

# Automated Detection of Mainstreamed Transphobic Content on YouTube

**Lydia Channon**

is a PhD student in the Computational Foundry at Swansea University. Their research focuses on far-right ideologies, gender dynamics, and countering violent extremism.

✉ [lydia.channon@swansea.ac.uk](mailto:lydia.channon@swansea.ac.uk)

**Nicola Mathieson**

is a Lecturer of International Relations in the Department of Politics at the University of Liverpool. Their research focuses on political violence and transnational extremism.

✉ [nicola.mathieson@liverpool.ac.uk](mailto:nicola.mathieson@liverpool.ac.uk)

How has transphobia entered mainstream discourse and media? And can the detection of mainstreamed transphobic content be automated? We argue that transphobia on social media platforms has undergone a process of “mainstreaming” that actively transformed historically fringe and extreme narratives so that content no longer appears taboo or inappropriate. Consequently, the mainstreaming of hate speech presents unique challenges for its detection and removal from social media platforms. This paper introduces a content classifier designed to identify transphobic content online and categorize the emergent mainstreamed narratives that currently typify transphobic content. To test the classifier, we utilize an original dataset of over 49,000 comments from five YouTube channels known to post transphobic content and identify nine mainstreamed transphobic narratives that fall below content moderation policies.

**KEYWORDS** transphobia; far right; content classifier; mainstreaming; YouTube

**DOI** [10.57814/49JZ-0663](https://doi.org/10.57814/49JZ-0663)

**LICENSE** Creative Commons Attribution-NonCommercial-NoDerivatives ([by-nc-nd](https://creativecommons.org/licenses/by-nc-nd/4.0/))

At the 2024 Paris Olympics, two cisgender women athletes—Imane Khalif from Algeria and Lin Yu-ting from Taiwan—were subject to a disinformation campaign accusing both of being transgender (BBC News 2024; Beacham 2024; Dunbar 2024; Li, Burke, and Abdelkader 2024). Both athletes had failed an unspecified “gender test” administered by the International Boxing Association at the World Championships the previous year, an organization that has since lost its recognition by the International Olympic Committee. Transphobic narratives of fairness and safety in women’s sport

quickly spread among competitors, media, and online despite neither of these women being trans. The media storm followed the concerning pattern of the year-on-year increase in recorded transphobia globally (ADL 2023; Higgins 2023; McLean 2021; Reid 2021; Squirrell and Davey 2023). The incidents in Paris also exemplified the intersectional dimensions of transphobic hate by targeting athletes of color. Transphobia disproportionately affects individuals with multiple marginalized identities along axes of race, gender, and socio-economic status (Colliver 2022; Gill-Peterson 2018; Koyama 2020; Upadhyay 2021). As the events in Paris highlighted, once the purview of extremist, fringe communities, transphobia has spread into mainstream media and conversation globally.

This article investigates how transphobia has entered the mainstream—in media, legislation, and wider society—by examining how transphobic narratives are framed online. The mainstream is defined by Brown, Mondon, and Winter (2023, 166) as “a normative, hegemonic concept that imbues a particular ideological configuration or system with authority to operate as a given or naturalize itself as the best or even only option, essential to govern or regulate society, politics and the economy.” Mainstreaming, then, is the process where movements on the margins attempt to move their ideologies from the fringes to the mainstream that results in content no longer being considered taboo or inappropriate as a means of “making their views more palatable to a wider audience” (Gallaher 2021, 224). Transphobia has become the focus of public discourse (Ashley 2020; Gambino 2023; Kessler 2022; Peterson 2021; Trinko 2019), legislation (ACLU 2024; The Trevor Project, n.d.; Trans Legislation Tracker 2024; Track Trans Legislation 2023) and increased targeting of trans people offline, including physical violence and murder (see, for example, Cardwell 2023; Compton 2019; Joseph 2022; Kelly and Rachwani 2023; Sasani 2024; Wareham 2023). While the LGBTQI+ community have always been a target of extremist hate and violence, there are relatively few studies that focus exclusively on the targeting of the trans community or how transphobic narratives are constructed and disseminated. Instead, transphobia is often noted as an element of gendered violent extremist rhetoric (Blee 2020; Hackworth 2018; Miller-Idriss 2020; True and Eddyono 2021) or a category of hate against minorities including misogyny, racism, homophobia, and antisemitism (Ryynänen, Kosonen, and Ylönen 2022).

To address this gap, this paper builds a content classifier using a Large Language Model (LLM) trained to identify transphobic content and measure the prominence of mainstreamed narratives. As transphobia often intersects with other forms of hate, we included other forms of targeted hate in our labels to capture this intersectionality, including Racism, Religion, Gender, Homophobia, and Xenophobia across a range of issue areas. As one of the main perpetrators of transphobic hate online are “right wing media commentators” (Squirrell and Davey 2023), we select five of these commentators known to post transphobic content on their YouTube channels. Prominent online right-wing figures are central in propagating anti-trans narratives, often framing trans people as a threat to societal values, under the guise of protecting societal norms, such as women’s rights and children’s safety (Ging 2019; Lewis 2018; Marwick and Lewis 2017; Mondon and Winter 2020). The selection of multiple influencers also highlights the transnational flow of ideas contributing to a global discourse on transphobia (Fielitz and Marcks 2019; O’Callaghan et al. 2013).

We utilized the language model DistilBERT to classify 49,000 comments. The content classifier identified 7,965 of the comments in our original dataset as likely containing transphobic content. We then conducted an analysis of these comments, identifying nine mainstreamed transphobic narratives: children, definitions, women's rights, mental health, misgendering and critiquing pronouns, transphobic terms, religion, detransitioning, and conspiracy. We find that the most common narrative within our dataset were related to children across three themes: child abuse, medical abuse, and education. The second most prominent narrative was defining the term "woman." The comments focused mostly on trans women and excluding them from being identified as women. We also conducted a comparative analysis of transphobic discourse being employed between comments and videos and network analysis of users in our dataset across channel to identify transphobic user communities and their online behavior identifying distinct differences between posting behaviors. In line with mainstreaming, comments rarely contained explicit hate speech or the use of derogatory terms. Instead, transphobic narratives were framed around concerns for women and children safety. The movement from explicit hate to threat-based narratives is a deliberate strategy by right-wing commentators to increase their appeal to a wider audience (Colliver, Coyle, and Silvestri 2019). Consequently, as seen in the case of the Paris Olympics, these narratives have crept into mainstream conversations, reporting and legislation and have become normalized in the public domain.

There are many practical applications of this project. First, this classifier contributes evidence of the scale of online transphobia and highlights the need to develop content moderation policies that are more nuanced than the current category of hate speech. Second, this classifier and dataset can help track and monitor transphobic narratives and identify how they are changing over time. Monitoring and deconstructing mainstreamed narratives are essential to curbing their impact, requiring both automated detection tools and public education to address the fear-based tactics employed (Squirrell and Davey 2023). Finally, this classifier may be a valuable tool to help researchers process large datasets that contain hate speech and other forms of extremism and reduce potential harm by reducing the amount of content researchers are exposed to.

The article proceeds as follows. First, we provide an overview of the concept of mainstreaming and its application in this project. Second, we outline the research design including case selection, data collection and labelling, and analysis. Third, we present the core findings of our analysis. Finally, we conclude with a discussion of our findings, their implications, and potential for future research. It is important to note that the purpose of this paper is to identify mainstreamed transphobic narratives and their dissemination online. While we do not draw on direct quotes from commenters throughout, this paper does contain terms and summaries of narratives that are transphobic in nature. This content may be distressing for readers.

## **MAINSTREAMING EXTREMIST NARRATIVES**

As content previously considered extreme increasingly moves from the fringe to everyday discourse, the concept of the mainstream and the subsequent process of mainstreaming have been the topic of growing academic research. In particular, the terms

are applied to the increasing presence of far-right actors and ideas in public discourse and elections (for example, see Acha Ugarte 2018; Bleakley 2023; Brown, Mondon, and Winter 2023; Cammaerts 2018; Cornell 2018; Ekman 2022; Hainsworth 2016; Klinger et al. 2023; Stern 2022). Despite the increasing usage of both concepts, there is limited conceptualization (Brown, Mondon, and Winter 2023; Kallis 2015). Instead, the terms have an assumptive quality that the mainstream is the center, and that mainstreaming is simply moving concepts from the extreme or fringe to the political middle. However, the mainstream is not the binary opposite of the extreme (Kallis 2015). Instead, the mainstream is “constructed, contingent and fluid” and not inherently good or bad (Brown, Mondon, and Winter 2023, 166). Brown Mondon, and Winter (2023, 166) define the mainstream as: “a normative, hegemonic concept that imbues a particular ideological configuration or system with authority to operate as a given or naturalize itself as the best or even only option, essential to govern or regulate society, politics and the economy.” Mainstreaming, then, is the process where actors on the margins attempt to move their ideologies from the fringes to the mainstream as a means of expanding their appeal and “making their views more palatable to a wider audience” (Gallagher 2021, 224). Mainstreamed extremist narratives can generate exclusionary public policy (Marwick and Lewis 2017; Mondon and Winter 2020) and increase violence and hate crimes (Baele, Brace, and Coan 2023; Perry 2012; Zuboff 2023), while making it harder to challenge or regulate discriminatory content in public discourse content (Klein 2012; Sunstein and Vermeule 2009).

While mainstreaming can be an unintentional consequence of societal change (Rothut et al. 2024), scholarship focuses on the intentional mainstreaming of discourse by far right and extremist actors by “silently shift[ing]... the public discourse towards more radical positions without being perceived as doing so” (Hohner et al. 2022, 310). For example, mainstreaming can dissociate discourse from radical or extremist ideologies to make it appear more innocuous or respectable (Gerstenfeld, Grant, and Chiang 2003; Graham 2016). Selepak and Sutherland (2012) demonstrate how the Ku Klux Klan have softened their language and symbolism to align their ideology with “white, religious and political conservatives.” Klinger et al. (2023) found that the fringe political party, Alternative for Germany (AfD), used intermediary networks to adopt the language of extreme identitarian movements without affiliating themselves with these actors. The process of mainstreaming can also lead to what is called normalization: what was previously borderline, or unacceptable discourse loses its taboo status and becomes common sense, a legitimate view or politically appropriate (Krzyżanowski 2020). The majority of this work has focused on anti-immigration discourse that has become more common as a result of Brexit, the immigration crisis in 2015, and the election of anti-immigrant populist political leaders particularly in Europe and North America (Bleakley 2023; Ekman 2022; Gallagher 2021; Graham 2016; Klinger et al. 2023). Anti-immigration sentiment, rather than at the fringes of discourse in liberal democracies, is now a core tenant of policy for populist political parties and actors and found in mainstream media and discourse, as well as increasing violence against migrant housing facilities (Bleakley 2023; Cammaerts 2018; Cornell 2018; Graham 2016; Kallis 2015; Klinger et al. 2023; Krzyżanowski 2020).

Far less is known about the mainstreaming of extremist discourse against other marginalized groups. While literature often notes the targeting of the trans com-

munity, there is limited research that explores the dynamics of this targeting separate from other minority identities including women, Black people, immigrants, and the LGBTQI+ community more broadly (Colliver 2023, 412; Ryyänen, Kosonen, and Ylönen 2022; Tsirbas and Zirganou-Kazolea 2025). However, with the rise of social media, there has been a consistent increase in research examining homophobia and transphobia on online platforms (Sánchez-Sánchez, Ruiz-Muñoz, and Sánchez-Sánchez 2024). Research on the impact of online targeting of the LGBTQI+ community include the experience of cyber-violence and cyberbullying (Colliver 2023; Garaigordobil and Larrain 2020; Mkhize, Nunlall, and Gopal 2020; Pescitelli 2018; Rivers, Daly, and Stevenson 2023) and the toll of exposure of homophobic and transphobic speech and acts on the wellbeing of the LGBTQI+ community (for example, see Billard 2024; Colliver 2022; Goldblum et al. 2012; Grant et al. 2011; Marzetti, McDaid, and O'Connor 2022; Mizock and Mueser 2014).

One more recent focus is on the quantitative measurement, or automated detection, of online hate speech towards the LGBTQI+ community across platforms. There are several projects that specifically seek to automate the detection of explicit transphobia online (Chakravarthi 2024; Lu and Jurgens 2022; Sharma, Gupta, and Singh 2023; Valerio 2022). While the automated models are assessed for accuracy, they often lack qualitative analysis of the themes and narratives or theoretical models for understanding how this content has become more prevalent. By focusing exclusively on identifying transphobic content, these models can also miss identifying the intersectional nature of online transphobia. For example, in 2018, Michelle Obama was often the target of a conspiracy that she was trans, intersecting race, conspiracy theories and transphobia (Brandwatch & Ditch the Label 2020; Lokmanoglu et al. 2023). To be most effective, LLMs need to include other forms of intersectional hate in the modelling and be capable of identifying coded language.

On the other hand, there is a more qualitative approach that looks at the nature of transphobic narratives online. Anti-trans ideologies frame trans people as a source infiltration, corruption and intimate danger to generate what is often referred to as a “moral panic” (Higgins 2023; Owen 2022; Slothouber 2020; Walker 2023). Moral panics can be understood as social anxiety or intensified concern towards an outgroup “that is disproportionate to the threat that it poses and is amplified further by the media” (Walker 2023, 995). Overall, these transphobic narratives are grounded in particular ideas about sex and gender that justifies trans exclusion (see, for example, Bassi and LaFleur 2022; Billard 2023; Dickey 2023; Earles 2019; Shaw 2019). While scholars have categorized the themes of transphobia in several ways, they are embedded in a narrative of harm. First, trans people are portrayed as a threat to women and children’s safety. Trans people, especially trans women, are portrayed as predatory, actively seeking access to women’s spaces to do harm (Colliver 2021; Colliver, Coyle, and Silvestri 2019; Conway 2023; Dickey 2023; 2022; Jones and Slater 2020; McLean 2021; Squirrell and Davey 2023; Thompson, 2022), a threat to fairness of women’s participation in sport (Billings et al. 2024; Fischer 2023; Lucas and Newhall 2019; Pape 2022), and a physical harm through the provision of medical care to trans people, especially trans youth (Ashley 2020; Billard 2024; Slothouber 2020).

The second overarching harm is perceived to be towards (Western) society. The existence and acceptance of the trans community is heralded as a sign of societal and

moral collapse which lead to acceptance of other unacceptable behaviors such as bestiality, pedophilia and necrophilia (Colliver 2021; Conway 2023; Walker 2023) and is sometimes framed through a religious lens such as equating being trans with “Satanism” and conceptions of degeneration (Dickey 2023; Shaw 2019; Squirrell and Davey 2023). There is also threat framing that portrays perceived “trans ideology” as a contagion that links back to the first harm of the safety of women and children (Ashley 2020; Thompson 2022; Walker 2023). As outlined by Bassi and LaFleur (2022, 312), regardless of the core arguments behind different movements” transphobia, they understand themselves as on one side of “a global battle of ideas” in defense of an “essentialist story of womanhood.” However, the ever-increasing levels of transphobic content online globally present challenges for these in-depth and more qualitative approaches that rely on researchers manually reviewing data.

In this paper, we propose bringing together these quantitative and qualitative approaches to be able to better manage the amount of data online research presents. We build on both approaches to analyze online transphobic content by building a LLM capable of identifying mainstreamed transphobic content. In turn, this allows us to systematically examine more transphobic content across social media channels and across issue areas. We also expand on existing models by taking an intersectional approach that interrogates how various forms of hate—such as racism, xenophobia, and homophobia—intersect with transphobic content.

## **RESEARCH DESIGN**

This project built and tested a content classifier capable of automating the detection and identification of mainstreamed transphobic content on YouTube. This section begins with an overview of the ethical approval for this project and the additional measures put in place to protect user identity. Second, we outline the choice of YouTube as a platform before outlining our selection of right-wing media commentators. Third, we outline our methodology including our labelling strategy and data analysis process using DistilBERT before introducing the analysis for our findings.

### **Ethics**

This project received ethics approval from Swansea University (ethics approval number 1 2024 9331 8296). We have taken additional measures to ensure anonymity and reduction of the harm of transphobic content. We have chosen not to publish identifying characteristics of commenters including commenter ID or video IDs in our final dataset. Further, while we identify shared narratives among commenters across videos, we do not directly quote from individuals. This decision is in part to protect the identities of users but also so not to direct people to hateful content. Instead, we only connect comments and themes to channels (Conway 2021).

### **YouTube Importance for Right-Wing Media**

We selected YouTube as a platform to analyze how transphobia has become mainstreamed. YouTube has emerged as a pivotal platform for right-wing commentators due to its vast reach, algorithmic amplification, and financial incentives. The platform’s low barriers to entry allow right-wing media to bypass traditional gatekeepers

and directly engage with global audiences, enabling the dissemination of ideologies often marginalized or challenged in mainstream outlets (Klein, Clutton, and Polito 2018; Lewis 2018; Marwick and Lewis 2017). Recommendation algorithms, which suggest videos based on user viewing history, can lead users toward increasingly extreme viewpoints, contributing to ideological echo chambers and further amplifying right-wing narratives (Ribeiro et al. 2020). Interactive features such as comments, likes, and shares foster highly engaged communities, reinforcing these messages and creating spaces where right-wing content is both consumed and propagated (Daniels 2018). Financially, YouTube monetization options—through advertising, sponsorships, and donations—allow right-wing content creators to sustain and expand their operations independently from traditional media funding (Tuters and Hagen 2020). This, combined with YouTube's international accessibility, enables transnational collaboration among right-wing groups, facilitating the global spread of ideologies (Klein, Clutton, and Polito 2018).

YouTube also has a practical benefit for online research. First, YouTube has an API that allows researchers to scrape data directly from their platform. This differs from other online platforms such as X that have revoked free researcher access to their API (Gotfredsen 2023). Second, YouTube allows the analysis of two forms of content—videos and comments—that allows researchers to examine the interaction of content creators and users (Google for Developers 2024). This allowed us to compare the content and nature of videos by our selected commentators and the comments left by users.

### **Commentator Selection**

We centered our case selection on right-wing media commentators known for disseminating transphobic content. There are, of course, many channels that could be selected that feature transphobic content. However, as we outline below in the data collection section, there are important API and data management limitations that affect the scope of this project. We selected cases based on two core factors. First, right-wing commentators have been widely recognized as key sources of transphobic narratives and their dissemination (Oliver 2022; Squirrell and Davey 2023). Second, these commentators operate as alternative media. Schroeder defines alternative media as non-mainstream platforms that challenge dominant ideologies and bypass traditional gatekeepers (Schroeder 2018). Alternative media, particularly in digital spaces, cater to marginalized or counter-hegemonic groups and can range from moderate to extreme positions. Right-wing commentators embody this alternative media role by positioning themselves as anti-elite, and as purveyors of “truth” and “reason” (Frischlich, Klaproth, and Brinkschulte 2020; Schroeder 2018). This framing allows them to legitimize their positions while disseminating transphobia and misinformation (Dowling, Johnson, and Ekdale 2022; Hameleers 2020). Consequently, these figures play a significant role in the global spread of disinformation and conspiracy theories, including those that target LGBTQI+ communities and promote transphobia (Devin 2023; DiMaggio 2022; GLAAD 2021; Mayerhöffer 2021; Ramirez 2023).

We selected five YouTube channels for our dataset: Kelly-Jay Keen (formally known as Posie Parker), Matt Walsh, Ben Shapiro, Candace Owens, and Blaire White. We selected these channels due to their diversity in gender identity, racial and ethnic

**Table 1. Summary of Characteristics of Selected YouTube Channels as of May 2024**

|                                | Gender Identity | Race and/or Ethnicity | Nationality | <i>The Daily Wire</i> affiliation | Followers (May 2024) | Videos (May 2024) |
|--------------------------------|-----------------|-----------------------|-------------|-----------------------------------|----------------------|-------------------|
| Posie Parker / Kellie Jay Keen | cis woman       | white                 | British     | N                                 | 120,000              | 580               |
| Matt Walsh                     | cis man         | white                 | American    | Y                                 | 2.86 million         | 4,000+            |
| Ben Shapiro                    | cis man         | Jewish                | American    | Y                                 | 6.86 million         | 5,900+            |
| Blaire White                   | trans woman     | mixed race            | American    | N                                 | 1.42 million         | 349               |
| Candace Owens                  | cis woman       | Black                 | American    | N (former)*                       | 3.19 million         | 1,900+            |

*Note.* Owen's joined Daily Wire in 2021 (Murray 2023) but confirmed her departure in March 2024 after clashes with Daily Wire's co-founder, Ben Shapiro. The feud reportedly centered around Owens antisemitism and opposition to US funding to Israel (Pengelly 2024).

background, as well as their relationship to the conservative media outlet, *The Daily Wire* summarized in Table 1. While four of the five commentators selected were based in the United States, we also chose to include Kelly Jane Keen based in the United Kingdom. First, as a so-called “trans exclusionary radical feminist” (TERF), Keen's online footprint is entirely transphobic, and all their content is centered on trans exclusion. Second, Keen has an interesting interaction with the American commentators selected, having either been interviewed by them on their channels, or their taglines, such as “adult human female” being cited on the other channels. Keen's inclusion is an acknowledgement of the overlap between the American far right and TERF ideologies and their intersecting claims on this issue (Dickey 2023). As right-wing commentators and transphobia are becoming increasingly networked and global, we were interested if content produced in the United Kingdom was using the same or distinct narratives then their US counterparts.

### Building the Content Classifier and Dataset

The mainstreaming of transphobia has led to an increase in online transphobic content (Brandwatch & Ditch the Label 2020; Lokmanoglu et al. 2023). The sheer increase in the amount of data that can be analyzed poses significant challenges for research design. Large datasets, especially those sourced from platforms like YouTube, demand extensive storage, computational power, and timely analysis (Nguyen and Liaw 2020). The potential vastness of data—including thousands of videos and millions of comments—can rapidly exhaust available resources, necessitating strategic data management (Bender and Friedman 2018; Menghani 2023; Zhou et al. 2017). Importantly, there are also daily limits on the amount of data that can be scraped using the YouTube API that need to be considered when determining the size of your dataset (Google for Developers 2024). In this methodology, we outline how we collected our data and determined the size of our original dataset.

### DistilBERT

We developed a content classifier utilizing the Large Language Model (LLM) program DistilBERT (Sanh et al. 2020) for text preprocessing and a Support Vector Machine

(SVM) for classification. A LLM is a machine learning model that can analyze and understand large amounts of text data, as well as generate new content. We used DistilBERT to conduct natural language processing (NLP) where we asked the model to label our comments according to content. We selected DistilBERT due to its efficiency and accuracy. It retains 97% of the previous language model BERT's language understanding while being faster and smaller, making it more effective at processing large datasets. DistilBERT also has a transformer-based architecture that enables it to capture subtle nuances within language, essential for detecting implicit transphobia (Salawu, Lumsden, and He 2021). Recent studies and work on multilingual and context-aware models highlight DistilBERT's success in detecting both explicit and subtle hate speech (C. Li et al. 2024; Rajendran et al. 2022; Sofat, Gill, and Bansal 2022). Additionally, benchmarks like HateXplain confirm its effectiveness in analyzing complex social media data, making it particularly suited to detecting nuanced transphobic narratives in this study (Mathew et al. 2022).

To improve the effectiveness of content classifiers, it is important to train them with labelled datasets. Fine-tuning classifiers with annotated data enhances model performance by enabling adaptation to specific tasks or domains, resulting in improved efficacy compared to using standalone pre-trained models. Moreover, this approach facilitates task specificity, allowing models to be tailored to particular domains or applications and accelerates convergence, expediting model adaptation to new tasks (Rajendran et al. 2022; Salawu, Lumsden, and He 2021). However, due to the intersectional nature of this project and the limited computational research on identifying hate speech, we built our own schema of labels and manually labeled training data. DistilBERT performs most effectively when the training data is relevant to the project. Our training data consisted of 5,450 comments from a wide range of YouTube channels with a focus on videos that were known to contain transphobic content.

### **Dataset Extraction**

To build our original dataset, we utilized YouTube API to scrape the top 1,000 comments from the top ten videos across the five selected channels (Google for Developers 2024). Where a video had less than 1,000 comments, all comments were scraped. This resulted in a dataset of 43,555 extracted via the YouTube API and 5,450 comments from our training data, resulting in a total of 49,005 comments in total. By prioritizing the top comments and their replies, the study focused on content with high levels of engagement, which is often indicative of deeper audience involvement and greater ideological influence (Google for Developers 2024). This approach aligns with previous research on engagement dynamics and content dissemination, which underscores the importance of high-engagement content in fostering online communities and amplifying extremist narratives (Mamié, Horta Ribeiro, and West 2021).

### **Label Selection**

In line with our intersectional understanding of transphobic content, we designed our data labels to capture multiple categories of hate drawing on social media regulatory guidelines, existing research, and expert guidance. First, a thorough examination of YouTube's user guidelines was conducted to identify prevalent themes within the platform's "violent or dangerous" content categories (YouTube, n.d.). This was followed

**Table 2. Summary of Data Labels According to Type**

| Nature                 | Target      | Category of Concern |
|------------------------|-------------|---------------------|
| Cyberbullying          | Gendered    | Anti-establishment  |
| Defamation             | Homophobic  | Conspiracy Theory   |
| Harassment             | Racist      | Counter Narrative   |
| Hate Group Promotion   | Religious   | Education           |
| Hate Speech            | Transphobic | Generic             |
| Incitement of Violence | Xenophobic  | Healthcare          |
| Stereotyping           |             | Political           |

by a comprehensive literature review on hate speech and extremism, which helped refine the label selection (Colliver 2023, 412; Ryyänen, Kosonen, and Ylönen 2022; Tsirbas and Zirganou-Kazolea 2025). Expert consultation with scholars in extremism research further validated the label choices, ensuring they reflect current online hate speech trends. Experts highlighted covert hate, such as dog-whistle rhetoric, and the compounded discrimination faced by marginalized groups like trans women of color (Haney-López 2014). This approach allowed the model to detect both explicit and subtle transphobia within broader discriminatory narratives (Collins and Bilge 2016; Frischlich et al. 2021; Lamble 2008; McCall 2005), enhancing its analysis of transphobic discourse through an intersectional lens.

We did not employ any strict definition of transphobia or set of particular transphobic terms but relied on the broader conception of transphobia as “any negative attitudes (hate, contempt, disapproval) directed toward trans people because of their being trans” (Bettcher 2014, 249). We included language and narratives that advocated for the removal of trans rights, misrepresentation and misgendering of trans people, abuse, violence, desired exclusion from social spaces, and other expressions of discrimination (TransActual, n.d.). We organized our labels into three main categories: the nature, target and category concern. The labels are summarized in Table 2 and defined in depth in Appendix A.

The model exhibited an F1 score of 0.922, indicating excellent accuracy comparable to other classification research (Nirbhik and Kumar 2023; Rajendran et al. 2022; Sofat, Gill, and Bansal 2022). The classifier identified 7,965 comments likely containing transphobic content. The following section categorizes the identified comments into the core mainstreamed narratives identified and measures their prominence in the dataset. To compare user comments against the content of videos, we also systematically reviewed each video. Of the fifty videos analyzed, thirty-three were identified as containing transphobic content, enhancing the accuracy of the broader findings.

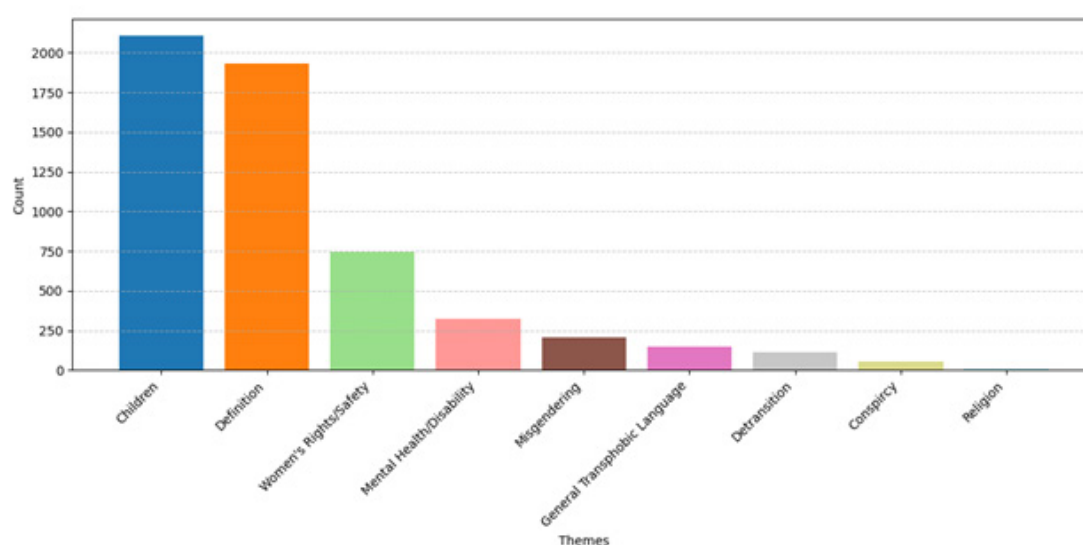
## **“JUST ASKING QUESTIONS”: MAINSTREAMED NARRATIVES OF TRANSPHOBIC CONTENT**

This section outlines the findings of our analysis. This section includes an overview of the transphobic narratives and language identified by our content classifier across our dataset. In the dataset of 49,005 comments, a total of 7,965 comments were iden-

tified as transphobic. We manually reviewed the transphobic comments to identify key themes associated with them. These themes are not exhaustive, nor mutually exclusive, as many comments encompassed multiple narratives. To determine the narratives prominence, we performed a simple count of the frequency of all identified terms (Appendix C) within our comments presented in Figure 1. Importantly, while the classifier was a useful tool, it was not flawless—some comments flagged as transphobic were not, and others with clear transphobic content went undetected. However, overall, the classifier has important potential for automating the detection of nuanced hate speech in large datasets (Poletto et al. 2021; Vidgen and Derczynski 2020).

Critically, the fact that the transphobic content identified by the classifier predominantly related to trans women reflects the broader offline discourse around gender (Marwick and Lewis 2017). This consistency between online and offline narratives suggests that online platforms, such as YouTube, play a crucial role in amplifying and normalizing transphobic discourse (Lewis 2018). A key observation was the lack of evidence used to substantiate claims in these comments, which mirrors well-documented patterns of prejudice where unfounded fears are used to marginalize groups (Mondon and Winter 2020).

While comments employed language that denigrates and invalidates the existence and experiences of trans people, there was clear evidence of mainstreaming through the absence of particular types of speech. Most notably, there were minimal calls for violence and reference to hate groups. Instead, there was a pattern in the use of the language of concern and care, often framed as users asking legitimate questions about safety. As demonstrated below, much of this care was embedded in ideas of gender essentialism. Gender essentialism—claims that gender is a category that is fixed, natural and discrete—requires an incorrect invocation of biology to claim that there are only two genders based on chromosomes and equating sex and gender (Atwood, Morgenroth, and Olson 2024). Gender essentialism is also grounded in stereotypes of gendered roles for men and women, including the need to protect perceived vulnerable groups. In the case of transphobic content, the perceived vulnerable groups are women and children.



**Figure 1. Summary of the number of transphobic narratives found in comments**

## Children

As expected, narratives around children, especially child safety, form the bulk of the transphobic content (over 36%). Children, as outlined by Dickey (2023, 36), act “as a powerful, deracialised, medicalised symbol, contrasted with the particular deviance ascribed to transfeminine people of colour.” It is also here that mainstreaming of language is most evident. While there is a section of comments that frame trans people as a threat to children and the provision of gender affirming care as akin to torture or mutilation, more common are comments that express opinions framed as concern—rather than hate or disgust—around children’s safety. These commenters usually seek to position themselves as concerned parents—for example, utilizing identity markers such “as a mother”—to legitimize their comments and do not identify their position as transphobic. Instead, as with comments surrounding women’s rights, commenters presented themselves as targeted and labelled transphobic for expressing what they claim as legitimate concerns (Colliver 2021). By positioning themselves as protectors, commenters avoid accusations of transphobia while perpetuating moral panic around gender identity (Mondon and Winter 2020). We identified three main themes among transphobic content including children: child abuse, medical abuse, and education.

### *Child Abuse*

The most common narrative focused on children was child safety from abuse. There are two narratives centering on child abuse. First, is that trans people pose a threat to children’s safety. As has been well reported, transphobic language often claims trans people are predators and a threat to children (Locantore and Wasarhaley 2020; Stone 2018; Williams 2020). Commenters drawing on these narrative deployed common tropes including grooming and fetishes. Second, was child abuse by parents. Comments framed parents that were supportive of gender affirming care as attention seeking, including accusations of having the medical condition Munchausen by proxy. The comments implied that parents were either seeking attention for themselves or that they were homophobic, preferring a trans straight child to a gay or lesbian child. Commenters frequently noted that parents should be reported to protective services for child abuse and children removed from their care.

These narratives reflect a strategic use of moral panic to frame trans individuals and supportive parents as societal threats. By invoking familiar tropes of child abuse, transphobic rhetoric taps into deep-seated fears about child safety, echoing historical patterns where marginalized groups are demonized (Cohen 2011). Accusations against parents of seeking attention or preferring a “straight” child exploit anxieties around parenting and medical authority, weaponizing these concerns to undermine the legitimacy of gender-affirming care (Hameleers 2020).

### *Medical Abuse*

Commentors conceptualized gender affirming medical care almost exclusively as surgical procedures or the use puberty blockers. While not all commenters were specifically referring to children when discussing medical care, they were the vast majority. Comments called surgery medical torture and utilized the language of mutilation and sterilization. In addition to criticizing doctors and parents for the provision of medical care, commenters also framed gender affirming care as a conspiracy by “big phar-

ma” for profits. Commenters reasoned that an increase in the number of children and adults accessing gender affirming care was a result of doctors wanting to profit from what was framed as lifelong medical care. Users described gender affirming care as an “industry”. Commenters also cited their opposition to puberty blockers and countered claims that the effects are reversible. There was no evidence to support these claims. Such disinformation not only delegitimizes gender-affirming treatments but also fuels real-world violence against medical facilities (Gzesh et al. 2024). Most recently, these mainstreamed narratives have resulted in the coordinated targeting of medical facilities that provide gender affirming care to children, including bomb threats (see, for example, Gupta et al. 2023; Gzesh et al. 2024; Yang 2022).

### ***Education***

As reflected in offline protests around drag story hours at libraries and the banning of books by school boards (Camiscoli et al. 2024; Coste 2024; Davis and Kettrey 2022; Ellis 2022; Squirrell and Davey 2023; Tylenda 2024), education was a consistent theme in transphobic comments including children. Comments centered on schools being a site of indoctrination and brainwashing, regardless of if there was any provision of education around gender or sexuality. Commenters accused schools of forcing teachers to accept children’s self-identification, often utilizing children’s playful self-identification as animals as evidence of children’s inability to understand their own identity. For parents that did not support any kind of education that featured discussions of gender, commenters threatened to send their children to religious schools or homeschool their children. There were also comments that sought to prevent anyone who is trans from teaching, working at, or being proximate to schools. However, comments also used this narrative against anyone that they perceived as progressive or non-conforming, often noting that someone’s colored hair or tattoos made them inappropriate to work with children.

### **Defining Sex and Gender**

The second dominant theme in comments (over 33%) focused on defining the term “woman” as a tactic to exclude trans women. It was in these comments where gender essentialism was most clearly deployed with incorrect invocations of biology that there are only two genders based on chromosomes and equating sex and gender (Atwood, Morgenroth, and Olson 2024). Many also quoted the google definition of woman weaponized by Keen—“Adult Human Female”—as means of exclusion. This focus on definitions was also likely heightened by the inclusion of comments from Matt Walsh’s channel as several of their top ten videos are excerpts from his documentary titled, “What is a Woman?” This narrative also focused on reproduction: users’ definition of women relies on the fact that women can fall pregnant. While many commenters claim that this does not exclude women who face fertility issues, they claim that it does exclude trans women. Comments also sought to make the distinction between cis people and trans people using terms such as “real men” or “real women” that reinforced narratives of exclusion.

Gender essentialism also extends beyond beliefs about biology to the roles of gender within society. Discussions of defining women were inherently entangled with stereotypes of women as mothers and carers. While trans women and their exclusion

were the core target of these comments, they also excluded anyone that commenters perceived as falling outside of these gendered stereotypes.

### *Women's Rights and Safety*

Trans exclusionary radical feminist (TERF) discourse was prevalent in the dataset at almost 13% of labelled comments, particularly through narratives that positioned trans women as a threat to cis women's rights. This reflects a broader strategy within TERF ideology, which frames trans inclusion as undermining the safety and rights of cis women (Hines 2018; Pearce 2018). The recurring use of Keen's slogan "Let Women Speak" in comments underscores how TERF discourse utilizes language of victimhood to claim that cis women are being silenced. In these comments, we observe what Thurlow describes as a "linguistic pivot from 'anti-trans' to 'pro-woman'" (Thurlow 2024). This tactic capitalizes on historical feminist struggles for women's spaces to justify exclusion (Lamble 2008). The division of comments into women's spaces, appropriation, and sports exemplifies how TERF rhetoric selectively focuses on areas where cis women's safety and identity are portrayed as most vulnerable (Mason 2023). As seen in the comments about children, rather than expressed through vitriol or hatred, trans-exclusion was made to appear more acceptable by justifying comments through the language of safety and fairness.

### *Women Only Spaces*

Transphobic comments frequently centered on the inclusion of trans women in women-only spaces, such as bathrooms, changing rooms, hospital wards, and prisons. These comments often portrayed trans women as predatory men attempting to infiltrate these spaces to harm cis women, a narrative rooted in fear and moral panic rather than evidence (Gwenffrewi 2022; Quatrini 2022). Citing instances of trans women in prisons convicted of sexual offenses, commenters framed women-only spaces as hard-fought protections that are now under threat from trans inclusion. This fear-based rhetoric, which has gained traction through prominent figures like J.K. Rowling, reinforces exclusionary ideologies (Pearce 2018). Such narratives, although disconnected from statistical reality, draw on misinformation or hypothetical scenarios to exacerbate the idea of an imagined threat.

### *Appropriation*

Another prominent narrative framed trans women as appropriating womanhood, with commenters arguing that trans women disrespected cis women's experiences, particularly in relation to reproductive health, motherhood, and gender-based discrimination. This narrative reinforced biological essentialism, the idea that womanhood is inherently tied to biology, by excluding trans women from the category of "real" women (Bettcher 2014; Hines 2018). Commenters also likened trans women to impersonators, reducing their identities to "fancy dress" or performance. Additionally, resistance to gender-inclusive terms like "menstruators" or "birthing people" illustrated a perception that expanding gender language diminishes cis women's experiences (Ging 2019). This rhetoric, which invalidates trans identities, fuels a zero-sum approach to gender rights, where any advancement for trans people is falsely seen as a threat to cis women.

## **Sport**

Unsurprisingly, the participation of trans women in sports is also a key narrative within our dataset. While some commenters noted that there could be a third category within sports to include trans women and maintain fairness, most saw trans women's participation as an attempt for men to infiltrate women's sports. Like the controversy around Khelif and Lin in the Paris Olympics, comments framed these discussions around issues of equality and fairness. A core element of gender essentialism as applied to trans participation in sport is the concept of "fairness." These arguments rely on perceived physical differences between cis men and women and the supposed "biological advantage" of male athletes (Atwood, Morgenroth, and Olson 2024). The infiltration of these ideas into the mainstream is clear in reporting and legislation. For example, as of 2023, there were at least 60 bills across the United States dedicated to excluding trans people from participating in sports with a focus on banning trans women from competing in women's sport (Atwood, Morgenroth, and Olson 2024).

## **Mental Health and Disability**

A clear theme within the comments was equating of being trans with mental health and disabilities (over 5% of labelled data). Comments that link being trans to mental health and disability pathologizes trans identities, framing them as psychological disorders and comments claim to be motivated by care and concern (Frischlich et al. 2021). Commenters often used derogatory language such as "crazy" or "insane" to imply that being trans was the result of a mental health condition. There were also frequent references to trans young people as "needing help" or being narcissists and attention seekers. The frequency of this theme indicates that it needs further exploration. For example, one channel connected autism spectrum disorder and being trans in both the video and comments. However, many of these comments, without the context of the video will not be identified as transphobic by a classifier.

## **Misgendering or Critiquing Pronouns**

Comments also deliberately misgendered trans people or mocked the use of pronouns. However, this was the most difficult form of transphobic language to label as it usually relies on the context of the video. Consequently, these comments only make up 3.5% of the data. When misgendering was explicit, it was usually the use of two pronouns as a single word—for example "heshe"—or the use of combination of pronouns to denote that the commenter disagrees with someone's preferred pronouns—"he/she" or "his/her." Commenters also mocked the use of pronouns by self-identifying as objects or people and making up pronouns. For example, one commenter claimed they identified as an ambulance and used the pronouns woo/woo or Michael Jackson and hee/haa. Misgendering serves as a direct and pervasive tool of transphobia, with the data reflecting the frequent use of incorrect or derogatory pronouns to dehumanize trans individuals (Lamble 2008).

Another common narrative was dismissing trans people's identity based on their conformity to gendered stereotypes. For example, Blair White, a trans woman, often criticizes trans people for their appearance as either cheap or uncommitted to appearing like a "real woman," or in the case of nonbinary people, not being trans. This view reflects contestations within certain parts of the trans community around who

counts as trans (Jacobsen, Devor, and Hodge 2022). There was also hyper fixation on body hair among comments, especially trans nonbinary people or trans women who choose to have facial hair. These comments were often mocking and aimed to discredit individuals gender identity. As outlined by Billard (2019, 464), this notion of “passing” presents a double-edged sword for the trans community: while “passing” legitimates a transgender person’s claim to their gender identity... [it] also renders them more malicious in their deception.”

## **Religion**

Religious-based transphobia often invokes theological justifications for rejecting trans identities (Campbell, Hinton, and Anderson 2019; Crenshaw 1989). There were two core themes around transphobia and religion within comments that featured in 3.45% of our frequency count. The first was the denial of the existence of trans people under the premise that God created people as man and woman. Commenters criticized those who “play god” or violate God’s plan by not identifying with the sex that they were assigned at birth. Narratives relied on the idea that God creates everyone perfectly and framed being trans as a violation of the bible. Most of these comments referenced a Christian God with only two references to Islam in the comments identified as transphobic. Second was the use of religious language to denigrate the trans community. Narratives included references to the trans community as a cult, a creation of the devil, Luciferian, and demons. There were also a limited number of references to activists and trans individuals belonging in hell and calls to “cast them out.” As highlighted by Blyth and McRae, grounding transphobia in theological and biblical languages creates a “dangerous power to grant divine mandate to transphobia and trans-exclusionary practices” (Blyth and McRae 2018, 113).

## **Transphobic Terms**

Comments also contained transphobia, however, these unexpectedly only appeared in around 150 comments (2.5%). There are, however, issues with using these terms to identify transphobic content. Many historically transphobic terms are used within the trans community as a form of reclamation in a way that is not necessarily derogatory (Edmondson 2021). The classification of these terms as transphobic risks restricting self-expression within the trans community. There are some terms—such as “trans-trender” or “transgenderism”—that were more clearly indicate transphobic content. One of the strongest indications of the mainstreaming of transphobic content is the lack of explicit transphobic content—especially in contrast to the other narratives identified.

## **Detransitioning**

Comments also discussed detransitioning (just under 2% of terms analyzed). These comments largely came from a video with Blaire White who interviewed someone who identified as having detransitioned. These comments were framed as concern for the safety and wellbeing of those who “make a mistake” in their decision to transition and that people—especially young people—will regret their decisions in the future. However, narratives were usually motivated by the perceived increase in people identifying as trans being driven by a “trend” and not gender dysphoria. These comments do

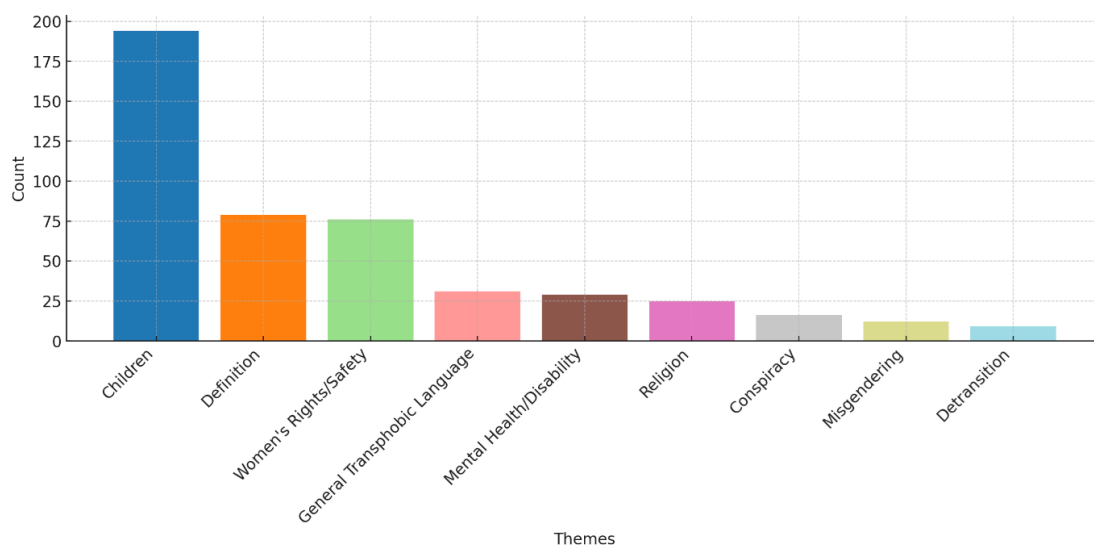
not usually deny the existence of gender dysphoria—especially those commenting on Blaire White’s videos who openly discusses her own experiences of gender dysphoria—but that most individuals are transitioning for social reasons including to fit in or because of social exclusion. One commenter compared being trans to goths: where social outcasts used to become goths, they were now choosing to become trans. Those who detransition are often used by commentators as evidence of the illegitimacy of trans people and the inherent risks of gender affirming care. These narratives contradict research on the experiences of individuals that receive gender affirming care that suggest an extremely low level of regret among trans patients after gender affirming surgery (Bustos et al. 2021), improved health outcomes for trans people (Swan et al. 2023), and that those who do detransition often do not regret accessing gender affirming health care (MacKinnon et al. 2022). What was interesting is that these comments did not use the language of “Rapid Onset Gender Dysphoria” or ROGD made popular in the much-critiqued paper published by Littman and later the book by Abigail Shrier that galvanized the anti-trans movement (Ashley 2020; Littman 2018; Jordan B Peterson Clips 2021; Restar 2020). In line with mainstreaming, this debate has moved away from an original source, but the ideas continue to be reflected in discourse.

## **Conspiracy**

Zuboff highlights that conspiracy theories serve to radicalize public discourse by linking marginalized groups to broader fears of societal collapse or manipulation (Zuboff 2019). Despite our labelled data containing a number of comments containing conspiracy theories including Joe Biden being the puppet of Obama, that China was responsible for political discontent within America, and a collection of references to global control, including the Great Reset Theory, the World Economic Forum, and antisemitic references to media and Hollywood, the classifier performed poorly in identifying comments that contained transphobia and conspiracy theories (falling under one percent in our count). Instead, the classifier identified comments that implied that being trans or the trans community was itself a conspiracy. The classifier identified comments containing terms such as “control”, “follow the money,” “wake up,” and “lie” as containing a conspiracy. There were a limited number of comments identified that contained references to actual conspiracy theories, including (but not limited to) references to the increased visibility of trans people and rights being a tool of distraction by global elites—a global syndicate, the UN, antifa, government, the left—who are pursuing depopulation, a ploy by big pharma for profits, and promoting transhumanism.

### **Comparative Analysis of Videos and Comments**

In addition to examining the comments of each video, we also reviewed the video content itself to see if video content and comments deviated in content. What we found was a more strategic deployment of transphobic content and language within the videos. Videos often use moderated language, such as “predator,” avoiding more explicit terms like “pervert” or “pedophile” commonly seen in comments. The absence of slurs in videos highlights creators’ efforts to evade deplatforming, while still engaging fringe followers through coded language and allowing the comment section to carry more direct hate speech (Ging 2019; Nagle 2017). In contrast, the anonymity of comment sections enables more explicit transphobia, creating spaces for unfiltered expressions of hate. This dual strategy—sanitized narratives in videos and more overt



**Figure 2. Frequency of transphobic narratives in selected videos**

hate in comments—facilitates the mainstreaming of transphobia, normalizing discriminatory discourse without immediate repercussions (Frischlich et al. 2021). This contributes to the broader embedding of transphobic ideologies into public discourse under a guise of legitimacy (Zuboff 2019).

This approach is likely motivated by a desire to broaden appeal and maintain monetization through sponsorships, given that overt hate speech risks demonetization or removal (Lewis 2018; Marwick and Lewis 2017). In many cases, transphobic content was reduced to a single, subtle phrase embedded within longer videos, allowing creators to convey their messages while minimizing the risk of censorship. Additionally, some creators pre-emptively limit the content they post on YouTube, opting to upload more explicit material to alternative platforms, framing this as a response to alleged censorship and suppression of free speech. This narrative of repression serves to galvanize their audience while maintaining plausible deniability on mainstream platforms (Frischlich et al. 2021).

## DISCUSSION

As stated by Sadjadi, “Transgender people are a lightning rod for fearmongering” (Sadjadi 2020). The transphobic comments and narratives identified within our dataset contain a spectrum of framing from care—that situate their concerns in the protection of other vulnerable groups—to explicit hate speech but almost all comments were based in mis- or disinformation. Most comments were framed in protective or care-centered language. Commenters claimed to just want to be able to ask questions and discuss the potential harm—even if based in disinformation—of the increased visibility of trans people. In line with mainstreaming, few comments were captured in our first category our labels—nature of speech—including cyberbullying, defamation, harassment, promotion of hate groups, hate speech or incitement of violence. The only label in this category that frequently applied was stereotyping. This stereotyping did not only apply to the trans community but also the role of women and men, femininity and masculinity.

This finding can be explained in one of two ways. First, YouTube is effectively moderating these types of speech or, second, commentators are actively self-censoring and using dog-whistles and coded language to avoid content moderation.

While the narratives identified were not surprising—these reflect broader narratives found in mainstream media and reporting on transphobia—what was notable was the limited comments that were identified as both transphobic and racist or xenophobic. It is possible that racist and xenophobic comments are also employing the same strategies to avoid detection and appeal to a wider audience and the classifier—being trained more prominently on identifying transphobia—struggled to identify more than one category of mainstreamed hate. As this data was collected in May 2024, it would also be interesting to see how the prominence of racist and xenophobic narratives has changed in response to comments made during the 2024 US Presidential campaign. Trump's claim that Vice President Harris “wants to do transgender operations on illegal aliens that are in prison” along with other anti-trans rhetoric risks increasing transphobia further (PBS 2024).

What was also notable is that while our labels covered some of the narratives including healthcare and education that we identified as part of our narratives, they did not match our overarching narrative themes. The identification of these narratives provides an opportunity to further refine our labels for more accurate comment categorization in the future.

## CONCLUSION

This paper makes important contributions to understanding the process of mainstreaming transphobia and understanding the spread of transphobic narratives in media, legislation and broader society. As aptly summarized by Aly, studies examining the impact and role of online terrorist and violent extremist content (TVEC) are “often based on the assumption that the violent extremist narrative works like a magic bullet to radicalize audiences” (Aly 2017, 62). What research on mainstreaming suggests is that extreme or fringe content in and of itself is not necessarily appealing to users but needs to be reframed to make it palatable to larger audiences. What is evident in our findings is that the narratives identified are not necessarily built around hate speech and incitement but the perceived threat of trans people, especially trans women, pose to emotive issues of another vulnerable group: women and children. It is evident throughout that the arguments and framing of this threat are not supported by evidence but have been socially constructed among commentators and framed to appear legitimate.

We identify three core implications from this project. First, our content classifier proved capable of automatically detecting transphobic content online. Social media companies and researchers can adapt this tool to improve content moderation policies, as well as to extract and examine transphobic content on a larger scale. The classifier was less accurate categorizing certain comments, namely, the over identification of any comment about trans people as transphobic and struggling to correctly identify conspiracy theories or deliberate misgendering. Interestingly, while the intersection of race and transphobia is well understood (Colliver 2022; Gill-Peterson 2018; Koyama 2020; Upadhyay 2021), the content classifier did not identify many cases of this inter-

action. The causes here may be two-fold. Racism online, like transphobia, has also undergone a process of mainstreaming. Commentators and their followers understand that explicitly racist content is likely to be removed and have developed a range of dog whistles and cloaked language that allows them to avoid detention. As we trained our model to detect nuanced incidences of transphobia, not racism, it is possible the model was not given enough information to accurately detect both forms of mainstreamed hate speech. With further development the classifier could be improved to increase detection of transphobic content online and research using large datasets that can be used for improved content moderation strategies

Second, as transphobic narratives, under the forces of mainstreaming and major political events, should be expected to change over time, this paper's dataset and model present a unique opportunity to track transphobic narratives. The case selection of the top comments against the top videos across the same channels can be replicated to identify how transphobia—and wider narratives of extremism—are evolving. Further, researchers can apply improved LLM models against the same data with the ability to compare effectiveness. This reduces challenges of online research where data is consistently changing, and the size of the potential data sources presents challenges for longitudinal and comparative analysis.

Finally, there are implications for research effectiveness and safety. The automated detection of transphobic content—and other forms of hateful and harmful content—presents an interesting opportunity for research ethics and safety. There is increasing focus on the potential harm to researchers from exposure to content during online research into fringe and extremist communities (for example, see recent research on the harm to researchers engaging in online research including Feuston et al. 2022; Gagnon and Mathieson 2023; Pearson et al. 2023;) We hope that the content classifier designed for this project can be expanded into an accessible tool that limits the exposure of researchers to hateful content as a part of their data collection. There is no way eliminate the harm of researching targeted hate—especially when the researcher is a member of the targeted community—but for researchers that seek to do the meaningful work of online content analysis at a large scale, there are ways of minimizing the risks.

There are however inherent flaws to pursuing queer and intersectional research using big dataset scrapped from social media. Feminist and queer scholars have noted that the disembodied and placeless nature of big data disconnects users and research from their language and context (Cooky, Linabary, and Corple 2018; Gieseeking 2018; Luka and Millette 2018; Leurs 2017). There will also always be a limit to the ability of content classifiers to capture the full reem of transphobic content. As experienced in the manual data labelling process, the researchers inherently knew some contents were transphobic due to context—intentionally misgendering people in the videos, referring to trans people as “crazy”—but could not be labelled as transphobic as without the context, there would be no way to know who the comment was directed towards. In this way, there will always be a need for in depth qualitative data analysis to complement big data projects. There is also a need to think carefully about how those members of the trans community, and the LGBTQI+ community more broadly, experience and are impacted by hateful online content. Their expertise needs to be acknowledged and centered in the construction of future research projects.

## REFERENCES

- Acha Ugarte, Beatriz. 2018. "The Far Right in Western Europe: 'From the Margins to the Mainstream' And Back?" *Cuadernos Europeos de Deusto*, no. 59: 75–97. <https://doi.org/10.18543/ced-59-2018pp75-97>.
- ACLU. 2024. "Mapping Attacks on LGBTQ Rights in U.S. State Legislatures in 2024." *American Civil Liberties Union*. <https://www.aclu.org/legislative-attacks-on-lgbtq-rights-2024>.
- ADL. 2023. "Year in Review: Anti-LGBTQ+ Hate & Extremism Incidents, 2022–2023." *Anti-Defamation League*. <https://www.adl.org/resources/report/year-review-anti-lgbtq-hate-extremism-incidents-2022-2023>.
- ADL. n.d. "Hate Group." *Anti-Defamation League*. <https://www.adl.org/glossary/hate-group>.
- Aly, Anne. 2017. "Brothers, Believers, Brave *Mujahideen*: Focusing Attention on the Audience of Violent Jihadist Preachers." *Studies in Conflict & Terrorism* 40 (1): 62–76. <https://doi.org/10.1080/1057610X.2016.1157407>.
- APA. 2023. "APA Dictionary of Psychology." *American Psychological Association*. <https://dictionary.apa.org/>.
- Ashley, Florence. 2020. "A Critical Commentary on "Rapid-Onset Gender Dysphoria."" *The Sociological Review* 68 (4): 779–99. <https://doi.org/10.1177/0038026120934693>.
- Atwood, S., Thekla Morgenroth, and Kristina R. Olson. 2024. "Gender Essentialism and Benevolent Sexism in Anti-Trans Rhetoric." *Social Issues and Policy Review* 18 (1): 171–93. <https://doi.org/10.1111/sipr.12099>.
- Baele, Stephane J., Lewys Brace, and Travis G. Coan. 2023. "Uncovering the Far-Right Online Ecosystem: An Analytical Framework and Research Agenda." *Studies in Conflict & Terrorism* 46 (9): 1599–1623. <https://doi.org/10.1080/1057610X.2020.1862895>.
- Bansal, Suhans, Naval Garg, Jagvinder Singh, and Freda Van Der Walt. 2024. "Cyberbullying and Mental Health: Past, Present and Future." *Frontiers in Psychology* 14. <https://doi.org/10.3389/fpsyg.2023.1279234>.
- Barr, Robert R. 2009. "Populists, Outsiders and Anti-Establishment Politics." *Party Politics* 15 (1): 29–48. <https://doi.org/10.1177/1354068808097890>.
- Bassi, Serena, and Greta LaFleur. 2022. "TERFs, Gender-Critical Movements, and Postfascist Feminisms." *TSQ: Transgender Studies Quarterly* 9 (3): 311–33. <https://doi.org/10.1215/23289252-9836008>.
- BBC News. 2018. "Woman Billboard Removed after Transphobia Row." *BBC News*, September 26. <https://www.bbc.com/news/uk-45650462>.
- BBC News. 2024. "JK Rowling and Elon Musk Reportedly Named in Imane Khelif Lawsuit." *BBC News*, August 14. <https://www.bbc.com/news/articles/cm2njjm4e2po>.
- Beacham, Greg. 2024. "Who Is Imane Khelif? Algerian Boxer Facing Gender Outcry Had Modest Success before Olympics." *AP News*, August 2. <https://apnews.com/article/olympics-2024-imane-khelif-boxing-paris-2e07d442ff-b29a61e09911884dcdaa9>.
- Ben Shapiro (@BenShapiro). n.d. *Ben Shapiro*. YouTube. <https://www.youtube.com/@BenShapiro>.
- Bender, Emily M., and Batya Friedman. 2018. "Data Statements for Natural Language

- Processing: Toward Mitigating System Bias and Enabling Better Science.” *Transactions of the Association for Computational Linguistics* 6: 587–604.
- Bettcher, Talia Mae. 2014. “Transphobia.” *TSQ: Transgender Studies Quarterly* 1 (1–2): 249–51. <https://doi.org/10.1215/23289252-2400181>.
- Billard, Thomas J. 2019. “‘Passing’ and the Politics of Deception: Transgender Bodies, Cisgender Aesthetics, and the Policing of Inconspicuous Marginal Identities.” In *The Palgrave Handbook of Deceptive Communication*, edited by Tony Docan-Morgan, 463–77. Cham: Springer. [https://doi.org/10.1007/978-3-319-96334-1\\_24](https://doi.org/10.1007/978-3-319-96334-1_24).
- Billard, Thomas J. 2023. “‘Gender Critical’ Discourse as Disinformation: Unpacking TERF Strategies of Political Communication.” *Women’s Studies in Communication* 46 (2): 235–243. <https://doi.org/10.1080/07491409.2023.2193545>.
- Billard, Thomas J. 2024. “The Politics of Transgender Health Misinformation.” *Political Communication* 41 (2): 344–52. <https://doi.org/10.1080/10584609.2024.2303148>.
- Billings, Andrew C., Leigh Moscovitz, Joshua R. Jackson, Emily Dirks, and Spencer M. Tomsett. 2024. “Trans Youth Sport Bans and the Facilitation of Moral Panic: A Cross-Platform Comparison of 2022 Media Narratives.” *Mass Communication and Society* 27 (6): 1555–79. <https://doi.org/10.1080/15205436.2024.2313109>.
- Billson, Chantelle. 2023. “Who Is Posie Parker, the Anti-Trans Activist Whose Protests Have Attracted Neo-Nazis?” *PinkNews*, March 22. <https://www.thepinknews.com/2023/03/22/posie-parker-kellie-jay-keen-minshull-anti-transgender-group/>.
- Bleakley, Paul. 2023. “Panic, Pizza and Mainstreaming the Alt-Right: A Social Media Analysis of Pizzagate and the Rise of the QAnon Conspiracy.” *Current Sociology* 71 (3): 509–25. <https://doi.org/10.1177/001139212111034896>.
- Blee, Kathleen. 2020. “Where Do We Go from Here? Positioning Gender in Studies of the Far Right.” *Politics, Religion & Ideology* 21 (4): 416–31. <https://doi.org/10.1080/21567689.2020.1851870>.
- Blyth, Caroline, and Prior McRae. 2018. “‘Death by a Thousand Paper Cuts’: Transphobia, Symbolic Violence, and Conservative Christian Discourse.” In *Rape Culture, Gender Violence, and Religion: Interdisciplinary Perspectives*, edited by Caroline Blyth, Emily Colgan, and Katie B. Edwards, 111–33. Cham: Springer. [https://doi.org/10.1007/978-3-319-72224-5\\_7](https://doi.org/10.1007/978-3-319-72224-5_7).
- Brandwatch & Ditch the Label. 2020. *Exposed: The Scale of Transphobia Online*. Brighton: Brandwatch. <https://www.brandwatch.com/reports/transphobia/>.
- Brown, Katy, Aurelien Mondon, and Aaron Winter. 2023. “The Far Right, the Mainstream and Mainstreaming: Towards a Heuristic Framework.” *Journal of Political Ideologies* 28 (2): 162–79. <https://doi.org/10.1080/13569317.2021.1949829>.
- Bustos, Valeria P., Samyd S. Bustos, Andres Mascaro, Gabriel Del Corral, Antonio J. Forte, Pedro Ciudad, Esther A. Kim, Howard N. Langstein, and Oscar J. Manrique. 2021. “Regret after Gender-Affirmation Surgery: A Systematic Review and Meta-Analysis of Prevalence.” *Plastic and Reconstructive Surgery Global Open* 9 (3): e3477. <https://doi.org/10.1097/GOX.0000000000003477>.
- Camiscoli, Sarah M., Paige Duggins-Clay, Maryam Salmanova, and Ibtihal Chamakh. 2024. “Youth Dignity Takings: Understanding and Restoring the Involuntary Property Loss of Book Bans and Trans Bans.” *Loyola Interdisciplinary Journal of Public Interest Law* 1 (1): 1.

- Cammaerts, Bart. 2018. "The Mainstreaming of Extreme Right-Wing Populism in the Low Countries: What Is to Be Done?" *Communication, Culture and Critique* 11 (1): 7–20. <https://doi.org/10.1093/ccc/tcx002>.
- Campbell, Marianne, Jordan D. X. Hinton, and Joel R. Anderson. 2019. "A Systematic Review of the Relationship between Religion and Attitudes toward Transgender and Gender-Variant People." *International Journal of Transgender Health* 20 (1): 21–38. <https://doi.org/10.1080/15532739.2018.1545149>.
- Candace Owens Podcast. 2024. "I'm DONE Being Threatened. The Michael Jackson & Diddy Connection EXPOSED." YouTube video. <https://www.youtube.com/watch?v=crpyOfjVcjw>.
- Cardwell, Hamish. 2023. "Spike in Online Hate toward Trans Community after Posie Parker Visit—Researchers." RNZ, April 4. <https://www.rnz.co.nz/news/national/487306/spike-in-online-hate-toward-trans-community-after-posie-parker-visit-researchers>.
- Chakravarthi, Bharathi Raja. 2024. "Detection of Homophobia and Transphobia in YouTube Comments." *International Journal of Data Science and Analytics* 18 (1): 49–68. <https://doi.org/10.1007/s41060-023-00400-0>.
- Cohen, Stanley. 2011. *Folk Devils and Moral Panics*. London: Routledge.
- Collins, Patricia Hills, and Sirma Bilge. 2016. *Intersectionality*. Malden, MA: Polity.
- Colliver, Ben. 2021. "Claiming Victimhood: Victims of the 'Transgender Agenda.'" In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*, edited by Jane Bailey, Asher Flynn, and Nicola Henry, 189–204. Leeds: Emerald. <https://doi.org/10.1108/978-1-83982-848-520211014>.
- Colliver, Ben. 2022. "Intersectional Oppression and Transgender People's Experiences of Discrimination." In *Contemporary Intersectional Criminology in the UK: Examining the Boundaries of Intersectionality and Crime*. Bristol: Bristol University Press.
- Colliver, Ben. 2023. "Responding to Transphobic Violence Online." In *The Routledge Companion to Gender, Media and Violence*, edited By Karen Boyle and Susan Berridge. New York: Routledge.
- Colliver, Ben, Adrian Coyle, and Marisa Silvestri. 2019. "The 'Online Othering' of Transgender People in Relation to 'Gender Neutral Toilets.'" In *Online Othering: Exploring Digital Violence and Discrimination on the Web*, edited by Karen Lumsden and Emily Harmer, 215–37. Cham: Springer. [https://doi.org/10.1007/978-3-030-12633-9\\_9](https://doi.org/10.1007/978-3-030-12633-9_9).
- Compton, Julie. 2019. "'Frightening' Online Transphobia Has Real-Life Consequences, Advocates Say." *NBC News*, November 22. <https://www.nbcnews.com/feature/nbc-out/frightening-online-transphobia-has-real-life-consequences-advocates-say-n1089456>.
- Conway, Maura. 2021. "Online Extremism and Terrorism Research Ethics: Researcher Safety, Informed Consent, and the Need for Tailored Guidelines." *Terrorism and Political Violence* 33 (2): 367–80. <https://doi.org/10.1080/09546553.2021.1880235>.
- Conway, Mekala. 2023. "Transgender Terror: An Analysis of Political Narratives in Tucker Carlson Tonight (2017–2022)." Master's thesis, Texas Tech University. <https://hdl.handle.net/2346/96363>.
- Cooky, Cheryl, Jasmine R Linabary, and Danielle J Corple. 2018. "Navigating Big Data

- Dilemmas: Feminist Holistic Reflexivity in Social Media Research.” *Big Data & Society* 5 (2): 2053951718807731. <https://doi.org/10.1177/2053951718807731>.
- Cornell, Svante E. 2018. “Erbakan, Kısakürek, and the Mainstreaming of Extremism in Turkey.” *Current Trends in Islamist Ideology* 23: 5–33.
- Coste, Françoise. 2024. “Drag Queen Story Hour and the Politics of Religious Outrage.” *American Fury: Essays on Moral Outrage in Culture and Politics*, edited by Myra Mendible, 172. Jefferson, NC: McFarland.
- Crenshaw, Kimberle. 1989. “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.” *University of Chicago Legal Forum* 1989 (1). <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>.
- Daniels, Jessie. 2018. “The Algorithmic Rise of the ‘Alt-Right.’” *Contexts* 17 (1): 60–65. <https://doi.org/10.1177/1536504218766547>.
- Davis, Alyssa J., and Heather Hensman Kettrey. 2022. “Clear and Omnipresent Danger: Digital Age Culture Wars and Reactions to Drag Queen Story Hour across Diverse Subreddit Communities.” *Social Currents* 9 (1): 25–44. <https://doi.org/10.1177/23294965211050019>.
- Devin, Leonard. 2023. “Can the *Daily Wire* Become a Right-Wing Entertainment Machine?” *Bloomberg*, May 3. <https://www.bloomberg.com/news/features/2023-05-03/daily-wire-tries-to-go-from-ben-shapiro-podcast-to-anti-disney>.
- Dickey, Briar. 2023. “Transphobic Truth Markets: Comparing Trans-Hostile Discourses in British Trans-Exclusionary Radical Feminist and US Right-Wing Movements.” *DiGeSt. Journal of Diversity and Gender Studies* 10 (2). <https://doi.org/10.21825/digest.85311>.
- DiMaggio, Anthony R. 2022. “Conspiracy Theories and the Manufacture of Dissent: QAnon, the ‘Big Lie’, Covid-19, and the Rise of Rightwing Propaganda.” *Critical Sociology* 48 (6): 1025–48. <https://doi.org/10.1177/08969205211073669>.
- Dowling, David O., Patrick R. Johnson, and Brian Ekdale. 2022. “Hijacking Journalism: Legitimacy and Metajournalistic Discourse in Right-Wing Podcasts.” *Media and Communication* 10 (3): 17–27. <https://doi.org/10.17645/mac.v10i3.5260>.
- Dunbar, Graham. 2024. “IOC Leader Condemns ‘Hate Speech’ Directed at Olympic Boxers.” *TIME*, August 3. <https://time.com/7007325/imane-khelif-lin-yu-ting-boxing-olympics-gender-fight-hate-speech/>.
- Durkheim, Emile. 2011. *Durkheim on Religion: A Selection of Readings with Bibliographies and Introductory Remarks*, edited by W.S.F. Pickering. Cambridge: James Clarke & Co.
- Earles, Jennifer. 2019. “The ‘Penis Police’: Lesbian and Feminist Spaces, Trans Women, and the Maintenance of the Sex/Gender/Sexuality System.” *Journal of Lesbian Studies* 23 (2): 243–56. <https://doi.org/10.1080/10894160.2018.1517574>.
- Edmondson, Daniel. 2021. “Word Norms and Measures of Linguistic Reclamation for LGBTQ+ Slurs.” *Pragmatics & Cognition* 28 (1): 193–221. <https://doi.org/10.1075/pc.00023.edm>.
- Ekman, Mattias. 2022. “The Great Replacement: Strategic Mainstreaming of Far-Right Conspiracy Claims.” *Convergence* 28 (4): 1127–43. <https://doi.org/10.1177/13548565221091983>.
- European Commission. n.d. *Identifying Conspiracy Theories*. <https://commission.europa.eu/anti-disinformation-centre/en/identifying-conspiracy-theories>.

[europea.eu/strategy-and-policy/coronavirus-response/fighting-disinformation/identifying-conspiracy-theories\\_en](https://europea.eu/strategy-and-policy/coronavirus-response/fighting-disinformation/identifying-conspiracy-theories_en).

- Feuston, Jessica L., Arpita Bhattacharya, Nazanin Andalibi, Elizabeth A. Ankrah, Sheena Erete, Mark Handel, Wendy Moncur, Sarah Vieweg, and Jed R. Brubaker. 2022. "Researcher Wellbeing and Best Practices in Emotionally Demanding Research." In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery. <https://doi.org/10.1145/3491101.3503742>.
- Fielitz, Maik, and Holger Marcks. 2019. "Digital Fascism: Challenges for the Open Society in Times of Social Media." Working Paper, Berkeley Center for Right-Wing Studies, University of California, Berkeley. <https://escholarship.org/uc/item/87w5c5gp>.
- Fischer, Mia. 2023. "Protecting Women's Sports? Anti-Trans Youth Sports Bills and White Supremacy." *Communication and Critical/Cultural Studies* 20 (4): 397–415. <https://doi.org/10.1080/14791420.2023.2267646>.
- Frischlich, Lena, Jens H. Hellmann, Felix Brinkschulte, Martin Becker, and Mitja D. Back. 2021. "Right-Wing Authoritarianism, Conspiracy Mentality, and Susceptibility to Distorted Alternative News." *Social Influence* 16 (1): 24–64. <https://doi.org/10.1080/15534510.2021.1966499>.
- Frischlich, Lena, Johanna Klapproth, and Felix Brinkschulte. 2020. "Between Mainstream and Alternative: Co-Orientation in Right-Wing Populist Alternative News Media." In *Disinformation in Open Online Media*, edited by Christian Grimme, Mike Preuss, Frank W. Takes, and Annie Waldherr, 150–67. Cham: Springer. [https://doi.org/10.1007/978-3-030-39627-5\\_12](https://doi.org/10.1007/978-3-030-39627-5_12).
- Gagnon, Audrey, and Nicola Mathieson. 2023. "Working with Extremist Content Online: Strategies to Improve Researcher Wellbeing." *RightNow!*, August 15. <https://www.sv.uio.no/c-rex/english/news-and-events/right-now/2023/working-with-extremist-content-online.html>.
- Gallaher, Carolyn. 2021. "Mainstreaming White Supremacy: A Twitter Analysis of the American 'Alt-Right.'" *Gender, Place & Culture* 28 (2): 224–52. <https://doi.org/10.1080/0966369X.2019.1710472>.
- Gambino, Lauren. 2023. "Anti-Trans Rhetoric Took Center Stage at CPAC amid Hostile Republican Efforts." *The Guardian*, March 7. <https://www.theguardian.com/us-news/2023/mar/07/cpac-anti-trans-rhetoric>.
- Garaigordobil, Maite, and Enara Larrain. 2020. "Acoso y ciberacoso en adolescentes LGTB: Prevalencia y efectos en la salud mental." *Comunicar: Revista Científica de Comunicación y Educación* 28 (62): 79–90. <https://doi.org/10.3916/C62-2020-07>.
- Gerstenfeld, Phyllis B., Diana R. Grant, and Chau-Pu Chiang. 2003. "Hate Online: A Content Analysis of Extremist Internet Sites." *Analyses of Social Issues and Public Policy* 3 (1): 29–44.
- Giesecking, Jen Jack. 2018. "Size Matters to Lesbians, Too: Queer Feminist Interventions into the Scale of Big Data." *The Professional Geographer* 70 (1): 150–56. <https://doi.org/10.1080/00330124.2017.1326084>.
- Gill-Peterson, Jules. 2018. *Histories of the Transgender Child*. Minneapolis: University of Minnesota Press.
- Ging, Debbie. 2019. "Alphas, Betas, and Incels: Theorizing the Masculini-

- ties of the Manosphere.” *Men and Masculinities* 22 (4): 638–57. <https://doi.org/10.1177/1097184X17706401>.
- GLAAD. 2021. “Ben Shapiro.” GLAAD, February 11. <https://glaad.org/gap/ben-shapiro/>.
- Gohal, Gassem, Ahmad Alqassim, Ebtihal Eltyeb, Ahmed Rayyani, Bassam Hakami, Abdullah Al Faqih, Abdullah Hakami, Almuhammad Qadri, and Mohamed Mahfouz. 2023. “Prevalence and Related Risks of Cyberbullying and Its Effects on Adolescent.” *BMC Psychiatry* 23 (1): 39. <https://doi.org/10.1186/s12888-023-04542-0>.
- Goldblum, Peter, Rylan J. Testa, Samantha Pflum, Michael L. Hendricks, Judith Bradford, and Bruce Bongar. 2012. “The Relationship between Gender-Based Victimization and Suicide Attempts in Transgender People.” *Professional Psychology: Research and Practice* 43 (5): 468–75. <https://doi.org/10.1037/a0029605>.
- Google for Developers. 2024. “YouTube Analytics and Reporting APIs.” *Google for Developers*. <https://developers.google.com/youtube/analytics>.
- Gotfredsen, Sarah Grevy. 2023. “Q&A: What Happened to Academic Research on Twitter?” *Columbia Journalism Review*, December 6. [https://www.cjr.org/tow\\_center/qa-what-happened-to-academic-research-on-twitter.php](https://www.cjr.org/tow_center/qa-what-happened-to-academic-research-on-twitter.php).
- Graham, Roderick. 2016. “Inter-Ideological Mingling: White Extremist Ideology Entering the Mainstream on Twitter.” *Sociological Spectrum* 36 (1): 24–36. <https://doi.org/10.1080/02732173.2015.1075927>.
- Grant, Jaime M., Lisa A. Mottet, J. J. Tanis, and D. Min. 2011. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey*. Washington, DC: National Center for Transgender Equality & National Gay and Lesbian Task Force.
- Gupta, Pranav, Ellis Barrera, Elizabeth R Boskey, Jessica Kremen, and Stephanie A Roberts. 2023. “Exploring the Impact of Legislation Aiming to Ban Gender-Affirming Care on Pediatric Endocrine Providers: A Mixed-Methods Analysis.” *Journal of the Endocrine Society* 7 (10): bvad111. <https://doi.org/10.1210/jendso/bvad111>.
- Gwenffrewi, Gina. 2022. “J. K. Rowling and the Echo Chamber of Secrets.” *TSQ: Transgender Studies Quarterly* 9 (3): 507–16. <https://doi.org/10.1215/23289252-9836176>.
- Gzesh, Ari S., Dana Prince, Scott K. Jelinek, Amy Hillier, Shanna K. Kattari, Jama Shelton, and Megan S. Paceley. 2024. ““Death Threats and Despair”: A Conceptual Model Delineating Moral Distress Experienced by Pediatric Gender-Affirming Care Providers.” *Social Sciences & Humanities Open* 9: 100867. <https://doi.org/10.1016/j.ssaho.2024.100867>.
- Hackworth, Lucy. 2018. “Limitations of “Just Gender”: The Need for an Intersectional Reframing of Online Harassment Discourse and Research.” In *Mediating Misogyny: Gender, Technology, and Harassment*, edited by Jacqueline Ryan Vickery and Tracy Everbach, 51–70. Cham: Springer. [https://doi.org/10.1007/978-3-319-72917-6\\_3](https://doi.org/10.1007/978-3-319-72917-6_3).
- Hainsworth, Paul. 2016. *The Politics of the Extreme Right: From the Margins to the Mainstream*. New York: Bloomsbury.
- Hameleers, Michael. 2020. “My Reality Is More Truthful Than Yours: Radical Right-Wing Politicians’ and Citizens’ Construction of ‘Fake’ and ‘Truthfulness’ on

- Social Media—Evidence from the United States and The Netherlands.” *International Journal of Communication* 14: 1135–52.
- Haney-López, Ian. 2014. *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*. New York: Oxford University Press.
- Higgins, Penelope. 2023. “Fascism and the Trans Villain: Historically Recurring Transphobia in Far-Right Politics.” *The Graduate History Review* 12:118–52. <https://doi.org/10.18357/ghr12202321578>.
- Hines, Sally. 2018. *Is Gender Fluid?* London: Thames & Hudson.
- Hohner, Julian, Heidi Schulze, Simon Greipl, and Diana Rieger. 2022. “From Solidarity to Blame Game: A Computational Approach to Comparing Far-Right and General Public Twitter Discourse in the Aftermath of the Hanau Terror Attack.” *Studies in Communication and Media* 11 (2): 304–33. <https://doi.org/10.5771/2192-4007-2022-2-304>.
- Jacobsen, Kai, Aaron Devor, and Edwin Hodge. 2022. “Who Counts as Trans? A Critical Discourse Analysis of Trans Tumblr Posts.” *Journal of Communication Inquiry* 46 (1): 60–81. <https://doi.org/10.1177/01968599211040835>.
- Jones, Charlotte, and Jen Slater. 2020. “The Toilet Debate: Stalling Trans Possibilities and Defending ‘Women’s Protected Spaces.’” *The Sociological Review* 68 (4): 834–51. <https://doi.org/10.1177/0038026120934697>.
- Jordan B Peterson (@JordanBPeterson). 2021. “Irreversible Damage? | Abigail Shrier | EP 159.” YouTube video. <https://www.youtube.com/watch?v=fSKQfATa-1I>.
- Jordan B Peterson Clips (@DrJordanBPetersonClips). 2021. “Gender Dysphoria Diagnosis and Treatment | Abigail Shrier & Jordan B Peterson.” YouTube video. <https://www.youtube.com/watch?v=5IbbcdXWxfA>.
- Joseph, Janice. 2022. “Multiple Invisibility of Black Victims of Transfemicide: An Intersectional Approach.” *Peace Review* 34: 501–17.
- Jubilee. 2023. “Trans Conservatives vs Trans Liberals | Middle Ground.” YouTube video. [https://www.youtube.com/watch?v=wSar7j\\_1MtM](https://www.youtube.com/watch?v=wSar7j_1MtM).
- Kallis, Aristotle. 2015. “When Fascism Became Mainstream: The Challenge of Extremism in Times of Crisis.” *Fascism* 4 (1): 1–24. <https://doi.org/10.1163/22116257-00401001>.
- Keen, Kellie-Jay (@KellieJayKeen). 2023. “This Is What Happens When Women Try to Speak #LetWomenSpeak #LetWomenSpeakAuckland #LetWomenSpeakNZ.” YouTube video. <https://www.youtube.com/watch?v=FsKn5XeJkWY>.
- Kelly, Cait, and Mostafa Rachwani. 2023. “What’s behind the ‘Terrifying’ Backlash against Australia’s Queer Community?” *The Guardian*, March 25. <https://www.theguardian.com/australia-news/2023/mar/25/whats-behind-the-terrifying-backlash-against-australias-queer-community>.
- Kesslen, Ben. 2022. “How a Theory about Transgender Contagion Went Viral.” *MIT Technology Review*, August 18. <https://www.technologyreview.com/2022/08/18/1057135/transgender-contagion-gender-dysphoria/>.
- Klein, Colin, Peter Clutton, and Vince Polito. 2018. “Topic Modeling Reveals Distinct Interests within an Online Conspiracy Forum.” *Frontiers in Psychology* 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00189>.
- Klein, Mark. 2012. “Enabling Large-Scale Deliberation Using Attention-Mediation Metrics.” *Computer Supported Cooperative Work* 21 (4): 449–73. <https://doi.org/10.1007/s10659-012-9241-1>.

- [org/10.1007/s10606-012-9156-4](https://doi.org/10.1007/s10606-012-9156-4).
- Klinger, Ulrike, W. Lance Bennett, Curd Benjamin Knüpfer, Franziska Martini, and Xixuan Zhang. 2023. "From the Fringes into Mainstream Politics: Intermediary Networks and Movement-Party Coordination of a Global Anti-Immigration Campaign in Germany." *Information, Communication & Society* 26 (9): 1890–1907. <https://doi.org/10.1080/1369118X.2022.2050415>.
- Koyama, Emi. 2020. "Whose Feminism Is It Anyway? The Unspoken Racism of the Trans Inclusion Debate." *The Sociological Review* 68 (4): 735–44. <https://doi.org/10.1177/0038026120934685>.
- Krzyżanowski, Michał. 2020. "Discursive Shifts and the Normalisation of Racism: Imaginaries of Immigration, Moral Panics and the Discourse of Contemporary Right-Wing Populism." *Social Semiotics* 30 (4): 503–27. <https://doi.org/10.1080/10350330.2020.1766199>.
- Lamble, Sarah. 2008. "Retelling Racialized Violence, Remaking White Innocence: The Politics of Interlocking Oppressions in Transgender Day of Remembrance." *Sexuality Research & Social Policy* 5 (1): 24–42. <https://doi.org/10.1525/srsp.2008.5.1.24>.
- Leurs, Koen. 2017. "Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-Cultural Research." *Feminist Review* 115 (1): 130–54. <https://doi.org/10.1057/s41305-017-0043-1>.
- Lewis, Rebecca. 2018. *Alternative Influence: Broadcasting the Reactionary Right on YouTube*. New York: Data & Society. <https://datasociety.net/library/alternative-influence/>.
- Lewis, Sophie. 2019. "How British Feminism Became Anti-Trans." *The New York Times*, February 7. <https://www.nytimes.com/2019/02/07/opinion/terf-trans-women-britain.html>.
- Li, Chong, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2024. "Improving In-Context Learning of Multilingual Generative Language Models with Cross-Lingual Alignment." arXiv. <http://arxiv.org/abs/2311.08089>.
- Li, David K., Minyvonne Burke, and Rima Abdelkader. 2024. "Boxer Imane Khelif Guaranteed a Medal as IOC Condemns Questions about Her Gender Identity." *NBC News*, August 5. <https://www.nbcnews.com/news/sports/totally-unacceptable-gender-identity-dispute-surrounding-boxers-lin-yu-rcna164995>.
- Littman, Lisa. 2018. "Rapid-Onset Gender Dysphoria in Adolescents and Young Adults: A Study of Parental Reports." *PloS One* 13 (8). <https://doi.org/10.1371/journal.pone.0202330>.
- Locantore, Dominic C., and Nesa E. Wasarhaley. 2020. "Mentally Ill, HIV-Positive, or Sexual Predator? Determining Myths Perceived as Representative of Transgender People." *Group Processes & Intergroup Relations* 23 (3): 378–401. <https://doi.org/10.1177/1368430219835032>.
- Lokmanoglu, Ayse D., Mowafak Allaham, Rod Abhari, Chloe Mortenson, and Esteban Villa-Turek. 2023. *A Picture Is Worth a Thousand (S)words: Classification and Diffusion of Memes on a Partisan Media Platform*. London: Global Network on Extremism and Technology. <https://doi.org/10.18742/pub01-117>.
- Lu, Christina, and David Jurgens. 2022. "The Subtle Language of Exclusion: Identifying the Toxic Speech of Trans-Exclusionary Radical Feminists." In *Proceed-*

- ings of the Sixth Workshop on Online Abuse and Harms*, edited by Kanika Narang, Aida Mostafazadeh Davani, Lambert Mathias, Bertie Vidgen, and Zeerak Talat, 79–91. Seattle: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.woah-1.8>.
- Lucas, Cathryn B., and Kristine E. Newhall. 2019. “Out of the Frame: How Sports Media Shapes Trans Narratives.” In *LGBT Athletes in the Sports Media*, edited by Rory Magrath, 99–124. Cham: Springer. [https://doi.org/10.1007/978-3-030-00804-8\\_5](https://doi.org/10.1007/978-3-030-00804-8_5).
- Luka, Mary Elizabeth, and Mélanie Millette. 2018. “(Re)Framing Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research.” *Social Media + Society* 4 (2). <https://doi.org/10.1177/2056305118768297>.
- MacKinnon, Kinnon R., Hannah Kia, Travis Salway, Florence Ashley, Ashley La-combe-Duncan, Alex Abramovich, Gabriel Enxuga, and Lori E. Ross. 2022. “Health Care Experiences of Patients Discontinuing or Reversing Prior Gender-Affirming Treatments.” *JAMA Network Open* 5 (7). <https://doi.org/10.1001/jamanetworkopen.2022.24717>.
- Mamié, Robin, Manoel Horta Ribeiro, and Robert West. 2021. “Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube.” In *Proceedings of the 13th ACM Web Science Conference 2021*, 139–47. New York: Association for Computing Machinery. <https://doi.org/10.1145/3447535.3462504>.
- Marwick, Alice, and Rebecca Lewis. 2017. *Media Manipulation and Disinformation Online*. Data & Society. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>.
- Marzetti, Hazel, Lisa McDaid, and Rory O'Connor. 2022. “Am I Really Alive?: Understanding the Role of Homophobia, Biphobia and Transphobia in Young LGBT+ People's Suicidal Distress.” *Social Science & Medicine* 298: 114860. <https://doi.org/10.1016/j.socscimed.2022.114860>.
- Mason, Gail. 2023. “Rethinking the Primacy of Consent: Community Education and Ethical Sex.” *Current Issues in Criminal Justice* 35 (2): 197–213. <https://doi.org/10.1080/10345329.2022.2152935>.
- Mathew, Binny, Punyajoy Saha, Seid Muhie Yimam, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2022. “HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection.” arXiv. <https://doi.org/10.48550/arXiv.2012.10289>.
- Mayerhöffer, Eva. 2021. “How Do Danish Right-Wing Alternative Media Position Themselves Against the Mainstream? Advancing the Study of Alternative Media Structure and Content.” *Journalism Studies* 22 (2): 119–36. <https://doi.org/10.1080/01461670X.2020.1814846>.
- McCall, Leslie. 2005. “The Complexity of Intersectionality.” *Signs* 30 (3): 1771–1800. <https://doi.org/10.1086/426800>.
- McClure, Tess, and Charlotte Graham-McLay. 2023. “Anti-Trans Activist Posie Parker Leaves New Zealand after Chaotic Protests.” *The Guardian*, March 26. <https://www.theguardian.com/world/2023/mar/25/anti-trans-activist-posie-parker-ends-new-zealand-tour-after-violent-protests-erupt>.
- McLean, Craig. 2021. “The Growth of the Anti-Transgender Movement in the United Kingdom. The Silent Radicalization of the British Electorate.” *International*

- Journal of Sociology* 51 (6): 473–82. <https://doi.org/10.1080/00207659.2021.1939946>.
- Menghani, Gaurav. 2023. “Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better.” *ACM Computing Surveys* 55 (12): 259. <https://doi.org/10.1145/3578938>.
- Miller-Idriss, Cynthia. 2020. *Hate in the Homeland: The New Global Far Right*. Princeton, NJ: Princeton University Press.
- Milner, H., and Tyrone Howard. 2013. “Counter-Narrative as Method: Race, Policy and Research for Teacher Education.” *Race Ethnicity and Education* 16 (4): 536–61. <https://doi.org/10.1080/13613324.2013.817772>.
- Mizock, Lauren, and Kim T. Mueser. 2014. “Employment, Mental Health, Internalized Stigma, and Coping with Transphobia among Transgender Individuals.” *Psychology of Sexual Orientation and Gender Diversity* 1 (2): 146–58. <https://doi.org/10.1037/sgd0000029>.
- Mkhize, Simangele, Reema Nunlall, and Nirmala Gopal. 2020. “An Examination of Social Media as a Platform for Cyber-Violence against the LGBT+ Population.” *Agenda* 34 (1): 23–33. <https://doi.org/10.1080/10130950.2019.1704485>.
- Mondon, Aurelien, and Aaron Winter. 2020. *Reactionary Democracy: How Racism and the Populist Far Right Became Mainstream*. New York: Verso.
- Murray, Conor. 2023. “Far Right Media Clash: The Ben Shapiro and Candace Owens Blowup Explained.” *Forbes*, November 16. <https://www.forbes.com/sites/conor-murray/2023/11/16/far-right-media-clash-the-ben-shapiro-and-candace-owens-blowup-explained/>.
- Nagle, Angela. 2017. *Kill All Normies: Online Culture Wars From 4Chan and Tumblr to Trump and the Alt-Right*. Washington, DC: Zero Books.
- Nguyen, Giang Thi, and Shu-Yi Liaw. 2020. “Big Data Application: From Identifying Barriers to Finding Solutions.” *International Journal of Economics, Business and Management Research* 4 (12): 186–203.
- Nirbhik, Nidhi, and Shailender Kumar. 2023. “Integrating Text and Emoticons for Detecting Extremist Affiliations on Twitter Using Deep Learning.” In 2023 *International Conference on Advanced Computing Technologies and Applications*, 1–6. Piscataway, NJ: IEEE. <https://doi.org/10.1109/ICACTA58201.2023.10392640>.
- O’Callaghan, Derek, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. 2013. “Uncovering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks.” In *Proceedings of the 5th Annual ACM Web Science Conference*, 276–85. New York: Association for Computing Machinery.
- Oliver, Indigo. 2022. “Transphobe of the Year: Matt Walsh.” *The New Republic*, December 27. <https://newrepublic.com/article/169609/transphobe-year-matt-walsh>.
- Owen, Leah. 2022. “‘Parasitically Occupying Bodies’: Exploring Toxifying Securitization in Anti-Trans and Genocidal Ideologies.” *Peace Review* 34 (4): 481–94. <https://doi.org/10.1080/10402659.2022.2129000>.
- Pape, Madeleine. 2022. “Something Old, Something New: Biofeminist Resistance to Trans Inclusion in Sport.” In *Justice for Trans Athletes*, edited by Ali Durham Greey and Helen Jefferson Lenskyj, 95–107. Leeds: Emerald. <https://doi.org/10.1108/978-1-80262-985-920221007>.

- PBS. 2024. "Fact-Checking the Harris-Trump Debate." *PBS News*, September 11. <https://www.pbs.org/newshour/politics/fact-checking-the-harris-trump-debate>.
- Pearce, Ruth. 2018. *Understanding Trans Health*. Bristol: Policy Press.
- Pearson, Elizabeth, Joe Whittaker, Till Baaken, Sara Zeiger, Farangiz Atamuradova, and Maura Conway. 2023. *Online Extremism and Terrorism Researchers' Security, Safety, and Resilience: Findings from the Field*. Dublin: VOX-Pol.
- Pengelly, Martin. 2024. "Candace Owens Leaves Daily Wire Site amid Israel and Antisemitism Tensions." *The Guardian*, March 22. <https://www.theguardian.com/media/2024/mar/22/candace-owens-leaves-daily-wire>.
- Perry, Barbara, ed. 2012. *Hate and Bias Crime: A Reader*. New York: Routledge.
- Pescitelli, Aynsley. 2018. "MySpace or Yours? An Exploratory Study of Homophobic and Transphobic Cyberbullying of Post-Secondary Students." In *Cyberbullying at University in International Contexts*, edited by Wanda Cassidy, Chantal Faucher, and Margaret Jackson, 52–65. New York: Routledge. <https://doi.org/10.4324/9781315189406-5>.
- Poletto, Fabio, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, and Viviana Patti. 2021. "Resources and Benchmark Corpora for Hate Speech Detection: A Systematic Review." *Language Resources and Evaluation* 55 (2): 477–523. <https://doi.org/10.1007/s10579-020-09502-8>.
- Puddifoot, Katherine. 2021. "Defining Stereotypes and Stereotyping." In *How Stereotypes Deceive Us*, edited by Katherine Puddifoot, 13–30. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780192845559.003.0002>.
- Quatrini, Amerigo. 2022. "On J. K. Rowling's Discourse on Transsexual Issues, An Analysis of the Language Used on Rowling's Twitter and the Sociolinguistic Implication of Hate Speech." *International Journal of Languages, Literature and Linguistics* 8 (2): 89–99. <https://doi.org/10.18178/IJLLL.2022.8.2.328>.
- R. Ellis, Justin. 2022. "A Fairy Tale Gone Wrong: Social Media, Recursive Hate and the Politicisation of Drag Queen Storytime." *The Journal of Criminal Law* 86 (2): 94–108. <https://doi.org/10.1177/00220183221086455>.
- Rajendran, Arundarasi, Vattikuti Sree Sahithi, Chhavi Gupta, Madhuri Yadav, Swati Ahirrao, Ketan Kotecha, Mayur Gaikwad, Ajith Abraham, Nada Ahmed, and Sarah M. Alhammad. 2022. "Detecting Extremism on Twitter During U.S. Capitol Riot Using Deep Learning Techniques." *IEEE Access* 10. <https://doi.org/10.1109/ACCESS.2022.3227962>.
- Ramirez, Ej Dickson, Nikki McCann. 2023. "The Right Boosted Trans Hate—And Ran Up Their Follower Counts." *Rolling Stone*, June 15. <https://www.rollingstone.com/culture/culture-features/trans-hate-follower-count-matt-walsh-1234770675/>.
- Reid, Graeme. 2021. "Global Trends in LGBT Rights During the Covid-19 Pandemic." *Human Rights Watch*, February 24. <https://www.hrw.org/news/2021/02/24/global-trends-lgbt-rights-during-covid-19-pandemic>.
- Restar, Arjee Javellana. 2020. "Methodological Critique of Littman's (2018) Parental-Respondents Accounts of 'Rapid-Onset Gender Dysphoria.'" *Archives of Sexual Behavior* 49 (1): 61–66. <https://doi.org/10.1007/s10508-019-1453-2>.
- Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and

- Wagner Meira. 2020. "Auditing Radicalization Pathways on YouTube." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–41. New York: Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372879>.
- Ring, Trudy. 2023. "Far-Right Trans YouTuber Said an Activist Drove a Woman to Suicide—But No One Died." *Advocate*, February 9. <https://www.advocate.com/transgender/transgender-2659396280>.
- Rivers, Ian, Jordan Daly, and Liam Stevenson. 2023. "Homophobic and Transphobic Online Harassment: Young People in Scotland during the COVID-19 Pandemic." In *Cyberbullying and Online Harms: Preventions and Interventions from Community to Campus*, edited by Helen Cowie and Carrie-Anne Myers. New York: Routledge.
- Rothut, Sophia, Heidi Schulze, Diana Rieger, and Brigitte Naderer. 2024. "Mainstreaming as a Meta-Process: A Systematic Review and Conceptual Model of Factors Contributing to the Mainstreaming of Radical and Extremist Positions." *Communication Theory* 34 (2): 49–59. <https://doi.org/10.1093/ct/qtae001>.
- Ryynänen, Max, Heidi Kosonen, and Susanne Ylönen. 2022. *Cultural Approaches to Disgust and the Visceral*. New York: Routledge. <https://doi.org/10.4324/9781003205364>.
- Sadjadi, Sahar. 2020. "The Vulnerable Child Protection Act and Transgender Children's Health." *TSQ: Transgender Studies Quarterly* 7 (3): 508–16. <https://doi.org/10.1215/23289252-8553202>.
- Salawu, Semiu, Jo Lumsden, and Yulan He. 2021. "A Large-Scale English Multi-Label Twitter Dataset for Cyberbullying and Online Abuse Detection: The 5th Workshop on Online Abuse and Harms." *Proceedings of the 5th Workshop on Online Abuse and Harms*, edited by Aida Mostafazadeh Davani, Douwe Kiela, Mathias Lambert, Bertie Vidgen, Vinodkumar Prabhakaran, and Zeerak Waseem, 146–56. Seattle: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.woah-1.16>.
- Sánchez-Sánchez, Ana M., David Ruiz-Muñoz, and Francisca J. Sánchez-Sánchez. 2024. "Mapping Homophobia and Transphobia on Social Media." *Sexuality Research and Social Policy* 21: 210–26. <https://doi.org/10.1007/s13178-023-00879-z>.
- Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. "Distil-BERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter." arXiv. <https://doi.org/10.48550/arXiv.1910.01108>.
- Sasani, Ava. 2024. "'Scared for Our Kids': Anger Mounts after Non-Binary Teen Dies Following School Fight." *The Guardian*, February 22. <https://www.theguardian.com/us-news/2024/feb/22/nex-benedict-death-oklahoma-owasso-lgbt-student>.
- Schroeder, Ralph. 2018. *Social Theory After the Internet: Media, Technology and Globalization*. London: UCL Press.
- Selepak, Andrew, and John Sutherland. 2012. "The Ku Klux Klan, Conservative Politics and Religion: Taking Extremism to the Political Mainstream." *Politics, Religion & Ideology* 13 (1): 75–98. <https://doi.org/10.1080/21567689.2012.659498>.

- Sharma, Deepawali, Vedika Gupta, and Vivek Kumar Singh. 2023. "Detection of Homophobia & Transphobia in Malayalam and Tamil: Exploring Deep Learning Methods." In *Advanced Network Technologies and Intelligent Computing*, edited by Isaac Woungang, Sanjay Kumar Dhurandher, Kiran Kumar Pattanaik, Anshul Verma, and Pradeepika Verma, 217–26. Cham: Springer. [https://doi.org/10.1007/978-3-031-28183-9\\_15](https://doi.org/10.1007/978-3-031-28183-9_15).
- Shaw, Daniel Odin. 2019. "“Something Old, Something New, Something Borrowed: The Alt-Right on Building Christendom Without Christ.” *Journal for the Study of Religions and Ideologies* 18 (54): 79–93.
- Siegel, Alexandra A. 2020. "Online Hate Speech." In *Social Media and Democracy*, edited by Joshua A. Tucker and Nathaniel Persily, 56–88. New York: Cambridge University Press.
- Slothouber, Van. 2020. "(De)Trans Visibility: Moral Panic in Mainstream Media Reports on de/Retransition." *European Journal of English Studies* 24 (1): 89–99. <https://doi.org/10.1080/13825577.2020.1730052>.
- Sofat, Chesta, Shabeg Singh Gill, and Divya Bansal. 2022. "Multi Class Classification of Online Radicalization Using Transformer Models." In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1034–38. Piscataway, NJ: IEEE. <https://doi.org/10.1109/CSCI58124.2022.00183>.
- Soken-Huberty, Emmaline. 2023. "What Is Gender Discrimination?" *Human Rights Careers*, January 12. <https://www.humanrightscareers.com/issues/what-is-gender-discrimination/>.
- SPLC. 2024. "Hate Crimes, Explained." *Southern Poverty Law Center*. <https://www.splcenter.org/hate-crimes-explained>.
- Squirrelrell, Tim, and Jacob Davey. 2023. *A Year of Hate: Understanding Threats and Harassment Targeting Drag Shows and the LGBTQ+ Community*. London: ISD. <https://www.isdglobal.org/wp-content/uploads/2023/06/Understanding-Threats-and-Harassment-Targeting-Drag-Shows-and-the-LGBTQ-Community.pdf>.
- Stern, Alexandra Minna. 2022. "Gender and the Far-Right in the United States: Female Extremists and the Mainstreaming of Contemporary White Nationalism." *Journal of Modern European History* 20 (3): 322–34. <https://doi.org/10.1177/16118944221110101>.
- Stieglitz, Stefan, and Linh Dang-Xuan. 2013. "Social Media and Political Communication: A Social Media Analytics Framework." *Social Network Analysis and Mining* 3 (4): 1277–91. <https://doi.org/10.1007/s13278-012-0079-3>.
- Stone, Amy L. 2018. "Gender Panics about Transgender Children in Religious Right Discourse." *Journal of LGBT Youth* 15 (1): 1–15. <https://doi.org/10.1080/19361653.2017.1395305>.
- Sunstein, Cass R., and Adrian Vermeule. 2009. "Conspiracy Theories: Causes and Cures\*." *Journal of Political Philosophy* 17 (2): 202–27. <https://doi.org/10.1111/j.1467-9760.2008.00325.x>.
- Swan, Jaime, Tania M. Phillips, Tait Sanders, Amy B. Mullens, Joseph Debattista, and Annette Brömdal. 2023. "Mental Health and Quality of Life Outcomes of Gender-Affirming Surgery: A Systematic Literature Review." *Journal of Gay & Lesbian Mental Health* 27 (1): 2–45. <https://doi.org/10.1080/19359705.2021.2016537>.

- The Trevor Project. n.d. "Bill Tracker Archive." *The Trevor Project*. <https://www.thetrevorproject.org/bill-tracker/>.
- Thompson, Mat A. 2022. "Choosing Threat, Embodying the Viral." *TSQ: Transgender Studies Quarterly* 9 (3): 480–87. <https://doi.org/10.1215/23289252-9836134>.
- Thurlow, Claire. 2024. "From TERF to Gender Critical: A Telling Genealogy?" *Sexualities* 27 (4): 962–78. <https://doi.org/10.1177/13634607221107827>.
- Track Trans Legislation. 2023. "2023 Anti-Trans Bills." *Track Trans Legislation*. <https://www.tracktranslegislation.com/>.
- Trans Legislation Tracker. 2024. "2024 Anti-Trans Bills: Trans Legislation Tracker." *Trans Legislation Tracker*. <https://translegislation.com>.
- TransActual. n.d. "Transphobia." *TransActual*. <https://transactual.org.uk/transphobia/>.
- Trinko, Katrina. 2019. "What It's Like to Lose Your Children to the 'Transgender Cult.'" *The Daily Signal*, October 30. <https://www.dailysignal.com/2019/10/30/what-its-like-to-lose-your-children-to-the-transgender-cult-from-a-mom-who-knows/>.
- True, Jacqui, and Sri Eddyono. 2021. "Preventing Violent Extremism—What Has Gender Got to Do With It? Gendered Perceptions and Roles in Indonesia." *European Psychologist* 26 (1): 55–67. <https://doi.org/10.1027/1016-9040/a000434>.
- Tsirbas, Yannis, and Lina Zirganou-Kazolea. 2025. "Hate Speech Mainstreaming in the Greek Virtual Public Sphere: A Quantitative and Qualitative Approach." *Communications* 50 (2): 392–414. <https://doi.org/10.1515/commun-2023-0051>.
- Tuters, Marc, and Sal Hagen. 2020. "(((They))) Rule: Memetic Antagonism and Nebulous Othering on 4chan." *New Media & Society* 22 (12): 2218–37. <https://doi.org/10.1177/1461444819888746>.
- Tylenda, Megan. 2024. "Banned Books & Banned Identities: Maintaining Secularism and the Ability to Read in Public Education for the Well-Being of America's Youth." *Indiana Journal of Law and Social Equality* 12 (1). <https://www.repository.law.indiana.edu/ijlse/vol12/iss1/5>.
- UNESCO. n.d. "UNESCO: Building Peace through Education, Science and Culture, Communication and Information." *UNESCO*. <https://www.unesco.org/en>.
- The University of Edinburgh. n.d. "What Is Transphobia?" *The University of Edinburgh*. <https://reportandsupport.ed.ac.uk/pages/what-is-transphobia>.
- UNOHCHR. n.d. "International Convention on the Elimination of All Forms of Racial Discrimination." *United Nations Office of the High Commissioner for Human Rights*. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial>.
- Upadhyay, Nishant. 2021. "Coloniality of White Feminism and Its Transphobia: A Comment on Burt." *Feminist Criminology* 16 (4): 539–44. <https://doi.org/10.1177/1557085121991337>.
- Valerio, Lizette Martínez. 2022. "Mensajes de odio hacia la comunidad LGTBQI+: análisis de los perfiles de Instagram de la prensa española durante la 'Semana del Orgullo.'" *Revista Latina de Comunicación Social*, no. 80: 364–88. <https://doi.org/10.4185/RLCS-2022-1749>.
- Vidgen, Bertie, and Leon Derczynski. 2020. "Directions in Abusive Language Training Data, a Systematic Review: Garbage in, Garbage Out." *PLOS ONE* 15

- (12): e0243300. <https://doi.org/10.1371/journal.pone.0243300>.
- Walker, Allyn. 2023. "Transphobic Discourse and Moral Panic Convergence: A Content Analysis of My Hate Mail." *Criminology* 61 (4): 994–1021. <https://doi.org/10.1111/1745-9125.12355>.
- Wareham, Jamie. 2023. "Beaten, Stabbed and Shot: 320 Trans People Killed In 2023—New Monitoring Report." *Forbes*, November 13. <https://www.forbes.com/sites/jamiewareham/2023/11/13/beaten-stabbed-and-shot-320-trans-people-murdered-in-2023/>.
- What Is a Woman?* 2022. Directed by Justin Folk. *The Daily Wire*. <https://www.dailywire.com/videos/what-is-a-woman>.
- White, Blaire. 2021. "Straight People Who Don't Date Trans People Are Now Calling Themselves 'Super Straight.'" *The Post Millennial*, March 8. <https://thepostmillennial.com/straight-people-who-dont-date-trans-people-are-now-calling-themselves-super-straight>.
- WHO. n.d. "Health Topics." *World Health Organization*. <https://www.who.int/health-topics>.
- Williams, Cristan. 2020. "The Ontological Woman: A History of Deauthentication, Dehumanization, and Violence." *The Sociological Review* 68 (4): 718–34. <https://doi.org/10.1177/0038026120938292>.
- Yang, Maya. 2022. "Boston Children's Hospital Receives Bomb Threat after Far-Right Harassment." *The Guardian*, August 31. <https://www.theguardian.com/us-news/2022/aug/31/boston-childrens-hospital-bomb-threat-far-right-lgbtq>.
- YouTube. n.d. "YouTube Hate Speech and Harassment Policy." *YouTube*. <https://www.youtube.com/howyoutubeworks/our-commitments/standing-up-to-hate/>.
- Zhang, Wei, Shiqing Huang, Lawrence Lam, Richard Evans, and Chengyan Zhu. 2022. "Cyberbullying Definitions and Measurements in Children and Adolescents: Summarizing 20 Years of Global Efforts." *Frontiers in Public Health* 10. <https://doi.org/10.3389/fpubh.2022.1000504>.
- Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. 2017. "Machine Learning on Big Data: Opportunities and Challenges." *Neurocomputing* 237: 350–61. <https://doi.org/10.1016/j.neucom.2017.01.026>.
- Zuboff, Shoshana. 2019. "Surveillance Capitalism and the Challenge of Collective Action." *New Labor Forum* 28 (1): 10–29. <https://doi.org/10.1177/1095796018819461>.
- Zuboff, Shoshana. 2023. "The Age of Surveillance Capitalism." In *Social Theory Re-Wired: New Connections to Classical and Contemporary Perspectives*, edited by Wesley Longhofer and Daniel Winchester. New York: Routledge.

## ACKNOWLEDGMENTS

A special thanks to Professor Anna Saunders and Dr Sarah Arens for inviting us to present this paper at the Global Far Right Studies network at the University of Liverpool. Thank you also to Vic Ge who shaped the early thinking on this project and its importance. We would also like to thank the reviewers for their expert and generous comments.