Hate Disguised as Debate:

Trans-Related Content in the Feeds of Nordic Youth



ABSTRACT

This research is a collaborative effort between the organisations NORDREF, Transfeminiinit ry, Ofbeldisforvarnaskólinn and Samtökin ’78, all of which have focused extensively on youth, equality, LGBTQI+ rights and digital environments in their work. Co-funded by the Nordic Gender Equality Fund and carried out in Finland, Iceland and Sweden, it is uniquely positioned to provide Nordic-specific data on what trans-related content children in the region encounter on social media and what messages they are exposed to – findings that can inform regional responses to the current backlash against LGBTQI+ rights rather than relying on general global trends.

Using new social media accounts with controlled search patterns, the research tracked what kind of anti-trans, trans-ambivalent and trans-affirming content is recommended to youth in Sweden, Iceland and Finland when they explore interest-based search terms. The research illustrates:

- Where gaps exist in platform policies, moderation and transparency
- Which measures are needed to strengthen protections for LGBTQI+ users, particularly youth
- What educators, parents and policymakers should know about the trans-related content young people are served, including hostile content framed as debate
- What obligations arise under the DSA and AI Act, and what this means for platform accountability and the protection of minors

The findings are situated within the EU’s dual regulatory framework, the Digital Services Act and the AI Act, and inform recommendations directed at platforms, regulators, educators, parents and policymakers working to address online hate and the hostile framing of minority rights.

TABLE OF CONTENTS

ABSTRACT 2

Executive Summary 5

About the Study 5

Key Findings 5

Regulatory Implications 7

Key Recommendations 7

Background and Nordic Relevance 8

Methodology & Research Design 12

Platform Selection 12

Account Creation 13

Search Terms & Engagement 13

Data Collection 15

Coding & Analysis 16

Limitations & Ethical Considerations 17

Overview of Current Algorithmic Governance 18

TikTok's Compliance: Transparency and Access Issues 19

YouTube's Compliance: Partial but Insufficient 19

Quantitative Conclusions: Discussion 21

Up to a Third of All Recommended Trans-Related Content is Anti-Trans or Ambivalent 21

Algorithms Do Not Actively Promote a "Trans Agenda" or "Gender Ideology" 21

TikTok Suggests Over 3 Times More Trans-Related Content Than YouTube 23

Differences Between a Search-Based Algorithm and a Recommendation Algorithm 24

Searching for J.K. Rowling Exposes Young Fans to Content Questioning Trans People's Rights 24

Rewatching Hostile Content: An Open Question with Serious Implications 25

Broader Challenges and Implications 27

Recommendations for Regulators 28

Qualitative Analysis: Sub-Category Frequency and Valence in Trans-Related Social Media Content 29
Overview 29
Overall Category Distribution 29
Sub-Category Frequency and Valence 29
Explication by Sub-Category 30
Political Debates & Arguments (n = 274, the most frequent sub-category) 30
Body, Identity & Health (n = 237) 32
Growing Up Trans (n = 145) 33
Law, Liberties & Access (n = 79) 34
Media Representation (n = 69) 34
Educational / Academic / Art (n = 61) 35
Discrimination (n = 49 — smallest but most contested) 35
Summary of Key Patterns 36
Methodological Note 36
Non–Trans–Related Qualitative Findings 36
The “Secret Account” Trend 37
Polarization by Social and Traditional Media 37
What’s Ahead 37
Find Your (Algo)Rhythm 37
Acknowledgments 38

Executive Summary

About the Study

This study, which had the working title #HateFYP, is the largest algorithm audit of its kind carried out in the Nordic region to date. Twelve new TikTok and YouTube accounts, registered as 15-year-olds, were operated on clean, factory-reset phones in Reykjavík, Stockholm and Helsinki between November 20-25, 2025. Each account searched for six terms. The first four were terms related to mainstream youth interests (video gaming or Harry Potter), and the last two were "gender identity" and "trans kids." After each search, researchers watched the top five results and then the next videos the platform recommended. In total, 5,406 videos were coded as trans-affirmative, trans-ambivalent, anti-trans or not trans-related, and every trans-related video was also coded by theme.

As a six-day controlled experiment with twelve accounts, the study reveals patterns in how the algorithms behave, but cannot measure what all Nordic children see. Anti-trans content was unevenly spread, with one Icelandic TikTok account producing 33 of the 49 anti-trans videos on TikTok.

Key Findings

1. Most of the trans-related content was affirming, but a significant share was not
Of the 471 trans-related videos, 71.8% were trans-affirmative, 16.8% ambivalent and 11.5% anti-trans. On TikTok, three in ten trans-related videos recommended to accounts registered to minors were anti-trans or ambivalent (30.6%), compared with two in ten on YouTube (20%). Exposing children repeatedly to content that questions trans people's existence or rights risks normalising those positions, even without illegal hate speech.

2. Anti-trans content is most often delivered as "debate"
Political debates and arguments were the most frequent theme in the trans-related content. Hostile content rarely looked like open hate. It typically took the form of street interviews, panel shows and "both sides" formats that presented trans people's right to exist or to access healthcare as an open question, while favouring the anti-trans position. Content categorized as *discrimination* (28.6% anti-trans) and *law, liberties and access* (25.3%) was where hostile framing was most likely to appear.

3. TikTok escalates far faster than YouTube
Once users searched for "trans kids," more than half of the TikTok feed (52.9%) became trans-related, compared with 10.4% on YouTube. On two individual TikTok accounts, the figure reached 73%. Overall, TikTok served more than three times as much trans-related content as YouTube (13.6% vs. 3.9%) and its recommendation system responds to user signals far more aggressively.

4. After neutral searches, TikTok's algorithm recommended most of the anti-trans content

The researchers used only neutral search terms ("gender identity" and "trans kids") and never searched for anti-trans material. Both platforms returned almost exclusively trans-related search results for these terms. The difference lay in what came next: of the 49 anti-trans videos recorded on TikTok, 45 appeared in the recommended feed rather than in search results, and trans-related content made up nearly half of TikTok's recommended feed after the "trans kids" search. On YouTube, by contrast, few trans-related videos appeared beyond the top five search results. The content brought by the recommendation algorithm therefore differs from that of the search algorithm, and platforms should report on the two separately.

5. Searching for J.K. Rowling exposes young fans to content questioning trans people's rights

The author of the Harry Potter series, J. K. Rowling, has campaigned prominently against gender self-identification and legal recognition of trans women as women in recent years. Searching for her name brought trans-related content into the feeds of all six Harry Potter accounts on both platforms. On YouTube, most of this content was ambivalent, presenting trans people's rights or identities as open questions; on TikTok, half was affirmative and half ambivalent or anti-trans. A search for PewDiePie, a comparable celebrity in the gaming world, produced no trans-related content.

6. There is no hidden "trans agenda" in the algorithms

Despite years of accusations from right-wing and religious groups, the study found no evidence that social media platforms push trans-related content or "gender ideology".[1][2] When the accounts searched only for topics related to gaming and Harry Potter (apart from its author), not a single trans-related video appeared on either platform. Young people meet this topic when they look for it or – more worryingly – when it is introduced to them through anti-trans or ambivalent commentary.

7. Content about trans young people themselves was overwhelmingly positive

Content about what it is like to grow up as a trans child was 93.1% affirmative, with no anti-trans videos, and educational content was 95.1% affirmative.

8. The findings extend beyond trans issues

On a TikTok account registered as belonging to a 15-year-old, up to 8% of recommended videos promoted keeping a "secret account" hidden from parents. On YouTube and TikTok, researchers also observed outrage-driven framing from mainstream news outlets, which presented complex issues such as climate change and human rights as "for or against" questions. Both patterns point

---

[1] Eckert, J. & McCloskey, M., The Heritage Foundation, 15 September 2022: How Big Tech Turns Kids Trans (originally published in The American Conservative).

[2] BizPac Review, 6 September 2022: Social media is influencing some young girls to become transgender.

to engagement-driven framing that can undermine child safety and democratic discourse more broadly.

Regulatory Implications

- **Digital Services Act:** Under Articles 34 and 35, very large platforms must assess and mitigate systemic risks, including risks to fundamental rights, civic discourse and the wellbeing of minors. Our findings show that, after neutral searches about trans topics, accounts registered as 15-year-olds were recommended content questioning trans people's rights and existence. This is the kind of risk these assessments should cover.

- **Transparency gap:** Neither TikTok nor YouTube provides enough transparency or researcher data access (Article 40) to verify why their algorithms behave as observed. In October 2025, the European Commission preliminarily found TikTok in breach of these obligations. It has since issued preliminary findings against TikTok on the protection of minors (July 2026) and on the 'rabbit hole' effect of its For You feed (November 2026).

- **AI Act:** In our study, both platforms recommended anti-trans content to accounts registered as 15-year-olds. The AI Act's Article 5(1)(b) prohibits AI systems that exploit age-related vulnerabilities in ways that cause significant harm. Recommendation systems that amplify hostile content to minors may fall within its scope.

Key Recommendations

1. **Platforms** should stop treating viewing time as a signal of approval. On 6 November 2026, the European Commission preliminarily found that when a TikTok user repeatedly watches content that may be harmful, "the system interprets these interactions as positive signs of interest and recommends more of the same," and concluded that TikTok should disable viewing time as an indicator for recommending content.[3] Our findings raise the same concern for hostile content about minority groups: young people may rewatch such content because it upsets them, not because they agree with it. Platforms should disable viewing time as a recommendation signal for users, especially those registered as minors, and rely instead on their explicit preferences. Repeat viewing alone of hostile content should not lead to more of it being recommended.

[3] European Commission: Commission preliminarily finds TikTok's 'For You' feed in breach of Digital Services Act for harmful 'rabbit holes' effect, press release IP/26/2361, Brussels, 6 November 2026.

2. **Platforms** should label or add context to debate-format content that presents a group's rights or existence as an open question, and should not prioritise such content in the feeds of users registered as minors.

3. **Regulators** should require platforms to be transparent about what minors are served. Platforms should be required to publish regular, comparable audits of the content recommended to users registered as minors, report separately on what reaches them through search and through recommendations, and give independent researchers real-time access to the data needed to verify these reports. Additional recommendations on enforcement, audits and transparency are set out in the section 'Recommendations for Regulators' (p. 28).

4. **Nordic authorities** should support independent monitoring. Nordic values are not encoded in the platforms, and as a result, the feeds of Nordic children are shaped by English-language culture-war debates imported from the US and UK that undermine the rights of Nordic citizens. Small countries cannot change this alone, but together and through European law they can. The Nordic Council of Ministers should support recurring, independent audits of what platforms show Nordic children, so that evidence doesn't depend on borrowed phones and on one-off projects. Authorities like Post- och telestyrelsen, Traficom and Fjölmiðlanefnd should coordinate with the European Commission and share cross-country evidence such as this study.

5. **Icelandic authorities** should bring the DSA into force in Iceland to ensure that Icelandic children are protected online on equal terms with their peers in Sweden and Finland. Iceland should press for swift incorporation of the DSA into the EEA Agreement, adopt national implementing legislation, and designate a well-resourced Digital Services Coordinator and Trusted Flaggers with expertise in children's rights.

6. **Educators and parents** should be supported with tools for open dialogue, critical thinking and media literacy, so young people can recognise manipulation presented as debate.

Background and Nordic Relevance

A backlash against LGBTQI+ rights is taking place in the Western World and beyond, with increased stigmatization and a scaling back of civil liberties in numerous European countries as well as the US. In Italy, Hungary, Georgia, Slovakia, Kazakhstan, the UK and in Russia, legal protections for LGBTQI+ people have been reversed, particularly in healthcare, legal recognition and family rights. Censorship has increased with bans on Pride marches, LGBTQI+ symbols and discussions in

schools. Moreover, conservative and religious groups, often funded from abroad, are actively lobbying against LGBTQI+ rights across Europe.[4567]

Executive orders issued by the Trump administration in 2025 and 2026 instructed federal agencies to remove and cease issuing materials or communications that "promote or inculcate gender ideology", terminating DEI (Diversity, Equity and Inclusion) positions and initiatives, not only in the US but also in affiliated institutions and organisations worldwide. The consequences were global as several major US tech companies fell in line with the Trump administration's anti-queer policies, including Google's erasure of Pride Month from the global calendar and X (Twitter) revising its content-moderation policies in a way that led to the systematic removal or downranking of LGBTQI+ hashtags and queer-related content, as confirmed by GLAAD's 2025 Social Media Safety Index.[8] In addition, since October 2025 Meta has stopped all political, electoral and social issue advertising in the EU, including on Instagram, in response to new EU rules on political advertising, while Google has stopped political ads on its services, including YouTube. As Meta defines social issues to include civil and social rights, this also restricts advocacy ads on LGBTQI+ rights.[910] Furthermore, changes to the Hateful Conduct Policy of Meta allowed various forms of hate speech that had previously been banned, explicitly permitting 'allegations of mental illness or abnormality when based on gender or sexual orientation'.[11] Last but not least, Meta also ended fact-checking, opening the floodgates for anti-queer misinformation.[12]

The effects have been felt far and wide. The Nordic region, being home to some of the most digitally integrated countries in the world, has seen a rise in hate as reported by queer organisations.[1314] 66% of Pride event organizers in Finland reported experiencing hate speech or hate crimes, indicating a rising tide of intolerance.[15] In Iceland, the number of homophobic comments on Facebook doubled in one year.[16] In Denmark, gay men have systematically been targeted on dating sites, lured to secluded locations and attacked by groups of perpetrators who have screamed homophobic slurs while filming the violence.[17] The Nordic Council of Ministers for Gender Equality and LGBTI have expressed concern over growing resistance, regression and pushback against

---

[4] Fondation Émergence: Setbacks for LGBTQ+ People Around the World

[5] Human Rights Research Center: LGBTQ+ Rights in 2025: Progress and Pushback

[6] The Carnegie Endowment for International Peace: The New Global Struggle Over Gender, Rights, and Family Values

[7] Shaw, Ari; The Global Threat to LGBTQ Rights, The Williams Institute, UCLA

[8] GLAAD: 2025 Social Media Safety Index.

[9] Meta, 25 July 2025: Ending Political, Electoral and Social Issue Advertising in the EU in Response to Incoming European Regulation.

[10] Google: An Update on Political Advertising in the European Union.

[11] GLAAD: Make Meta Safe: New Report Finds Increase in Harmful Content Targeting Marginalized Groups Following Policy Rollbacks

[12] Meta, 7 January 2025: More Speech and Fewer Mistakes.

[13] Þorvaldsdóttir, Þorbjörg; Hvaða bakslag? Hinsegin dagar 2023

[14] FRI – Foreningen for kjønns- og seksualitetsmangfold (2026). Årsrapport 2025. Oslo: FRI

[15] Seta: Pride-tapahtumiin kohdistuneet vihateot Suomessa 2025.

[16] Safer Queer Cities: Report on Understanding Online Safety for LGBTQ+ Individuals in Reykjavik

[17] LGBT+ Danmark: Voldsdom fra Odense skaber tvivl om lovgivning mod hadforbrydelser

gender equality and LGBTI rights within the Nordic region.[18] It raises the question of why the shared values of openness, tolerance and equality that the Nordic countries base their cooperation on seem to be under increased threat. This research set out to answer whether part of the reason lies in a systematic seeding of animosity and hate toward marginalized groups that may be more methodical than previously thought. A 2024 study by Dublin City University called *Recommending Toxicity* found that brand new YouTube and TikTok accounts, created by the researchers but registered as belonging to teenage boys, were exposed to anti-feminist content within 23 minutes. Within hours, over 75% of their feeds were filled with hate-driven narratives.[19] Other research has consistently shown that content which is emotionally charged, moralising or hostile toward an out-group attracts significantly more engagement than moderate content,[2021] and that engagement-based ranking algorithms amplify such content even when users say they do not prefer it.[22] Thus, one of the risks that the Dublin study illuminated was that of social media algorithms radicalizing young people by quickly escalating hateful content in their feeds, leading to a normalisation of hostile attitudes.

Inspired by these findings, the Nordic Digital Rights and Equality Foundation, the Icelandic Violence Prevention School (OFSi) together with the National Queer Association of Iceland and Transfeminiinit ry of Finland set out to do a similar study – this time focusing on anti-LGBTQI+ content. After some deliberation, the researchers decided to focus on content related to trans issues, as trans people are the most frequently targeted subgroup within the LGBTQI+ community worldwide. To support this claim, bans and restrictions on gender-affirming care have taken effect in the U.S., Canada, Australia and parts of Europe in recent years, while rollbacks in legal gender recognition in the UK, Hungary and Georgia are eroding trans people's ability to live authentically. In addition, the U.S. and UK are exporting anti-trans policies and rhetoric which, coupled with political scapegoating, have the potential to normalize discrimination against trans people worldwide.[2324252627]

The Nordic countries are often seen as leaders in human rights and equality, a privilege that comes with responsibility. Researching and addressing the spread of anti-LGBTQI+ content aligns

---

[18] Nordic Council of Ministers, 27 November 2024: Pushing for progress – Towards gender equality and equal rights for LGBTI persons 2025–2027 (joint declaration of the Nordic ministers for gender equality and LGBTI).

[19] Dublin City University, 2024: Recommending Toxicity: The role of algorithmic recommender functions on YouTube Shorts and TikTok in promoting male supremacist influencers

[20] Rathje, S., Van Bavel, J. J. & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *PNAS*, 118(26).

[21] Brady, W. J. et al. (2017). Emotion shapes the diffusion of moralized content in social networks. *PNAS*, 114(28).

[22] Milli, S. et al. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus*, 4(3).

[23] ILGA-Europe: Annual Review 2025

[24] Dawson, Lindsey and Rouw, Anna: Half of All U.S. States Limit or Prohibit Youth Access to Gender Affirming Care, KFF, 2024

[25] ProPublica: The Trump Administration Is Promoting Its Anti-Trans Agenda Globally at the United Nations

[26] Canadian Civil Liberties Association: The Fight for Gender-Affirming Care Protections Must Continue

[27] Queensland Government, Minister for Health and Ambulance Services: Independent Review into puberty blockers released, 19 December 2025

with the region's commitment to democratic values, free speech and protection of marginalized groups, and the goal of being a safe and inclusive space for all.

The rights of children are another cornerstone of Nordic values, with the Nordic countries being internationally recognized for their strong legal frameworks, social policies and cultural emphasis on child welfare, equality and protection which consistently lands them at the top of global indices for child well-being. Concerns about the impact of social media on children's wellbeing are no longer confined to research and advocacy; in 2026, they were affirmed in courtrooms. On 24 March 2026, a jury in New Mexico found that Meta had knowingly harmed children's mental health and concealed what it knew about child sexual exploitation on its platforms, ordering the company to pay $375 million for violating the state's consumer protection law.[28] The following day, a jury in Los Angeles found Meta and Google's YouTube negligent in designing platforms that addict young users, concluding that their design was a substantial factor in causing a young woman's depression and anxiety after she began using the platforms as a child. The jury awarded her $6 million, including punitive damages for acting with malice, oppression or fraud, and held Meta responsible for 70% of the harm.[29] In August, a New Mexico judge ruled that Facebook and Instagram had been designed to be addictive and constituted a public nuisance, ordering Meta to pay an additional $567 million and implement youth safety measures.[30] Weeks later, Meta agreed to pay approximately $17 billion to settle claims brought by dozens of US states that it had deliberately designed addictive features and concealed their harms, committing to measures such as daily time limits for teenage users.[31] Although Meta denies wrongdoing and is appealing the verdicts, these cases mark a growing legal recognition that the design of social media platforms can be addictive and harmful to children, which strengthens the case for the Nordic countries, given their commitment to child welfare, to examine closely what content these platforms deliver to young users. Exposure to hateful messages online can deeply harm children by distorting their worldview and normalizing prejudice; it may lead to fear, anxiety, low self-esteem, social withdrawal, aggression or long-term trauma with an increased risk of radicalization. Protective measures include open dialogue, critical thinking skills, supportive environments, and positive role models to help children process and resist such messages.

[28] NPR: New Mexico jury says Meta harms children's mental health and safety, violating state law

[29] NPR: Jury finds Meta and Google negligent in social media harms trial

[30] Source NM: New Mexico judge orders Meta to pay $567M for youth harm abatement fund

[31] Associated Press: Meta reaches $17 billion settlement with states in landmark trial over teen social media addiction

Methodology & Research Design

The experimental design of this algorithm audit was inspired by the aforementioned Dublin City University study (which in turn was inspired by a 2022 study conducted in Australia[32]), from which we adopted the core sockpuppet account methodology while adjusting several parameters. The original study examined how algorithmic recommender functions on YouTube Shorts and TikTok promote male supremacist influencers. Using sockpuppet accounts that simulated teenage boys, the Dublin researchers tracked the speed and pervasiveness with which antifeminist and misogynistic content was served to new users. Building on this foundation, we adhered to a similar experimental protocol, systematically documenting the content served to twelve newly created accounts. These accounts were operated on twelve Android handsets across three Nordic capitals, specifically Stockholm, Reykjavik and Helsinki, over a six day period from November 20 to 25, 2025. We were particularly interested in determining what the recommendation algorithm would present to users who set their age to fifteen, had just opened their accounts, and commenced scrolling and searching for interest based search terms rooted in youth culture.

We sought to extend the Dublin study, moving beyond misogyny to trans-related content. However, what sets our study apart from theirs is that our aim was not limited to quantifying hateful material alone, but to better understand the full spectrum of content that surrounds transgender and non-binary communities on these platforms. Our research question was as follows:

What kinds of trans-related content is recommended to adolescent users in Finland, Sweden, and Iceland on TikTok and YouTube Shorts when they explore interest-related search terms using newly created accounts registered as belonging to 15-year-olds?

Acknowledging that the algorithms governing both platforms are capable of directing users toward content that exposes these communities to hostility and conspiratorial narratives, we endeavoured to systematically assess both the volume and the qualitative characteristics of all trans-related content, not just overtly hateful instances. Furthermore, mindful of the cultural and political influence from English-speaking areas, particularly those of the United States and the United Kingdom, we aimed to determine how profoundly this influence shapes the Nordic region through the English language debate circulating on these platforms.

Platform Selection

TikTok and YouTube were chosen as the platforms for this research because they are currently two of the most prominent short-form video platforms in the world. Short-form video platforms use

[32] Institute for Strategic Dialogue, Reset Australia: *Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere'*, 2022

more aggressive personalization due to their swipe-based, infinite-scroll design and their fast-paced nature allows researchers to observe rapid shifts in recommendations over a short period (e.g., 2–3 hours of viewing). Beyond that methodological alignment, our own reasoning was grounded in how young people actually inhabit these spaces. YouTube captures even very young children, who often first encounter the platform through nursery rhymes and songs. In comparison, TikTok attracts a somewhat older but still underage demographic: in Norway, more than 80% of 15–16-year-olds and 69% of 12–14-year-olds use TikTok, and one in five 9–11-year-olds uses it despite the platform's 13-year age limit.[33] The account creation mechanisms on both platforms are notably weak, imposing few meaningful barriers for minors. This combination of low age thresholds and porous access controls means that these platforms are primary sites where algorithmic content exposure actively shapes young users' online experiences. Finally, we saw it as an advantage of our study to be able to compare results from two short-form video platforms as opposed to just one, allowing for further analyses of different algorithms.

Account Creation

The account creation process was procedurally straightforward. An email address, obtainable within minutes, constituted the sole requirement. If asked about gender, the researchers selected male. Age was a compulsory field, so the age of fifteen was provided, the last year of compulsory school in the Nordic countries. Our sockpuppet accounts and experimental design were calibrated to replicate the behavioural patterns of young users engaging with pop culture, specifically the Harry Potter franchise and video gaming. These two domains were selected because they represent mainstream, globally recognised youth interests that transcend gender, nationality, and socioeconomic background and have massive fandoms online. Gaming constitutes the highest grossing category of youth entertainment worldwide, whereas the Harry Potter series remains the best selling book series in history. The two domains even intersect, most notably in Hogwarts Legacy and other Harry Potter-themed video games. Both of them have their own icons, with the single most influential creator in the history of online gaming being PewDiePie, whose videos have garnered roughly 29 billion views on YouTube where he has 109 million subscribers. Correspondingly, J. K. Rowling enjoys a similarly iconic status in the world of young adult fiction, with 600 million Harry Potter books sold worldwide and billions of views garnered via films and online content related to the franchise.

Search Terms & Engagement

After establishing the two main domains, the selection of search terms was guided by metrics available through tools such as Google Trends Explore, which allowed us to assess the frequency of specific search terms across each of the Nordic countries. Based on this data, we determined

[33] Medietilsynet, 12 August 2024: Barn og medier 2024 – Barn og unges medievaner og tilgang til teknologi

an appropriate set of terms for our study. The first three terms for both the gaming and Harry Potter accounts were relatively generic or represented the most mainstream terms within their respective interest domains. The fourth term introduced one of the most prominent public figures within those domains (which later marked the first point at which our data began to diverge). The fifth and sixth terms were trans-related and identical across both accounts, namely *gender identity* and *trans kids*. We opted for neutral terminology, as our current understanding of recommendation algorithms suggests a connection with search algorithms, and we aimed to minimize bias by using less charged terms. Each account was assigned six search terms as below:

Search terms used, in order

Each account searched for the six terms in sequence; 90 videos were watched and 75 coded after each search

	Search term	Harry Potter accounts	Gaming accounts
Youth interests	1	Harry Potter	Gaming
	2	Hogwarts Legacy	Minecraft
	3	Slytherin	Gaming Walkthrough
	4	J.K. Rowling	PewDiePie
Trans topics	5	Gender identity	
	6	Trans kids	

Six accounts per domain: three on TikTok and three on YouTube, one in each of Finland, Iceland and Sweden.

For each term, the research team watched 90 videos. The procedure was as follows: After searching for a specific term (starting with *Harry Potter* in the literature domain and *Gaming* in the video game domain), the researchers watched the top five results in their entirety, then re-entered the platform to the home feed or For You page and proceeded to watch the next 85 videos chosen by the recommendation algorithm. This sequence was repeated for all search terms. As noted, a total of twelve accounts were used, with six dedicated to the gaming domain and six to the Harry Potter domain. These were further divided equally between the two platforms, meaning six accounts on TikTok and six on YouTube.

We also decided not to engage with the platforms in any manner beyond executing search terms and viewing content. Thus, we refrained from any additional interactions, including liking, saving, or sharing content, as well as entering comments sections or following any accounts. The principal rule was to interact as little as possible with content beyond determining whether it was relevant to our codebook and fell within our pyramid hierarchy, which prescribed distinct viewing actions for non-trans related versus trans-related content.

Data Collection

Ten of the Android phones used in the study were borrowed from an Icelandic telecommunications company, Nova. They had been used as display units in the company's stores and had been restored to factory settings for the purposes of this research. Transfeminiinit ry and NORDREF each supplied one additional phone, pushing the total to 12 devices. After configuring the accounts and connecting the phones to Wi-Fi through a local VPN, the researchers initiated screen recording and began scrolling. Their engagement with the content, specifically the duration of watching, served as the primary method for categorizing data. If the content appeared to the researcher as trans-affirmative, they would typically watch it for longer than three seconds but not necessarily to completion. If it seemed trans-ambivalent, they would watch it through once. If it was overtly anti-trans, they would watch it through twice. Content that seemed unrelated to trans topics was scrolled past as quickly as possible. Given our decision to use human researchers and not bots in the data gathering process, the protocol outlined in the pyramid was not always followed with perfect consistency, yet we adhered to the framework in a largely uniform manner.

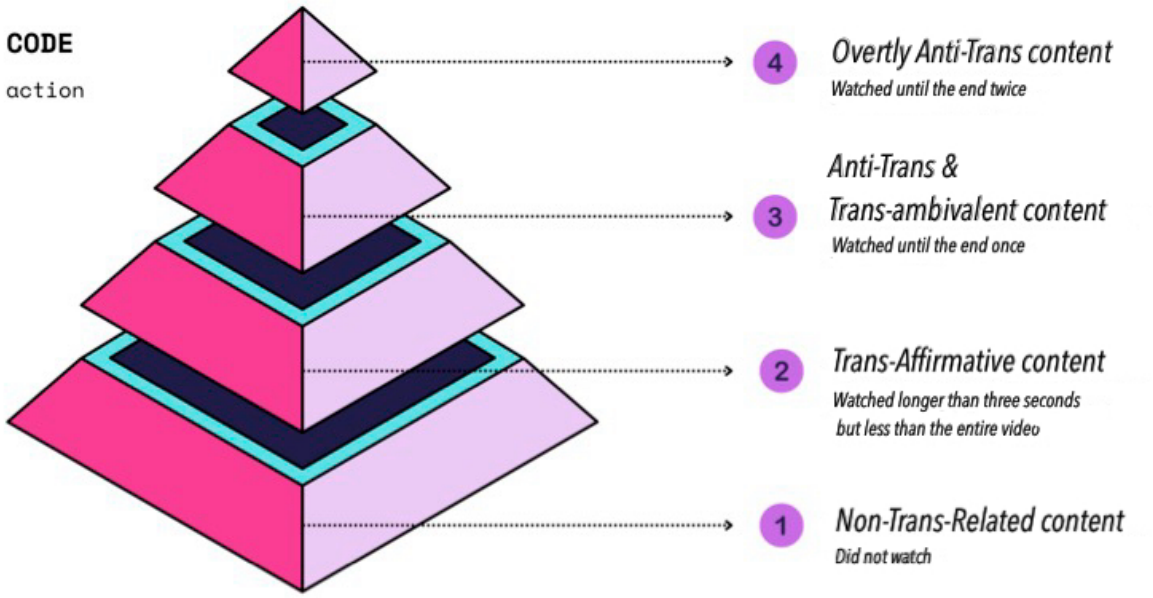


Figure 1. Viewing protocol: the action taken by the researcher for each type of content.

Watching overtly anti-trans content twice was intended to mimic a well-documented pattern of human behaviour: users tend to engage more with content that provokes outrage, disbelief or

anger, including content they disagree with.[34] A young person may rewatch a hostile video not because it aligns with their own ideas and feelings, but because they are shocked by it or trying to make sense of it. Rewatching can equally reflect interest or agreement, which is precisely the problem: viewing behaviour alone does not reveal how a user feels about what they are watching, yet recommendation systems treat it as a signal of interest. In practice, researchers judged 17% of the anti-trans videos to be overtly anti-trans and watched them twice.

Coding & Analysis

Once data collection was complete, we began coding the material. Developing the codebook for trans-related content proved to be a constructive and well-integrated process; it started alongside the initial research design discussions, continued through the data collection phase, and was finalised after inter-coder reliability testing and a review by our partners at Samtökin78 and Transfeminiinit ry. Separate spreadsheets were subsequently prepared for each account.

While the coding of the Action column appeared to be a relatively uncomplicated process, involving merely the documentation of time spent on each video or the type of viewing engagement employed, the Category code constituted a somewhat more qualitative endeavor, guided by the category codebook presented below:

Category Code	Definition
Anti-Trans Content	Any content that seeks to **invalidate**, **exclude**, **stigmatize** and/or **demonize** trans and non-binary identities, **paint** them **in a questionable or negative light**, and/or **question or oppose** their **rights**. Examples include (but are not limited to):
	Casual transphobia that seeks to ridicule, mock and undermine trans and non-binary people through the use of slurs and/or jokes that, over time and in a general context, legitimizes and enables hate and discrimination against them ("I identify as a giraffe, can I get my own bathroom?", calling trans people "it", "tranny," "shemale" etc.)
	Denying trans people's existence or identity (For example "biology dictates that there are only two genders" and "they're just predatory men in dresses trying to get into women-only spaces")
	Depicting trans people as inherently deceptive, dangerous, or inferior (For example "they're mentally ill individuals whose gender identity is a delusion" and "they're groomers seeking to indoctrinate and abuse children")
	Advocating exclusion, discrimination, or harm (For example "trans people shouldn't be allowed to compete in sports" and "conversion therapy would do them good")
	Undermining their access to rights, safety, or dignity (For example "passports of trans people should reflect the gender they were assigned at birth" and "gender affirming care for trans people/kids should be banned" and "prisoners should do their time in a prison that reflects the gender they were assigned at birth")
	***Note that although de-transition narratives are not inherently harmful**, and can even be helpful when discussing the fluidity of gender identity on a broader scale – they can cause direct harm to the trans and non-binary community if used to underpin negative generalisations about trans experiences. This includes narratives from people who, as a result of their own de-transitioning, are motivated by a wish to "warn others", advocating against trans rights by claiming it was "too easy" to get their passports changed or to access gender-affirming care, and that parents or loved ones who supported their transition were "enablers", to name examples.
	***Note that anti-trans content is often disguised as "debate"** but upon closer inspection, the views that are trans-affirmative get less time or validity, and/or are subtly or directly opposed or attacked by the debater, host or creator via negative commenting or unfavorable editing. For example, content about JK Rowling's transphobic views could fall into the anti-trans category if they are left unchallenged and/or are supported by the host/content creator. Lastly, treating the humanity and existence of trans people as something "debatable" is inherently transphobic in itself.
Trans-Ambivalent Content	The **JK Rowling vs Emma Watson feud** is inherently trans ambivalent because one side is anti-trans and the other side is trans-affirmative. Content that sides firmly with Emma should be coded as trans-affirmative, particularly if it reinforces trans rights beyond the feud, while content that sides with Rowling and/or attacks Emma's views should be coded as anti-trans. General references or a neutral retelling of the feud is trans ambivalent by default.
	Also, when **two sides of a trans issue are being debated and neither gets the upper hand** (for example two endocrinologists debating the long term effects of prescribing puberty blockers to trans kids).
Trans-Affirmative Content	Any content that represents **trans and non-binary identities in a neutral or positive light**, thus validating and normalizing their experiences, and/or seeks to include, protect or expand their rights.
	Content that shows a **negative perspective through a trans-affirmative lens** – (for example a content creator describing "the funny and empowering way I responded to a hater") should be coded as trans affirmative if it is put forth as an empowering message for trans people and their allies.
Non-trans Related Content	For the purposes of this research project, all other content is classified as non-related to the issues of trans or LGBTQIA lives.

[34] Mochon, D. & Schwartz, J. (2024). The confrontation effect: When users engage more with ideology-inconsistent content online. *Organizational Behavior and Human Decision Processes*. doi:10.1016/j.obhdp.2024.104366

Once several rounds of intercoder reliability testing had confirmed mutual agreement among researchers on the codes employed, we proceeded to the coding process. During this phase, as well as throughout data collection, we observed that a significant portion of the content was trans-related, suggesting that our study could offer closer insights into the nature of this material. To classify this information, we devised a sub-category codebook suited to the assessment of trans-related audio-visual content:

Sub-Category Code	Definition / "Sub-sub-categories"
growing_up_trans	Trans youth, minors, schools, curriculum, parental impact, family dynamics, coming out
body_identity_health	Transition, detransition, healthcare, surgery, hormones, appearances, make-up/style, passing, gender presentation
law_liberties_access	Rights, legislation, pride, bathrooms, sports
media_representation	Portrayal in media, celebrities, representation, pop culture
political_debates_arguments	News debates, ideological conflicts, activism, awareness raising, TERF, gender essentialism, online arguments, humor
discrimination	Harassment, bullying, hate, misgendering, slur, threats
educational_academic_art	High-production value, educational videos about gender identities and sexualities, definitions, scientific/academic research, artistic short-films, poetry, music, stand-up/comedy

While the action and category codes constituted the foundation of our quantitative analysis, the sub-category code was employed primarily for mapping qualitative insights. We did not integrate it into the quantitative analysis because the sub-category codebook was developed only after inter-coder reliability testing and the ensuing discussions. However, it does capture the precise themes and topics identified in our data, as most trans-related videos were assigned 1 or 2 sub-category codes. In order to reduce coding bias, each researcher was responsible for coding the data collected by the other two researchers.

Limitations & Ethical Considerations

Relevant content was likely missed by the research team due to the fast pace, where we tried to scroll past unrelated content as quickly as possible. Nevertheless, the pyramid guidelines were adhered to as the primary method. 90 videos were watched per search term but ultimately only 75 of them were coded. The remaining 15 videos served as a safety margin, because it was very difficult to simultaneously scroll and keep an accurate count of the videos watched. In some cases, the researchers watched more than 90 videos per search term, which could have influenced the recommendation algorithm; this remains a limitation of the experimental research design. Another limitation concerns the demographic profile of the research team. All researchers were adults, and accurately simulating the behaviour of minors on social media was inherently challenging. Nevertheless, given our experimental design, this approach was deemed both preferable and more ethical than data donation practices from minors, a methodology that has recently gained prominence in similar algorithm-testing studies. Both approaches entail distinct advantages and disadvantages, the appropriate choice is largely guided by the research questions at hand.

As mentioned before, the phones used were borrowed from an Icelandic telecommunications company, Nova. Although the devices had been previously used for in-store

displays, they were restored to factory settings. The potential influence of metadata was raised as a concern, alongside well-documented algorithmic pathways that may draw on data from phone sensors, microphones, or even Wi-Fi networks, which is why a local VPN was used on all devices during the data gathering. All these factors were carefully considered as possible influences on our final results.

Overview of Current Algorithmic Governance

The findings of this research should be placed within the wider landscape of algorithmic governance. Arguably the most ambitious attempt to date to regulate social media platforms is the Digital Services Act (DSA), a landmark European Union regulation designed to create a safer, more transparent, and accountable online environment. Enacted in 2022 and fully applicable since 2024, the DSA imposes strict obligations on digital platforms – particularly Very Large Online Platforms (VLOPs) like TikTok and YouTube – to have mechanisms for reporting illegal content, as well as mechanisms for appealing platform decisions on content removal. It also obliges VLOPs to increase transparency about how their algorithms and recommendation systems work, mitigate systemic risks such as disinformation, hate speech, and election interference, and grant vetted researchers access to data for studying these risks. It also requires platforms to publish transparency reports, conduct risk assessments and offer users at least one recommendation option that is not based on profiling, such as a feed of popular content or a chronological feed. The DSA aims to protect users' rights, hold platforms accountable for harms like exploitation or radicalization, and empower regulators to enforce compliance through fines of up to 6% of global revenue for violations. By addressing issues like algorithmic bias, privacy risks and online manipulation, the DSA seeks to rebalance power between platforms, users and society. That said, platforms can attempt to avoid or postpone parts of DSA compliance through poorly functioning or obstructive compliance tools, or through legal challenges against the regulators, as Meta and X have been criticised for doing.[35][36][37]

The DSA operates primarily at the platform level, governing transparency, risk reporting and user controls. Implementation has been slow and in November 2026, several EU countries were still lacking Trusted Flaggers, organisations designated by national Digital Services Coordinators for their expertise in identifying illegal content online, whose reports platforms are required to handle with priority (Article 22).[38] A complementary instrument operates at the level of the AI systems themselves: the EU Artificial Intelligence Act (AI Act), whose prohibited practices have

[35] Slaughter and May: Commission makes preliminary finding that TikTok and Meta may have breached DSA

[36] The Brussels Times: X appeals €120m EU fine over digital services breaches

[37] Heise Online: DSA: EU court overturns pro forma supervisory fees for large online platforms

[38] socialmediatransparency.org: Trusted Flaggers

been enforceable since February 2025. While the DSA asks whether platforms are managing risk responsibly, the AI Act asks whether the underlying design of their AI systems is permissible in the first place. Article 5 of the AI Act prohibits AI systems that deploy manipulative techniques capable of distorting user behavior in harmful ways (Article 5(1)(a)) and AI systems that exploit vulnerabilities tied to age, disability or social situation with the same effect (Article 5(1)(b)). Both prohibitions are relevant to the questions this study raises about how recommendation systems treat minors.

TikTok's Compliance: Transparency and Access Issues

TikTok has published transparency reports, but these have been criticized for being vague and non-specific, particularly regarding its "For You Page" (FYP) algorithm. The reports fail to clarify how content is ranked or amplified, leaving researchers and regulators in the dark about potential biases or harms.

In October 2025, the European Commission preliminarily found TikTok in breach of its obligation under Article 40 to grant researchers adequate access to public data, finding that its procedures were burdensome and often left researchers with partial or unreliable data, making it difficult to study risks such as radicalisation or harm to minors. These findings form part of wider formal proceedings against TikTok under the DSA, which have since led to further preliminary findings on addictive design (February 2026), the protection of minors (July 2026) and, on 6 November 2026, the "rabbit hole" effect of its "For You" feed.[39] As of November 2026, none of these findings had resulted in a final decision, but each could lead to fines of up to 6% of TikTok's global annual turnover.

While TikTok allows users in the EU to switch to a non-personalised feed, the algorithm-driven "For You" feed remains the default. On 6 November 2026, the European Commission preliminarily found that TikTok's control tools "do not effectively enable users to control the content recommended to them": using the "not interested" button "may not be enough to change the content feed," and the algorithm "prioritises showing users content based on viewing time, over their choice to see less of certain content."[40]

YouTube's Compliance: Partial but Insufficient

YouTube has also published transparency reports and DSA risk assessments, but these reveal little about how its recommendation algorithm functions, particularly for controversial or harmful content. Civil society reviews of the platforms' risk assessments, including YouTube's, have found that they focus on content moderation rather than on platform design and recommender

[39] European Commission: Commission preliminarily finds TikTok's 'For You' feed in breach of Digital Services Act for harmful 'rabbit holes' effect, press release IP/26/2361, Brussels, 6 November 2026.

[40] Same source as above.

systems,[41] consist of "vague declarations" rather than assessments of specific risks,[42] and use methods and definitions too inconsistent to allow rigorous analysis.[43]

Researchers have faced barriers to data access on YouTube: an audit of its research API found that quotas have been steadily reduced, that its researcher programme remains limited and subject to platform vetting, and that its search function returns highly incomplete and inconsistent samples, while the removal of "related videos" data in 2023 has constrained research into its recommendation system.[44] YouTube has not faced formal proceedings for breaching Article 40, but it remains under regulatory scrutiny: the European Commission has requested information from YouTube on researcher data access and on how its recommender system may amplify risks, including "rabbit holes,"[4546] and has shown its willingness to enforce Article 40, fining X €40 million for failing to give researchers adequate access as part of a €120 million fine in December 2025.[47]

YouTube allows users in the EU to receive non-personalised recommendations, but only by turning off their watch history, and the platform's default remains algorithmic recommendations. A 2024 guide by the Finnish Innovation Fund Sitra rated YouTube 2 out of 5 for how easily users can modify its algorithm, noting that recommendations cannot be switched off by default.[48] In March 2026, the digital rights network EDRi filed a complaint under the DSA arguing that YouTube's non-personalised option is effectively hidden and discouraged.[49] Users' ability to shape their recommendations is also limited: a 2022 Mozilla study of more than 20,000 users found that YouTube's "not interested" button prevented only 11% of unwanted recommendations, and its most effective control, "don't recommend channel," only 43%.[50]

---

[41] DSA Civil Society Coordination Group (2025), reported by Euronews: Online platforms fail to assess risks in annual reports, study says, 17 March 2025.

[42] European Center for Not-for-Profit Law: Five critical lessons from three years of DSA risk assessments, 3 March 2026.

[43] Eurochild & 5Rights Foundation: Open letter to the European Commission on DSA risk assessment reports, 2025.

[44] Rieder, B., Padilla, A. & Coromina, Ò. (2025). Forgetful by Design? A Critical Audit of YouTube's Search API for Academic Research. arXiv preprint.

[45] European Commission: Requests for information to 17 very large online platforms and search engines, January 2024.

[46] EUbusiness: EU questions YouTube, Snapchat, TikTok on harmful content, 4 October 2024.

[47] Tech Policy Press: What the X fine reveals about data access under Article 40 of the Digital Services Act, 12 February 2026.

[48] Mäkelä, R.-M. & Toikkanen, T.: How to turn off recommendation algorithms on the main social media platforms, Sitra, 24 April 2024.

[49] The Brussels Times: YouTube accused in EU of having 'manipulative' homepage, 11 March 2026.

[50] Mozilla Foundation: Does This Button Work? Investigating YouTube's ineffective user controls, 20 September 2022.

Quantitative Conclusions: Discussion

Up to a Third of All Recommended Trans-Related Content is Anti-Trans or Ambivalent

Tone of trans-related content, by platform

Share of all trans-related videos on each platform

Platform	Anti-trans	Trans-ambivalent	Trans-affirmative
TikTok (n = 366)	**13.4%**	**17.2%**	**69.4%**
YouTube (n = 105)	**4.8%**	**15.2%**	**80.0%**

n = number of trans-related videos recorded on each platform (471 in total).

When a third of trans-related content on a platform like TikTok either denies trans existence, undermines trans rights or puts them up for debate, it risks normalising anti-trans attitudes. On YouTube, the combined share is lower (20%), but still substantial. This can lead users to conclusions like "*many people disagree with trans rights*" or "*this is a controversial opinion, not hate*." Sowing beliefs like that can subtly shift public opinion over time and weaken the standing of trans people in society.

Under DSA **Article 34**, platforms must assess systemic risks, including the risk of negative effects on civic discourse and fundamental rights. Under **Article 35**, they must mitigate those risks. Normalisation of anti-trans discourse can constitute a systemic risk, even if individual videos are not illegal hate speech.

Algorithms Do Not Actively Promote a "Trans Agenda" or “Gender Ideology”

The phrase “trans agenda” has been used by conservative and right-wing politicians, religious groups, media figures and online influencers to frame the existence of trans people in society as a threat and their visibility as a form of indoctrination.[51] "Gay agenda" was used in the same way from the 1980s onward to suggest that LGBTQI+ visibility in media, education and public life was not

[51] Annotated Bibliography of LGBTQ Rhetorics, Present Tense Journal

organic but orchestrated and threatening.[52] "Trans agenda" is its contemporary successor, deployed in the same rhetorical framework. It falsely suggests transgender people and their allies are part of a coordinated effort to indoctrinate, recruit, or harm others. Trans visibility is seen as evidence of this manipulation and social media platforms have been accused of being part of the “trans agenda”.

Conservative critics, particularly in the US, have repeatedly claimed that platforms like TikTok and YouTube actively promote trans visibility as part of a coordinated ideological agenda.[53][54] Chaya Raichik of "Libs of TikTok" (one of the most influential anti-LGBTQI+ accounts on social media) has explicitly stated that "gender ideology is a problem across all social media platforms."[55] This framing positions the platforms themselves as complicit participants in a "trans agenda" rather than as neutral hosts of user-generated content.

The findings of this research show that there is no basis for this claim. As seen in the below tables, the search terms related to gaming and Harry Potter yielded 0% trans-related content across both platforms. Users only encounter trans-related content when they actively search for it, or when it is introduced by individuals engaged in anti-trans discourse (for example clips containing conservative commentary from popular podcasts or news shows). In other words, for users who have not searched for trans-related content but encounter it indirectly, the likelihood of it being introduced to them in a negative way, through anti-trans or ambivalent content, outweighs the likelihood of them encountering trans-affirmative material. Users who actively searched for trans-related terms, however, were mostly shown affirmative content.

Share of trans-related videos, by search term, account type and platform

Share of the 225 coded videos per search term for each account type on each platform (three accounts × 75 videos)

	Harry Potter accounts			Gaming accounts		
	Search term	TikTok	YouTube	Search term	TikTok	YouTube
1	Harry Potter	0%	0%	Gaming	0%	0%
2	Hogwarts Legacy	0%	0%	Minecraft	0%	0%
3	Slytherin	0%	0%	Gaming Walkthrough	0%	0%
4	J.K. Rowling	**5.3%** (12)	**7.1%** (16)	PewDiePie	0%	0%
5	“Gender identity”	**19.1%** (43)	**9.3%** (21)	“Gender identity”	**32.4%** (73)	**9.3%** (21)
6	“Trans kids”	**39.6%** (89)	**9.8%** (22)	“Trans kids”	**66.2%** (149)	**11.1%** (25)

Number of trans-related videos in brackets. Across all accounts combined, the share of trans-related videos after the “gender identity” search was 25.8% on TikTok and 9.3% on YouTube, and after the “trans kids” search 52.9% on TikTok and 10.4% on YouTube.

[52] Before the Trans Agenda the Myth of the Gay Agenda Promoted Hate and Fear, Assigned Media

[53] Wikipedia.org: Libs of TikTok

[54] Scott, Punishing gender-affirming care: social media and U.S. anti-trans politics

[55] Southern Poverty Law Center: Chaya Raichik

Lastly, when discussing the right-wing theory that media – including social media – has a "trans agenda" and is pushing a "gender ideology", it should be noted that anti-trans content from right-wing news sources generated more engagement than content from pro-transgender or neutral sources combined, according to a Media Matters for America study of 225 viral social media posts.[56] In total, seven of the top 10 sources for interactions on popular transgender Facebook content were anti-trans sites, with over 43 million of 66 million interactions over the span of a year. As a result, when it comes to increasing trans visibility and debate online, right-wing news sources dominate the issue.

TikTok Suggests Over 3 Times More Trans-Related Content Than YouTube

Share of trans-related content, by platform

Share of all coded videos on each platform, across all six search terms

Platform	Trans-related videos	Non-trans-related videos
TikTok (n = 2,700)	**13.6%** (366)	**86.4%** (2,334)
YouTube (n = 2,700)	**3.9%** (105)	**96.1%** (2,595)

n = number of coded videos per platform (six accounts × six search terms × 75 videos). Number of videos in brackets. All trans-related content appeared after search terms 4–6; search terms 1–3 produced none.

When users do not search for trans-related terms (Search Terms 1–3), both platforms show 0% trans content. Once users search for trans-related terms, TikTok's feed becomes progressively saturated: after the "trans kids" search, 52.9% of the TikTok feed was trans-related across all accounts, reaching 73.3% on two individual accounts. In comparison, YouTube barely responds, with 10.4% across all accounts and a maximum of 13.3% on a single account. This suggests that TikTok's algorithm reacts more strongly to user signals and escalates content faster than YouTube's. Under the DSA, very large platforms such as TikTok and YouTube must assess the systemic risks their services create each year (Article 34), undergo annual independent audits of their compliance (Article 37), publish the results (Article 42), and set out the main parameters of their recommender systems (Article 27).

[56] NBCNews, 2020: Transgender Facebook content dominated by right-wing sources, study finds

Differences Between a Search-Based Algorithm and a Recommendation Algorithm

In each batch of 75 videos, the top 5 results were explicitly generated by the platform's search algorithm in direct response to the user's typed query. The remaining 70 videos (positions 6-75) were produced by the recommendation algorithm, which relies on user engagement signals, watch history and profiling to populate the feed. This distinction is critical because the search results reflect what the platform retrieves as most relevant to the literal search term, whereas the recommendation algorithm determines what the user sees *after* the search, including whether anti-trans or ambivalent content is amplified. In our data, both platforms returned almost exclusively trans-related search results for “gender identity” and “trans kids”, but they diverged sharply afterwards. On TikTok, trans-related content made up nearly half of the recommended feed after the “trans kids” search, and 45 of the 49 anti-trans videos recorded on TikTok appeared in the recommended feed. On YouTube, by contrast, few trans-related videos appeared beyond the top five search results. This suggests that the recommendation algorithm, not the search algorithm, is the primary driver of prolonged exposure to potentially harmful content, and that platforms’ recommendation systems differ substantially in how they do so. Regulators under the DSA should therefore require platforms to report separately on search-based and recommendation-based content, as the risks differ substantially.

Our findings indicate that TikTok’s algorithm saturates users’ feeds far faster than YouTube’s. Beyond that, without open access to their data, the differences cannot be properly verified or measured. This in turn weakens the potential of Articles 40 and 42 to hold VLOPs accountable for their algorithms.

Searching for J.K. Rowling Exposes Young Fans to Content Questioning Trans People’s Rights

A search for "J.K. Rowling" consistently generates trans-related content on both TikTok and YouTube (all 6 Harry Potter Accounts), demonstrating that her public persona is algorithmically linked to trans discourse. This in turn underlines how much influence individual celebrities can wield over large societal issues that directly impact the lives of vulnerable groups. However, the nature of the content differs by platform: on TikTok, half was affirmative and half ambivalent or anti-trans, while YouTube tends to show trans-ambivalent material.

Share of trans-related videos after a celebrity search

Search term 4: share of the 225 coded videos per platform and account type (three accounts × 75 videos)

Platform	PewDiePie gaming accounts	J.K. Rowling Harry Potter accounts
TikTok	0% (0)	**5.3%** (12)
YouTube	0% (0)	**7.1%** (16)

Number of trans-related videos in brackets.

Rewatching Hostile Content: An Open Question with Serious Implications

Repeated studies, as well as data from the social media platforms themselves, show that people linger longer on content that offends them or provokes strong emotions such as anger or hostility towards an outgroup. It is human nature to react with disbelief when confronted with content that flies in the face of facts or society's accepted values. It should therefore come as no surprise that users are more likely to react to, linger on, or even rewatch content that displays extremism in views or behaviour and provokes strong emotional reactions, even if they completely disagree with the content. Across Facebook field studies and several experiments, people engaged more with posts that *contradicted* their political views than with posts they agreed with. In some cases it was up to four times as much. Users commented more on opposing content, often with anger or disgust, even though they were less likely to follow the source. The effect appeared on socially important topics such as vaccines and gun control, but not on trivial ones like smartphone brands.[57]

Recommendation algorithms on short-form video platforms are widely understood to treat watch time and repeat viewing as signals of interest and respond by increasing similar content in the user's feed. TikTok itself has stated that completed viewing is a strong signal of interest and that replays inform its recommendations,[58] and independent investigations have found that how long users linger on or rewatch a video is among the most important signals its algorithm uses.[5960] This raises a specific concern for young users who are likely to rewatch content that shocks or upsets them, not because they agree with it but because they are trying to process it. An algorithm that reads that reaction as interest is designed to serve more of the same content, even to a young

[57] Mochon, D. & Schwartz, J. (2024). The confrontation effect: When users engage more with ideology-inconsistent content online. *Organizational Behavior and Human Decision Processes*. doi:10.1016/j.obhdp.2024.104366

[58] TikTok Newsroom: How TikTok recommends videos #ForYou, 18 June 2020

[59] The Wall Street Journal: "Inside TikTok's Algorithm: A WSJ Video Investigation," 21 July 2021 (summarised by 9to5Mac)

[60] Smith, B., "How TikTok Reads Your Mind," The New York Times, 5 December 2021 (summarised by Gizmodo)

person who found it distressing. This risks severely impacting youth who are particularly at risk of internalizing hateful messages to the detriment of their mental health and wellbeing.

Under our viewing protocol, videos that researchers judged to be overtly anti-trans were watched twice, but otherwise not engaged with in any other way (no likes, shares, comments or saves that could indicate to the algorithm that the user would appreciate more of this type of content). On TikTok, rewatching occurred on two of six accounts, and the consequences differ. On one of them (an account in Iceland searching for gaming-related terms), rewatching took place in the last two search batches, and the share of anti-trans content among trans-related videos rose from 34.1% to 46.2%. On an account in Sweden searching for Harry Potter related terms, by contrast, rewatching occurred in a batch where half of the trans-related videos were anti-trans, but the share dropped to 0% in the following batch and reached only 8.3% in the last. The results are therefore inconclusive and a causal link cannot be established.

Anti-trans content and rewatching on TikTok, by account and search term

Search term 4 = PewDiePie (gaming accounts) / J.K. Rowling (Harry Potter accounts); 5 = “gender identity”; 6 = “trans kids”

Search term	Measure	Gaming accounts			Harry Potter accounts		
		Finland	Iceland	Sweden	Finland	Iceland	Sweden
4 PewDiePie / J.K. Rowling	Anti-trans share of trans-related videos	n/a	n/a	n/a	**14.3%**	**0%**	**50%**
	Videos watched twice	–	–	–	–	–	12.5%
5 “gender identity”	Anti-trans share of trans-related videos	**13.3%**	**34.1%**	**7.1%**	**6.7%**	**13.6%**	**0%**
	Videos watched twice	–	9.4%	–	–	–	–
6 “trans kids”	Anti-trans share of trans-related videos	**1.8%**	**46.2%**	**0%**	**10%**	**0%**	**8.3%**
	Videos watched twice	–	3.4%	–	–	–	–

n/a = no trans-related content in the batch. – = no video watched twice. “Videos watched twice” is the share of all watched videos in the batch that were watched to the end twice. Search terms 1–3 produced no trans-related content and are not shown.

Nonetheless, the question matters. If rewatching a hostile video increases a minor’s exposure to similar content, the effect would fall squarely within the systemic risks that Articles 34 and 35 of the DSA require platforms to assess and mitigate, including risks to the mental well-being of minors and to fundamental rights. Adolescents are cognitively and emotionally more susceptible to influence than adults. An algorithm that systematically exploits that susceptibility to drive engagement, at the cost of exposing minors to escalating volumes of content that denies their peers' existence or right to healthcare, would not merely fall short of best practice; it could also fall within Article 5(1)(b) of the AI Act, which prohibits AI systems that exploit age-related vulnerabilities in ways that cause significant harm.

With these findings in mind, we recommend that regulators require TikTok and YouTube to disclose how repeat viewing is weighted in their recommendation systems for users registered as minors, and that future research test this mechanism directly through controlled designs in which rewatching is varied independently of content type and search terms.

How much of the YouTube Shorts feed was watched, by search term

All six YouTube accounts combined; 450 videos per search term, 2,700 in total

Search term	Watched to the end twice	Watched to the end once	Watched >3 seconds, not to the end	Total watched	Share of videos watched
1	1	29	3	**33**	7.3%
2	0	31	8	**39**	8.7%
3	0	30	1	**31**	6.9%
4 PewDiePie / J.K. Rowling	0	30	12	**42**	9.3%
5 “gender identity”	0	32	35	**67**	14.9%
6 “trans kids”	1	36	46	**83**	18.4%
Total	**2**	**188**	**105**	**295**	**10.9%**
Share of all 2,700 videos	0.1%	7.0%	3.9%	10.9%	

The top five search results in each batch (30 per search term, 180 in total) were watched to the end, which accounts for almost all videos watched once. One of the two videos watched twice was the result of a technical glitch. Search terms 1–3 were Harry Potter, Hogwarts Legacy and Slytherin on the Harry Potter accounts, and Gaming, Minecraft and Gaming Walkthrough on the gaming accounts.

On YouTube, only two videos were watched twice, one of them because of a technical glitch. Apart from the top five search results in each batch (180 videos in total), which were always watched to the end, only 115 videos were watched for more than three seconds. Since both platforms use watch time as a signal, this limited engagement may partly explain why YouTube’s recommendations changed less than TikTok’s, and the comparison between the two platforms should be read with this in mind.

Broader Challenges and Implications

Both TikTok and YouTube have published reports and established data access procedures, but these have been widely criticised as inadequate. The European Commission has preliminarily found that TikTok’s procedures are burdensome and frequently result in partial or unreliable data,[61] an audit of YouTube’s research API found that it returns highly incomplete samples,[62] and civil society reviews have found the platforms’ risk assessments to consist of “vague declarations”.[63] Legal challenges have also slowed enforcement: X is contesting its €120 million DSA fine in court, and Meta and TikTok successfully challenged the method used to calculate the DSA supervisory fee.[6465] While both platforms offer the user controls the DSA requires, these have been found to be

[61] European Commission preliminary findings on TikTok and Meta, 24 October 2025, reported by Society for Computers & Law.

[62] Rieder, B., Padilla, A. & Coromina, Ò. (2025). Forgetful by Design? A Critical Audit of YouTube’s Search API for Academic Research. arXiv preprint.

[63] ECNL: Five critical lessons from three years of DSA risk assessments, 3 March 2026.

[64] The Brussels Times: X appeals €120m EU fine over digital services breaches.

[65] Heise Online: DSA: EU court overturns pro forma supervisory fees for large online platforms, September 2025.

ineffective or hard to use, and default settings still favour algorithmic feeds: the Commission has preliminarily found that TikTok's algorithm prioritises viewing time "over [users'] choice to see less of certain content",[66] and a DSA complaint filed in March 2026 argues that YouTube's non-personalised option is effectively hidden.[67]

Recommendations for Regulators

As of November 2026, the European Commission has preliminarily found TikTok in breach of the DSA on researcher data access, addictive design, the protection of minors and the "rabbit hole" effect of its recommender system, while YouTube faces sustained criticism and requests for information but no findings to date. Both platforms have published explanations of their main ranking signals, reports and data access procedures, but these offer limited insight into how recommendations are ranked, and have been widely criticised as inadequate and restrictive. The EU's enforcement actions are a step forward, but stronger measures are needed to ensure meaningful transparency and accountability.

In addition to the key recommendations at the start of this report, the European Commission and Nordic authorities such as Post- och telestyrelsen (Sweden), Traficom (Finland) and Fjölmiðlanefnd (once Iceland implements the DSA) should collaborate to:

1. **Enforce existing rules in full**, including fines of up to 6% of global turnover where breaches are confirmed, real-time data access for vetted researchers under Article 40, and the Commission's call to disable viewing time as an indicator for recommending content.
2. **Strengthen independent audits** under Article 37 and require that they be published in full, so that compliance claims can be verified.
3. **Clarify guidelines for data sharing and reporting** to prevent "transparency washing": publishing reports and tools that appear to provide transparency without enabling meaningful scrutiny.
4. **Support civil society, academia and the media** in documenting platform failures and holding platforms to account.

Nordic governments, for their part, should ensure that their national Digital Services Coordinators have the funding, staff and expertise in children's rights needed to carry out these tasks.

[66] European Commission: Commission preliminarily finds TikTok's 'For You' feed in breach of Digital Services Act, IP/26/2361, 6 November 2026.

[67] The Brussels Times: YouTube accused in EU of having 'manipulative' homepage, 11 March 2026.

Qualitative Analysis: Sub-Category Frequency and Valence in Trans-Related Social Media Content

Coded content from TikTok and YouTube recommendation feeds — Finland, Sweden, Iceland

Overview

This analysis draws on 5,406 coded content entries collected from 12 recorded TikTok and YouTube sessions across three researchers and three countries (Finland, Sweden, Iceland). Each entry was assigned a Category Code describing its stance toward trans and non-binary identities (anti-trans, trans-ambivalent, trans-affirmative, or non-trans related) and, where relevant, one or more Sub-Category Codes describing its thematic content. Of the 5,406 coded entries, 471 were trans-related. These received a total of 914 sub-category codes: most videos (420) were assigned two codes, 35 received one and 13 received 3. Three videos received no code as they simply didn't fit within any of the subcategories.

The focus of this analysis is the frequency of the sub-category codes and, critically, how each sub-category's content skews across the affirmative–ambivalent–anti-trans spectrum. This matters because raw topical frequency alone does not tell us whether a recommendation system is amplifying trans people's own voices or hostile commentary about them. The valence breakdown within each theme is where that distinction becomes visible.

Overall Category Distribution

Before turning to sub-categories, the overall picture: the substantial majority of recommended content was unrelated to trans issues. Of the content that was trans-related, affirmative content was the clear majority.

Category Code	Count	% of Total	% of Trans-Related Only
Non-trans related	4,935	91.3%	—
Trans-affirmative content	338	6.3%	71.8%
Trans-ambivalent content	79	1.5%	16.8%
Anti-trans content	54	1.0%	11.5%

Sub-Category Frequency and Valence

The table below ranks the seven sub-category codes by total frequency and shows how each breaks down across the three trans-related category codes. Percentages are calculated within each sub-category row (i.e., of the videos coded with that theme, what share were affirmative vs. ambivalent vs. anti-trans).

Sub-Category Code	Total	Affirmative	Ambivalent	Anti-Trans
Political debates & arguments	274	172 (62.8%)	56 (20.4%)	46 (16.8%)
Body, identity & health	237	193 (81.4%)	27 (11.4%)	17 (7.2%)
Growing up trans	145	135 (93.1%)	10 (6.9%)	0 (0.0%)
Law, liberties & access	79	47 (59.5%)	12 (15.2%)	20 (25.3%)
Media representation	69	28 (40.6%)	28 (40.6%)	13 (18.8%)
Educational / academic / art	61	58 (95.1%)	2 (3.3%)	1 (1.6%)
Discrimination	49	15 (30.6%)	20 (40.8%)	14 (28.6%)

Two things are worth noting in how to read this table. First, ranking by total frequency (top to bottom: political debates, body/identity/health, growing up trans) surfaces the topics the algorithm serves most often regardless of tone. Second, ranking by anti-trans share (discrimination at 28.6%, law/liberties/access at 25.3%, media representation at 18.8%) surfaces which themes are most contested, where hostile content is disproportionately likely to appear when the topic comes up at all.

Explication by Sub-Category

Political Debates & Arguments (n = 274, the most frequent sub-category)

Political debates and arguments was the single most common thematic code, appearing in 274 trans-related videos. This category was also one of the more contested categories: while 62.8% of debate-framed content was trans-affirmative, a meaningful 16.8% was anti-trans and a further 20.4% was ambivalent. The codebook's definition of anti-trans content anticipates this, as debate-format videos often present "both sides" while overtly or covertly favoring one.

The researchers repeatedly flagged this dynamic in their qualitative notes. One example of these notes is a researcher describing a street-interview video in which a man stopped people on the street to argue that transgender children do not exist, while he is opposed by a young woman countering his argument. Despite this example being "an argument" it was classified as anti-trans because the framing favored the position of the content creator, the man arguing against the right of trans people to exist.

Another entry describes a video featuring a debate over whether "gender is not a social construct but a biological reality". This was coded anti-trans despite seemingly being framed as a debate, since it inherently mobilizes "biological reality" as some sort of shared universal understanding that even overlooks gender variance in biology itself. The frequency data shows that political

debate is the largest thematic category in the dataset, and that more than a third of it (37.2%) was ambivalent or anti-trans.

Implications for Platform Policies

The prevalence with which the debate format is used to further an anti trans discourse points to a need for platforms to craft policies that recognize the harm in this format. Recognizing transphobic content as a violation of community guidelines is vital, even if the content does not meet the legal threshold for hate speech. Content presented as debates are not mere debates when their premise is arguing whether certain groups of people have the right to exist. It is harmful to trans people and their allies who are exposed to it, and it has the potential to influence other users to adopt transphobic attitudes. Content presented as a debate over the right to exist, and over the right to have rights, produces a systemic risk against trans people and as such it should be flagged in community guidelines, monitored, and lead to repercussions.

All three countries involved in this study (Finland, Iceland and Sweden) prohibit harassment and discrimination on the grounds of gender identity in their equality legislation, and Sweden and Iceland also explicitly include gender identity in their criminal hate speech provisions. Content that systematically questions or disparages trans people's identities may therefore be unlawful under national law in the Nordic countries, even where it does not meet the threshold of a crime.

Affirmative Content Within the Political Debates & Arguments Theme

While the debate/argument format is where anti-trans content is most likely to appear disguised as a (mere) argument of two balanced sides, the majority of content coded within this sub-category (62.8%) was coded as trans-affirmative.

The most common pattern is content coders described as "awareness raising", a phrase that recurs across dozens of entries and functions almost as shorthand in this dataset. This ranged from myth-debunking (a trans creator "refuting myths about detransitioning and detransitioners") to explainer content on hormone therapy, pronoun use, and gender terminology. Much of this content was not adversarial in any conventional sense. It was coded as "political" because engaging publicly with gender identity related themes is inherently entering the political conversation, regardless of whether the video contains an opposing viewpoint at all.

A second major pattern is first-person narrative and humor. Many entries feature trans creators simply describing their own lives: transition timelines, coming-out stories, or self-deprecating jokes about gender dysphoria. Given the political nature of the public discourse about trans issues, jokes and humor were coded as political content, irrespective of whether the creator was trans or not.

A smaller cluster of the data was noted to directly rebut anti-trans claims. For example, the qualitative notes of coders logged creators "refuting the claim that LGBT people are mentally ill" or "responding to allegations that trans people are dangerous, mixing sarcasm with facts to push back against the narrative." Dr. Helen Webberley, who is a UK-based doctor, gender specialist and medical educator, and creates content on trans rights, recurs here specifically in this rebuttal

capacity. In videos featuring her, she uses her position as an expert in the field to challenge anti-trans claims, which is important because in the examples noted above about “debates” about trans rights, the anti trans debaters are often not in any way qualified to speak on trans issues. They simply hold anti-trans opinions, yet this might not be obvious to casual content consumers, leading to the false perception that the rights to exist for marginalized communities is something that can be debated.

Finally, a recurring sub-thread ties the Political Debates and Arguments theme to the Media Representation theme: references to the public dispute between J.K. Rowling and Emma Watson, which was timely during the data gathering phase. Of the 27 trans-related videos referring to the feud, nearly half (13) were ambivalent, typically news and celebrity outlets retelling the dispute without taking a clear position. A third were coded as trans-affirmative, while five were coded as anti-trans.

A Note on Coding Uncertainty

Several coder notes in this sub-category flag genuine classification uncertainty, for example, "I'm not sure we discussed how to code content like this," and multiple instances of coders unsure whether a video should be filed under political debate, representation, or awareness-raising. Given that political_debates_arguments is the single largest sub-category in the dataset (n=274), this pattern of self-reported coder uncertainty can be seen as a limitation, since it suggests some conceptual overlap or under-specification in the codebook at the boundary between "awareness raising," "representation," “humor”, and "debate."

Body, Identity & Health (n = 237)

The second most frequent theme, and considerably more affirmative in tone: 81.4% of this content was trans-affirmative, with only 7.2% anti-trans. Much of this content consists of first-person identity narratives. Coders noted young creators discussing gender dysphoria, transition, or gender-fluid identity directly to camera.

The 11.4% of content coded as trans-ambivalent sometimes involved creators expressing genuine emotional conflict alongside their identity content (e.g., a caption reading "I hate being gender fluid, no one sees me as a boy"), which coders classified as ambivalent rather than affirmative specifically because of the mixed emotional register, even though the content originated from within trans experience.

The anti-trans share of this theme was concentrated in detransition-narrative content, which the codebook explicitly flags as an anti-trans risk: detransition stories are not inherently harmful, but become anti-trans coded when used to generalize against transition access broadly. Detransition narratives also stood out disproportionately to the actual rate of trans people who do choose to detransition, thus framing it as more common than it is.

As such, detransition narratives serve to crumble the foundation of work towards trans rights, framing trans identities and bodily autonomy over transition as unstable and thus illegitimate. Trans rights organizations and activists also flag the promotion and high visibility of detransition narratives as key political tools in furthering transphobia. This does not mean that the trans rights community is attempting to "silence" those who identify as detransitioners (as was claimed by some content creators in the videos). Detransition narratives become harmful when they are poorly contextualized, making it difficult for content consumers to grasp the distinctly low percentage of those whose quality of life has decreased after beginning gender transition. In a survey of more than 90,000 trans adults in the United States, 98% of those receiving hormone therapy said it had made them more satisfied with their lives, only 0.36% went back to living as their sex assigned at birth because transition wasn't right for them.[68] As a result, transitioning improves the lives of an overwhelming majority of trans people.

Growing Up Trans (n = 145)

This theme was the only sub-category with no anti-trans content: 93.1% affirmative, 6.9% ambivalent, and zero anti-trans logs in the dataset. It covers content about trans youth, family dynamics, school, and coming out, and the complete absence of anti-trans-coded content suggests that when the algorithm surfaces content specifically about trans children and adolescents, it is largely first-person or sympathetically framed content. An alternative theory is that hostile content involving children is rapidly detected and removed.

The content included educational, high production value animated shorts, coming-out stories, and youth-perspective narratives, rather than adversarial commentary about trans youth. This is a notable finding given that debates about trans youth are often the most politically charged in the broader public discourse. The coded data suggests the weaponization of the rights of trans children for an anti-trans stance is not visible when the topic is coded as being about the lived experience of growing up trans specifically.

However, these findings do not include the vitriol and anti-trans sentiments shared in the comment field or possibly attached to the video when it is reposted, shared, stitched or used to create AI generated content in the future. The trans affirmative skew also does not account for the possible hateful and threatening messages shared to the creators of these accounts, and the very real threat they pose to their safety and sanity.

An example from the data set: A notable share of the content coded as *Growing up Trans* comes from a single recurring source: a well-known family vlogging channel documenting a young trans girl's transition (logged by coders under the recurring character "Edie" / "Edie and family"). Their content appeared 20 times across the dataset, spanning all three countries and both platforms. Nearly every instance was coded trans-affirmative with the subcategory code growing_up_trans, with content ranging from milestone moments (the child seeing the trans flag colors for the first

[68] Rastogi, A., Menard, L., Miller, G. H., Cole, W., Laurison, D., Caballero, J., Murano-Kinney, S. & Heng-Lehtinen, R.: *Health and Wellbeing: A Report of the 2022 U.S. Transgender Survey*, Advocates for Transgender Equality, June 2025.

time) to a parent narrating the family's support journey ("Five years ago, my daughter came to me and told me she is trans, and this is the story of why I supported her") to lighter, non-trans-related family content from the same channel.

The recurrence of “Edie’s” channel across independently coded accounts and platforms indicates the channel functions as a consistent node in the recommendation algorithm's trans-affirmative content pipeline, when search terms such as ‘trans kids’ were fed to the algorithm. This example lifted from the data set is a concrete illustration of what "affirmative" content about trans youth looks like in practice within this dataset: parent accompanied, milestone-oriented, and framed around family support rather than direct engagement in debate or controversy. The tone of the videos is upbeat and aims to normalize the experience of trans joy in young children rather than respond directly to political debates around the rights of trans children. As such, they provide positive and accepting representation for other trans youth potentially seeking out trans content online.

Law, Liberties & Access (n = 79)

This sub-category shows the second-highest anti-trans share of any theme (25.3%), alongside 59.5% affirmative and 15.2% ambivalent content. Coded examples include content about bathroom bans, legislative restrictions on gender-affirming care, and sports-eligibility debates. This is one of the two themes (alongside discrimination) where the anti-trans share is high enough to suggest that legal and rights-based content functions as a flashpoint: a category where hostile framing is disproportionately likely to surface relative to how often the topic appears at all.

Media Representation (n = 69)

Content about trans representation in mainstream media and celebrity culture (e.g., the Rowling–Watson dispute) was the most evenly split theme in the dataset: 40.6% affirmative and exactly 40.6% ambivalent, with 18.8% anti-trans. At the time of data gathering (November 2025), the author of the Harry Potter series, J.K. Rowling, and former Harry Potter movie franchise star Emma Watson, were prominently featured in a public feud. A snapshot in time, this feud, on-going for years due to Rowling’s anti-trans activism, and Watson’s support for trans rights, took on new momentum as Watson had, shortly before the gathering of the data, been featured on a popular podcast hosted by Jay Shetty. In the podcast, Shetty asks Watson what she thinks of the “extremely hurtful” comments J.K. Rowling had made about her, including that Watson ruined the movies for Rowling. Watson’s response on the podcast marked a departure from her prior, more adamant stance on trans rights (read as direct opposition to Rowling), to a more conciliatory tone towards Rowling. Several content creators coded in the data noted that Watson ended up talking for a long time while saying very little, avoiding taking a stance against Rowling while simultaneously not signaling withdrawal from her own stance for trans rights. The feud, documented by content creators, then continued with Rowling responding to Watson on her X account:

> *I wasn't a multimillionaire at fourteen. I lived in poverty while writing the book that made Emma famous. I therefore understand from my own life*

> *experience what the trashing of women's rights in which Emma has so enthusiastically participated means to women and girls without her privileges.*
>
> *The greatest irony here is that, had Emma not decided in her most recent interview to declare that she loves and treasures me - a change of tack I suspect she's adopted because she's noticed full-throated condemnation of me is no longer quite as fashionable as it was - I might never have been this honest.*[69]

Reflecting the genuinely contested nature of celebrity-adjacent discourse, some of the content sided with Watson (33%), some with Rowling (19%), while content that neutrally retold the feud was coded as ambivalent (48%) per the codebook's guidance. This is the theme where the specific person or outlet being covered mattered most to the code assigned, more than the topic itself, showing that a single celebrity search can bring content questioning trans people's rights into young users' feeds.

Educational / Academic / Art (n = 61)

The most affirmative-skewing theme (95.1% affirmative, 3.3% ambivalent, 1.6% anti-trans). This code captures high-production-value explainer content, short films, and academic material. Its near-total affirmative skew suggests that formally educational content about gender identity is rarely used as a vehicle for anti-trans messaging in this dataset. The one anti-trans-coded example involved a detransition narrative framed as "escaping trans ideology," again illustrating the codebook's point that detransition narratives are rarely neutral or trans affirmative. Rather they are used as a tool to undermine trans rights. This also raises a flag especially when a detransition narrative is framed as educational but is anti-trans. Can young viewers decipher the harmful tone when it is framed as educational?

Discrimination (n = 49 — smallest but most contested)

Though the least frequent sub-category, discrimination has the highest anti-trans share of any theme (28.6%) and the highest ambivalent share (40.8%), leaving only 30.6% affirmative, the lowest affirmative share in the dataset. For every theme except discrimination and media representation, affirmative content is the majority. Coders' notes show this code was applied both to content documenting discrimination against trans people (affirmative) and to content where a speaker dismissed or pathologized trans identity in ways coders read as discriminatory in themselves (anti-trans). The same sub-category label captures both the reporting of harm and the enactment of it, which helps explain why its valence is so mixed.

[69] Rowling, J.K. [@jk_rowling]: post on X, 29 September 2025.

Summary of Key Patterns

1. Political debates & arguments and body/identity/health are the two most frequent themes by volume, together accounting for over half of all sub-category codes.
2. Discrimination and law/liberties/access carry the highest anti-trans shares (28.6% and 25.3% respectively), despite being lower-volume themes, suggesting these topics are disproportionately likely to attract hostile framing and content when they do appear.
3. Growing up trans and educational/academic/art are the most reliably affirmative themes (93.1% and 95.1% respectively), with growing up trans showing zero anti-trans-coded instances.
4. Media representation and discrimination are the only themes where affirmative content is not the majority. Media representation splits evenly between affirmative and ambivalent content (40.6% each, with 18.8% anti-trans), while discrimination has the lowest affirmative share in the dataset (30.6%), with 40.8% ambivalent and 28.6% anti-trans. This suggests these are sites of active discursive conflict rather than settled framing.
5. Across the dataset, the debate/argument format recurs as a mechanism through which anti-trans content is delivered under the appearance of balance, a pattern coders flagged qualitatively and which the frequency data supports: 37.2% of debate-format content was ambivalent or anti-trans.

Methodological Note

Sub-category codes are non-exclusive (a single video could receive multiple sub-category labels), so the percentages in the Sub-Category Frequency and Valence table are calculated relative to each sub-category's own total rather than to the overall trans-related content pool. The 914 sub-category codes therefore exceed the 471 trans-related videos, and readers should interpret the frequency counts as counts of thematic instances rather than counts of unique videos. All figures are drawn from the coded "Coding Table" sheets across the 12 source files, the four repeated Category Code definitions are provided in each file's Codebook sheet.

Non–Trans–Related Qualitative Findings

While gathering data from thousands of videos on two of the largest social media platforms today, the researchers noticed various trends unrelated to trans rights, some of which deserve a mention in the context of youth safety and social cohesion.

The “Secret Account” Trend

As many as 8% of all videos recommended on one of our TikTok sockpuppet accounts, which were registered as belonging to 15 year old children, promoted the idea of a “secret account”. It is a well known trend among youth to escape parental scrutiny by having an “official” social media account that they show to their parents, while having another account that is known only by their friends and peers (and potentially discoverable by bad actors and predators). This creates a false sense of safety among parents, who think they are aware of what their child is doing online, and enables an unmonitored digital existence for youth. It is a cause for concern how many videos actively encouraged the use of secret accounts and the bypassing of the safeguards that guardians are able to provide for young users in their care. The fact that the recommendation algorithm would promote it so aggressively prompts the question whether social media platforms are prioritizing children’s online engagement over their safety, thus looking the other way in favor of youth having multiple accounts – even when it poses a risk to their wellbeing.

Polarization by Social and Traditional Media

Galvanizing users’ emotions with shock or outrage leads to more engagement, as previously stated. Recommendation algorithms that reward controversial content and increase the visibility of misinformation, fringe behaviors and beliefs including conspiracy theories (with the return of measles as a result of vaccine misinformation being one example[70]), are thereby polarizing by design. During the data gathering, the researchers noticed traditional media outlets employing similar tactics on their social media accounts, presenting complex topics such as climate change or human rights as an “either or” issue that users are “for or against”, pitting people into two opposing camps instead of hosting the nuanced debate that is fundamental to the wellbeing of our democracies.

What’s Ahead

Find Your (Algo)Rhythm

To respond to the need for education and awareness as outlined in this research, NORDREF and Transfeminiinit in collaboration with Ofbeldisforvarnaskólinn are launching **Find Your (Algo)Rhythm,** a project that empowers LGBTQ+ youth through algorithmic literacy in the Nordic region. Find Your (Algo)Rhythm addresses how harmful stereotypes and misinformation affecting

[70] WHO/Europe & UNICEF, 13 March 2025: European Region reports highest number of measles cases in more than 25 years.; KFF, 23 April 2025: KFF Tracking Poll on Health Information and Trust: The Public’s Views on Measles Outbreaks and Misinformation.

vulnerable youth can get amplified on social media. It combines research findings with interactive workshops, educational toolkits, and evidence-based policy advocacy targeting youth, parents, educators, and policymakers.

By creating multilingual resources, the initiative promotes digital rights, platform accountability and long-term integration of algorithmic literacy in Nordic schools. More information will be shared on the websites and social media of NORDREF and Transfeminiinit ry – stay tuned!

Acknowledgments

This research was co-funded by the Nordic Gender Equality Fund, administered by NIKK – Nordic Information on Gender – on behalf of the Nordic Council of Ministers. We are grateful to Nova for lending us ten phones in November 2025, allowing for a successful data gathering process. We owe our gratitude to Christian Mogensen for advising us on the research design and preparation process. Samtökin ‘78 provided valuable advice and assistance with the development of the codebook and with the launch event. Finally, we thank the President of Iceland, Halla Tómasdóttir, for her kind support of this project.

This report was authored by Thordis Elva for NORDREF, Nikola Maljkovic for Ofbeldisforvarnaskólinn and Nelli Ruotsalainen for Transfeminiinit ry, who also carried out the data collection and coding. It was published on 11.11.2026.