DHQ: Digital Humanities Quarterly

Preview
2026
Volume 20 Number 1

Infrastructures of Listening: The ManoWhisper Podcast Analysis Pipeline

Nick Ruest <ruestn_at_yorku_dot_ca>, York University, Digital Scholarship Infrastructure https://orcid.org/0000-0003-1891-1112

Brianna I. Wiens <biwiens_at_uwaterloo_dot_ca>, University of Waterloo, English Language & Literature https://orcid.org/0009-0003-9942-3862

Karmvir K. Padda <karmvirkaur_dot_padda_at_uwaterloo_dot_ca>, University of Waterloo, Sociology and Legal Studies https://orcid.org/0000-0002-3390-3789

Shana MacDonald <shana_dot_macdonald_at_uwaterloo_dot_ca>, University of Waterloo, Communication Arts https://orcid.org/0000-0001-5828-6345

DOI: pending

Abstract

Podcasting has become a central infrastructure for the circulation and normalization of misogynistic and extremist ideologies, yet the scale, length, and affective density of audio content pose significant challenges for critical qualitative research. This paper introduces ManoWhisper, a feminist computational research infrastructure designed to support the large-scale analysis of misogynistic podcast ecosystems while preserving the contextual depth required for interpretive and ethical engagement. ManoWhisper combines automated audio acquisition, transcription, sentence-level classification, indexing, and visualization within a searchable web-based interface that enables researchers to move between computational pattern detection and close qualitative reading. Grounded in a feminist methodology of dwelling, the tool is designed to slow analysis, foreground emotional labour, and support collaborative research across varying levels of technical expertise. It allows for an in-depth consideration of more extremist media ecosystems across a variety of key factors. This paper documents ManoWhisper's end-to-end pipeline, from content collection and transcription to classification, indexing, and interface design. Further we demonstrate its application across multiple peer-reviewed and public-facing research projects examining misogyny, masculinity, and gender-based extremism in podcasting, as well as how it is being used in policy and government institutions. We position ManoWhisper as methodological infrastructure that redistributes analytic capacity, makes repetition and scale visible, and enables ethically grounded engagement with harmful media. We conclude by reflecting on the tool's limitations, its implications for feminist digital methods, and its relevance for understanding how misogynistic content circulates not only across media platforms but into emerging domains such as AI training data.

Introduction

Digital media has fundamentally transformed how extremist content circulates and gains cultural legitimacy. Podcasting, in particular, has emerged as a potent medium for the dissemination of misogynist ideologies, reaching hundreds of millions of listeners globally and increasingly serving as a primary source of political commentary for younger audiences [Rizwan et al. 2025] [Singh 2025]. The intimate, conversational format of podcasts combined with their accessibility and algorithmic amplification creates ideal conditions for normalizing harmful rhetoric while cultivating parasocial relationships — the sense of personal intimacy listeners develop with hosts they have never met — between podcasters and their audiences [Berry 2016] [Llinares, Fox, and Berry 2018]. As scholars have increasingly warned, these same infrastructures that enable civic engagement and education are also routinely weaponized to circulate misogynistic, racist, and authoritarian ideologies [Ging 2019] [Allchorn 2024] [Miller-Idriss 2025]. Figures such as Andrew Tate, Nick

Fuentes, Charlie Kirk, Ben Shapiro, Matt Walsh, Jeremy Mackenzie, and Gavin McInnes exemplify this phenomenon. Through their shows, these influencers construct what recent scholarship has described as male supremacist masculinities: ideological frameworks that naturalize men's dominance over women and frame overt resistance to gender equality as both justified and necessary [Miller-Idriss 2025]. Their content blends self-help narratives with grievance politics, packaging patriarchal ideology and gender hierarchies as “common sense” while vilifying feminism, LGBTQ2IA+ rights, and anti-racist movements. The real-world consequences are measurable, with teachers reporting boys mimicking this content in schools, domestic violence organizations connecting it to offline harm, and in extreme cases, perpetrators of violence consuming this material immediately before committing acts of femicide [Czerwinsky 2024] [Padda, Ruest, and MacDonald, in review].

2 Within this context, the scale of this media ecosystem defies traditional research methods. Through our work with the SIGNAL (Strategies for Intersectional Gender-Justice, Networked Action, and Liberation) Network, a network that authors Ruest, Wiens, and MacDonald co-direct alongside colleague Jada Watson (University of Ottawa), we have assembled a collection exceeding 90,000 podcast episodes, representing over 100,000 hours of audio from the Intellectual Dark Web, conspiracy networks, QAnon circles, the Alt-Right, White Supremacist movements, and the “manosphere.” The Intellectual Dark Web refers to a loose network of commentators, including figures like Jordan Peterson, Sam Harris, Joe Rogan, and others, who frame contrarian views on gender, race, and identity as suppressed common sense, while the Manosphere is an umbrella term for online communities organized around male grievance, including incel (involuntary celibate) forums, pick-up artist networks, and men’s rights movements [Ging 2019]. While this collection offers unprecedented insight into the circulation of extremist gender discourse, its sheer volume renders manual analysis impractical and methodologically insufficient.

3 This development of ManoWhisper emerged from this tension between scale and interpretive depth. This project originated in a conversation between two of our researchers grappling with a deceptively simple question: *What can we do with podcasts at this scale?* At the time, the lead author (Ruest) was preparing a sabbatical project focused on building transcription pipelines using OpenAI’s Whisper model for audiovisual materials held in cultural heritage repositories. This work involved sustained engagement with emerging approaches to large-scale transcription and the integration of machine-learning tools into archival and library workflows [Rodriguez and Brown 2023] [Corall 2024] [Miller 2024] [Simonsen 2025]. Simultaneously, other SIGNAL researchers were wrestling with the methodological impossibility of manually analyzing the massive volume of manosphere podcast content our team had been tracking, recognizing both its importance for understanding contemporary extremism and the sheer impracticality of traditional qualitative approaches at this scale. The conversation revealed an opportunity: Technical infrastructure designed for one purpose could be adapted to address an urgent research need. What began as a discussion about transcription technology evolved into a larger methodological intervention of creating a tool that would combine computational efficiency with feminist research ethics, enable collaboration across varying levels of technical expertise, and support the kind of careful, contextual analysis our team's dwelling methodology requires.

4 The development of computational tools for humanities-based inquiry has generated substantial methodological debate within digital humanities, particularly around the relationship between scale, interpretation, and ethical responsibility. Scholars have interrogated how “distant reading” and large-scale text analysis, while enabling new forms of pattern recognition across corpora that would be impossible to read manually, risk flattening the interpretive complexity and contextual specificity that humanistic inquiry demands [Jockers 2013] [Moretti 2013] [Underwood 2019]. Feminist and critical DH scholars have pushed this critique further, arguing that the choice of what to digitize, what to analyze, and what tools to build are never neutral decisions but reflect and reproduce existing structures of power, access, and epistemic authority [McPherson 2012] [Losh and Wernimont 2018] [Risam 2019]. This tension between computational efficiency and interpretive depth is more than a technical problem to be solved — it is a constitutive challenge of the field that requires, as scholars have argued, explicit methodological frameworks for moving between quantitative pattern detection and qualitative close engagement [Drucker 2011] [Gold and Klein 2016]. ManoWhisper and our methodological approach to the tool emerges from this debate directly.

5 Responding to these ongoing debates, ManoWhisper is heavily inspired by Erlend Simonsen’s Knowledge Fight Interactive Search Tool [Simonsen 2025], developed in the context of documenting and analyzing right-wing conspiracy

media. Our work was undertaken in close dialogue with colleagues across the SIGNAL Network and, while it overlaps with the Knowledge Fight tool in its emphasis on large-scale transcription and searchability, ManoWhisper extends this approach in several key ways. In particular, it foregrounds feminist methodological commitments, incorporates classifications of misogyny and hate speech, and situates the collecting criteria within Canadian and transnational gender discourse ecosystems, including the Manosphere, femosphere (ecosystems of female-hosted content that reinforces and responds to manosphere ideologies), and incel communities. These extensions reflect the authors' broader research goals of understanding how digital gender discourses shape local institutions, political cultures, and everyday life. We encourage all to explore the Knowledge Fight Interactive Search Tool as it is regularly updated with new shows, episodes, features, and insights.

In what follows, we document how ManoWhisper works, from initial content acquisition through transcription, classification, visualization, and web interface design. We situate the tool within our dwelling methodology, demonstrating how it enables hybrid research that combines computational and interpretive approaches. Our discussion section reflects on practical research applications, collaborative affordances, limitations, and future possibilities, including the tool's potential for investigating how extremist content may shape AI training data. We also situate this work within critical AI and Science and Technology Studies scholarship that has fundamentally challenged computational neutrality, documenting how algorithms and training data encode existing social hierarchies and reproduce structural inequity [Eubanks 2017] [Noble 2018] [Benjamin 2019] [Bender et al. 2021]. ManoWhisper builds on this foundation while extending it in a specific direction to trace how extremist media ecosystems become source material for machine-learned common sense. Throughout the paper, we aim to maintain transparency about what this technology can and cannot accomplish, acknowledge the emotional labor this research demands, and foreground the feminist ethics that guide our work. 6

How it all Works - The Pipeline

Acquisition/Collection

We begin by identifying shows (i.e., podcasts, Rumble channels, YouTube channels) associated with the Intellectual Dark Web, conspiracy networks, QAnon, the Alt-Right, White Supremacist/Nationalist movements, or the Manosphere. Once identified, each show's RSS feed or channel URL is added to a custom bash utility, téléchargeur [Ruest and Padda 2025], which uses cron to regularly check feeds and download newly released episodes. Episodes are stored within a standardized directory structure organized at the series level. 7

An example directory structure for a single show appears as follows: 8

```
America First – Nicholas J. Fuentes
├── descriptions
├── downloaded_files.txt
├── hate-speech
├── metadata.txt
├── metadata.json
├── misogyny
├── mp3
├── summarizations
└── vtt
```

Newly downloaded episodes are deposited in the mp3 directory. At the time of download, basic metadata, including episode title, air date, and duration, is automatically extracted and recorded in a metadata.txt file, establishing the foundation for subsequent transcription, classification, and analysis stages in the pipeline.

Transcription & Classification

The next stage of the pipeline is les-observateurs [Ruest 2025a], a Python utility built on celery (a distributed task queue) and RabbitMQ (a message and streaming broker). This utility continuously monitors designated directories for newly downloaded audio files. When a new mp3 file appears, it is automatically queued, transcribed with OpenAI's Whisper large-v3-turbo model [Radford et al. 2022], and saved as a WebVTT transcript in the vtt directory. Although platforms such as Apple and YouTube do provide transcriptions, these are not always easily accessible programmatically and were not readily available when the project began. As such, we opted to transcribe episodes on our own computer infrastructure using OpenAI's Whisper after evaluating it against alternatives such as Whisper.cpp and WhisperX. 9

Once transcripts are generated, a suite of downstream utilities collectively referred to as manowhisper-cli [Ruest and Padda 2025] performs several derivative processing tasks, including text classification using transformer-based language models such as BART (a sequence-to-sequence model) and BERT (a bidirectional encoder). These include automated summarization using bart-large-finetuned-filtered-spotify-podcast-summ [Clifton et al. 2020], and generation of classification scores for misogyny [Attanasio et al. 2022] and hate speech [Vidgen et al. 2021]. 10

These classifiers were selected following evaluation against multiple candidates, where we prioritized classification models with published methodologies and documented training data. The misogyny and hate speech classifiers are BERT-based, fine-tuned (as described in their associated publications), for sentence-level classification tasks, and produce probabilistic scores. In the case of the misogyny classifier, it includes an extension to BERT, incorporating entropy-based attention, "forc[ing] the model to build token representations by attending to a wider context, i.e., consider a larger number of tokens from the rest of the sentence" [Attanasio et al. 2022].The models operate on surface linguistic features and learned contextual patterns rather than explicit definitions, and provide consistent and reproducible approximation of misogynistic and hateful language as defined in their respective training corpora. 11

During the classification process, transcripts are segmented into individual sentences, with each sentence classified independently. The resulting classification outputs are stored as csv files, with separate files generated for misogyny and hate speech for each episode. By design, the models we use can be swapped out or updated as definitions and training data evolve, allowing classification scores to be re-generated as needed. In addition to classification, we generate statistics for each episode using manometrics [Ruest 2025c], including word count, character count, sentence count, and words per minute. These measures provide baseline metrics that enable comparison across shows with different formats and lengths. 12

To support exploratory and longitudinal analysis, the pipeline also includes a side set of visualization utilities that generate time-series plots of misogyny and hate-speech scores, word clouds, and charts based off of the statistics. These visualizations make it possible to identify trends, peaks, and shifts in rhetorical intensity across episodes and over time. The utility names for these scripts are deliberately tongue-in-cheek, reflecting the use of dark humour and satire as a counterweight to the emotional weight of sustained engagement with misogynistic and extremist content. 13

Indexing

In the final step, episode transcripts are indexed in Apache Solr before becoming available in the web application. Solr is a search platform widely used in many GLAM applications, including institutional and cultural heritage repositories, and library or archival catalogues. This process is handled by two Python scripts that run several times per day via cron. The first script generates a `metadata.json` file for each show [Ruest 2025c]. Within this file, each episode is represented by a set of key-value pairs capturing episode title, air date, duration, number of sentences, word count, vocabulary size, number of characters with and without spaces, and speaking rate. The second script [Ruest 2025b] parses the `metadata.json` file alongside each episode's transcript and associated hate speech and misogyny classification csv files, generating a Solr document for each episode. Once indexing is complete, episodes and the associated metadata, transcripts, and classification outputs are immediately available in the web application for search, filtering, and analysis. 14

The Webapp

All of the transcripts and derivative data generated through the pipeline, outlined above, are brought together in a single web application [Ruest 2025b]. The impetus came from a need in our research team for a canonical source for the data, a consistent way to query it, and above all, an effective and efficient research tool. The webapp lowers that barrier by offering a robust and accessible keyword search across the full corpus, while also enabling researchers to explore a very large body of transcripts quickly and reproducibly. This functionality is particularly significant for qualitative researchers without the computational background, who can now perform comprehensive corpus-wide queries that would otherwise require substantial technical expertise. 15

The application is straightforward, built with Flask and Solr (Flask is a lightweight web application framework). Its landing page presents the podcasts currently in the corpus, along with the episode counts and a global search box for querying the entire dataset (Figure 1). From there, users can move into individual shows. A show page displays aggregated statistics and text classification results alongside a sortable list of episodes (Figure 2). Episodes can be ordered by date, misogyny count, hate speech count, or episode length depending on the research need. Clicking on a specific episode opens its dedicated page, where metadata and classification results generated by manowhisper-cli are presented together with an interactive time-series chart of misogyny and hate speech scores (Figure 3). Most importantly, the full transcript of the episode is made available. Each line of the transcript carries its own time code, allowing direct links to specific moments, and also has highlighted lines of misogyny or hate speech. The time-series chart itself is also interactive; clicking on any point in the graph takes a user directly to the relevant segment of the transcript. 16

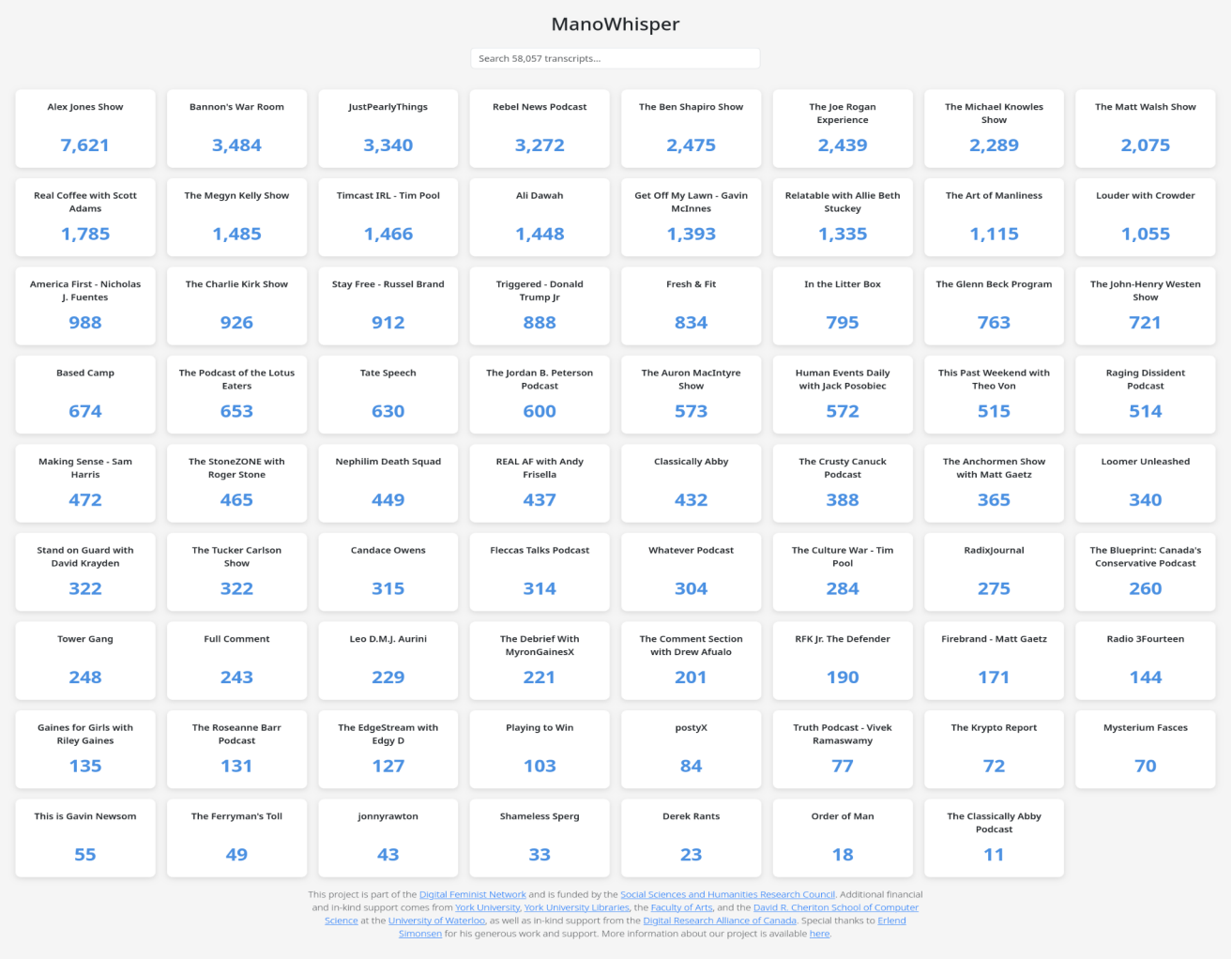

Figure 1. The ManoWhisper homepage.

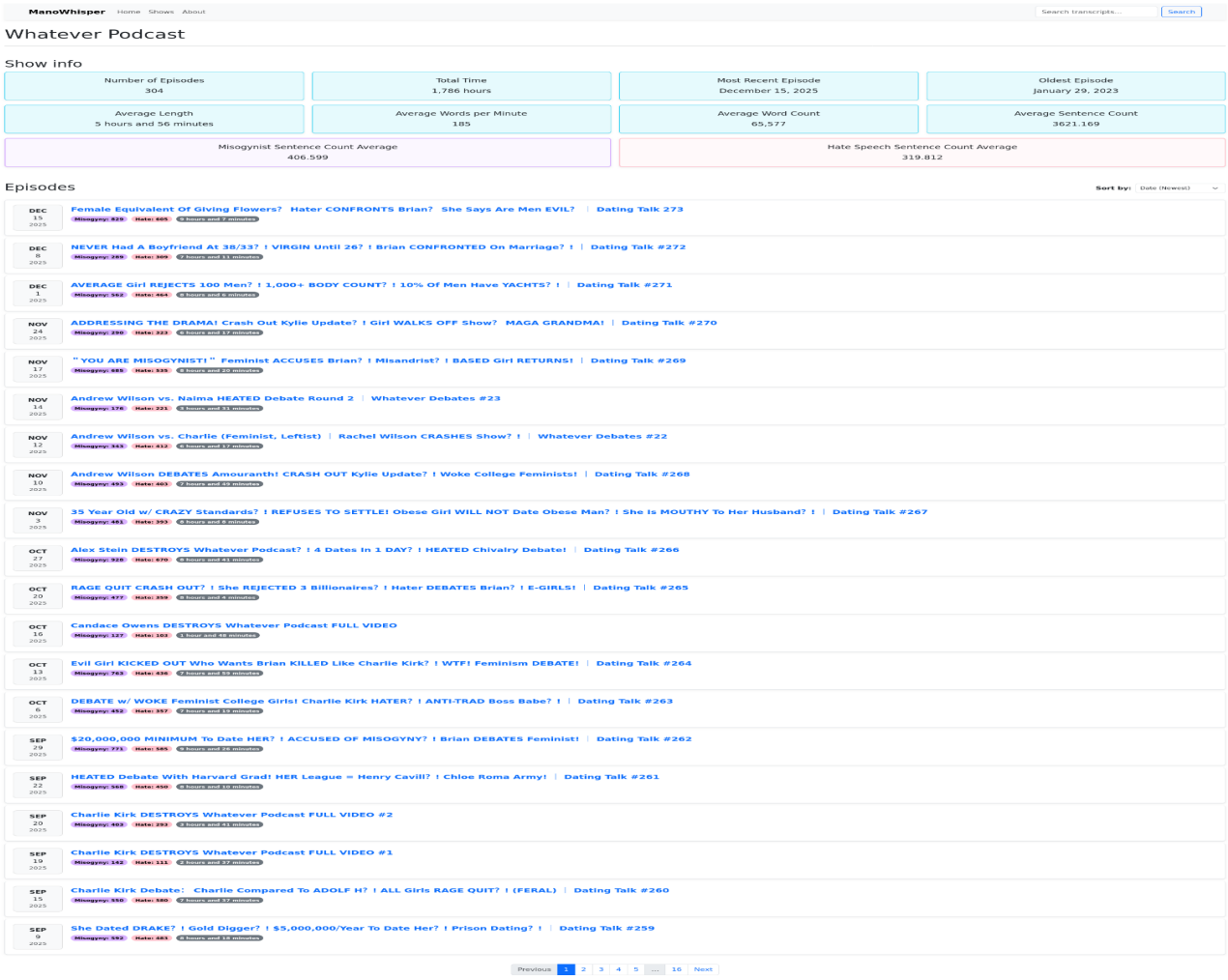

Figure 2. Example of a show or podcast series in ManoWhisper.

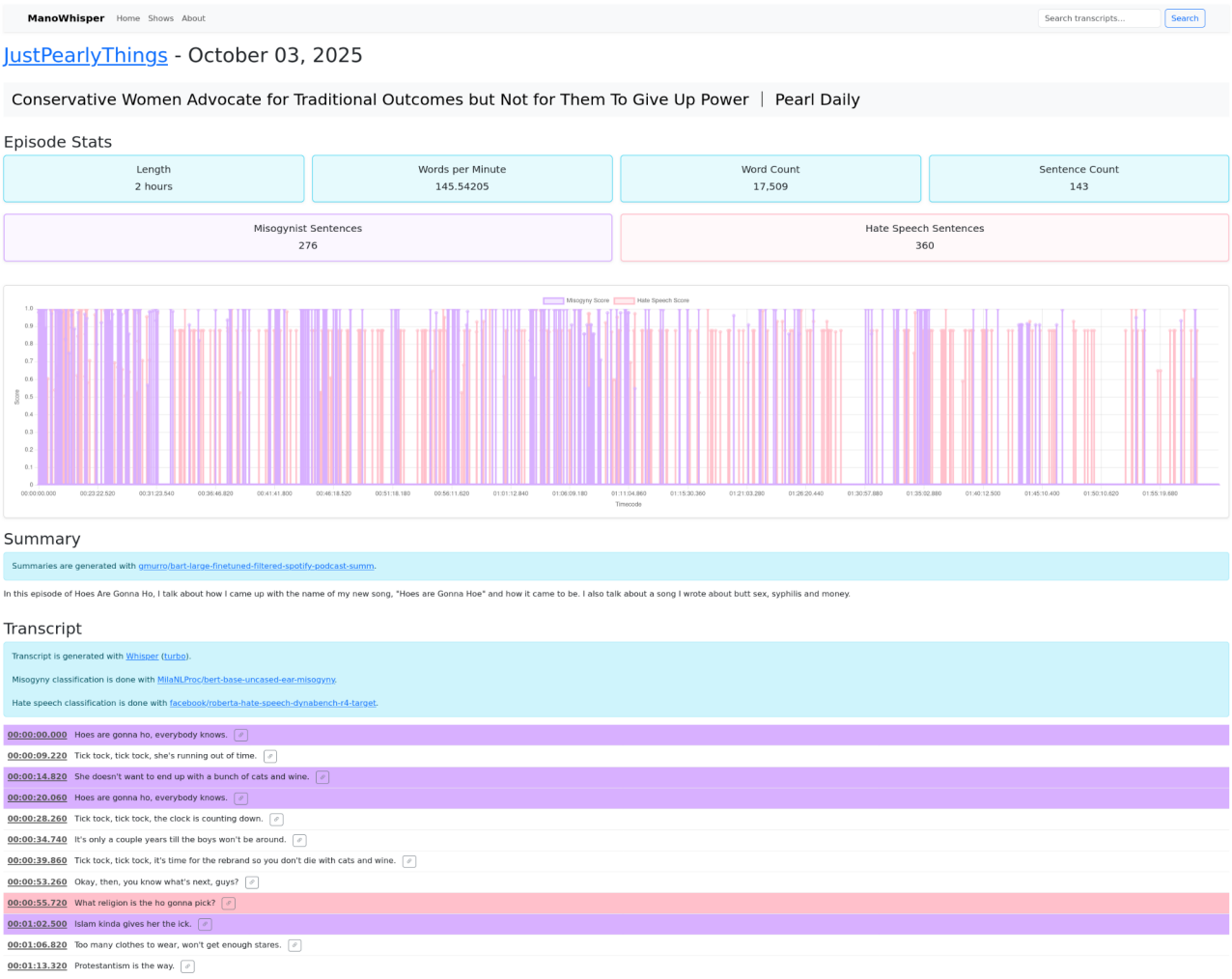

Figure 3. Example of a podcast or episode in ManoWhisper.

As mentioned above, search functionality extends across the entire corpus. Keyword, phrase, and advanced searching is supported, with results highlighted in context and linked directly to the corresponding time codes in the transcripts (Figure 4). Results can be narrowed by podcast or across multiple podcasts at once with a facet. The results can also be sorted by date, misogyny count, and hate speech count.

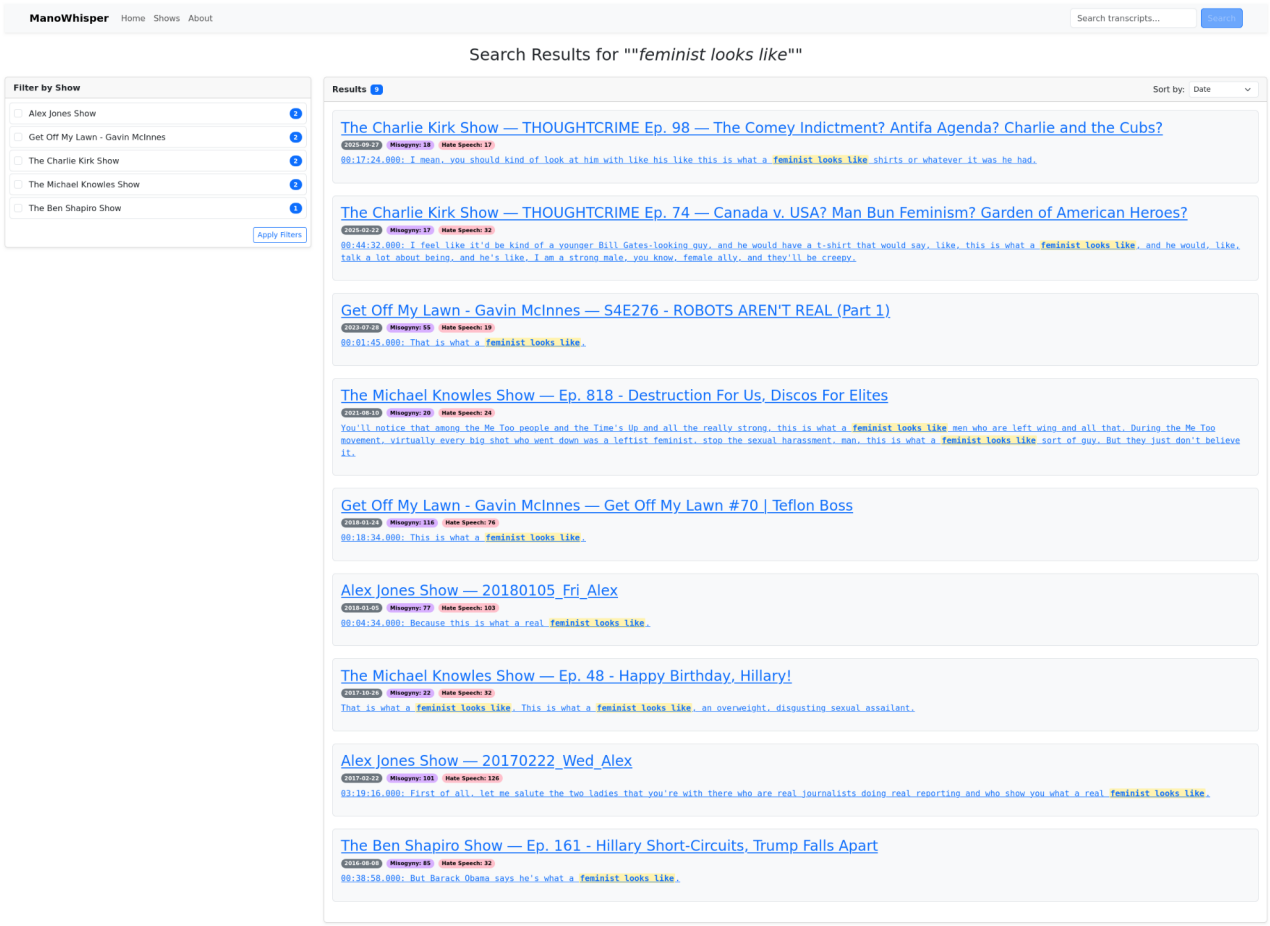

Figure 4. Visualization of the ManoWhisper architecture.

Alongside the web interface for human exploration, the webapp also exposes an API, making the same data accessible programmatically. Endpoints are available for search queries, show lists, show information, ans episode-level metadata and transcripts. Finally, the dataset itself is refreshed daily. New transcripts, their associated metadata, and text classification results are added in regular batches, ensuring the corpus is current as episodes are released or new podcasts are identified (Figure 5). Each of these design choices, from the interactive time-chart series to highlighted transcript lines, linked time codes, and corpus-wide searches, was made in deliberate service of more ethical, reflexive commitments. As such, the pipeline shown below is both technical infrastructure and the material instantiation of an approach to research that asks scholars to slow down, dwell, and engage with data contextually.

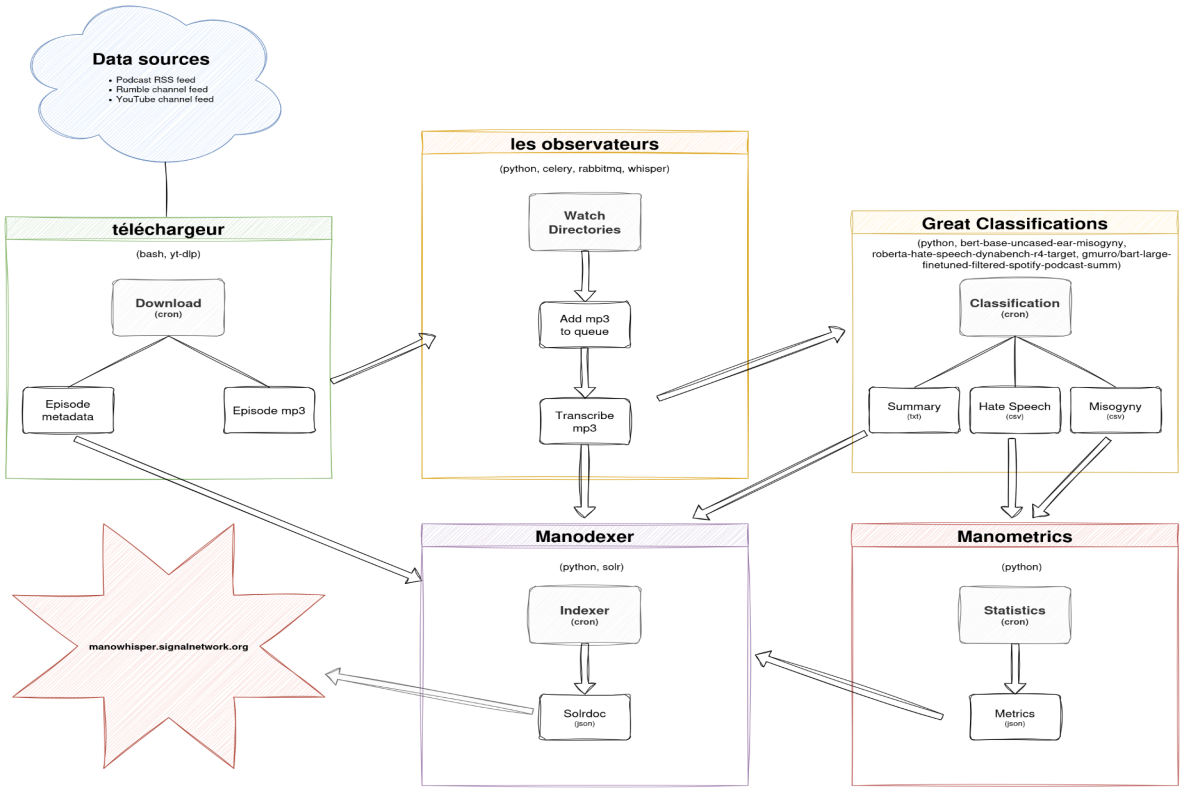

Figure 5. Example of a keyword search in ManoWhisper.

Dwelling with ManoWhisper: A Methodological Approach

ManoWhisper, as part of our larger methodological toolkit housed under SIGNAL, is designed not only to classify and quantify misogyny and hate speech but to support digital dwelling — a feminist methodological practice that we, as a team, have developed over several years of research [Wiens 2021] [Wiens 2022] [Wiens and MacDonald 2024] [MacDonald, Wiens, and Ruest 2025]. As a method, dwelling asks "that we linger in online spaces to sit with ideas, find out how tools work, how different tactics can be tools, and how they can be used in counter hegemonic ways to center marginalized voices and bring forth new ways of engaging in the world" [Wiens 2021, p. 86]. Dwelling allows us as researchers to focus on "the relationship between or 'intra-actions' [Barad 2003] of the researcher, research scene, participants, data, affects, and sociopolitical context" to attend to the "individual stories found through these data" in order to better contextualize "the broader themes or trends of the aggregate" and to more closely consider the "interconnected domains of influence between individual spheres and their relationship to collective and then structural levels" of power [Wiens 2021, p. 86]. Such intra-actions, as [Barad 2003] lays out, speak to the mutually constitutive entanglements of the research scene, rather than assuming pre-existing and stand-alone entities. 19

In our recent work on feminist media archives, we argue that despite the popularity of and reliance on big data, "there is a pressing need to slow down and grapple with an expanded sense of what counts as data and consider what can be lost within standard practices of big data collection such as the automated scraping of web archives" [MacDonald, Wiens, and Ruest 2025, p. 199]. Unlike computational methods that rely on distant reading and aggregate patterns alone, dwelling, especially here with ManoWhisper, asks researchers to sit with smaller cohorts of data — to move between the zoomed-out perspective that time-series visualizations and classification scores provide, and the zoomed- 20

in close reading of individual transcript segments, their contexts, and their affective impacts. This approach responds to what we identify as a central methodological challenge. Indeed, “many of the methodologies for exploring [web archive] data at scale utilize a cycle of high-level reading and zooming in close reading” [MacDonald, Wiens, and Ruest 2025, p. 205]. The classification scores and statistics generated by ManoWhisper (e.g. misogyny counts, hate speech counts, speaking rates) can reveal important patterns across episodes and shows. However, these numbers alone risk flattening the lived experiences and power dynamics embedded within the rhetoric of extremist content. As such, dwelling with ManoWhisper outputs means reading transcripts alongside classification scores — not just noting that an episode has a high misogyny count, but examining which sentences were classified as misogynistic, how that language functions rhetorically, and what ideological work it performs. It means attending to temporal dynamics by using time-series visualizations to identify spikes in harmful rhetoric, then dwelling with those specific moments to understand their triggers and contexts. It involves recognizing both patterns and anomalies, looking for what is typical across a show's episodes and what stands out as exceptional or particularly revealing. As we argue, this is “an extended autoethnographic effort to explore how our actions affect the scene and what this suggests for the entangled practices and digital communities we find ourselves in” [Wiens and MacDonald 2024, p. 32]. As such, this requires situating data within broader contexts by connecting individual episodes to larger media events, political moments, and movement dynamics.

In practice, the webapp's design facilitates dwelling by providing interactive time-series charts where clicking on data points takes users directly to the relevant transcript segment, by highlighting lines classified as misogyny or hate speech within full transcripts, by including time codes that allow researchers to trace specific moments across different data layers, and by enabling both corpus-wide searches and episode-specific deep dives. This architecture supports what we have elsewhere called a hybrid research approach [MacDonald, Wiens, and Ruest 2025]: Researchers can identify patterns computationally through search and classification, then dwell with specific content to understand how that pattern manifests, what it means, and what contexts shape it. As we write, such hybrid approaches that combine “analysis from both institutional and curated collections alongside those on social media” help us “elucidate the structure and scope of different kinds of collections, as well as an exploration of what they both offer and elide” [MacDonald, Wiens, and Ruest 2025, p. 201]. Dwelling, then, is particularly important for research on harmful content because not only does it center researcher reflexivity, acknowledging that one's own affective responses, social position, and situated knowledge shape what one notices and how one interprets it; it also offers a mediating screen that provides necessary distance when dwelling. Giving the researcher a chance to zoom in and out within the larger textual interface of transcripts helps researchers protect their own affective responses and ensure their well-being. 21

Dwelling resists what we might call algorithmic flattening — the tendency of computational analysis to reduce complex ideological and rhetorical systems to aggregate scores, erasing the contextual and affective dimensions that give language its meaning. It makes space for care, recognizing the emotional labor of engaging with misogynist and hateful content and building in time for processing and reflection, and it keeps focus on the power structures of podcast media and its normative rhetoric which upholds the inequalities and violence perpetuated by these male supremacist discourses beyond their formal features. This resistance is not unique to our approach, speaking to a broader methodological current within feminist, Black, and Indigenous methodologies that have pushed back against the extractive tendencies of big data methods and the academy at large. Indeed, such approaches have insisted that bodies, labours, and power relations embedded in data must remain visible, relational, and present throughout the research process and beyond [Losh and Wernimont 2018] [Risam 2019] [D'Ignazio and Klein 2020]. In addition, we take inspiration from Robin Wall Kimmerer’s (2013) description of pond restoration work, where the labour is marked by learning from mistakes, and also by slowness, care, and attentiveness. It is not a matter of moving fast and breaking things but remaining with the uncomfortable parts and learning them closely. For Kimmerer, this kind of ecological tending requires what she calls reciprocal attention. The researcher does not stand apart from what she studies but enters into a relationship with it, changed by the encounter as much as she changes it. This orientation is explicitly anti-extractive. Knowledge is not taken from a system but grown within it, through patient presence and accountability to what is found there. We understand our dwelling practice in analogous terms. Here, the data is not a resource to be mined for findings but a site to be inhabited, one that makes ethical and affective demands on those who enter it. 22

This sensibility resonates, too, with adrienne maree brown's (2017) framework of emergent strategy, which draws on principles of complexity and interdependence to imagine social change as something that unfolds through small, iterative, relational acts rather than through top-down master plans. brown's insistence that "small is good, small is all" [brown 2017] and that transformation happens at the level of the granular and the local speaks directly to what we are trying to do methodologically, resisting the gravitational pull of scale and aggregation and instead stay close to the particular utterance, the specific rhetorical move, and the individual moment of harm. Her principle that "what we practice at the small scale sets the patterns for the whole" [brown 2017] offers a way of thinking about dwelling not as methodological inefficiency but as deliberate prefigurative practice. 23

Similarly, Alexis Pauline Gumbs's practice of Black feminist study extends this logic into the archive and into the body. As she writes in *M Archive: After the End of the World* (2018), Gumbs refuses the extractive impulse of conventional archival method to not mine the Black feminist record for data points but instead inhabit fragments, sitting with their incompleteness, and allows meaning to accumulate through sustained proximity rather than synthesis. In her work *Undrowned: Black Feminist from Marine Animals* (2020), this durational attentiveness extends to marine mammals as teachers and as figures whose strategies of survival, breath, and collectivity can only be learned by staying with them long enough to be changed. Together, these works model what it might mean to study something without consuming it, to remain present to difficulty without resolving it prematurely into findings. This resonates with what we mean by dwelling, where the way we handle one podcast episode, one transcript, one slur, is continuous with the kind of scholarship, and the kind of world we are trying to build. Where much DH scholarship has focused on what computational methods can reveal, this work has been equally insistent on attending to what they risk concealing to ask whose speech is classified and whose experiences are flattened. Dwelling, as we practice it, is our response to that insistence — it is our commitment to staying with the data long enough that those concealments become visible once again. 24

In our work with visualization tools, we have found that spending time dwelling with the visualization, be it image or text is highly productive within big data sets. Moving between a zoomed-out version of the entire dataset and a controlled zooming in to a more specific, curated section of the data, allows researchers to engage in both high-level and close-level "reading" from a single-entry point. This mode of dwelling reveals what we have called "emergent intertextualities" [Wiens and MacDonald 2024]: the ways that meaning emerges from moving slowly through data, attending to how individual pieces relate to each other and to broader political and cultural contexts. Dwelling with ManoWhisper means treating each transcript segment, each spike in classification scores, and each anomaly in speaking rates as potentially meaningful and as part of the larger story of how extremist movements use media to spread harmful ideologies. Following adrienne maree brown's argument that within feminist activist frameworks, "It's all data" [brown 2017, p. 14], ManoWhisper's design supports researchers in practicing what we call dwelling as a method of question-asking: a slower, more deliberate engagement with all forms of data that honors the complexity and stakes of studying extremist rhetoric while maintaining the researcher's ethical and political commitments. 25

Discussion

Within both our collaborative and independent work through our network, SIGNAL, ManoWhisper has proven to be an invaluable tool for our research, enabling insights that would be nearly impossible to achieve through manual analysis alone. The sheer scale of the data — thousands of hours of podcast content spanning multiple shows, hundreds of thousands of sentences, and millions of words — presents not only a technical challenge but also a significant human and emotional burden. Creating computational tools like ManoWhisper helps ease that burden while enabling the deep research and critical insights necessary for understanding how extremist movements weaponize media platforms to spread misogynist and hateful ideologies. Indeed, ManoWhisper has been foundational to multiple publications that map the landscape of online misogyny, trace its circulation across platforms, and document feminist counter-responses. In this section, we reflect on ManoWhisper's practical applications in our work, the scale and complexity of the data that it helps us analyze, and the ways it supports our team's research and that of collaborators working to counter online extremism. 26

Research Applications: From Data to Publication

The following examples illustrate how the tool's specific affordances shaped the research questions it was possible to ask. Early versions of ManoWhisper directly supported four up-coming peer-reviewed publications, as well as a public-facing piece in *The Conversation*. Across these projects, ManoWhisper functioned as a shared research infrastructure, combining large-scale podcast transcription, sentence-level misogyny classification, time-series visualization, and searchable corpus navigation. Rather than treating automated outputs as analytic endpoints, we consistently used these tools to guide and contextualize feminist discourse analysis across large and other inaccessible audio corpora. 27

In "Mapping Online Misogyny and Feminist Digital Counterprotest in the Post-Pandemic Landscape," an article in *Women's Studies in Communication*, we engaged with ManoWhisper's collection of over 11,000 manosphere podcast episodes to trace how misogynist discourse circulates across platforms and becomes laundered into mainstream conservative talking points [MacDonald et al. 2026]. Sentence-level misogyny and hate speech classification allowed us to identify patterns in how podcasters such as Nick Fuentes and Ben Shapiro responded to Kansas City Chiefs kicker Harrison Butker's 2024 commencement speech. Delivered at Benedictine College, the address attracted widespread controversy for urging women to prioritize homemaking and motherhood over professional careers. Time-series visualizations enabled us to map spikes in misogynistic rhetoric and connect them to specific media events, demonstrating how manosphere discourse reinforces patriarchal gender hierarchies while providing ideological groundwork for initiatives like Project 2025. This policy blueprint was developed by the Heritage Foundation ahead of the 2024 U.S. election that proposes sweeping the rollback of reproductive rights, LGBTQ+ protections, and gender equity programs in favour of growing the American family — language that seeks to hide its pronatalist, anti-feminist themes. Our piece ultimately emphasized that what happens online is not merely cultural noise but a coordinated effort by conservative political organizations, media outlets, and right-wing influencers to shape gender norms and roll back feminist progress, and that digital platforms like podcasts (as evidenced through ManoWhisper's transcripts) function as battlegrounds where power is context and seemingly trivial content carries real political weight. 28

Similarly, in "Policing 'What a Feminist Looks Like': Interrogating the Scraps of Misogyny Across the Visual and Discursive Cultures of Social Media, Podcasts, and Generative AI," submitted to a special issue of *Big Data & Society*, ManoWhisper enabled longitudinal phrase-tracking across manosphere content from 2016 to 2025 [MacDonald, Wiens, and Ruest, in review]. By tracing the phrase "this is what a feminist looks like" we identified 19 podcast episodes that weaponized the phrase to disparage feminists. Misogyny scores helped us quantify the concentration of anti-feminist rhetoric in these episodes, while close readings of the full transcripts revealed how the phrase functioned as both mockery and a tool of what we call "the disinformation of misogyny" — digital content that seeks to invalidate feminists as credible speaking subjects. 29

In "Regulating Femininity in the Manosphere: An Exploration of Femmephobic Discourse in Andrew Tate and Nick Fuentes Podcasts," ManoWhisper supported a mixed-method analysis of over 1,300 podcast episodes from *America First* and *Tate Speech* by enabling large-scale transcription, corpus navigation, and targeted retrieval of gendered discourse [Padda, MacDonald, and Ruest 2026]. Systematic keyword-based searches (e.g., motherhood, rape, pro-choice, sexuality) across the full corpus, situated close qualitative readings of selected episodes within a broader discursive landscape. Sentence-level misogyny classification was used as a heuristic device to flag moments of heightened gendered hostility and guide feminist interpretive analysis rather than as a standalone analytic endpoint. This integration of computational filtering with close discourse analysis enabled us to trace how femininity is regulated through recurring binaries of hegemonic and pariah femininity, and how femmephobic rhetoric is normalized through podcasting's parasocial intimacy and long-form narrative structure. 30

ManoWhisper also supported "Mic Check for Misogyny: Podcasting, Male Supremacist Masculinities, and the Radicalization of Gendered Violence," a forthcoming peer-reviewed article examining podcasting as cultural and affective infrastructure for gender-based radicalization and extremism. Drawing on 1,834 podcast episodes from Andrew Tate, Nick Fuentes, and Gavin McInnes [Padda, Ruest, and MacDonald, in review], the project relied on ManoWhisper's transcription archive to identify relevant episodes and return to specific moments within long-form conversations. This approach made it possible to examine how male supremacist masculinities are constructed and normalized over time through everyday talk, humour, and grievance narratives, demonstrating how podcasts function as affective infrastructures that legitimize coercive masculinity and gendered violence through repetition, intimacy, and parasocial 31

authority.

Beyond our uses of ManoWhisper in recent and forthcoming publications, the site is being used by organizations that study extremism and hate, and also in provincial and federal government consultations and research, including Public Safety Canada. Owing to the nature of this work, we are hesitant to explicitly identify these other applications due to their sensitivity and individuals involved. In addition, through our emerging network collaborations with international partners, ManoWhisper is also being explored by scholars and policy groups in Europe and the UK. We admit that the majority of the work highlighted here is our own given the newness of the database. We hope that others can use this as a platform to drive and enable research questions. 32

The Scale and Scope of the Challenge

The sheer volume of data contained within manosphere podcast ecosystems makes ManoWhisper not just useful but necessary. As noted in the introduction, the corpus now exceeds 90,000 transcripts and more than 100,000 hours of content encompassing hundreds of thousands of sentences and millions of words. The diversity of actors, messaging strategies, and degrees of explicit versus coded language presents an analytical challenge that would be insurmountable through traditional qualitative methods alone. Within this ecosystem, we find a staggering breadth of rhetorical tactics. Some podcasters deploy overt calls to violence and explicitly dehumanizing language toward women and marginalized communities. Others employ dog whistles, that is, coded language that signals extremist meaning to in-group listeners while maintaining plausible deniability to outsiders, and what [Rothermel 2023] calls “evidence-based misogyny” — pseudo-rational arguments that mimic academic discourse to legitimize deeply gendered hierarchies. ManoWhisper's classification scores help us identify these patterns and track how rhetoric evolves over time. For instance, we can note when a show's misogyny scores spike in response to particular political events, when certain guests introduce new themes that subsequently spread across the network, or when coded language becomes more explicit in the wake of deplatforming efforts. Beyond the technical challenge, there is a profound human and emotional cost to engaging with this material. Listening to hours of content that advocates for violence against women, denies the existence of rape culture, and demonizes feminists as enemies of civilization takes a significant toll on researchers. ManoWhisper helps ease this burden in several ways. First, the webapp's searchable interface allows researchers to identify relevant segments without having to listen to entire episodes. Second, the visualization tools provide high-level overviews of content, enabling pattern recognition without sustained immersion in harmful rhetoric. Third, the misogyny and hate speech highlighting in transcripts helps researchers quickly locate and contextualize problematic content. While these affordances do not eliminate the emotional labor of this work, they do make it more manageable, sustainable, and ultimately “do-able” to engage in this kind of work. 33

Supporting Collaborative and Distributed Research

One of ManoWhisper's most significant contributions has been democratizing access to this data across our research team and collaborators. Not every researcher needs to be fluent in command-line tools, comfortable with Python, or skilled in RegEx pattern matching (and, certainly, two of the four researchers on this paper are not!). The webapp provides a shared, reproducible interface that team members with varying technical backgrounds can use to explore the corpus, test hypotheses, and conduct both exploratory and targeted analysis. For the members of our team who work at the intersections of feminist media studies, digital culture, and computational methods, ManoWhisper has enabled forms of hybrid analysis that combine the scale of computational approaches with the nuance of close reading. Researchers can search for specific terms across the entire dataset, filter results by podcast or time period, examine episode-level statistics, and then drill down into individual transcripts to understand how patterns manifest in context. 34

This flexibility supports the kind of feminist dwelling described above, allowing researchers to move between computational pattern detection and close rhetorical engagement with the text. Beyond our immediate team, ManoWhisper has been adopted by other researchers and organizations working to understand and counter online extremism. While we remain deliberately circumspect about some of these partnerships (both to protect ongoing work and to allow researchers and advocacy groups to operate without drawing unwanted attention from the communities they study), we can note that the tool has found applications in academic research, civil society organizations focused 35

on countering hate and disinformation, and educational initiatives aimed at building media literacy. The open-source nature of the ManoWhisper codebase, combined with clear documentation and the webapp's accessible interface, has facilitated this wider adoption while maintaining the feminist ethics of care that guide our approach.

Caveats and Limitations

36 While no researcher wants to admit this about their tool, ManoWhisper's outputs are not perfect, and the tool should be understood as an aid to human analysis rather than a replacement for it (much like our method of dwelling calls for). The transcripts generated by OpenAI's Whisper model, while remarkably robust, can struggle with poor audio quality, heavy accents, overlapping speakers, technical jargon, or names. The BART summarization model may occasionally miss key points, hallucinate, or introduce significant inaccuracies. Most significantly, the misogyny and hate speech classifiers — while trained on large datasets and validated against human annotations — produce both false positives (flagging content that is not actually problematic) and false negatives (missing content that should be flagged). Researchers must approach ManoWhisper's classifications critically, treating them as indicators rather than definitive judgments. In our own work, we always verify computational findings through close reading of transcripts, situate patterns within broader political and cultural contexts, and remain attentive to the limitations of any automated classification system (aka: dwelling). The classification scores are most useful as a way to identify potentially significant episodes or moments for further investigation, not as standalone evidence of harms.

37 Despite these limitations, ManoWhisper represents a significant methodological intervention in the study of online extremism and digital misogyny. By combining computational analysis with feminist commitments to dwelling, care, and situated knowledge, the tool enables research that is both rigorous and ethically grounded. It makes visible patterns that would otherwise remain hidden, provides evidence for the scale and sophistication of extremist media operations, and supports the kind of collaborative, interdisciplinary work necessary to understand and ultimately counter these harmful movements. For our team and our collaborators, ManoWhisper has transformed how we engage with manosphere content, making the work more sustainable, more systematic, and more shareable, while never losing sight of the human stakes involved in studying digital hate.

38 At the same time, though, it is worth acknowledging that the tool's design does not force or require researchers to engage with it through the methodology of dwelling. ManoWhisper's technical architecture cannot enforce the feminist commitments that motivated its development — a researcher could, for instance, run the pipeline quickly, extract classification scores or summaries, and move on without close reading, contextual situating, or critical reflection, and the tool would continue to function. This limitation warrants naming directly, as the ethical grounding comes from the researcher who wields the tool, not from the software itself. In this sense, ManoWhisper is susceptible to the same extractive, decontextualized uses that feminist methodology critiques. As feminist science and technology studies scholarship reminds us, tools are never simply neutral, but neither do they fully determine their own use [Winner 1980] [Wajcman 2004] [Eubanks 2017] [D'Ignazio and Klein 2020].

39 The values embedded in ManoWhisper, from its design choices to documentation and framing of outputs as indicators rather than verdicts, represent an invitation to dwell rather than a technical mandate. We have tried to build that invitation into every layer of the tool, but we cannot guarantee it will be accepted. As [D'Ignazio and Klein 2020] argue, data tools, while they can be used to challenge power, also always risk being used to reinforce it, depending on the choices and commitments researchers bring to their work. Researchers using ManoWhisper outside of a critical framework risk producing exactly the kind of rapid, large-scale, decontextualized analysis of marginalized communities' speech that feminist digital methods exist to contest. As such, we flag this as a limitation and encourage future users to attend carefully to the methodological commitments the tool was designed to support.

Considerations for Future Work: ManoWhisper and the Politics of Training Data

40 Beyond its immediate research applications, we see ManoWhisper as serving a critical function in documenting and analyzing how manosphere content may be implicated in the training of large language models themself (LLMS) and generative AI systems. As researchers on the team have demonstrated with their concept of “machine learning

misogyny" [Wiens and Kodituwakku 2025], computational systems increasingly "reproduce, amplify, and legitimize gender-based discrimination" (1) not merely through biased outputs but through the very data they are trained on. The massive scale of podcast content in the Manosphere (over 100,000 hours of transcribed audio representing millions of words) constitutes a significant portion of the openly accessible speech data available on the Internet. This raises urgent questions about whether and how this content becomes part of the datasets used to train conversational AI systems, chatbots, and other generative technologies.

41 ManoWhisper's corpus provides a unique resource for investigating this concern. By making manosphere podcast transcripts searchable and analyzable, this tool enables researchers to trace specific rhetorical patterns, misogynistic tropes, and hate speech formations that may be absorbed into LLM training data. For instance, our classification of misogyny and hate speech at the sentence level allows us to identify not only what harmful content exists, but how it is linguistically structured — that is, the specific phrases, argumentative moves, and discursive strategies that characterize this ecosystem. When LLMs are trained on large-scale web harvests that include podcast transcripts, YouTube captions, or social media content linking to these shows, they risk encoding, whether accidentally or not, these patterns into their generative outputs. The result, as researchers [Wiens and Kodituwakku 2025] on the team have argued, is not simply bias in a technical sense, but the structural reproduction of misogyny as a form of automated cultural logic. This concern is not abstract. Recent investigations into LLM training datasets have revealed widely used uncurated corpora like Common Crawl, which scrapes billions of web pages, inevitably include extremist content, hate speech, and conspiracy theories alongside more benign material [Bender et al. 2021] [Dodge et al. 2021] [Luccioni and Viviano 2021]. Manosphere podcasts, often hosted on platforms like YouTube, Rumble, and Spotify, are accompanied by automatically generated transcripts that are indexed by search engines and thus accessible to web-scraping tools.

42 ManoWhisper's documentation of this content allows us to ask: How much manosphere rhetoric is currently embedded in the datasets training systems like ChatGPT, Claude, Gemini, Grok, or Meta AI? What happens when conversational agents learn language patterns from podcasters who frame feminism as a civilizational threat, advocate for the execution of women who report assault, or describe women as inherently duplicitous? The implications extend beyond representation to function. If LLMs trained on manosphere content reproduce gendered discourse that dismisses survivors, naturalizes misogyny, or frames feminist critique as irrational, these systems become not neutral tools but active participants in what we have called the disinformation of misogyny: the systematic invalidation of feminists and women as credible speaking subjects [MacDonald, Wiens, and Ruest]. Scholars in critical AI studies have long contested the notion that these systems are or could be neutral, documenting how training data encodes existing social hierarchies, how model design reflects the priorities of those who build and fund it, and how outputs reproduce and amplify structural inequities [Noble 2018] [Benjamin 2019] [Gordon 2019] [Bender et al. 2021]. Our argument here is not that LLMs fail to achieve a neutrality they were designed to possess, but rather that the specific pipeline from manosphere podcast ecosystems into LLM training data constitutes a particular and traceable mechanism by which misogynist ideology enters these systems, one that has not yet been adequately documented or interrogated. In this sense, ManoWhisper offers not just a critique of AI neutrality in the abstract, but an evidentiary method for tracing how and from where specific harmful discourses are encoded. This is particularly concerning given the increasing deployment of LLMs in content moderation, mental health support, educational chatbots, and dating apps. As researchers on the team have articulated elsewhere, AI companions and conversational agents already exhibit troubling patterns of engaging in sexual conversations with minors, reinforcing gendered subservience, and automating feminized emotional labor [Wiens and Kodituwakku 2025]. If these systems are trained in part on manosphere transcripts, they risk not only reflecting existing misogyny but amplifying and legitimizing it through cultural authority these systems accrue that, as scholars have shown, often derives precisely from the false perception of computational objectivity [Pasquale 2016] [Noble 2018] [Benjamin 2019] [D'Ignazio and Klein 2020].

43 ManoWhisper thus functions as both a research tool and an evidentiary archive. It provides a way to trace the circulation of misogynist discourse from niche podcast networks into mainstream AI systems, making visible a pipeline that would otherwise remain hidden. Our ongoing work includes analyzing specific linguistic markers — phrases like "this is what a feminist looks like," "false accusations," "evidence-based misogyny," and "tradwife" (shorthand for "traditional wife," used in manosphere communities to idealize women who embrace domestic subservience) — to

determine how frequently they appear in our corpus and how they might be reproduced in LLM outputs. By identifying these patterns in the source material, we can better understand how they manifest in generative AI, and we can advocate for more rigorous content filtering and ethical training practices. This work aligns with broader calls within critical AI studies for transparency in training data, accountability for harmful outputs, and feminist oversight of the datasets and models shaping digital futures. Ultimately, ManoWhisper's value extends beyond mapping the Manosphere as a contemporary phenomenon. It also equips researchers, technologists, and policymakers to interrogate how today's extremist content becomes tomorrow's machine-learned common sense. In an era where LLMs are increasingly embedded in everyday technologies, from search engines to virtual assistants to educational tools, understanding what these systems have learned, and from whom, is essential to challenging machine learning misogyny and building more just, equitable, and perhaps feminist digital infrastructures.

Conclusion

For us, reflections on our past and future uses of ManoWhisper the tool paired with our uses of dwelling represents more than a technical solution to a data problem: It embodies a methodological intervention that brings together computational scale with feminist ethics of care. Throughout this paper, we have demonstrated how the tool functions simultaneously as a pipeline for processing extremist content, a webapp for collaborative research, and an archive that makes visible the scale and sophistication of manosphere discourse. By combining OpenAI's Whisper transcription model with classification algorithms for misogyny and hate speech, we can situate these computational outputs within a framework of dwelling. Here, we want to emphasize that feminist ethics and computational scale do not have to be in tension; by engaging reflexively with ManoWhisper, care and methodological diligence can co-exist, which stands as a reminder that the tools for studying harm need not enact those same harms. The research applications we have detailed, from mapping online misogyny in the post-pandemic landscape to tracing the weaponization of feminist language across platforms, illustrate how ManoWhisper facilitates work that would, as we have discussed, be quite impossible through manual scraping and analysis alone. The corpus cannot be read, heard, or analyzed by individual researchers working in isolation. However, what we want to emphasize once again here is that the tool does not and cannot replace human judgement; rather, it complements it, helping researchers to "zoom in" and identify moments worthy of sustained attention while never losing sight of the human stakes involved in studying digital harms. 44

As we look toward future applications, ManoWhisper's potential as an evidentiary archive for understanding how extremist content shapes cultural, political, social, and technical data becomes increasingly urgent. As mentioned above, the same transcripts that document contemporary misogynist rhetoric may already be (and likely are) circulating within the datasets that train LLMs, encoding these harmful patterns into the conversational agents, content moderation systems, and educational tools that mediate everyday life. By making this content searchable and analyzable, ManoWhisper equips researchers to trace the pipeline from extremist podcasts to machine-learned common sense — a pipeline that must be interrupted if we are to build more just and equitable digital futures. Ultimately, ManoWhisper reflects our team's commitment to making accessible yet technically sophisticated research tools that are open to collaborators with varying skillsets, while maintaining the depth necessary for scholarly work. It demonstrates that computational methods need not be extractive or decontextualized. Indeed, ManoWhisper reminds us that scale and care can coexist and that tools for studying harmful content do not have to themselves enact harm but can be built with attention to the emotional labor of studying such phenomena. As extremist movements continue to weaponize digital platforms to spread misogynist, racist, and homophobic ideologies, tools like ManoWhisper become crucial not only for documentation and analysis but for the broader project of feminist resistance in digital spaces. 45

Acknowledgements

This project is funded by the Social Sciences and Humanities Research Council. Additional financial and in-kind support comes from York University and York University Libraries; the Faculty of Arts and the David R. Cheriton School of Computer Science at the University of Waterloo; and the Digital Research Alliance of Canada. 46

Works Cited

Allchorn 2024 Allchorn, D.W. (2024) “A digital megaphone: The far right's use of podcasts for radicalisation”, *GNET*, 22 October. Available at: https://gnet-research.org/2024/10/22/a-digital-megaphone-the-far-rights-use-of-podcasts-for-radicalisation/ (Accessed: 23 April 2025).

Attanasio et al. 2022 Attanasio, G., Nozza, D., Hovy, D., and Baralis, E. (2022) “Entropy-based attention regularization frees unintended bias mitigation from lists”, in *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics, pp. 1105–1119. Available at: https://doi.org/10.18653/v1/2022.findings-acl.88.

Barad 2003 Barad, K. (2003) “Posthumanist performativity: Toward an understanding of how matter comes to matter”, *Signs: Journal of Women in Culture and Society*, 28(3), pp. 801–831. Available at: https://doi.org/10.1086/345321.

Bender et al. 2021 Bender, E.M., Gebru, T., McMillan-Major, A., and Smitchell, S. and (2021) “On the dangers of stochastic parrots: Can language models be too big?”, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Virtual Event, Canada: ACM, pp. 610–623. Available at: https://doi.org/10.1145/3442188.3445922.

Benjamin 2019 Benjamin, R. (2019) *Race after technology: Abolitionist tools for the new Jim code*. Available at: https://www.ruhabenjamin.com/race-after-technology (Accessed: 2 May 2025).

Berry 2016 Berry, R. (2016) “Podcasting: Considering the evolution of the medium and its association with the word 'radio'”, *Radio Journal: International Studies in Broadcast & Audio Media*, 14(1), pp. 7–22. Available at: https://doi.org/10.1386/rjao.14.1.7_1.

Clifton et al. 2020 Clifton, A., Pappu, A., Reddy, S., Yu, Y., Karlgren, K., Carterette, B., and Jones, R. (2020) “The Spotify podcast dataset”. arXiv. Available at: https://doi.org/10.48550/arXiv.2004.04270.

Corall 2024 Corall, J. (2024) “scyllaridae-whisper”. Lehigh University. Available at: https://hub.docker.com/r/lehighlts/scyllaridae-whisper.

Czerwinsky 2024 Czerwinsky, A. (2024) “Aggressive communication online”, in S.L. Connaughton and S. Pukallus (eds) *The Routledge handbook of conflict and peace communication*, 1st edn. New York: Routledge, pp. 160–169. Available at: https://doi.org/10.4324/9781003392002-19.

D'Ignazio and Klein 2020 D'Ignazio, C. and Klein, L. (2020) *Data feminism*. Cambridge, Massachusetts: MIT Press.

Dodge et al. 2021 Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groenveld, D., Mitchell, M., and Gardner, M. (2021) “Documenting large webtext corpora: A case study on the Colossal Clean Crawled Corpus”. arXiv. Available at: https://doi.org/10.48550/arXiv.2104.08758.

Drucker 2011 Drucker, J. (2011) “Humanities approaches to graphical display”, *Digital Humanities Quarterly*, 5(1).

Eubanks 2017 Eubanks, V. (2017) *Automating inequality: How high-tech tools profile, police, and punish the poor*, 1st edn. New York: St. Martin's Press.

Ging 2019 Ging, D. (2019) “Alphas, betas, and incels: Theorizing the masculinities of the manosphere”, *Men and Masculinities*, 22(4), pp. 638–657. Available at: https://doi.org/10.1177/1097184X17706401.

Gold and Klein 2016 Gold, M.K. and Klein, L.F. (eds) (2016) *Debates in the digital humanities 2016*. University of Minnesota Press. Available at: https://doi.org/10.5749/j.ctt1cn6thb.

Gordon 2019 Gordon, F. (2019) “Virginia Eubanks (2018) Automating inequality: How high-tech tools profile, police, and punish the poor”, *Law, Technology and Humans*, pp. 162–164. Available at: https://doi.org/10.5204/lthj.v1i0.1386.

Gumbs 2018 Gumbs, A.P. (2018) *M archive: After the end of the world*. Durham: Duke University Press.

Gumbs and brown 2020 Gumbs, A.P. and brown, a.m. (2020) *Undrowned: Black feminist lessons from marine mammals*. Chico: AK Press (Emergent strategy series, no. 2).

Jockers 2013 Jockers, M.L. (2013) *Macroanalysis: Digital methods and literary history*. Urbana: University of Illinois Press (Topics in the digital humanities).

Kimmerer 2013 Kimmerer, R.W. (2013) *Braiding sweetgrass*, 1st edn. Minneapolis: Milkweed Editions.

Llinares, Fox, and Berry 2018 Llinares, D., Fox, N. and Berry, R. (eds) (2018) *Podcasting: New aural cultures and digital media*. Cham: Springer International Publishing. Available at: https://doi.org/10.1007/978-3-319-90056-8.

Losh and Wernimont 2018 Losh, E.M. and Wernimont, J. (eds) (2018) *Bodies of information: Intersectional feminism and*

digital humanities. Minneapolis: University of Minnesota Press (Debates in the digital humanities).

Luccioni and Viviano 2021 Luccioni, A.S. and Viviano, J.D. (2021) “What's in the box? A preliminary analysis of undesirable content in the Common Crawl corpus”. arXiv. Available at: https://doi.org/10.48550/arXiv.2105.02732.

MacDonald et al. 2026 MacDonald, S., Wiens, B.I., Ruest, N., and Padda, K. (2026) “'All you have to do is publicly execute a few women who have lied': Mapping online misogyny and feminist digital counterprotest in the post-pandemic landscape”, *Women's Studies in Communication*. Available at: https://doi.org/10.1080/07491409.2025.2579089.

MacDonald, Wiens, and Ruest 2025 MacDonald, S., Wiens, B. and Ruest, N. (2025) “Dwelling with feminist media archives in the age of big data”, *Internet Histories*, 9(1–2), pp. 198–215. Available at: https://doi.org/10.1080/24701475.2024.2431799.

MacDonald, Wiens, and Ruest MacDonald, S., Wiens, B.I. and Ruest, N. (in review) “Policing 'what a feminist looks like': Interrogating the scraps of misogyny across the visual and discursive cultures of social media, podcasts, and generative AI”, *Big Data & Society*.

McPherson 2012 McPherson, T. (2012) “Why are the digital humanities so white?”, in M.K. Gold (ed.) *Debates in the digital humanities*. University of Minnesota Press.

Miller 2024 Miller, M. (2024) “Lomax & Whisper.cpp: Automatic transcription of Alan Lomax's 1938 Midwest Folk Song Collection using Whisper.cpp”, 12 September. Available at: https://thisismattmiller.com/post/lomax-whisper/.

Miller-Idriss 2025 Miller-Idriss, C. (2025) *Man up: The new misogyny and the rise of violent extremism*. Princeton: Princeton University Press.

Moretti 2013 Moretti, F. (2013) *Distant reading*. London: Verso.

Noble 2018 Noble, S. (2018) *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press. Available at: https://nyupress.org/9781479837243/algorithms-of-oppression/ (Accessed: 2 May 2025).

Padda, Ruest, and MacDonald Padda, K., Ruest, N. and MacDonald, S. (in review) “Mic check for misogyny: Podcasting, male supremacist masculinities, and the radicalization of gendered violence”, *Men and Masculinities* [Preprint].

Padda, MacDonald, and Ruest 2026 Padda, K.K., MacDonald, S. and Ruest, N. (2026) “Regulating femininity in the manosphere: An exploration of femmephobic discourse in Andrew Tate and Nick Fuentes podcasts”, *Journal of Femininities*, pp. 1–29. Available at: https://doi.org/10.1163/29501229-bja10032.

Pasquale 2016 Pasquale, F. (2016) *The black box society: The secret algorithms that control money and information*, 1st paperback edn. Cambridge, Massachusetts: Harvard University Press.

Radford et al. 2022 Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I.(2022) “Robust speech recognition via large-scale weak supervision”. arXiv. Available at: https://doi.org/10.48550/arXiv.2212.04356.

Risam 2019 Risam, R. (2019) *New digital worlds: Postcolonial digital humanities in theory, praxis, and pedagogy*. Evanston: Northwestern University Press.

Rizwan et al. 2025 Rizwan, N., Deb, N., Roy, S., Solanki, V.S., Garimella, K., and Mukherjee, A. (2025) “Dynamics of toxicity in political podcasts”. arXiv. Available at: https://doi.org/10.48550/arXiv.2501.12640.

Rodriguez and Brown 2023 Rodriguez, D. and Brown, B.J. (2023) “Comparative analysis of automated speech recognition technologies for enhanced audiovisual accessibility”, *Code4Lib Journal* [Preprint], (58). Available at: https://journal.code4lib.org/articles/17820.

Rothermel 2023 Rothermel, A. (2023) “The role of evidence-based misogyny in antifeminist online communities of the 'manosphere'”, *Big Data & Society*, 10(1), p. 20539517221145671. Available at: https://doi.org/10.1177/20539517221145671.

Ruest 2025a Ruest, N. (2025) “les-observateurs”. Zenodo. Available at: https://doi.org/10.5281/ZENODO.18075828.

Ruest 2025b Ruest, N. (2025) “manodexer”. Zenodo. Available at: https://doi.org/10.5281/ZENODO.18075870.

Ruest 2025c Ruest, N. (2025) “manometrics”. Zenodo. Available at: https://doi.org/10.5281/ZENODO.18075892.

Ruest and Padda 2025 Ruest, N. and Padda, K. (2025) “Digital-Feminist-Network/ManoWhisper: 1.2.0”. Zenodo. Available at: https://doi.org/10.5281/zenodo.15287350.

Simonsen 2025 Simonsen, E. (2025) *Knowledge Fight interactive search tool*. Available at: https://fight.fudgie.org/

(Accessed: 22 January 2025).

Singh 2025 Singh, S. (2025) *How many podcasts are there? (2025 growth stats)*. Available at: https://www.demandsage.com/podcast-statistics/.

Underwood 2019 Underwood, T. (2019) *Distant horizons: Digital evidence and literary change*. Chicago: University of Chicago Press.

Vidgen et al. 2021 Vidgen, B. Thrush, T., Waseem, Z., and Kiela, D. (2021) “Learning from the worst: Dynamically generated datasets to improve online hate detection”, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, pp. 1667–1682. Available at: https://doi.org/10.18653/v1/2021.acl-long.132.

Wajcman 2004 Wajcman, J. (2004) *TechnoFeminism*. Cambridge: Polity.

Wiens 2022 Wiens, B. (2022) *How to use creative and embodied methods*, ed. by K. Gregory. London: SAGE Publications. Available at: https://doi.org/10.4135/9781529608359.

Wiens, Ruest, and MacDonald2025 Wiens, B., Ruest, N., and MacDonald, S. (2025) “Misogyny has become a political strategy — here's how the pandemic helped make it happen”, 18 June. Available at: https://doi.org/10.64628/AAM.yqefqv4qc.

Wiens and MacDonald 2024 Wiens, B. and MacDonald, S. (2024) “Dwelling as method: Lingering in/with feminist curated data sets on Instagram”, *Journal of Digital Social Research*, 6(2), pp. 27–45. Available at: https://doi.org/10.33621/jdsr.v6i2.211.

Wiens 2021 Wiens, B.I. (2021) “Virtual dwelling: Feminist orientations to digital communities”, in *Networked feminisms: Activist assemblies and digital practices*. Lanham: Rowman & Littlefield.

Wiens and Kodituwakku 2025 Wiens, B.I. and Kodituwakku, A. (2025) “The future of digital feminism: Surviving surveillance, misinformation, and machine learning misogyny”, *IEEE Transactions on Technology and Society*, pp. 1–14. Available at: https://doi.org/10.1109/TTS.2025.3615509.

Winner 1980 Winner, L. (1980) “Do artifacts have politics?”, *Daedalus*, 109(1), pp. 121–136.

brown 2017 brown, a.m. (2017) *Emergent strategy: Shaping change, changing worlds*. Chico: AK Press.