 'alternative news'. **What can platforms do to limit the impact of the cross-platform ecosystem approach for boosting content, luring users to alternative ecosystems and rebuilding back-up accounts?** (Target group of recommendations: platforms)

3. Implicit hate and identity politics-based narratives are less obvious and more palatable to individuals. **Besides counter- and alternative narratives, how to make vulnerable individuals and/or societies more resilient against identity politics-based frames and implicit hate?** (Target group of recommendations: policymakers, practitioners)

Guidelines for recommendations
- **which specific problem** that needs to be addressed?
- **who** should **act**?
- **what action** should be taken?
- **how** to implement this? **How** does it solve the issue?
- **why**? (what is the objective and expected outcome?)
- **what indicates success?**
- what are potential **challenges/downsides** of this recommendation?

Comment / Notes Section

RECO_DAR
EXECUTIVE SUMMARY

IN BRIEF

- A variety of German-speaking far-right figures use TikTok. Some users counter account suspensions by using back-up accounts, content mirroring, and a cross-platform ecosystem.

- Links from Telegram to TikTok are significantly more common (3:1) than vice versa within the far-right ecosystem, likely to boost users and content, e.g. after account removals.

- Thus, the far-right currently seems to prioritize TikTok presence. Content dissemination on TikTok is sufficiently successful, reducing incentives to lure users to fringe platforms.

- Most links from TikTok lead to 'alternative' news to grant legitimacy to far-right narratives – often based on mis- and disinformation – by providing a 'source'.

- Within the far-right ecosystem, explicit hate speech is minimal, while implicit hate speech ('borderline content') extensively appears, evading content moderation. Surprisingly, overt hate is more prevalent on TikTok than on Telegram or fringe platforms.

- The discursive strategy focuses on identity politics, which is hierarchical, adversarial, binary and manipulative, differentiating among others between 'genuine' autochthons and others seen as an existential threat. This actively pushes polarisation and hate towards perceived others.

CONTEXT

The RECO_DAR[1] project analyses a **German-speaking right-wing extremist ecosystem's** strategies for recruitment and hate speech online, focusing on TikTok while including Telegram and fringe platforms. It draws on **142 German-speaking far-right accounts** (12,433 followers and an average of 18,269 views). Posts published on these accounts in the period from February 2020 and November 2023 were scraped between September and November 2023. The project performs manual and computational analysis of content and a computational analysis of the audience and the follower networks.

FINDINGS – ACTORS AND USER BASE

- Most **major public figures** from the Austrian and German **far-right** are present on TikTok.

- 19% of users have posts containing **conspiracy theories**, 8% posted **pro-Russian propaganda**

- While a **significant share of** far-right accounts was **suspended** shortly after we identified them, **back-up accounts**, in some cases continuous **content mirroring**, and the ecosystem approach across platforms allowed some of them to maintain a **sustained presence** on the platform.

- Common, overarching features of the user base: few but influential **female users**; high prevalence of **right-wing populism** and **anti-mainstream** views, along with **subcultural actors** and **conspiracy theorists**.

---

[1] "Right-wing extremist eco-systems driving hate speech: dissemination and recruitment strategies"

FINDINGS – ECOSYSTEM

- Far-right users are **highly interconnected** by their **audiences** (followers, commenters) but **form distinct communities** (online 'bubbles') with unique defining features in terms of ideology, content strategies, and topics. This ecosystem can be categorized into four communities (see figure below).

- **Far-right public figures' audiences** considerably **overlap** with fringe, more **radical users' audiences**.

- There are significant generational differences in terms of online behavior within the far-right. **Young users** are **concentrated** in a single community. **Austrian users** are also mostly **clustered** in a single community, questioning the assumption of a truly borderless German-speaking scene.

- Despite the presence of a variety of ideologies within communities and even within individual users' content, **each community has a dominant ideology**. This supports the 'salad bar theory' of extremism while also demonstrating its limitations and suggests that **audiences (online 'bubbles') are formed around specific ideologies** even within the far-right scene.
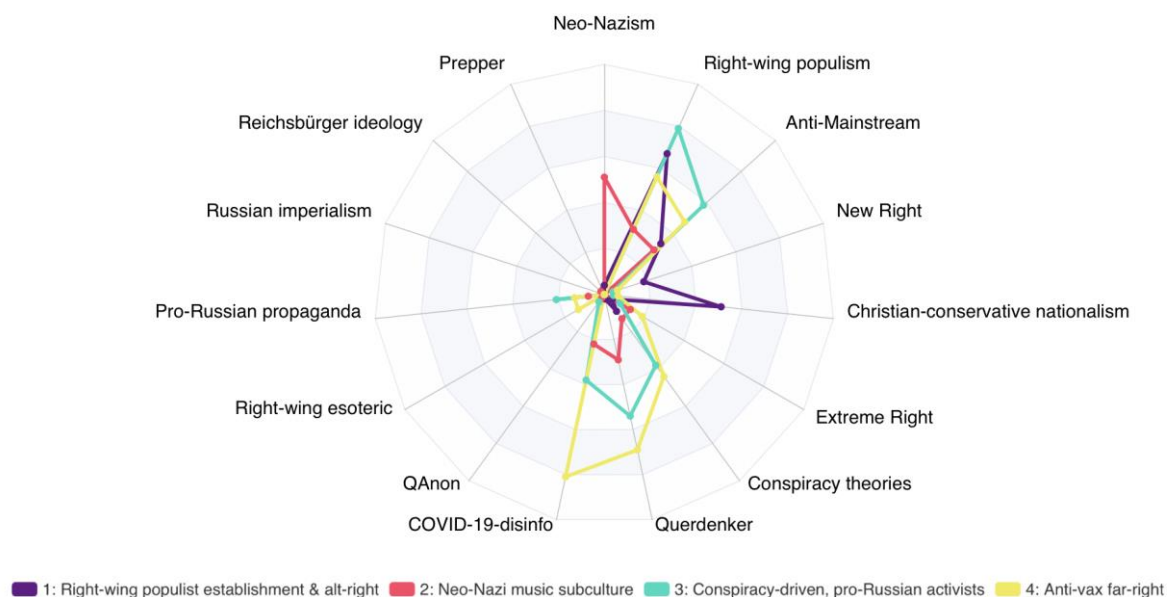

Ideologies under the far-right umbrella in four TikTok communities: 'salad bar of extremism' - to a limited extent

FINDINGS – RECRUITMENT AND LINKAGES

- **Outlinking** from TikTok is however still utilised for **strategic purposes**. Links mostly lead to **'alternative' news pages** to grant further legitimacy to arguments made on TikTok with 'sources', countering what is perceived as mainstream media, and potentially also to evade content moderation by spreading problematic content indirectly on other platforms. These links are accompanied by the **negative framing** of mainstream institutions (e.g., media, governments) to strengthen **in-group solidarity** and create a sense of opposition to external threats.

- **No large-scale, systematic recruitment via links from TikTok** to less moderated networks was found. On the contrary, the ecosystem appears to be used the other way around: **linking from**

Telegram to TikTok, presumably to boost far-right accounts and content, outnumbers links from TikTok to Telegram **3:1**.

- These findings suggest that **maintaining an active presence on TikTok is a strategic priority**, with other platforms across the ecosystem playing a supportive role in **mainstreaming far-right narratives and ideologies** among a broad target group.
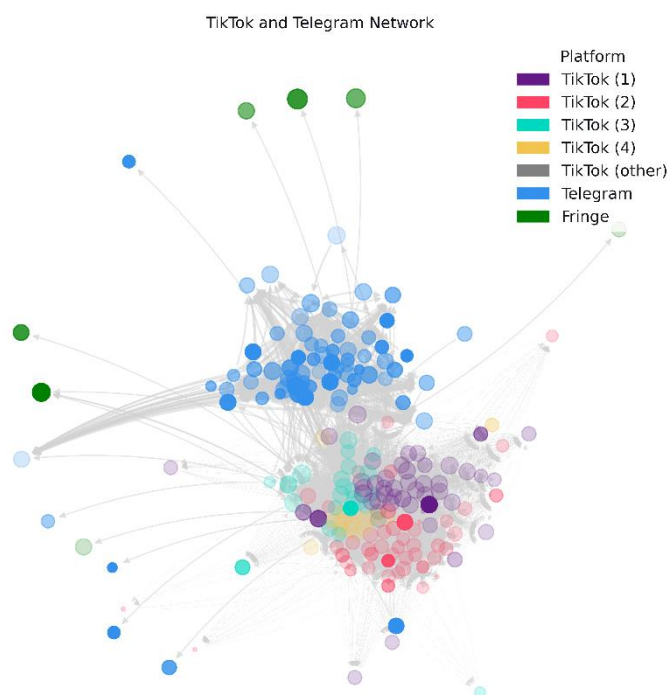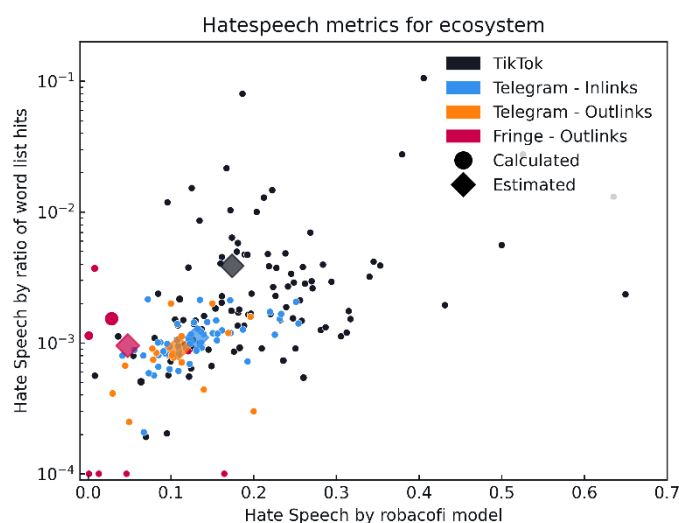

Linkages across the ecosystem's platforms.
The higher the transparence of nodes, the more hateful its content is.

- **Explicit hate speech** or calls for violence are **rare** in the ecosystem, but not absent. TikTok posts contain more overt hate on average than posts on Telegram or fringe platforms. Across the entire ecosystem, right-wing populists and Neo-Nazis have the most openly toxic content. While major accounts are somewhat less toxic on average than fringe accounts, the difference is minimal.

- **Implicit hate speech** and **incendiary content** remain a **major issue** across the entire ecosystem, with a variety of sophisticated strategies to evade content moderation. While these are less grave than open calls for violence or hatred, they **can inflict harm** (e.g. by spreading hateful stereotypes) and can even motivate individuals to engage in violence.

Comparison of overt hate speech across the ecosystem. Fringe platforms include VK, Odyssee, Rumble, GloriaTV.

FINDINGS – CONTENT AND FRAMING STRATEGIES (TIKTOK, TELEGRAM, ODYSEE)

- **Identity politics** is the dominant strategy in the ecosystem, focusing on manipulating the audience into adopting a **polarized, adversarial, binary and manipulative world view**, where every person and entity is either a victimized, genuine autochthon or part of the enemy alliance. The framing starts off with an initial situation of **multiple crises** creating **existential threats**, to which solutions rooted majorly in the autochthonous **identity and way of life**, including the **glorification of the past** and specific values are presented. These are conscious attempts to actively polarize society by providing scapegoats for complex problems.

- The conscious tactic of **framing perceived enemies as an existential threat** to the viewers' identity, way of life, and community is often combined with the targeting of specific groups/individuals portrayed as responsible for these threats. This can mobilize the audience, **incite hatred and potentially even violence**, while **maintaining plausible deniability**.

- No major differences were found in terms of framing strategies across platforms.
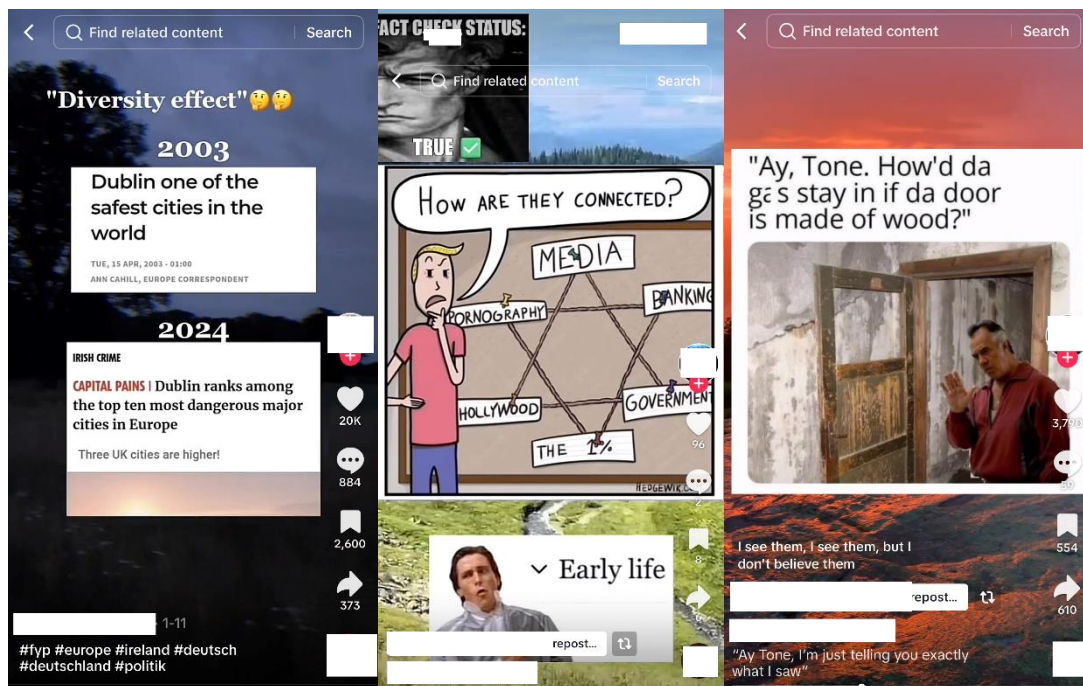
Examples of implicit hate to evade content moderation on TikTok: suggestive anti-migration narratives, anti-Semitic conspiracy theory and Holocaust denial in memes