POLICY BRIEF
February 2025

HIDDEN EXTREMISM
Countering the Far-Right's Online Ecosystems

The German-speaking far-right has been actively trying to polarise societies and make their radical ideas accepted and even supported by mainstream society. They have successfully used sophisticated, manipulative communication strategies on popular social media platforms, such as TikTok, where broader audiences can be reached. These efforts are supported by a decentralised network across multiple online platforms and appear to be fruitful, given the far-right's recent electoral successes.

The RECO_DAR project (Right-wing extremist eco-systems driving hate speech: dissemination and recruitment strategies) systematically mapped a German-speaking far-right online ecosystem. It unravelled several new strategies far-right actors employ on TikTok and other social media platforms to reach and influence existing and new audiences. The following evidence-based findings and related recommendations offer policymakers, social media platforms, and practitioners an account of these developments and provide options for dealing with them.

How has TikTok and social media been successfully exploited by the far-right?

CAMOUFLAGING TOXICITY

The far-right's narratives on TikTok are mostly not openly violent or hateful. Instead, they use a combination of videos, symbols, and coded language to (implicitly) spread hate and extremist messages without being easily noticed or flagged, thus amounting to a 'hidden extremism'.

CREATING NEGATIVE EMOTIONS TO POLARISE SOCIETIES

They blame certain groups for social problems, presenting them as urgent threats to the viewers to inspire fear, anger, and action. They push an artificially created autochthonous identity, pictured as under attack by an equally artificially created 'progressive' identity that includes several groups, such as foreigners, 'the Left', 'the Greens', international and European institutions, LGBTQI+, media, etc.

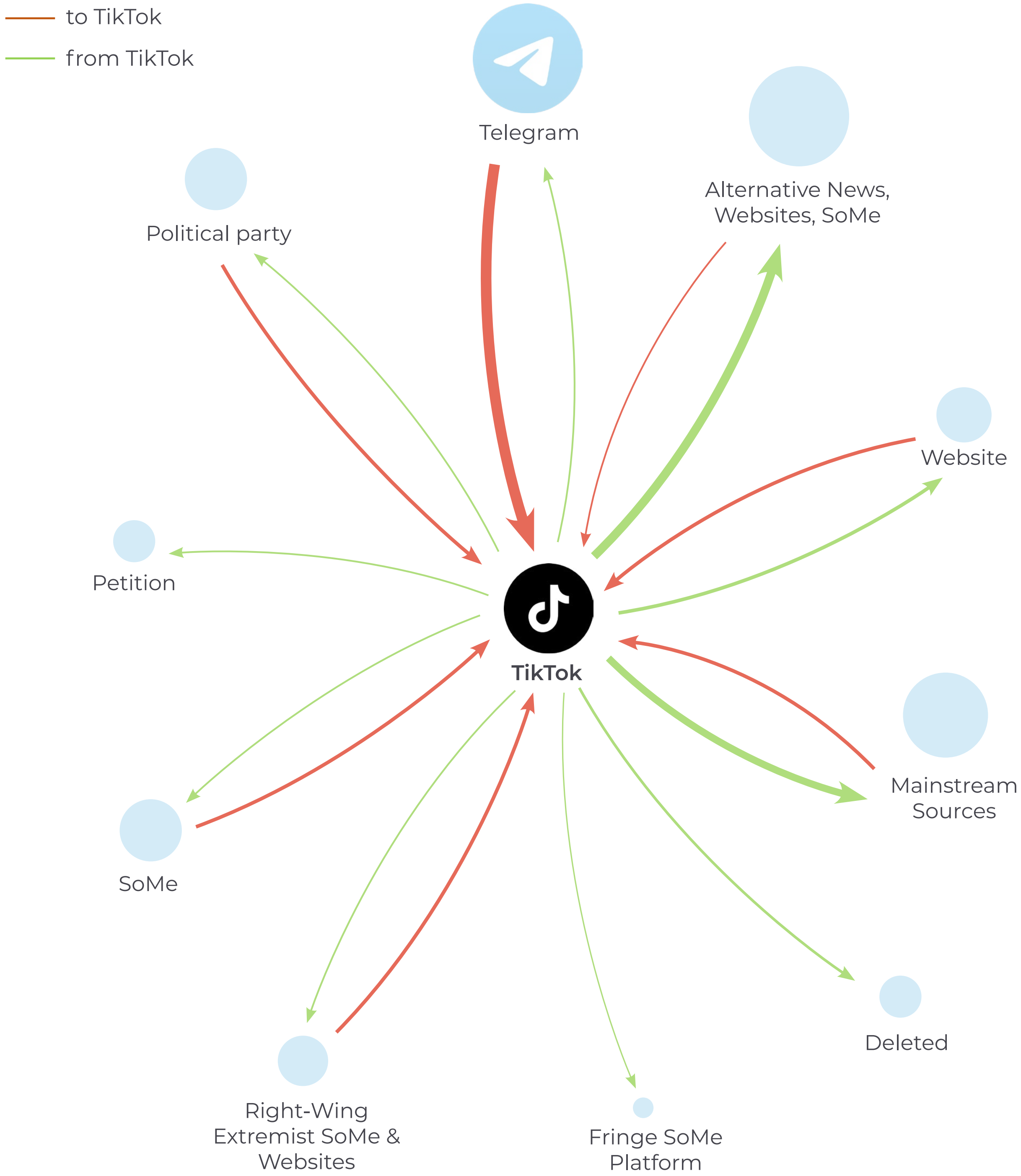
RECO_DAR
RESEARCH PROJECT

LEGITIMIZING MISINFORMATION

They make their arguments seem credible on TikTok by linking to 'alternative' news outlets - unreliable websites that look like credible news sources but spread dis- and misinformation. Simultaneously, they delegitimise established institutions and sources of information by portraying them as hostile, dishonest, and acting against society's interests.

BOOSTING CONTENT

They organize campaigns on other platforms (e.g. Telegram) to make far-right posts and users on TikTok more popular. If far-right accounts are removed, these networks on different platforms are also used to increase backup accounts' reach on TikTok.



Why does it matter?

ERODING DEMOCRACY FROM WITHIN

The careful formulation of radical narratives allows the far-right to make their ideas seem less extreme and more acceptable to a broader, moderate audience. It can convince viewers to adopt far-right positions by exaggerating (often existing) problems and framing them as existential threats, triggering protective instincts. This helps the far-right succeed in elections and gain support for laws that weaken democracy and human rights, such as limiting the free coverage of news in the press, undermining the independence of courts, and taking away equal rights from religious and LGBTQ minorities.

BUILDING ALTERNATIVE REALITIES

Spreading misleading news via links to 'alternative' websites undermines trust in reliable media and democratic systems. Finding common solutions will become impossible if people do not know what to believe and can no longer agree about basic facts.

CREATING PARALLEL SOCIETIES

The negative identity discourse about certain societal groups and state institutions establishes an atmosphere of tension, lack of trust, and a split in society. This harms community cohesion and can lead to more conflicts. Long-term, polarised societies often struggle to find a shared vision for living together and addressing common problems.

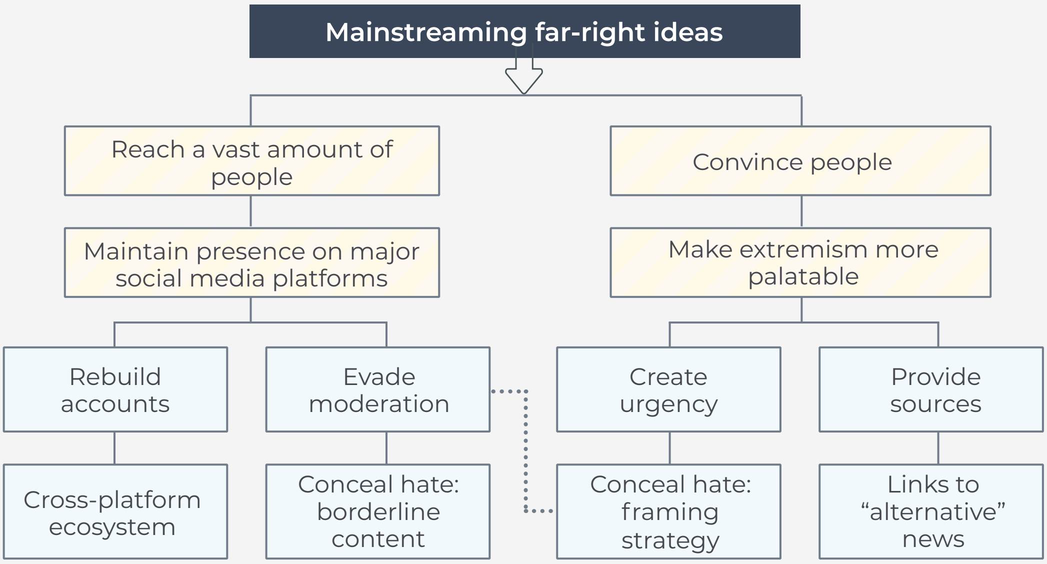
REACHING MANY PEOPLE BY BEATING MODERATION

Using 'hidden extremism' and a network across platforms, the far-right stays active on major online platforms and evades moderation efforts despite spreading implicitly hateful views. Boosting content makes the algorithm more likely to recommend it to new viewers. By doing so, the far-right's harmful messages reach many people and have a more substantial impact.

RECOMMENDATIONS

Considering the severe risks that the far-right's actions online pose to democracy and human rights, relevant stakeholders must take action to counter these efforts and limit their impact. The following presents recommendations, including their potential risks and associated challenges.

The German-speaking far-right's framework for mainstreaming extremism online.


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

Options for platforms

1. **To counter the far-right's ecosystem-driven sustainment strategies, social media platforms should adapt their content moderation approaches by adding more emphasis on 'pre-empting' ecosystem-driven content evasion strategies. Before sanctioning an account, (1) look for and pre-flag back-up accounts, then (2) remove or shadow-ban pre-flagged back-up accounts.**

   - **Implementation:** look for almost identical usernames, especially among tagged profiles in bios, captions, comments, and posts of accounts suspected of violating community guidelines. In addition, consider metadata (e.g. IP addresses), which can also help counter bot networks and state-sponsored disinformation campaigns. Verify by comparing content and audience.

   - **Risks:**

   - False positives can lead to user dissatisfaction and complaints. To mitigate, offer an option to appeal.

   - Accusations of overreach/censorship/mass surveillance, with potential legal consequences for infringing on freedom of expression. To mitigate, be transparent about the rationale, e.g. by referring to independent research on the issue.

   - **Challenges:**

   - Identifying back-up accounts is an additional burden on resources, requiring personnel when done manually, or development, maintenance and computing power if done by AI.

   - Metadata-based identification is easy to bypass using a VPN or public wireless network and might violate GDPR.

2. **To facilitate the enforcement of community guidelines related to removing implicit hate, social media platforms should work with subject-matter experts to diversify and continuously update training datasets for AI tools.**

   - **Implementation:** partner with domain-specific extremism researchers and representatives from affected communities to annotate datasets for dog whistles, codewords, and borderline content.

   - **Risks:**

   - Accusations of censorship. To mitigate, explain the extent of the issue and its harmfulness in simple language with concrete examples.

   - False positives by AI can lead to user dissatisfaction and complaints. To mitigate, offer an appeals process.

   - Challenges:

   - Dog whistles constantly change and adapt, requiring continuous monitoring and regular resource-intensive dataset updates.

   - Data bias may lead to particular groups or religions being disproportionately impacted.

   - The financial burden to continuously annotate datasets and train AI.

Options for platforms

3. **To counter the far-right's ecosystem-driven disinformation and propaganda, social media platforms should establish or expand the screening mechanism of external links beyond bios to include URLs found in comments, captions, and posts, and warn users about these links if their content violates community guidelines.**

- **Implementation:** Partner with subject-matter experts from extremism and disinformation research to establish a database of the most frequent websites known to spread extremist ideas and disinformation, including indicators. Train AI models on detecting such content. Have operators verify URLs flagged by AI before adding a warning label. Prioritize screening links associated with accounts or networks already flagged for violating community guidelines.

- **Risks:**

- Accusations of overreaching censorship and/or political bias.
- False positives might lead to user dissatisfaction and substantial economic losses, given that content creators and brands rely on external links. To mitigate, add a human-in-the-loop to verify flagged URLs made by AI and a fast-track appeals process for content creators and brands.
- Legal challenges for infringing on freedom of expression and commerce.
- Legal limitations on cross-platform tracking, data scraping, and analysis.

- **Challenges:**

- Easy to evade by masking links, e.g. via URL shorteners, QR codes, and leet.
- Links in posts are resource-intensive to detect as they require OCR.
- Significant computational, operational, and financial resources required to process data, train an AI system, and employ operators that review flagged URLs and handle appeals.
- Technical difficulties with scraping data from a variety of platform types (e.g. websites, other social media platforms, etc.)

4. **To counter the far-right's ecosystem-driven disinformation and propaganda efforts more efficiently, social media platforms should allow users to report external links for review.**

- **Implementation:** add a report function for URLs, like reporting users and specific posts. Screen these links using AI for apparent violations of community guidelines or a selection of these, e.g. promoting violence. If a URL is flagged, have an operator verify the decision and add a warning label to users.

- **Risks:**

- Accusations of overreaching censorship and/ or political bias. To mitigate this, include an assessment of this measure in regular transparency reports.
- Legal challenges for infringing on freedom of expression and commerce.
- Legal limitations on cross-platform tracking, data scraping and analysis.
- Abuse by mass reports, straining resources. To mitigate this, track reporting patterns to penalize bad-faith actors.

- **Challenges:**

- Significant computational, operational and financial resources required to screen and flag URLs.
- Technical difficulties with scraping data from various platform types (e.g. websites, other social media platforms, etc.) for screening.

Options for Policymakers (EU)

1. **To counter the far-right's ecosystem-driven online strategies, policymakers should support incorporating threat intelligence on cross-platform extremist networks and propaganda campaigns into Global Internet Forum to Counter Terrorism (GIFCT) data-sharing mechanisms.**

   - **Implementation:** a platform for preventing terrorists and violent extremists from exploiting digital platforms already exists (GIFCT), but the data-sharing cooperation mainly revolves around content, not ecosystems. Data-sharing mechanisms should be amended to include ecosystem-focused threat intelligence, such as cross-platform content boosting or account sustainment campaigns.

   - **Risks:**

   - Some relevant data may be sensitive, including trade secrets and vulnerabilities, making information sharing unfeasible or competitively unbeneficial for some platforms.
   - Shared data can be leaked, abused, and exploited to damage competitors.

   - **Challenges:**

   - Limits on what data can be shared by GDPR.
   - Difficult to get platforms involved without clear incentives given the risks.
   - The additional operational burden for GIFCT.
   - Protocols and formats for sharing data might be complex, given the different systems used by platforms.

2. **To proactively counter backup accounts and other account sustainment strategies of the far-right, policymakers should support knowledge-sharing among social media companies on content moderation, e.g. TrustCon. Policymakers should fund research on such methods and facilitate exchange between platforms and experts.**

   - **Implementation:** support existing initiatives and incentivise social media companies to participate, e.g. by dispatching government officials and law enforcement representatives to establish connections, discuss needs and issues, and initiate public-private partnerships.

   - **Risks:**

   - Sharing solutions and technology may disincentivise platforms from 'thinking outside the box' and investing in resolving moderation-related issues. To mitigate, focus on non-proprietary issues, e.g. trends in evasion strategies, to which researchers can contribute.

   - **Challenges:**

   - Patented technology and other solutions that offer a competitive advantage to a platform are legally protected and are unlikely to be shared.
   - Given that each platform is different, solutions might not be transferable.
   - Technological solutions are cost-intensive to develop and test. Platforms are not incentivised to share their expensive solutions with competitors, as they do not benefit from it, yet competitors would gain an additional advantage.

Options for Policymakers (EU)

3. **To facilitate the timely removal of borderline content, policymakers should amend the Digital Services Act (DSA) by enacting obligatory guidelines for service providers to simplify mechanisms for reporting. The current reporting processes are too long and complicated for laypeople, especially youth.**

   - **Implementation:** simplify the language of the reporting process by partnering with linguists, communication experts and youth if needed; reduce the predefined reporting categories; educate users by providing examples and/or tutorials; and use AI to suggest which category the content is most likely to fit.

   - **Risks:**

   - Potential operational overload by inaccurate overreporting.

   - **Challenges:**

   - Balancing simplicity and precision: maintaining the rationale of pre-sorting reports based on what precisely the reported post/user violates.

4. **To boost social media platforms' efforts to counter borderline content, policymakers should amend transparency report requirements under the DSA by mandating platforms to include specific measures taken against borderline content/implicit hate speech.**

   - **Implementation:** mandate reporting on detection methods, user education efforts, the volume of flagged content, type of violation, and measures taken.

   - **Risks:**

   - Further financial and operational burden on platforms, especially smaller ones.
   - Disclosed measures might be sensitive and thus help bad-faith actors exploit and circumvent systems and rules in place.
   - Competitors and critiques could weaponize reports by putting public pressure on platforms for not doing enough or too much.

   - **Challenges:**

   - Defining borderline content and implicit hate.
   - Finding a balance between public accountability and providing exceptions to prevent the forced disclosure of sensitive details.

RECOMMENDATIONS

Options for practitioners

1. **To limit the impact of borderline content and identity politics-based framing, practitioners should focus on edutainment (educational entertainment) and fact-based posts that facilitate opinion shaping.**

   - **Implementation:** include young people in creating contemporary, age-appropriate formats, such as (animated) infographics and short videos. Take precautions to avoid appearing to be taking sides and focus on facilitating critical thinking based on facts instead.

   - **Risks:**

   - Reactance: if content is perceived as being told what to do/think, it is likely counterproductive. To mitigate, present hard facts and avoid taking sides in debates.

   - **Challenges:**

   - It requires training, skills, and time to create consistently high-quality material and build an audience.

   - Oversaturation: difficulty gaining traction with content on serious topics, competing with other content creators and brands.

   - Finding the balance between being too trivial and entertainment-focused about serious topics and being too complex.

2. **To make societies more resilient against identity politics-based frames and borderline content, practitioners should establish joint monitoring projects with researchers to create an early warning system. These projects should provide frequent (e.g. weekly) briefings on content and trends to frontline practitioners (teachers, youth workers).**

   - **Implementation:** establish a collaboration between front-line professionals and subject-matter experts to keep track of what youth is talking about. Monitor offline discussions and online developments related to extremism and disseminate these in short written briefings in plain language. Provide examples explaining current topics, dog whistles, context, background, implications, and corresponding counter/alternative narratives. Optionally, provide periodic Q&A sessions or ad hoc counselling.

   - **Risks:**

   - Practitioners may feel overwhelmed by integrating this into their already stretched daily routines. Work with practitioners to find easily digestible and practical formats to mitigate this.

   - Accusations of mass surveillance, primarily if funded by the state. To mitigate, provide transparency reports about methods, objectives, activities, and ethical guidelines.

   - **Challenges:**

   - Additional financial burden at a time when public resources are stretched thin in the EU.

   - Potential lack of local context if monitoring is done on a higher level.

This report was produced as part of the EU-funded RECO_DAR project: Right-wing extremist eco-systems driving hate speech: dissemination and recruitment strategies
RECO_DAR is a collaboration between the organisations SCENOR and modus|zad.
More information: https://www.scenor.at/recodar

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.