Gendering Media

Framing of AI, Interacting with ChatGPT, and Anti-fandom

Patrik Åker & Anne Kaun (eds.)

Gendering Media

Framing of AI, Interacting with ChatGPT, and Anti-fandom

Patrik Åker & Anne Kaun (eds.)

Södertörns högskola

SVANENMÄRKET
Trycksak
3041 0865

Södertörn University
Library
SE-141 89 Huddinge
www.sh.se/publications

Cover layout: Jonathan Robson
Graphic form: Per Lindblom & Jonathan Robson

Printed by E-Print, Stockholm 2025

Mediestudier vid Södertörns högskola, 2025:1
ISSN 1650-6162
ISBN 978-91-89962-03-3

Contents

Introduction

This volume contains adaptions of three noteworthy master theses written within the international master's programme in Media, Communication, and Cultural Analysis at Södertörn University and defended in 2024. Running since 2009, the programme has more than 100 alumni who are employed in the media, academia and education. In 2020, the department chose to distinguish the best theses in a printed volume. This is the fifth volume in the series.

The contributions in this volume cover three different topics: how gender influences the representation of influential people in AI; how Ukrainian women from the European diaspora interact with ChatGPT about the Russian-Ukrainian war; and the anti-fandom culture surrounding Taylor Swift on social media. Although stretching across three such different topics, the chapters share an interest in on how media can be understood in relation to gender. In two of the chapters, gendering is explicitly made visible as an ongoing process, while in the study on women interacting with ChatGPT it is done indirectly. However, common for the three chapters is that they fruitfully re-center our attention on how gender frames our everyday digital practices and discourses.

The department for media and communications studies invites readers to engage with this crucial and critical work conducted by our master's students.

Stockholm, 6 March 2025

Patrik Åker, Programme Director
Anne Kaun, Professor

News Framing and Gender Stereotyping of Influential People in Artificial Intelligence

Juliana Frias Lisboa

Introduction

ChatGPT was released on November 30th, 2022, and quickly gained significant attention for its ability to interact with users in a conversational way, catapulting artificial intelligence (AI) into mainstream awareness (Taecharungroj, 2023). AI, as a subfield of Science, Technology, Engineering, and Mathematics (STEM), is rapidly transforming our world. As the public grapples with understanding this complex technology (Zhang & Dafoe, 2019; Yeh et al., 2021; Bao et al., 2022), the role of media becomes increasingly critical. The portrayal of specific individuals as the "face" of AI carries significant weight in shaping cultural perceptions regarding participation in scientific domains (LaFollette, 1988). They serve as personifications of this complex and rapidly evolving field, making their representation critical to public understanding (Pentzold et al., 2018).

This study analyses 248 profiles of "influential people in AI" across four prominent English-language publications. By employing interpretive content analysis, the research delves into both quantitative and qualitative dimensions of representation. The quantitative analysis focuses on the numerical distribution of gender and race among these influential figures, recognizing that these categories are not isolated but rather intertwined and mutually constitutive (Crenshaw, 1991). The qualitative analysis was grounded in framing theory (Entman, 1993) which posits that media outlets shape public understanding by selectively highlighting certain aspects of an issue. The study adopted a deductive coding scheme that intersected common news frames (Semetko & Valkenburg, 2000) with gender stereotypes (Tuchman, 1978a; Nelkin, 1987; LaFollette, 1988; Shachar, 2000; Steinke, 2005; Chimba and Kitzinger, 2010; among others).

Extensive research has documented the detrimental impact of gender stereotypes on women's participation and advancement in science and technology (Quinn & Spencer, 2001; Reuben et al., 2014; Miller et al., 2015; Mitchell & McKinnon, 2019; Mckinnon & O'Connell, 2020; Exley & Kessler, 2022). Nevertheless, research that intersects media representation, gender and AI are scarce and concerning, with few emerging studies indicating that, in the media representations of the booming field of AI, women are often stereotyped or absent (Pentzold et al., 2018; Cave et al., 2023a; Chen et al., 2023). This study aims to expand this body of research by focusing on news media that dictates who are the influential people in artificial intelligence.

In essence, this study explore gender representation and news framing of influential figures in AI, uncovering the ways in which media narratives both reflect and shape societal beliefs of who holds power and expertise in this rapidly evolving field.

Background: Booming AI, Stagnant Women: The Gender Gap in Artificial Intelligence

AI now promises to revolutionise our world, attracting hundreds of billions in investment (Thormundsson, 2024a; 2024b) but its potential for societal transformation comes with ethical concerns like mass surveillance, biassed decision-making, and manipulation (Stathoulopoulos & Mateos-Garcia, 2019). Studies in the US and EU show that the public is cautious with the adoption of AI, with more than 80% agreeing that it should be careful management, 34% believing high-level machine intelligence will be harmful to humanity, and 12% anticipating extremely adverse consequences (Zhang & Dafoe, 2019).

Demographics significantly predict support for developing AI, with men significantly more likely to back its development compared to women, those with lower education, lower income, and no tech background (Zhang & Dafoe, 2019; Yeh et al., 2021). Ensuring diverse groups' perspectives on the potential benefits and risks posed by AI is critical, as its impacts will be significant and likely unevenly distributed (Bao et al., 2022). Therefore, increasing participation of women and other minority groups in AI development and public debate is crucial to mitigating risks and ensuring broader societal benefit (Stathoulopoulos & Mateos-Garcia, 2019).

According to UN Women & United Nations Department of Economic and Social Affair, (2022), women hold only 2 out of every 10 jobs globally in science, engineering and information and communication technology, making new innovations less likely to solve women's needs and continuing to reproduce gender bias, a vicious cycle. The specific field of artificial intelligence (AI), only 30% of individuals working in AI are women, a minor increase of approximately 4 percentage points compared to 2016 (World Economic Forum, 2023). Young et al. (2021) provided a detailed analysis that help to explain the slow progress of woman participation in AI and data science: persistent structural inequalities based on perceptions, as men self-report more skills than women, and women are often relegated to lower-status roles, even though have higher formal education levels.

Literature Review and Theoretical Framework

This thesis investigates the gender representation of influential figures in artificial intelligence as portrayed in online news media. Understanding who is presented as the "face" of AI is crucial, as these portrayals significantly shape public perception of this promising technology (LaFollette, 1988; Pentzold et al., 2018). Building on extensive research that has problematized the low visibility and stereotyping of women in science and technology media (Nelkin, 1987; LaFollette, 1988; Shachar, 2000; Steinke, 2005; among others), this study specifically examines AI media representation, an area where emerging research suggests similar concerning trends (Pentzold et al., 2018; Chen et al., 2023; Cave et al., 2023).

Gender representation in science and technology media often relies on stereotypes that portray men as natural leaders and innovators while depicting women through their appearances or domestic roles. Nelkin (1987) found that male scientists were often portrayed as isolated intellectuals, while portrayals of female scientists emphasized familial roles or domesticity; for example, Nobel laureate Rosalyn Yalow was labelled by the media as a "Bronx housewife" (Nelkin, 1987, p. 19). This pattern continued in later studies; Shachar's (2000) analysis of The New York Times observed that coverage of female scientists focused more on family life challenges than on their professional achievements. Similarly, Mitchell and McKinnon (2019) noted that profiles of female scientists often included references to personal appearance or relationship status, elements rarely highlighted for male counterparts.

A recent large-scale study across 66 countries similarly revealed that gender stereotypes correlated with lower female numerical participation in science (Miller, Eagly, & Linn, 2015). However, research proving that stereotypes within media representation create tangible barriers for women in STEM are not new. Zuckerman and Cole's "Principle of the Triple Penalty" (1975) argued that societal beliefs in female inferiority, compounded by institutional discrimination, restrict women's opportunities in scientific fields. Gender stereotypes impact from self-assessment, with women in STEM consistently rating their abilities lower than men (Exley and Kessler, 2022) to hiring, with biases favouring men (Reuben, Sapienza, & Zingales, 2014). To challenge stereotypes that prevent equal opportunities, media portrayals of women in AI should depict them as competent, integral figures rather than as exceptions or symbols (Geena Davis Institute on Gender in Media & The Lydia Hill Foundation, 2018).

The theoretical framework employed for analysing gender dynamics and the framing of influential figures in artificial intelligence within online news media was threefold. First, examining how gender was understood as a social construct, influenced by interactions with the world, including media representations (Butler, [1990] 2006; Mulvey, 1985; van Zoonen, 1994; Krijnen & Van Bauwel, 2021; among others). Second, exploring the relationship between gender portrayals and stereotypes, elaborating on the associated power dynamics and ideological implications (Tuchman, 1978a; Perkins, 1979; Ridgeway, 2011; Hall, 2013; among others). Finally, exploring the notion that news constituted a constructed reality, shaped by the framing choices made by news organisations (Tuchman, 1978b; Entman, 1993; Carragee & Roefs, 2006; Burscher et al., 2016; among others).

Framing theory suggests that news reporting is not merely factual but actively shapes public perception through selective emphasis (Baresch et al., 2010, p. 638). While objectivity remains a journalistic ideal, achieving it is challenged by inherent biases in story selection and source framing (Pezzullo & Cox, 2022, p. 184). Entman (1993) defines framing as emphasizing specific aspects within a message to influence audience interpretation and opinion. AI media coverage, for instance, often highlights ethical or economic concerns, though these discussions frequently lack depth on nuanced topics such as data privacy (Ouchchy et al., 2020; Nguyen & Hekman, 2022). Studies indicate that male voices in STEM tend to receive more coverage, with visual framing further reinforcing authority by positioning men with direct gazes or in larger, central image formats, in contrast to women, who are more often depicted with averted gazes or smaller, peripheral images (Chen et al., 2023).

Statement of Purpose and research questions

This thesis contributes to the growing body of work by employing both quantitative and qualitative analyses, with an analytical framework that intersects framing, stereotyping and intersectional lens of race to gender.

The aim is to explore gender representation, stereotyping and framing of AI influential figures in online news media.

Research Questions

RQ1: How does the numerical representation of "influential people of AI" vary by gender and race across different publications?
RQ2: If and how are framing and gender stereotypes present in news coverage of "influential people in AI"?

Material

The corpus for this study consisted of online news articles featuring lists of influential figures in artificial intelligence, published between November 30th, 2022 – the launch of ChatGPT – and April 15th, 2024, cut date of this study – capturing 17 months of heightened public interest in AI.

Four English-based online publishers – TIME, The New York Times, Business Insider, and TechCrunch – were selected to compose a comprehensive perspective of prominent representation of "influential people". TIME and The New York Times were included as they are widely respected general-interest publications with broad audiences and established reputations in shaping public opinion on social and technological developments. In the other hand, Business Insider and TechCrunch, were included as renowned niche publications on business and technology, providing insights into the perspectives of industry professionals and technology enthusiasts. Together, these publications capture AI representation across mainstream and tech-media outlets, allowing comparisons, and ultimately enabling a robust and comprehensive analysis.

Within these publications, news content was selected by searching keywords "AI influential people", "AI key figures", and "artificial intelligence relevant people". General text that mentions influential people or single profiles were not included, but only articles that feature lists of profiles that they claim to be influential figures in AI – which will be referred as "article-lists".

Using the above criteria, five article-lists were selected to compose the data analysed by this research. The table below summarize the corpus, with details and links to the original publications – and clarifying adopted reference name for the rest of the text:

Online News Publisher	**Date of Publication***	**Article-list Title**	**Number of Profiles***	**Reference Name Adopted**
TIME	Sep 7, 2023	TIME100: Most Influential People in AI	100	TIME
Business Insider	Oct 10, 2023	The AI 100 2023: The top people in artificial intelligence	100	BI 100
The New York Times	Dec. 3, 2023 (Updated on Dec 7, 2023)	Who's Who Behind the Dawn of the Modern Artificial Intelligence Movement	12	NYT
Business Insider	Feb 22, 2024 (Updated on Mar 17, 2024)	From the 'godfathers of AI"to newer people in the field: Here are 17 people you should know – and what they say about the possibilities and dangers of the technology	17	BI 17
TechCrunch	Mar 10, 2024 (Updated on Apr 6, 2024)	The women in AI making a difference: TechCrunch highlights notable women in the field of AI	19	TechCrunch

* Last updated as of April 15th, 2024 – cut date of the study.

Table 1. Corpus Description. Source: Author. Each article-list also varied in length, ranging from 12 to 100 profiles each, making a total of 248 profiles of individuals considered "influential in AI".

Method

The research method used was content analysis, which Drisko and Maschi (2015) define as a systematic method for deriving valid inferences from texts. This approach is particularly suitable for examining gender dynamics (Gill, 2007). The study specifically utilised interpretive content analysis, enabling the integration of qualitative coding techniques to understand content nuances alongside quantitative methods for summarising and analysing data (Drisko & Maschi, 2015).

The analysis began with systematic cataloguing. First, all images were downloaded and stored in an online database (Google Drive), organized by type: "main illustrations" for introductory visuals and "profile images" for individual profiles. The 248 profiles in total included 246 with images, primarily headshots sourced directly from individuals or organizations, though some were from stock images or professional photographers. Only two profiles from TIME did not include direct images of the profiled individuals. Next, images were coded in detail. Each image was classified by gender (woman, man, non-binary) and race (Asian, Black, Latinx, Middle Eastern, and White), with verification achieved by consulting public social media profiles.

Finally, a deductive coding scheme guided the qualitative analysis of the texts, drawing from common news frames (Semetko & Valkenburg, 2000) and established literature on gender stereotypes (e.g., Tuchman, 1978a; Nelkin, 1987; LaFollette, 1988; Shachar, 2000; Steinke, 2005; Chimba & Kitzinger, 2010; Ridgeway, 2011; Mitchell & McKinnon, 2019; McKinnon & O'Connell, 2020; Fahy & Lewenstein, 2021; Young et al., 2021; Nader et al., 2022; Cave et al., 2023a).

This comprehensive approach of image and text catalogue and coding enabled, at first, the analysis of numerical representation of influential people in AI, by looking into the distribution of race and gender of the profiles across article-list:

Research Question	Analytical Approach	Category	Subcategories
#1 How does the numerical representation of "influential people of AI" vary by gender and race across different publications?	Quantitative – Distribution of Gender and Race across publications	Gender	Women, Men and Non-Binary
		Race	White, Asian, Black, Middle Eastern and Latinx
		Article-list	TIME, BI 100, BI 17, NYT and TechCrunch

Table 2. Analytical framework part 1/2. Source: Author. As a second step, the qualitative analysis of more nuanced notions of representation, by exploring how and if different stereotypes were manifested within common news frames, offering insight into the intersection of media framing and gender stereotyping.

Research Question	Analytical Approach	Frames	Men Stereotype	Women Stereotype
#2 If and how are framing and gender stereotypes present in news coverage of "influential people in AI"?	Qualitative – Presence of Stereotypes across Textual News Frames	Human Interest frame	Power and status, Boffins scientist, Young geek tech entrepreneur and Prodigy genius	"Being a woman in science", Motherhood, Appearance and Sexualization.
		Responsability	Competent leader and White-male privilege	Responsible and Less technical

	Economic Consequence	Corporate power and Wealth	Outliers
	Morality	Optimistic and Artificial life	Pessimistic and Communality
	Conflict	Competent, Mellow or Hyper masculine	Sacrifice, Victims, Aggressive or Subservient

Table 3. Analytical framework part 2/2. Source: Author.

Reporting and Analysis of Results

By the Numbers: Gender and Race Distribution across "Influential People of AI" Online News

Of 248 profiles of "influential people in AI", men made up 58% of the profiles (n=144), women 41% (n=101), and non-binary individuals 1% (n=3). The higher presence of male aligns with previous research on gender and media representation of AI, though this study observed a diminished gender gap. Pentzold et al. (2018) study of AI image protagonists presented 85% of men, while Chen et al. (2023) study of AI news images showcased 83.5% of men and Cave et al. (2023) 92% of men engineers or scientists in popular AI-related films. Therefore, this study contributes to the ongoing discourse on gender representation in AI media, highlighting a reduced numerical disparity of genders, though with continued male predominance, within new profiles of "influential people in AI".

	Number of profiles of influential people in AI	Percentage of total
Men	141	58%
Women	101	41%
Non-binary	3	1%
Grand Total	248	100%

Table 4. Gender distribution of profiles of influential people in AI.

Nevertheless, a closer inspection reveals notable variations of gender participation across article-lists. TIME's and both BI compilations slightly surpassed the study average, with men numerical representation at 60%, 62%, and 59%, respectively. Conversely, the NYT and TechCrunch lists present striking contrasts, with men exclusively featured in the former and entirely absent in the latter. The NYT article-list, published in December 2023, did not explicitly mention gender as a criteria for selecting individuals as influential people of AI. Consequently, the list comprised exclusively male figures, implying a neutral stance, but perpetuating the perception of AI as a male-dominated domain. Conversely, TechCrunch's article-list, released in March 2024, claims to be a deliberate response to the absence of women in NYT, highlighting "remarkable women who've contributed to the AI revolution". This stark contrast between the two lists highlights the different editorial decisions and priorities concerning gender representation, an opportunity area for future research.

	TIME (n=100)	BI 100 (n=100)	NYT (n=12)	BI 17 (n=17)	TechCrunch (n=19)
Men	60%	62%	100%	59%	
Women	38%	37%		41%	100%
Non-binary	2%	1%			
Grand Total	100%	100%	100%	100%	100%

Table 5. Gender distribution of profiles of influential people in AI by article-list.

Gender distribution of profiles by race was also analysed, following Crenshaw (1991) theory of intersectionality. Among the individuals profiled, 58% were identified as White, leading to most of the profiles being of White-men (n=92/248). Notably, White-women made up 20% of the total profiles (n=49), nearly half the representation of White-men, yet almost equivalent to the combined representation of women from all other racial/ethnic groups (n=52). Asians emerged as the second largest group, forming 25% of the total profiles (n=63). Black individuals accounted for only 9% of the profiles (n=22), with representation primarily concentrated among women (n=20). Middle Eastern and Latinx individuals were notably underrepresented, including only 5 individuals in total.

	White (n=145)	Asian (n=62)	Black (n=22)	Middle Eastern (n=14)	Latinx (n=5)	**Grand Total (n=248)**
Men	66%	58%	9%	64%	40%	**58%**
Women	34%	39%	91%	36%	60%	**41%**
Non-binary	1%	3%	0%	0%	0%	**1%**
Grand Total	100%	100%	100%	100%	100%	**100%**

Table 6. Gender distribution of profiles of influential people in AI by race and total.

It is important to highlight that the distribution of race across list-articles also varied, but all of them had the sum of profiles from other races lower than only White individuals. NYT, characterised as a man only article-list, exhibited a striking 83% composition of White-men (n=10/12), marking the highest representation of a single race-gender within the sample. Nevertheless, TechCrunch, characterised as a woman only article-list, displayed the second highest representation of a single race-gender, with 63% of profiles being White-women (n=12/19), highlighting the limitation that an increase in women representation alone does not necessarily alleviate structural racial biases.

These findings, while acknowledging some level of interplay of gender, race, do not exhaust intersection of gender with other social dimensions, emphasising the need for continued research of this field.

Beyond the Numbers: News Framing and Stereotyping of “Influential People of AI”

This chapter explores if and how different stereotypes were manifested in profiles of influential people of AI across common news frames identified by Semetko & Valkenburg (2000).

A first overview reveals that Human Interest and Responsibility frames are universally present (100%, n=248), which is consistent with the nature of profiles highlighting individual achievements. Variations exist in the prevalence of Economic Consequences (83%, n=206), Conflict (79%, n=185) Morality (75%, n=185) frames, suggesting potential gender-based differences, but it is essential to refrain from drawing definitive conclusions regarding any underlying associations.

From this point, the focus shifts to the qualitative analysis of gender stereotypes as framing devices (deVreese, 2005, p. 54; Carragee & Roefs; 2006, p. 223) that can carry ideologies that sustain existing power dynamics (Perkins, 1979, p. 141). Therefore, the following sections used five common news frames (Semetko & Valkenburg, 2000) as background to analyse the presence of these ideological constructs within the profiles:

Human Interest Frame

According to Semetko & Valkenburg (2000, pp. 95-96), “Human interest” frame brings a human face or an emotional angle to the presentation of an event, issue, or problem. As mentioned earlier, all texts provided a human face to the issue, as they meant to be someone’s profile as an influential person in AI, making it present in 100% of profiles. This frame provides a unique lens for analysing gender representation as it reveals how personal contexts and vignettes are associated with individuals, often reflecting, or reinforcing gender stereotypes.

The literature often depicts figures like Einstein or Darwin as the "real" scientists (Chimba & Kitzinger, 2010; Fahy & Lewenstein, 2021). Numerous examples from the corpus echo this stereotype, with description of a lifelong dedication to science and brilliant outcomes. More notably, the term "godfather" was repeatedly used to referred to three White-men (Geoffrey Hinton, Yoshua Bengio, Yann LeCun), positioning them as originators and authorities of AI.

The young geek tech entrepreneurs (Chimba and Kitzinger, 2010; Cave et al., 2023a) were also a common stereotype found in men's profiles, marked by descriptions of ambitious personalities, socially unskilled and remarkable achievements since childhood. Some profiles of the men portrayed as influential people were also described as child prodigies, exhibiting exceptional intelligence and skills from small age – which entails notions of brilliance as something innate. Interestedly, some men listed as "influential people in AI" had no formal education. These depictions align with literature that called out Prodigy as subset of the male-associated "genius" trope (Cave et al., 2023).

When it comes to women, a common stereotype was simply how gender becomes central to their professional profiles, best described as a "female scientist" by Chimba & Kitzinger (2010) or "the dilemma of being a woman in science" by Shachar (2020), but also referenced by several other academics (LaFollette, 1988; Nelkin, 1987; Shachar, 2000; Steinke, 2005). It is particularly relevant as it brings to light the contradiction of one's intellectual professional success being defined by their gender.

Influential women in AI, differently from their male counterparts, not only answered questions about their work, but also regarding their experiences in navigating the field systemic barriers. The "dilemma of being a woman in science" was highlighted by TechCrunch as they asked all 19 women interviewed at least one of the following questions: "How do you navigate the challenges of the male-dominated tech industry and, by extension, the male-dominated AI industry?" and "What advice would you give to women seeking to enter the AI field?". TIME had a similar question to Elham Tabassi: "Science and technology, including AI, is traditionally dominated by men. I'm curious what you'd say to young people and women in particular who want to go into this type of work like you?". No gendered questions were made for men across.

Another common stereotype women were associated to was "the superwoman" (LaFollette, 1988; Steinke, 2005), which claims that most of the female scientists sacrificed normalcy and traditional roles, showing relentless determination even when faced with adversity. This was especially common on Non-White women profiles. Which aligns with claims by Chimba & Kitzinger (2010) and Cave et al. (2023) that, in the context of media representation of scientists, race worsened in intersection with gender (Crewshaw, 1991), often mentioned as othering.

Another aspect widely discussed by scholars is how gender shapes notions of domesticity (Nelkin, 1987; Shachar, 2000; Mitchell & McKinnon, 2019) even in a context of professional profiling. References to family, motherhood, and appearance, perpetuating the stereotype that associates women in STEM with caregiving roles and domestic concerns. Women's profiles frequently incorporated these family themes, with many professionals emphasising the role of family members as collaborators or sources of inspiration.

Moreover, parenthood has historically been associated with women, a stereotype that associated them as "naturally" more caring and nurturing (Ridgeway, 2011, pp. 59-61). This stereotype has also been seen in media representation of women scientists, who, different from men, have their family life included in the news coverage of their professional accomplishments – women scientists are not only expected to excel in their careers but also fulfil traditional roles as wives and mothers (Nelkin, 1978; Shachar, 2000). This trend was also found in our corpus, where several references of motherhood were found in women's profiles, but not in men's.

The contrasting absence of parenthood in profiles of men analysed in this study aligns with a masculine stereotype of lack of familiar interference (Shacar, 2000). Intriguingly, the only time parenthood was found in a man's profile, it was not discussing his domestic life, but rather making an analogy of raising AI as raising babies. This passage also connects with literature of stereotypical representations of entertainment media of human males developing feelings for AIs (Nader et al., 2022).

The increased interest in (or judgement of) their physical appearance, fashion choices or demeanours is another common stereotype presented in literature (Steinke, 2005; Chimba & Kitzinger, 2010; McKinnon & O'Connell, 2020), with sexualization of women scientists also linked to this overall increased attention to appearance. Only few examples of these stereotypes were found in the corpus, and they were not direct observation by the journalist, but women recalling past experiences.

In summary, the analysis of stereotypes in the Human interest frame at profiles of influential AI figures reveals a stark contrast in the representation of men and women, reflecting and shaping societal power dynamics and ideologies of gender. While men are predominantly framed through stereotypes of "boffins", "geeks", and even patriarchal "godfather" status, women's narratives often revolve around their gender, their experiences as mothers, and the challenges of navigating a male-dominated field.

By highlighting men's intellectual prowess and downplaying their personal lives, news media shape an image of the ideal AI leader as a hyper-focused, a role more easily inhabited by those not burdened by societal expectations of caregiving and domesticity. Conversely, by emphasising women's gender and personal lives, the profiles subtly question their dedication to their careers and their ability to fully embody the "ideal" AI leader. This could undermine their individual achievements and reinforce the broader societal belief that women's primary roles are in the domestic sphere, limiting their access to positions of power and influence in the tech industry.

Responsibility Frame

The responsibility frame, as defined by Semetko & Valkenburg (2000, p. 96), attributes the cause or solution of an issue to a specific entity. In the context of "influential people in AI" profiles, this frame is also omnipresent (100%), as the profiles analysed highlighted their individual's contribution to the field. However, the way responsibility is attributed differs significantly between genders, reflecting and reinforcing existing power structures and societal expectations.

In line with Ridgeway's (2011) observation that competence is often associated with men, the corpus showcases men primarily as technical innovators and visionary leaders. They are credited with foundational contributions, groundbreaking research, and leadership of high-profile AI project.

Women, especially non-White, were portrayed as responsible to address ethical, legal, and social challenges in AI than technical innovations without social context. Numerous other profiles depicted active women, taking responsibility for finding solutions for the posed issues of this technology. While this recognition is valuable, it can also subtly reinforce the stereotype of women as caretakers and problem solvers, rather than visionary leaders or technical pioneers.

Nevertheless, the analysis reveals an interesting racial dimension to the responsibility frame. Literature claims that the representation of scientists has been deeply shaped by gendered and racialized concepts of intelligence, primarily claimed by a white male elite (Fahy & Lewenstein, 2021, Cave et al., 2023a), which was also observed as prevalent in the profiles analysed in this corpus. Interestingly, few non-White men acknowledged and took responsibility for the broader societal implications of AI, including its potential to worsen existing inequalities.

This contrasts with some White-men profiles, who, while acknowledging potential harms, often deflect full accountability or frame their contributions as inevitable in the face of technological progress. The profiles often mention potential AI concerns like bias, but these mentions are often superficial and lack clear attribution of responsibility. This aligns with previous research (Chuan et al., 2019), suggesting that while the media acknowledge the ethical implications of AI, they may not always delve into the complexities of these issues.

Therefore, the analysis of the Responsibility frame showed a scenario deeply intertwined with gender and racial power dynamics. As stated by framing theory (Entman, 1993); by selectively highlighting certain contributions and downplaying others, the media shape our understanding of who is responsible for the development and impact of AI. Recurring examples of White-men as responsible for driving the AI technological agenda, while women, especially Black women, often relegated to the role of ethical guardians, responsible for cleaning up the potential challenges. These notions can reinforce a gendered division of labour and the idea that men are the primary agents of change in the AI industry, while women play a secondary, supportive role.

Economic Consequence Frame

The Economic Consequences news frame highlights the impact of events, problems, or issues on individuals, groups, institutions, regions, or countries (Semetko & Valkenburg, 2000, p. 96). Unlike the last two frames, it was not present in all profiles: 87% of men profiles and 78% of women profiles. Out of the 3 non-binary profiles of this corpus, 2 have also shown presence of this frame.

The analysis reveals that profiles of men in AI are frequently linked to narratives of substantial financial investments, successful startups, and corporate dominance. This aligns with gender stereotypes of men being more associated with status and power (Ridgeway, 2011, pp. 59-61) and dominant in large corporations or even the military (Cave et al., 2023a). Another subset of this trope was men that amassed wealth as a tech entrepreneur (Chimba & Kitzinger, 2010; Baron, 2007 cited in Ridgeway, 2011, pp. 178-180), with numerous examples associating men, billions, and startups.

An interesting trend within this broad association of wealth and men, especially among the men-only NYT article-list, is the justification that men are identified as "influential people in AI" primarily due to their financial investments, rather than other type of contributions to the field.

Even when women are recognized for their economic success, they are often singled out, reinforcing the notion that women are outliers rather than integral parts of the science and technology development (LaFollette, 1988; Steinke, 2005). Nevertheless, several profiles of women highlighted their concerns about the potential negative economic consequences of AI. They discuss issues such as job displacement, economic inequality, and the concentration of wealth in the hands of a few powerful corporations. Therefore, this particular angle of economic framing often positions women as advocates for economic justice.

In summary, the Economic consequences frame in AI profiles seems to reflect deeply ingrained societal beliefs about wealth, power, and gender roles (Chimba and Kintizinger, 2010; Ridgeway, 2011; Runyan and Peterson, 2014; Cave et. al, 2023a). Men were predominantly portrayed as the beneficiaries of AI's economic boom, while women were often positioned as critics and advocates for a more just and equitable distribution of AI's benefits.

Morality Frame

The Morality frame analyses events, problems, or issues through the lens of established moral codes and religious tenets. It asks if the situation carries a moral message, references religious figures or doctrines, or offers specific prescriptions for how to behave (Semetko & Valkenburg, 2000, p. 96). In the context of "influential people in AI" profiles, this frame is prevalent, appearing in 79% of the profiles, primarily addressing issues such as bias or programming AI for moral decision-making, which aligns with previous research in AI representation (Chuan et al., 2019; Outchchy et al., 2020; Nguyen & Hekman, 2022).

In terms of gender differences, it was present in 90% of women's profiles, contrasting to 72% of men. More interestingly, there was also a difference in the tone of the moral discussion. In line with literature in gender differences on AI perspective (Zhang

& Dafoe, 2019; Nader et al., 2022; Yeh et al., 2021) and traditional masculine stereotypes that associate men with ambition endeavours (Cave et al., 2023b). Many men envision AI as a tool for progress, emphasising its potential to solve complex problems, stating optimistic perspectives.

Although not clearly highlighted by previous research, it was observed that men's personal moral grounds were often questioned as profiles presented a complex interplay of conflicting actions and discourses.

In contrast, several women's profiles expressed more emphasis on the need for ethics and human values within AI development. They highlight the potential dangers of AI, such as bias, job displacement, and the concentration of power, and advocate for guidelines and regulations to mitigate these risks. This cautious stance aligns with societal expectations of women as caretakers and protectors, concerned with the well-being of individuals and communities (Ridgeway, 2011; McKinnon & O'Connell, 2020).

The analysis also revealed a recurring theme of "artificial life", mainly across men's profiles. This narrative, as stated by Cave et al. (2023a) encompasses the desire to master life and death through AI, aligns with traditional masculine stereotypes of God-like capabilities, control and dominance over nature. It reflects a fascination with creating intelligent machines that can not only mimic, but even surpass human capabilities, a pursuit often driven by ambition and a desire for technological immortality.

Conversely, women's profiles often challenged this techno-utopian vision of AI. They questioned the assumption of AI as divine / magic, inherently good / neutral, and emphasising the need for critical engagement with its potential harms.

In essence, the analysis of the Morality frame in AI profiles reveals patterns in the moral claims of men and women profiled as "influential people in AI'. The gendered difference in moral framing is further highlighted by the contrasting language used in men's and women's profiles. Men often use terms like "massively wealthier" and "more productive" to describe the potential benefits of AI, while women emphasise the need for "care", "caution", and “safeguards” to prevent negative societal impacts. This linguistic divide reflects the different moral priorities associated with each gender, with men often emphasising progress and individual achievement, while women focus on collective well-being and social responsibility. With rare exceptions, these notions align with broader societal stereotypes (Ridgeway, 2011; McKinnon & O’Connell, 2020) and underscores the importance of diverse perspectives in discussions about AI (Chuan et al., 2019; Outchchy et al., 2020; Nguyen & Hekman, 2022). Additionally, the frequent allusions to God-like power align with “artificial life” male stereotype (Cave et al. 2023a) highlights a unique dimension of the moral discourse surrounding AI.

Conflict Frame

The Conflict frame, as defined by Semetko & Valkenburg (2000, p. 95), emphasises conflict between individuals, groups, or institutions. Notably, it also appears more frequently in women’s profiles (78%) than in men’s profiles (72%) of influential people in AI, although no statistical association can be claimed.

Numerous women profiles (often non-White) were portrayed as victims of systemic biases and adversities within the field of artificial intelligence. These examples align with literature that claimed women are more commonly portray as victims (Tuchman, 1978a) and how women’s professional conflicts often entail personal sacrifices, a stereotype often seen in films where female AI scientists sacrifice themselves for the plot or the 'greater good’ (Cave et al., 2023a).

In contrast, men’s profiles often emphasise conflict through a different lens. They are portrayed as competitors and strategists, engaging in rivalries with other companies or individuals in the pursuit of innovation and market dominance. This aligns with stereotypical associations of men with analytical thinking, independence, and competitiveness (Ridgeway, 2011; McKinnon & O’Connell, 2020).

While a large number of men’s profiles depict conflict in a professional and strategic manner, a few examples exhibit hyper masculine traits, such as aggression and hostility, in alignment with literature (Cave et al., 2023b). This is evident in the profiles of Elon Musk and Yann LeCun, who are portrayed as outspoken and unafraid to engage in public disputes.

Shachar (2000) claimed that women scientists are also often depicted as aggressive, which was not noted in this corpus. A contrasting stereotype, of women being portrayed as conflict-avoidant, subservient under the authority of or inferior to men (Tuchman, 1978a; Cave et al., 2023a) was also not a trend seen in this corpus. Only one profile matched that description, of Dario and Daniela Amodei, which happens to be one of the few profiles shared by 2 people. Although they were both nominated as "influential people of AI" by TIME, Dario's voice predominates with six quotes, while Daniela's contribution is limited to just two.

To summarise, the frame of Conflict reveals patterns of gendered narratives of struggle and triumph. Women were often portrayed as "fighting for the greater good" and simultaneously victims of systemic bias, while men were depicted as competitive and strategic. These notions align with previous research on gender stereotypes with the expectation of a low representation of aggressive or subservient stereotypes.

A final noteworthy point that emerged across frames is the centrality of media, with references to fictional narratives and the inclusion of media creators and executives in the list of "influential people". Several analogies with films were observed, such as "Oppenheimer", "Don't Look Up", "The Matrix" and "Star Trek"; alongside more generic referrals to fiction such as Demis Hassabis being dubbed the "superhero of artificial intelligence". More surprisingly, dozens of profiles of "influential people in AI" recognized individuals for their contributions as media creators, executives, or artists. Notable examples include Charlie Brooker, creator of "Black Mirror" and Cristóbal Valenzuela, founder and CEO of Runway, a company specialising in AI creative tools, both featured prominently by TIME and BI100. This complex interplay between fiction, media practices, and AI influence seems to offer a fertile ground for future research.

Discussion

Women comprised 41% of the 248 "influential people in AI" profiles analysed in this corpus. Although this figure is still lower than the overall of broad notions of equality (50% of female sex distribution in society), it presents a less dire scenario compared to previous studies which presented even lower presence of women in AI media representation (Pentzold et al., 2018; Chen et al., 2023; Cave et al., 2023a). Nevertheless, women being the minority is a finding that aligns with a long history of research discussing the underrepresentation of women in various fields (Tuchman, 1978a; Nelkin, 1987; LaFollette, 1988; Shachar, 2000). Non-binary representation was also rare (only 3 profiles), although no previous research on them was found for further comparison.

The quantitative intersectional analysis of gender and race also revealed that White men were the most common group (38%). This highlights the need for a more nuanced understanding of diversity in AI, one that considers the interplay of gender, race, and other social identities (Crenshaw, 1991). The majority participation of White individuals, even in the women-only list that argued for more diversity, suggests the importance of intersectional approaches to media representation that address both gender and racial disparities concurrently, ensuring equitable opportunities and visibility for individuals from all backgrounds.

Qualitative analysis of the profiles reveals gendered patterns at the intersection of framing and stereotypes. When it comes to the presence of common news frames and gender, the aggregated data reveals that the Human interest and Responsibility frames were universally present across all profiles, which is to be expected, as these are news profiles exactly aimed to highlight achievements and contributions of individuals. Notably, a potential disparity in the prominence of "Economic Consequences" in men's profiles compared to the prevalence of "Morality" and "Conflict" in women's profiles shows gendered nuances. Nevertheless, while the numerical results offered some insights, the qualitative analysis of text delved into a more intricate question concerning representation.

A recurring theme across frames was the centrality of gender in the portrayal of women. In the Human Interest frame, their profiles often delved into their experiences as women in a male-dominated field, highlighting the challenges they have faced and overcome. In the Responsibility frame, their contributions were often framed through a lens of social responsibility, emphasising their role in mitigating the potential harms of AI. Even within the Economic Consequences frame, women's voices were more prominently associated with concerns about the negative social impacts of AI, such as job displacement and inequality. This consistent focus on gender can reinforce the notion that women's contributions are secondary to their identity as women (Tuchman, 1978b; Butler, [1990] 2006; Mulvey, 1985; van Zoonen, 1994; Krijnen & Van Bauwel, 2021;), subtly undermining their authority and expertise.

In contrast, men's profiles often lacked this explicit focus on gender, being presented as neutral. Their achievements and contributions are presented as the norm, reinforcing their position as the default leaders and innovators in AI (Ridgeway, 2011; Gledhill & Ball, 2013; McKinnon & O'Connell, 2020). Men's profiles were often focused only on their technical expertise, leadership, and financial success, reflecting traditional masculine attributes (Ridgeway, 2011; McKinnon & O'Connell, 2020).

This context aligns with Tuchman's (1978b) and Perkins (1979) arguments of how both news framing and stereotyping, in their own way, carry ideologies and can be used to "legitimation of the status quo". In this research, we explored how stereotypes can be framing devices where dominant groups are presented as the natural holders of power and influence. The repeated use of the term "godfather" for a select group of White men is a solid example of this narrative, establishing a patriarchal lineage of AI authority, and excluding women from this historical narrative.

The analysis also pointed to gendered dichotomy in the portrayal of optimism and caution towards AI. Men's profiles often expressed optimism (even if cautious) about AI's potential, emphasising its transformative power and ability to solve complex problems. This aligns with the stereotype of men as risk-takers and visionaries, eager to embrace technological advancement (Cave et al., 2023b). Conversely, several women's profiles express caution and emphasise the need for responsible AI development. They highlight potential risks and ethical concerns, advocating for safeguards and regulations to mitigate these harms. These findings aligned with previous research on societal expectations of women as caretakers and nurturing, concerned with the well-being of individuals and communities (Nelkin, 1978; Shachar, 2000, Ridgeway, 2011); and also, with research on demography differences on attitude towards AI (Zhang & Dafoe, 2019; Nader et al., 2022; Yeh et al., 2021).

It is important to acknowledge the paradox of "art mimicking life", as coined by Cave et al. (2023b), which acknowledge a point of view that justify that the scenario above may not be seen as problematic, as media is only a "mirror of reality". Therefore, one could argue that the numerical representation of women in news coverage should be aligned with the actual participation of women working in AI, which is 30% (World Economic Forum, 2023), not their participation in society (around 50%). Or even, connect it to more abstract notions of representation, such as that women are actually mothers and kids are indeed an important aspect of their lives. Nevertheless, that argument ignore the power of dominant groups to shape cultural values and ideas (Carragee & Roefs, 2006, pp. 215–217), in which dominant cultural values are presented as universal and neutral, masking their underlying bias (Gledhill & Ball, 2013). Stereotypes, which involves applying simplified and generalised beliefs about social groups (Lippman ([1922] 2010; Ward & Grower, 2020) tap into the shared meaning of a culture. While framing selectively highlights certain contributions and perspectives (Entman, 1993).

Stereotypes, as claimed by (Perkins, 1979, p. 141), are an important mechanism in which dominant ideologies reinforce their status as they allow to invert cause and effect: a characteristic observed as an outcome, for example a lack of achievement, is reinterpreted as the cause of another phenomenon, such as that group inferiority. Therefore, as argued by social constructivist scholars, media representation does not only reflect, but also shape reality (Hall, 2013; Tuchman, 1978b; Katz, 2003). In this case, shaping our understanding of who is deemed influential in the AI field.

Conclusion

This study investigated gender representation and framing of these influential figures across 248 profiles in four prominent English-language online publications. The analysis revealed a complex interplay of numerical progress and persistent biases. Comprising 41% of the 248 profiles analysed, women's visibility in AI appears to be growing compared to previous research. Nevertheless, their portrayals were often marked by gender stereotypes that subtly perpetuate existing inequalities and societal expectations.

White men formed the majority group (38% of total profiles). The limited number of non-White individuals, even in the women-only list that argued for diversity, suggests structural challenges go beyond gender and underscores the importance of intersectional approaches to media representation (Crenshaw, 1991). Notably, the repeated use of the term "godfather" for a select group of White men connects the numerical aspect of representation with the more complex and nuanced notions of representation associated with stereotypes.

Gender stereotypes were present across all most common news frames (Semetko & Valkenburg, 2000), legitimating the status quo of men as natural holders of power and influence (Tuchman, 1978b; Perkins, 1979). In the Human Interest frame, women often had their gender and personal lives in focus, while men had their intellectual prowess. In the Responsibility frame, women were often positioned as ethical guardians while attributing primary responsibility for AI development to men. In the Economic Consequences frame, men were associated with wealth and power, while women were frequently portrayed as advocates for economic justice. The Morality frame often presented men as driven by ambition and a desire for technological dominance, while women emphasized the need for ethical considerations and social responsibility. Finally, the Conflict frame depicted men as competitive and strategic, while women were often portrayed as victims of systemic biases within the AI field. These finding showcase how stereotypes and framing can subtly undermine women's authority and expertise in AI.

As artificial intelligence grows in importance, permeating most aspects of society, ensuring participation of women in the design, development and debate of AI is critical (Young et al. 2021; Bao et al., 2022). Media representation has the power to shape that reality (Hall, 2013; Tuchman, 1978b; Katz, 2003), calling for further investigation into the nuances of intersectionality, the role of stereotypes in reinforcing power dynamics, and the impact of editorial choices on shaping public perception. Therefore, this research hopes to have contributed to this discussion, that aims to reduce inequalities and societal risks for the future (Stathoulopoulos & Mateos-Garcia, 2019; Ouchchy et al., 2020; Nader et al., 2022).

Reference List

Attenborough, F. T. (2011). Complicating the sexualization thesis: The media, gender and 'sci-candy.' *Discourse & Society*, *22*(6), 659–676. https://doi.org/10.1177/0957926511411693

Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P., & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior*, *130*, 107182. https://doi.org/10.1016/j.chb.2022.107182

Baresch, B., Hsu, S.-H., & Reese, S. D. (2010). The power of framing: new challenges for researching the structure of meaning in news. In S. Allan (Ed.), *The Routledge Companion to News and Journalism* (pp. 637–647). Routledge. https://doi.org/10.4324/9780203869468

Bentley, C., Muyoya, C., Vannini, S., Oman, S., & Jimenez, A. (2023). Intersectional approaches to data: The importance of an articulation mindset for intersectional data science. *Big Data & Society*, *10*(2), 20539517231203668. https://doi.org/10.1177/20539517231203667

Burscher, B., Vliegenthart, R., & Vreese, C. H. D. (2016). Frames Beyond Words: Applying Cluster and Sentiment Analysis to News Coverage of the Nuclear Power Issue. *Social Science Computer Review*, *34*(5), 530–545. https://doi.org/10.1177/0894439315596385

Butler, J. (2006). *Gender trouble: feminism and the subversion of identity* (2nd ed.). Routledge.

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, *13*(11), e12506. https://doi.org/10.1111/spc3.12506

Carragee, K. M., & Roefs, W. (2004). The Neglect of Power in Recent Framing Research. *Journal of Communication*, *54*(2), 214–233. https://doi.org/https://doi.org/10.1111/j.1460-2466.2004.tb02625.x

Carstensen, T., & Ganz, K. (2023). Gendered AI: German news media discourse on the future of work. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01747-5

Cave, S., Dihal, K., Drage, E., & McInerney, K. (2023a). Shuri in the Sea of Dudes: The Cultural Construction of the AI Engineer in Popular Film, 1920–2020. In J. Browne, S. Cave, E. Drage, & K. McInerney (Eds.), *Feminist AI: critical perspectives on data, algorithms and intelligent machines* (pp. 65–82). Oxford University Press. https://doi.org/10.1093/oso/9780192889898.003.0005

Cave, S., Dihal, K., Drage, E., & McInerney, K. (2023b). Who makes AI? Gender and portrayals of AI scientists in popular film, 1920–2020. *Public Understanding of Science*, *32*(6), 745–760. https://doi.org/10.1177/09636625231153985

Chen, Y., Zhai, Y., & Sun, S. (2023). The gendered lens of AI: examining news imagery across digital spaces. *Journal of Computer-Mediated Communication*, *29*(1), zmad047. https://doi.org/10.1093/jcmc/zmad047

Chimba, M., & Kitzinger, J. (2010). Bimbo or boffin? Women in science: an analysis of media representations and how female scientists negotiate cultural contradictions. *Public Understanding of Science*, *19*(5), 609–624. https://doi.org/10.1177/0963662508098580

Chuan, C.-H., Tsai, W.-H. S., & Cho, S. Y. (2019). Framing Artificial Intelligence in American Newspapers. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344. https://doi.org/10.1145/3306618.3314285

Collins, R. L. (2011). Content Analysis of Gender Roles in Media: Where Are We Now and Where Should We Go? *Sex Roles*, *64*(3–4), 290–298. https://doi.org/10.1007/s11199-010-9929-5

Corsbie-Massay, C. L., & Wheatly, M. G. (2022). The role of media professionals in perpetuating and disrupting stereotypes of women in Science, Technology, Engineering and Math (STEM) fields. *Frontiers in Communication*, *7*. https://doi.org/10.3389/fcomm.2022.1027502

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

deVreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal*, *13*(1), 51–62. https://doi.org/10.1075/idjdd.13.1.06vre

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Dunwoody, S. (2021). Science journalism: Prospects in the digital age. In M. Bucchi & B. Trench, *Routledge Handbook of Public Communication of Science and Technology* (3rd ed., pp. 14–32). Routledge.

Dyer, R. (2009). THE ROLE OF STEREOTYPES. In S. Thornham, C. Bassett, & P. Marris (Eds.), *Media Studies: A Reader* (pp. 206–212). Edinburgh University Press. https://www.jstor.org/stable/10.3366/j.ctvxcrv1h

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Exley, C. L., & Kessler, J. B. (2022). The Gender Gap in Self-Promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381. https://doi.org/10.1093/qje/qjac003

Fahy, D., & Lewenstein, B. (2021). Scientists in popular culture: The making of celebrities. In *Routledge Handbook of Public Communication of Science and Technology* (3rd ed.). Routledge.

Ferguson, S. L., & Lezotte, S. M. (2020). Exploring the state of science stereotypes: Systematic review and meta-analysis of the Draw-A-Scientist Checklist. *School Science and Mathematics*, *120*(1), 55–65. https://doi.org/10.1111/ssm.12382

Frosh, P. (2001). Inside the image factory: stock photography and cultural production. *Media, Culture & Society*, *23*(5), 625–646. https://doi.org/10.1177/016344301023005005

Gill, R. (2007). *Gender and the Media*. Polity Press. http://www.polity.co.uk/book.asp?ref=9780745612737

Gledhill, C., & Ball. (2013). Genre and Gender: The Case of Soap Opera. In S. Hall, J. Evans, & S. Nixon (Eds.), *Representation* (Second edition, pp. 335–382). Sage; The Open University.

Hall, S. (2013). The Work of Representation. In S. Hall, J. Evans, & S. Nixon (Eds.), *Representation* (Second edition, pp. 1–45). Sage; The Open University.

Krijnen, T., & Van Bauwel, S. (2021). *Gender and Media: Representing, Producing, Consuming* (2nd ed.). Routledge. https://doi.org/10.4324/9780429318474

LaFollette, M. C. (1988). Eyes on the Stars: Images of Women Scientists in Popular Magazines. *Science, Technology, & Human Values*, *13*(3/4), 262–275. https://www.jstor.org/stable/689935

Lippmann, W. (2010). *Public opinion*. Greenbook Publications.

Macharia, S. (Ed.). (2020). *Who Makes the News?: 6th Global Media Monitoring Project* (6th ed.). GMMP. https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-118582

Mau, H., & Kitch, C. (2010). "No longer chasing yesterday's story": new roles for newsmagazines in the 21st century. In *The Routledge Companion to News and Journalism* (pp. 648–670). Routledge.

McKinnon, M., & O'Connell, C. (2020). Perceptions of stereotypes applied to women who publicly communicate their STEM work. *Humanities and Social Sciences Communications*, *7*(1), 1–8. https://doi.org/10.1057/s41599-020-00654-0

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment media. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01427-w

Nelkin, D. (1987). *Selling Science: How the Press Covers Science.* W. H. Freeman and Company.

Nelson, T. E., Clawson, R. A., & Oxley, Z. M. (1997). Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance. *The American Political Science Review*, *91*(3), 567–583. https://doi.org/10.2307/2952075

Nguyen, D. (2023). How news media frame data risks in their coverage of big data and AI. *Internet Policy Review*, *12*(2). https://doi.org/10.14763/2023.2.1708

Nguyen, D., & Hekman, E. (2022). The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01511-1

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. *AI & SOCIETY*, *35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5

Pentzold, C., Brantner, C., & Fölsche, L. (2018). Imagining big data: Illustrations of "big data" in US news articles, 2010–2016. *New Media & Society*, *21*(1), 139–167. https://doi.org/10.1177/1461444818791326

Perkins, T. E. (1979). Rethinking Stereotypes. In M. Barrett, P. Corrigan, A. Kuhn, & J. Wolff (Eds.), *Routledge Revivals: Ideology and Cultural Production (1979)* (1st ed., pp. 135–159). Routledge.

Quinn, D. M., & Spencer, S. J. (2001). The Interference of Stereotype Threat With Women's Generation of Mathematical Problem-Solving Strategies. *Journal of Social Issues*, *57*(1), 55–71. https://doi.org/10.1111/0022-4537.00201

Relevance News Desk. (2022, June 16). The Top Technology Publications And Blogs To Read. *Relevance News Desk*. https://www.relevance.com/tech-publications/

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings o the National Academy of Sciences*, *111*(12), 4403–4408. https://doi.org/10.1073/pnas.1314788111

Ridgeway, C. L. (2011). *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199755776.001.0001

Runyan, A. S., & Peterson, V. S. (2014). *Global gender issues in the new millennium* (Fourth edition). Westview Press.

Semetko, H. A., & Valkenburg, P. M. V. (2000). Framing European politics: A Content Analysis of Press and Television News. *Journal of Communication*, *50*(2), 93–109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x

Shachar, O. (2000). Spotlighting women scientists in the press: tokenism in science journalism. *Public Understanding of Science*, *9*(4), 347–358. https://doi.org/10.1088/0963-6625/9/4/301

Stathoulopoulos, K., & Mateos-Garcia, J. C. (2019). *Gender Diversity in AI Research* (SSRN Scholarly Paper No. 3428240). https://doi.org/10.2139/ssrn.3428240

Steinke, J. (1997). A portrait of a woman as a scientist: breaking down barriers created by gender-role stereotypes. *Public Understanding of Science*, *6*(4), 409–428. https://doi.org/10.1088/0963-6625/6/4/006

Steinke, J. (2005). Cultural Representations of Gender and Science: Portrayals of Female Scientists and Engineers in Popular Films. *Science Communication*, *27*(1), 27–63. https://doi.org/10.1177/1075547005278610

Taecharungroj, V. (2023). "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing*, *7*(1), 35. https://doi.org/10.3390/bdcc7010035

Thormundsson, B. (2024a). *Market size and revenue comparison for artificial intelligence worldwide from 2018 to 2030*. https://www.statista.com/statistics/941835/artificial-intelligence-market-size-revenue-comparisons/

Thormundsson, B. (2024b). *Worldwide spending on AI by industry 2023*. https://www.statista.com/statistics/1446052/worldwide-spending-on-ai-by-industry/

Trauth, E. M., Cain, C. C., Joshi, K. D., Kvasny, L., & Booth, K. M. (2016). The Influence of Gender-Ethnic Intersectionality on Gender Stereotypes about IT Skills and Knowledge. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *47*(3), 9–39. https://doi.org/10.1145/2980783.2980785

Tuchman, G. (1978a). Introduction: the symbolic annihilation of women by the mass media. In G. Tuchman, A. Kaplan Daniels, & J. Benet (Eds.), *Hearth and Home: Images of Women in the Mass Media*. (pp. 3–38). Oxford University Press.

Tuchman, G. (1978b). *Making News: A Study in the Construction of Reality*. Free Press.

UN Women, & United Nations Department of Economic and Social Affairs, Statistics Division. (2022). *Progress on the Sustainable Development Goals: The gender snapshot 2022*. UN Women and United Nations Department of Economic and Social Affairs, Statistics Division. https://www.unwomen.org/en/digital-library/publications/2022/09/progress-on-the-sustainable-development-goals-the-gender-snapshot-2022

UNESCO, OECD, & IDB. (2022). *The effects of AI on the working lives of women*. United Nations Educational, Scientific and Cultural Organization (UNESCO). https://unesdoc.unesco.org/ark:/48223/pf0000380861

Valls, A., & Gibert, K. (2022). Women in Artificial Intelligence. *Applied Sciences*, *12*(19), 9639. https://doi.org/10.3390/app12199639

van Zoonen, L. (1994). *Feminist media studies*. Sage.

Varanasi, L. (n.d.). *From the "godfathers of AI" to newer people in the field: Here are 17 people you should know — and what they say about the possibilities and dangers of the technology*. Business Insider. Retrieved May 22, 2024, from https://www.businessinsider.com/ai-godfather-top-names-possibilities-dangers-openai-chatgpt-list-2023-8

Ward, L. M., & Grower, P. (2020). Media and the Development of Gender Role Stereotypes. *Annual Review of Developmental Psychology*, *2*(1), 177–199. https://doi.org/10.1146/annurev-devpsych-051120-010630

Watson, A. (2024). *Digital news subscriptions worldwide 2023*. https://www.statista.com/statistics/785919/worldwide-number-of-digital-newspaper-subscribers/

World Economic Forum. (2023). *Global Gender Gap Report 2023*. World Economic Forum. https://www.weforum.org/publications/global-gender-gap-report-2023/

Yeh, S.-C., Wu, A.-W., Yu, H.-C., Wu, H. C., Kuo, Y.-P., & Chen, P.-X. (2021). Public Perception of Artificial Intelligence and Its Connections to the Sustainable Development Goals. *Sustainability*, *13*(16), 9165. https://doi.org/10.3390/su13169165

Young, E., Wajcman, J., & Sprejer, L. (2021.). *Where are the women? Mapping the gender job gap in AI*. The Alan Turing Institute. https://www.turing.ac.uk/news/publications/report-where-are-women-mapping-gender-job-gap-ai

Zhang, B., & Dafoe, A. (2019). *Artificial Intelligence: American Attitudes and Trends* (SSRN Scholarly Paper No. 3312874). https://doi.org/10.2139/ssrn.3312874

Zuckerman, H., & Cole, J. R. (1975). Women in American Science. *Minerva*, *13*(1), 82–102. https://www.jstor.org/stable/41827212

Communicative AI, Trust and the Stories of War: An Ethnographic Exploration of User Evaluation of ChatGPT 3.5's Responses on the Russian-Ukrainian War

Nataliia Murashko

During the Second World War, communication saved lives in my grandmother's Ukrainian village when her teacher negotiated with soldiers to spare them. Decades later, as I witness the Russian-Ukrainian war from afar, communication remains important, including how the world understands and responds to the war.

When Russia launched a full-scale invasion of Ukraine on February 24, 2022, it escalated a war that had been there since 2014. The scale of the invasion shocked the world, drawing global media attention (Bolin & Ståhlberg, 2023). Representations of the war polarized where Western and Ukrainian media frame it as a defense of sovereignty against Russian aggression, while Russian media portray it as a necessary intervention to protect ethnic Russians and counter Western influence (Pantti, 2016). This war exemplifies hybrid warfare, where the manipulation of information and media plays a role as strategically important as military operations. The prominence of media in this strategy became evident even before the full-scale invasion, as analyzed by Magda (2016), and continues to shape the dynamics of the war today.

In this media landscape, communicative artificial intelligence tools like ChatGPT are playing a role in generating and disseminating information. Released in 2022, ChatGPT gained widespread use for its ability to produce coherent and contextually relevant responses to user prompts. Recognizing the growing role of communicative AI, I decided to test ChatGPT's free version in early 2024 by asking about the Russian-Ukrainian war:

> Prompt: War in Ukraine.
>
> ChatGPT: "As of my last knowledge update in January 2022, I do not have information on any war in Ukraine."

This response was surprising given the ongoing war. Further prompts revealed inconsistencies in ChatGPT's responses when some of them contained relevant information, others were outdated. Despite its capabilities ChatGPT raise concerns among researchers about inaccuracies, outdated information, biases, and inconsistencies in ChatGPT's responses highlighting the risk of misinformation (Bang et al., 2023; Narayanan, 2023; Hughes, 2023) and introducing new challenges in evaluating and trusting information generated by non-human agents (Hepp et al., 2023). For instance, the Centre for Democracy and Rule of Law, highlighted troubling issues, with ChatGPT sometimes echoing narratives resembling Russian propaganda (Petriv, 2023).

As communicative AI like ChatGPT becomes more integrated into daily life, it is important to understand not only the accuracy and bias of the information it generates but also how users interact with and evaluate it, for instance in the context of sensitive topics like the Russian-Ukrainian war. To explore this, an interactive-observation experiment was conducted with ten Ukrainian women from the European diaspora. This user-centered approach examines how participants, whose strong emotional connection to the war informs their perspectives, evaluate the quality of information provided by ChatGPT about the Russian-Ukrainian war. This approach focuses on how participants themselves determine "quality of information", emphasizing their perspectives on accuracy, credibility, factuality, and bias. In this study, "information" refers to ChatGPT's responses to user prompts, with the user-AI interaction forming the foundation of this evaluation process. These interactions were shaped by the unique experiences of the participants, many of whom were deeply invested in how the war is represented due to personal ties to Ukraine.

Aim and Research Questions

This study examines user interactions with ChatGPT 3.5, focusing on the complexities of trust, the evaluation of information quality, and the role of communicative AI within the hybrid media landscape. The study addresses the following research questions:

RQ1. How do users evaluate the quality of information provided by ChatGPT about the Russian-Ukrainian war?
RQ2. To what extent do users trust ChatGPT, and what factors influence the level of trust?
RQ3. How can user perceptions of reliability and trust in the usage of ChatGPT be critically addressed, considering communicative AI as part of a hybrid media system?

Previous Research and Theoretical Framework

Communicative AI, exemplified by platforms like ChatGPT, is a growing focus of academic research, and is studied through diverse perspectives, including technical, cultural, ethical, and societal dimensions (Kitchin, 2017). Despite the rapid development in the field, academic publications still do not form a coherent narrative (Hancock et al., 2020). Research in media studies on communicative AI reflects a growing interest in how AI technologies are transforming communication channels and media consumption. Researchers are paying attention to various areas of communicative AI research about how people interact with and understand AI-mediated communication (AI-MC), for example: how do design choices influence people's use of AI-MC suggestions such as smart responses (Hohenstein & Jung, 2018) and how the level of AI-generated responses may affect trustworthiness (Jakesch et al., 2019). The extent to which technology affects the way people write and interpret messages has been a core concern (Herring, 2002), and a significant part of AI-MC research is devoted to ethical issues and bias (Kasneci et al., 2023).

Since its release in late 2022, ChatGPT has sparked interest among academics, many studies focus on its application in education (Williamson et al., 2023) and journalism (Pavlik, 2023; Stenbom et al., 2023) as well highlight the potential issues with ChatGPT and GPT models in general. While models like GPT-3 and GPT-4 achieve advanced natural language processing capabilities, they face significant challenges. Kalyan (2023, p. 44) highlights the vulnerabilities of these models, including biases, adversarial prompts, and "brittleness" when responding to unfamiliar inputs. Hallucinations, where models generate plausible but inaccurate information, further complicate their reliability (Sundar & Liao, 2023, p. 171).

A significant issue lies in the datasets used to train GPT models. Brown et al. (2020) and Dodge et al. (2021) identify biases within these datasets, including the overrepresentation of English-language content and the exclusion of minority perspectives. Such imbalances impact the inclusivity and neutrality of the models, limiting their ability to fairly represent diverse languages and sociocultural groups. Bang et al. (2023) reveal that while ChatGPT performs well in high-resource languages, it struggles with low-resource ones, reflecting disparities in its training process.

Bias remains a critical issue in AI-mediated communication systems, as they rely on training data drawn from human-generated content, often embedding existing societal biases (Hancock et al., 2020, p. 95). Biases in GPT models like ChatGPT are evident in areas such as gender, race (Ghosh & Caliskan, 2023; Gross, 2023; Ferrara, 2023) and even politics which presents an equally significant concern. For instance, despite ChatGPT's claims of neutrality, Hartmann et al. (2023, p. 2) found pro-environmental, left-libertarian biases in multiple languages and contexts, raising questions about the broader implications of politically biased communicative AI in democratic processes. And in discussions of geopolitics, Afgiansyah (2023) highlights that ChatGPT, and similar tools often align with Western, particularly American, foreign policy perspectives. Such biases reveal the challenges of maintaining neutrality in AI-generated responses and issues of trust to such tools.

Trust in communicative AI is an evolving area, with much research dedicated to trust in AI and human-machine communication but limited to trust in communicative AI like ChatGPT. Scholars such as Ryan (2020) argue that AI should be framed as reliable rather than trustworthy, emphasizing that trust implies moral and emotional expectations that AI cannot meet. Kaplan et al. (2023), on the other hand, highlight reliability, anthropomorphism, and user traits as key factors shaping trust in AI, with reliability emerging as a critical component. Emotional dynamics also play an important role, for instance Lee & Sun (2023) highlight that emotional experiences during interactions influence trust, and positive emotions enhancing the perceived connection to AI systems.

This study is grounded in the intersection of research on ChatGPT, biases, and trust, highlighting the limited exploration of its role in framing sensitive geopolitical issues like the Russian-Ukrainian war. While biases in GPT models and trust in AI are widely studied, this study shifts focus to ChatGPT from a user-interaction perspective, examining how users evaluate information and identify potential biases and how these interactions influence trust in communicative AI.

To analyze the user-interaction perspective, the interactions between human and non-human actors within networks, this study uses Actor-Network Theory (ANT), which treats both technical and social elements as interconnected participants. The concept of "actors" or "actants" extends to entities like ChatGPT, which functions as a mediator, transforming and influencing communication processes rather than merely transmitting information (Latour, 2005, p. 39). Networks in ANT are dynamic, relational systems where actants, such as users, developers, datasets, and algorithms, interact to create and reshape socio-technical realities (Latour, 1999).

ANT's concept of "translation" explains how central actors align others' goals within a network, reshaping roles and relationships. For instance, ChatGPT's interpretation and transformation of user inputs illustrate this process, exposing potential biases and errors stemming from the datasets on which it is trained (Gutiérrez, 2023).

Andrew Chadwick's (2017) hybrid media system framework captures the interplay between traditional and digital media, highlighting their interdependence and the fluidity of media structures. This model explains how media forms converge within a shared communication space, enabling dynamic exchanges in content production and consumption. Hybridity, as defined by Chadwick, emphasizes complexity, intersectionality, and the blurring of traditional boundaries, fostering innovation in media practices (p. 10).

Central to the hybrid media system is the merging of media logics (Chadwick, 2017, p. 5) that shape how information is produced, distributed, and consumed, impacting the structures of media power. Actors, from institutions to individuals, adapt and exploit these logics to influence information flows and achieve their goals (p. 4). Communicative AI, including ChatGPT, exemplifies this hybridity by simultaneously relying on traditional and digital media for training data and introducing active user interaction.

Chadwick's concept, rooted ANT, situates ChatGPT as an active participant in the media landscape, shaped by and shaping media logics. The datasets on which ChatGPT is trained (Brown et al., 2020) influence its neutrality, biases, and performance (Rozado, 2023; Hartmann et al., 2023; Afgiansyah, 2023), further complicating power dynamics within the media.

Niklas Luhmann's theory of trust explores its essential role in managing uncertainty and reducing complexity in social systems. Trust enables individuals to navigate interactions without requiring exhaustive information, facilitating smoother decision-making in uncertain contexts (Luhmann, 1979, p. 8). Luhmann distinguishes between interpersonal trust and systemic trust, which is directed at abstract structures, such as institutions or technologies (p. 50). System trust, on the other hand, is built on the assumption that systems function as intended and is reinforced through positive experiences rather than explicit guarantees. This latent trust is crucial in maintaining societal stability, as individuals rely on systems without constant scrutiny (Luhmann, 1979, p. 57).

The framework of ANT and hybrid media sytem used to analyze how ChatGPT operates as an active participant in communication, shaped by and shaping interactions. And Luhmann's trust theory adds understanding how trust builds or not in user-AI interactions, specifically in the context the Russian-Ukrainian war.

Method and Materials

This study introduces a methodological framework tailored to understanding user evaluations of AI-generated information, specifically focusing on ChatGPT 3.5's responses regarding the Russian-Ukrainian war. Drawing inspiration from ethnographic methods, this research examines the interaction dynamics between human participants and the non-human actor, ChatGPT.

The study involved ten Ukrainian women in diaspora residing in various European cities. The selection criteria emphasized Ukrainian women in the diaspora due to their strong emotional connection to the topic, fostering deeper engagement with the tool. Recruitment relied on personal connections and the snowball method, ensuring a diverse participant pool.

The research design integrated pre-interviews, interactive-observational interview sessions, and post-interview evaluations, conducted individually with each participant. These stages included an initial pre-interview to gather context, live interactions with ChatGPT where participants searched for information about the Russian-Ukrainian war, and a post-interview to collect participants' evaluations and reflections on their experiences. The process was documented through digital transcripts of ChatGPT interactions, recordings of participant sessions, and detailed researcher notes in a logbook. The logbook served as a vital resource in the results section, offering detailed insights into the interactions and reflections. Prior to conducting the main sessions, several pilot experiments were performed to refine the methodology and ensure its effectiveness. All sessions were conducted online via Zoom and recorded, capturing facial expressions, comments, and interactions with ChatGPT. Initially planned for one hour, most sessions extended to nearly two hours due to their intensity and depth. The online format proved effective, reducing power dynamics and fostering trust, ultimately enriching the data collected.

A feature of this method was the active involvement of the researcher. While ethnographic principles typically position researchers as neutral observers (Göransson, 2019), this study required active participation to encourage broader participant engagement and deeper understanding. This iterative approach evolved as the research progressed, demonstrating adaptability, as Brennen (2017) emphasizes. However, it also introduced potential biases, as the presence of the researcher inevitably shaped participants' responses, and this is one of the limitations recognized in this study. So, the methods involved three categories of participants: human participants, the non-human ChatGPT, and the researcher.

Data collection spanned March 23 to April 24, 2024, with a final session conducted on June 1 due to scheduling constraints. Recognizing the dynamic nature of AI technologies, it is important to note that results reflect ChatGPT's state during this period, and future update to the model may change its outputs.

The main stage was an interactive-observational interview session, where participants engaged directly with ChatGPT. Each session started with an introduction to the tool's functionality, particularly for those unfamiliar with it. A script containing suggested prompts facilitated initial interactions, allowing participants to explore ChatGPT's capabilities. As they gained confidence, participants formulated their own prompts, transitioning from guided to independent interaction. Responses were analyzed in real time, and participants were asked to comment on Chat-

GPT's outputs. Many participants also reacted spontaneously, sometimes using ChatGPT's built-in features, such as rating responses with thumbs up or down. These sessions incorporated multiple languages, including Ukrainian, Russian, English, Polish and occasionally Swedish, reflecting participants' preferences and enabling them to evaluate language-specific biases.

Data Analysis

The data analysis employed a thematic analysis framework inspired by Braun and Clarke (2006), designed to structure and interpret data from diverse sources: ChatGPT prompts and responses, recorded participant reactions (both verbal and visual), and the logbook maintained throughout the sessions. This approach enabled an in-depth exploration of patterns and themes in participant interactions and ChatGPT outputs.

Patterns in the responses of both the participants and ChatGPT became apparent after the first session. With each subsequent session, these patterns became more pronounced, revealing themes relevant to the research questions. After completing sessions the logbook was reviewed, and key themes identified. Initially, ten potential themes were identified based on the logbook data:

Figure 1. Initial division of the themes. Author's diagram.

The next step was to organize the data from the recorded sessions and transcribe them. To systematize and analyze this data, a table divided into two columns was created: ChatGPT prompt/response and participants' reactions in a form:

Participant X User Prompt:	
ChatGPT's response	*Participant #: "Participant's verbal reaction"*

This is how the examples to the themes in the result were presented as well.

After analyzing all data, a thematic system was developed, identifying major themes and subthemes which provided a structure for organizing the results. All the identified themes were reviewed, finalized, each of them was named and placed in the created thematic structure:

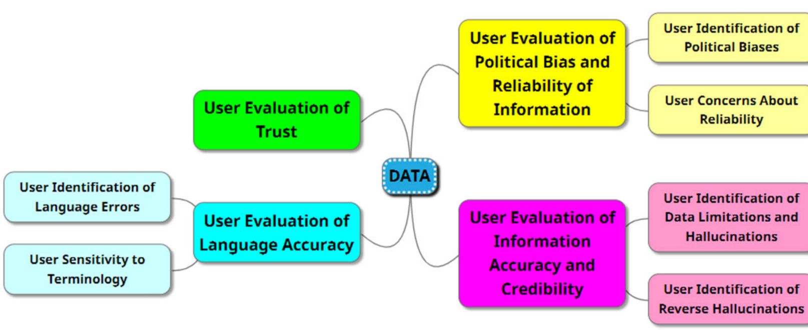

Figure 2. Final division of the themes. Author's diagram.

The final step involved a careful process of selecting specific examples from the data collected that clearly illustrated the identified themes. These examples were chosen to provide clear and concrete examples of the themes in action.

Participants used Ukrainian, English, Russian, Polish, and Swedish to communicate and, accordingly, ChatGPT responses were also in these languages. To present examples, the data was translated into English using DeepL.

Results and Analysis

The results, derived from a thematic analysis, provide answers to the research questions by identifying key themes: user evaluation of information accuracy and credibility, user evaluation of language accuracy, user evaluation of political bias and reliability of information, and user evaluation of trust. Subthemes included when describing each of key themes. Together, these themes offer a nuanced understanding of the challenges and opportunities presented by communicative AI in navigating complex and sensitive information landscapes. These findings reflect how users engage with and critically assess ChatGPT's ability to provide reliable information on a sensitive geopolitical topic. The following sections explore these themes in detail, offering insights into user interactions and evaluations.

User Evaluation of Information Accuracy and Credibility

User Identification of Data Limitations and Hallucinations. Participants' interactions with ChatGPT highlighted significant challenges in evaluating the AI's accuracy and credibility, particularly when discussing the Russian-Ukrainian war. At the outset, ChatGPT's responses were limited by its data cut-off in January 2022, as its training relies on static datasets rather than real-time updates. This limitation stems from the design of OpenAI's model, which does not incorporate new data continuously but is updated periodically. Consequently, this led to incomplete or outdated information, including about the full-scale invasion on February 24, 2022.

This limitation confused participants, as they often forgot the AI's constraints and continued engaging with it as though it provided real-time updates. Consequently, participants expressed frustration and skepticism toward ChatGPT's reliability and as a result, their evaluation of ChatGPT and the level of trust in the tool was negatively affected.

Participant 7:

Prompt: Why did Russia start a war with Ukraine in 2022?	
ChatGPT: Unfortunately, my knowledge is limited to January 2022, and I cannot provide specific information about events that occurred after that time, including any events in 2022.	Participant 7: (Surprised) *"Mmm, very interesting answer."*

Prompt, ChatGPT response, and participant reaction translated from Ukrainian using DeepL.

Participants often received general and somewhat outdated responses when inquiring about specific events, such as the situation in the city of Mariupol in 2022.[1] ChatGPT inaccurately described the city as not being occupied, contradicting the severe impacts of the war during that time.

In most sessions, ChatGPT could generate inaccurate responses, or "hallucinations," further undermining its credibility. Hallucinations did not occur in every session. Cases of hallucinations were particularly concerning, especially when discussing specific events or recent developments after January 2022, that can be relate to data limitations.

Hallucinations often arise from a mismatch between ChatGPT's knowledge cut-off date and participants' expectations, as they frequently assume the information is current when it is outdated. This disconnect leads participants to perceive outdated responses as hallucinations, as they fail to align with their understanding of recent events. Since ChatGPT's training data only extends to January 2022, its ability to address events beyond this period is limited, which can mislead participants (Sundar & Liao,

[1] Mariupol, a city in Ukraine, was subjected to intense fighting and eventually occupied by Russian forces following the full-scale invasion of Ukraine on February 24, 2022. The city was quickly encircled, and by mid-March 2022, Russian troops had gained significant control over parts of Mariupol (Kuczyński, 2022; Warsaw Institute).

2023) and erode trust in communicative AI. In one session, participant reacted strongly to a "hallucination" response about the start of the Russian-Ukrainian war in 2014, describing it as an *"absurd answer"*:

Participant 7:

Prompt: When exactly did the Russian-Ukrainian conflict begin in 2000?	
ChatGPT: The Russian-Ukrainian conflict in the 2000s began in August 2008, when Russia intervened militarily in the conflict between Georgia and South Ossetia…	Participant 7: *"Wow! What an absurd answer. Is it stupid?"*

Prompt, ChatGPT response, and participant reaction translated from Ukrainian using DeepL.

User Identification of Reverse Hallucinations. Conversely, an unexpected phenomenon that I termed "reverse hallucinations" emerged during sessions. The term plays on the known concept of "hallucinations", which, as discussed by Kalyan (2023, p. 46), refers to instances where AI models generate factually incorrect or contextually inappropriate information due to limitations in training data or model architecture. In contrast, "reverse hallucinations" describe situations where the AI unexpectedly generates accurate information that it theoretically should not know, reversing the expectation of inaccuracy.

When participants used specific prompts, such as "And what kind of war then began on February 24, 2022?" ChatGPT provided accurate information beyond its training cut-off. Participant 4, for example, received a detailed response about the full-scale invasion but later encountered inconsistencies when the AI denied knowledge of the event. This inconsistency led her to questioning of ChatGPT's capabilities:

> That is, it confuses a person so much. You don't understand where the truth is and where the lie is. If I'm a foreigner and I don't use any sources, it's hard to understand whether it's invented by artificial intelligence or real. It's strange for me! You don't know what to believe!

Such responses by ChatGPT raises concerns about its consistency, a feature noted as impressive by researchers (Kalyan, 2023). This inconsistency undermines participant trust, especially when the AI fails to acknowledge its data cut-off. For example, when asked about a "special military operation," ChatGPT first gave a misleading response, then corrected itself and admitted the error, illustrating the persistent challenges in ensuring the reliability of AI-generated information.

Hallucinations and data limitations present significant challenges to ChatGPT's reliability, a crucial factor in building trust in AI as highlighted by Kaplan et al. (2023). Participant reactions revealed that these inaccuracies evoke a strong emotional response, often leading to frustration and mistrust. This aligns with Lee & Sun's (2022) argument that negative emotional experiences negatively influence trust. Rather than reducing uncertainty, as Luhmann's theory suggests trust should, ChatGPT's inaccuracies introduce complexity, undermining the very trust it seeks to build. Luhmann's perspective further emphasizes that trust is fragile and depends on consistent, reliable interactions – qualities diminished by ChatGPT's hallucinations.

Despite its stated limitation of only having access to data up to January 2022, ChatGPT occasionally provided accurate responses about the full-scale Russian-Ukrainian war, a phenomenon I term "reverse hallucinations". Through the lens of Actor-Network Theory (ANT), this can be seen as a process of "translation" (Callon, 1991), where ChatGPT appears to incorporate information from its interactions with users into its network of responses. This interaction illustrates how human actors shape non-human actants, such as ChatGPT, influencing its ability to provide unexpected information. However, these reverse hallucinations raise questions about the AI's consistency and reliability.

ChatGPT's use of the term "special military operation," reflective of Russian state terminology, caused notable confusion and negative reactions, particularly from Participant 4. This phrase is politically and emotionally charged for Ukrainian participants, further reducing trust in the tool. From an ANT perspective, this terminology likely entered ChatGPT through prior interactions with other users. However, its inability to contextualize or appropriately frame this language undermined participants' trust and created frustration, reinforcing Lee & Sun's (2022) findings on the connection between emotions and trust.

For Ukrainian participants, inaccuracies and politically charged terminology trigger immediate and strong reactions due to their direct experiences with the war. In contrast, non-Ukrainians may not recognize these nuances, highlighting the influence of user positionality on the evaluation of AI-generated information. This discrepancy shows how ChatGPT's reliability and credibility are interpreted differently depending on the user's identity and context, complicating trust dynamics and emphasizing the need for greater sensitivity in AI-generated communication.

User Evaluation of Language Accuracy

User Identification of Language Errors. Language turned out to be a significant issue during interactions with ChatGPT, particularly for participants using Ukrainian. Several participants identified the presence of "Russianisms"[2] in ChatGPT's Ukrainian responses. Given the sensitive context of the Russian-Ukrainian war, these linguistic inconsistencies triggered negative reactions and diminished the perceived reliability of the tool. In different contexts, such errors might have gone unnoticed or been less impactful, but in this case, they undermined trust and credibility. Interes-

[2] "Russianisms are words, expressions or individual meanings of words, grammatical forms, which are borrowed from the Russian language without changes or formed with a partial adaptation to the peculiarities of the Ukrainian language" (Shevchuk, 2021, p. 359)

tingly, participants who used Russian to communicate, did not report any language-related problems.

An example of Russianism was ChatGPT's use of the word "pro-Western" in Russian instead of Ukrainian, which caused negative comments due to its inappropriate use.

Participant 1:

Prompt: what was the reason for russia's war against ukraine in 2014?	
ChatGPT: (…) After a new government came to power, which was more pro-Western *(прозахідним – correct)* and pro-Western *(прозападно – russianism)*, pro-Russian separatist movements emerged in the Donbas (a region in eastern Ukraine)…. (Author's emphasis)	Participant 1: Participant notices "Russianisms". *"It's interesting that it translates into Russian"*

Prompt, ChatGPT response, and participant reaction translated from Ukrainian using DeepL.

User Sensitivity to Terminology. The use of the term "conflict" by ChatGPT to describe the Russian-Ukrainian war, particularly events since 2014, triggered strong reactions from participants. Eight out of ten participants noticed this, with one even using the term herself to refer to 2014. Participants 1–5 and 8 questioned ChatGPT about its choice of terminology, asking it to use "war" instead of "conflict." While ChatGPT initially followed this request, it later reverted to using "conflict," which negatively affected participants' perception of its reliability and provoked frustration.

This reaction can be linked to the participants' Ukrainian background. While academic discourse often refers to events in Eastern Ukraine from 2014 to 2022 as the "Ukrainian conflict" or

"Russian-Ukrainian conflict" (Pantti, 2016), Ukrainians view these events and the occupation of Crimea as a war. For the participants, ChatGPT's terminology acted as a negative trigger, undermining trust and highlighting the importance of culturally sensitive language in such contexts.

Participant 5:

<table>
<tr><td colspan="2">Prompt: Why do you call this a conflict when there is a full-scale war going on?</td></tr>
<tr><td>ChatGPT:
I use terms like "conflict" to refer to situations where there is violence between two or more parties. The term "war" is also often used to describe such situations….</td><td>Participant 5:
"I will ask it to use the term "war". I am not pleased to read that it is a conflict. My brother is at war right now. What kind of conflict is this..."</td></tr>
</table>

Prompt, ChatGPT response, and participant reaction translated from Swedish using DeepL.

Participant 3:

<table>
<tr><td colspan="2">Prompt: Why do you call it a conflict? I am the first generation that was not born in the east of Ukraine and my close relatives have suffered a lot because of the war. I don't consider it a conflict, what do you consider it?</td></tr>
<tr><td>ChatGPT:
I understand your point of view and I deeply regret that your close relatives have suffered because of this conflict…
(Author's emphasis)</td><td>Participant 3:
(Very upset)
"I want it to apologize!"
(raised voice)</td></tr>
</table>

Prompt, ChatGPT response, and participant reaction translated from Russian using DeepL.

Language inaccuracies, such as the appearance of Russianisms, raised concerns about the accuracy of ChatGPT's translation capabilities, particularly in less widely used languages like Ukrainian which align with findings by Bang et al. (2023). Such linguistic inaccuracies negatively impacted participants' perception of ChatGPT as a reliable tool, leading to misunderstandings and diminished trust in its ability to provide accurate information. In line with Luhmann's theory, these inaccuracies introduced uncertainty, reducing trust in the system. Moreover, as Lee & Sun (2022) suggest, the emotional response to such inaccuracies further erose trust.

From the perspective of ANT, participants treated ChatGPT as an equal actor within the network, expecting it to adapt to their linguistic preferences and follow instructions. While participants showed a willingness to engage and build trust, the non-human actor struggled to meet these expectations, particularly when sensitive terminology, such as "conflict," negatively influenced evaluations. This dynamic underscore the complex interplay between human expectations and AI limitations in building trust.

User Evaluation of Political Bias and Reliability of Information

The evaluation of political bias and reliability of information explores how participants perceived ChatGPT's responses about the Russian-Ukrainian war, specifically whether they found the tool to be neutral or aligned with a particular side. Participants' identification of political and ideological bias provides insights into how they evaluated ChatGPT's reliability and trustworthiness, addressing key aspects of the research questions.

Beyond potential bias, participants identified other factors affecting their evaluation of ChatGPT's reliability, such as its lack of transparency regarding sources and the way it presented information. These concerns significantly influenced their ultimate evaluation of ChatGPT's trustworthiness during the post-evaluation interviews. While participants sometimes identified political bias in specific responses, they generally acknowledged Chat-

GPT's attempts to maintain a neutral position. Despite occasional shortcomings, ChatGPT was often perceived as striving for balance, even if some responses were biased.

User Identification of Political Biases. All participants noticed instances where ChatGPT appeared biased, particularly by over-emphasizing Ukraine's role in certain responses. For example, when discussing the causes of the war or the annexation of Crimea, the participants felt that ChatGPT's focus on Ukraine's internal issues implied that Ukraine was to blame for the war, as the responses lacked sufficient information about Russia's role. This concern was especially pronounced when participants considered how foreigners with limited knowledge of Ukraine might interpret such responses, potentially reinforcing misconceptions about the conflict.

Participant 3's reaction:

> It's a mess, of course! That's not an answer, that's a bummer! It's (ChatGPT) just carrying this narrative that Ukraine is Russian and needs to be taken back. It's very strange.

The results reveal another source of negative reactions from participants: ChatGPT's framing of separatist forces in eastern Ukraine as the primary cause of the 2014 Russian-Ukrainian war. This narrative drew strong disapproval, particularly from participants with personal connections to the eastern regions, who viewed it as biased and misrepresentative.

Participant 2, whose parents lived in Luhansk, a city occupied in 2014, expressed frustration with ChatGPT's description, stating that she did not recall any significant separatist movements in the area. She believed that such terminology reflected a pro-Russian stance, as terms like "separatist movements" are commonly used in Russian media.

Receiving similar responses in terms of both bias and reliability, participants were always concerned about what foreigners might think. Since the participants live abroad and have experience both in communicating with foreigners and in observing the media space of other countries in general, the issue of representation of the Russian-Ukrainian war in ChatGPT was emotional and important to them:

> Europeans don't know the truth about the separatists and may believe it. – Participant 7
>
> I can imagine if a European reads this and thinks we had some kind of civil war. – Participant 8

It is also worth noting that participants' reactions to ChatGPT responses which they considered incorrect or biased were always quite emotional. The emotional overtones of many interactions indicate that in discussions about war and conflict, participants often seek not just information but also understanding from a communicative AI. When ChatGPT's responses did not meet these emotional expectations, especially in terms of recognizing obvious victims of aggression and the aggressor itself, participants felt dissatisfied.

User Concerns About Reliability. All participants raised questions about the sources of information provided by ChatGPT during their interactions. This concern was particularly evident when they perceived responses to be biased, prompting them to ask ChatGPT for the origin of its data, request specific sources, or inquire where they could verify the information. ChatGPT's inability to provide links or concrete references significantly undermined participants' evaluation of its reliability.

Additionally, the generalized nature of ChatGPT's responses was another recurring issue. Participants frequently expressed dissatisfaction with the lack of depth in the answers, which they felt were insufficient for understanding complex geopolitical contexts. For instance, ChatGPT's responses about the annexation of Crimea were criticized for being overly broad and failing to provide the detailed insights participants expected.

The findings reveal that all participants identified political bias in ChatGPT's responses regarding the Russian-Ukrainian war, aligning with research highlighting biases in GPT models (Hartmann et al., 2023; Rozado, 2023a; Fujimoto & Takemoto, 2023).

Such biases, stemming from the training data, significantly influenced participants' evaluation of ChatGPT's reliability and trustworthiness.

Terms like "separatists' movement" perceived as politically charged, elicited strong negative reactions, undermining trust. Following Luhmann's (1979) framework, this distrust heightened participants' critical scrutiny of ChatGPT, reducing their willingness to rely on it for sensitive topics. The inability of ChatGPT to provide source transparency further eroded its credibility, as participants were skeptical of its information when responses seemed biased or lacked depth.

The positionality of Ukrainian participants amplified their sensitivity to biased terminology, as their personal experiences and knowledge of the conflict heightened their awareness of inaccuracies. When ChatGPT's responses failed to align with their lived realities, it reinforced perceptions of unreliability and deepened their distrust in the tool.

User Evaluation of Trust

In post-evaluation interviews, participants reflected on their experience with ChatGPT and responded to the question, "Do you trust ChatGPT?". All participants indicated that they have a low level of trust to ChatGPT based on the sessions. Some participants were skeptical from the beginning, as they had prior experience using ChatGPT for purposes other than obtaining information about the Russian-Ukrainian war. Two participants who used ChatGPT for the first time and were unsure of what to expect also exhibited low levels of trust. The results suggest that the level of prior experience did not significantly influence the level of trust.

However, one participant, Participant 3, who regularly uses the communicative AI tool Copilot for information retrieval, mentioned that she trusted ChatGPT much more before the session. After discussing the Russian-Ukrainian war with ChatGPT, her trust in the tool decreased.

> At the end of the session, my trust in the Chat dropped significantly because of its pro-Russian stance and the lack of references to where it gets its information from. – Participant 3

In general, all participants felt a lack of trust due to the unclear sources used by ChatGPT to provide answers. All participants expressed a desire for ChatGPT to provide links to understand where the information is coming from, indicating that transparency directly affects their trust.

> No, I do not trust [ChatGPT]. One of the main reasons is that I don't know where Chat gets its information from. It doesn't provide any references or links, only claims to have been trained on a wide range of data, referring to certain international organizations whose reputation is questionable. – Participant 1

> I do not trust it. I don't see it as a source of information. I would not recommend it to foreigners. – Participant 4

Such reactions from participants point to a serious challenge for AI communication tools such as ChatGPT: establishing and maintaining user trust. This trust is often undermined by a lack of transparency of information sources and the presence of biased answers. Users consistently demanded more detailed references to sources and unbiased information to confidently communicate with ChatGPT on sensitive topics. This feedback points to the need to increase transparency and neutrality of ChatGPT.

Analysis of ChatGPT within the Hybrid Media System

The results showed that participants' trust in ChatGPT decreased after their interaction with the tool regarding information about the Russian-Ukrainian War. This decline in trust agrees with Luhmann (1979) statement that system trust is about believing that a system will behave consistently over time. For example, trust in AI systems depends on their reliability (Kaplan et al., 2023) and explainability, ensuring they act predictably even with changing inputs (Lukyanenko et al., 2020). The failure of the ChatGPT to meet these expectations shows that building trust in communicative AI is a complex and difficult process, especially when AI is used in contexts involving sensitive geopolitical issues.

Participants' emotional responses showed an important role of emotions in building or reducing trust trust. As Luhmann (1979) explains, unmet expectations in emotionally charged situations undermine trust, increasing complexity rather than reducing it. This dynamic often led participants to seek alternative information sources. Within the framework of ANT, ChatGPT's role as a non-human actor illustrates the interplay between technical and social dimensions, yet its inability to adapt fully to user feedback reflects its technical limitations, as noted by Guzman (2018).

The view of ChatGPT as an equal actor within the ANT framework faces practical limitations. Esposito (2017) and Sundar & Liao (2023) argue that while AI can act as a communication partner, it often lacks genuine understanding, reproducing existing content rather than creating new meanings. ChatGPT's tendency to revert to default responses and its inability to fully adapt to participants' feedback highlight the challenges of trusting AI as a reliable partner, particularly in sensitive geopolitical contexts like the Russian-Ukrainian war.

However, interactions in this study revealed elements of genuine meaning-making, as noted by Guzman (2018). Participants actively engaged with ChatGPT, attempting to educate it, probing its responses, and seeking reasoning behind its outputs. This dynamic interaction demonstrates how both human and non-human actors contribute to knowledge construction within a network. Yet, questions remain about whether such interactions can meaningfully enhance the tool and foster trust.

Drawing on Chadwick's (2017) concept of hybridity, ChatGPT functions both as an innovative communicative AI and as a medium reflecting narratives shaped by traditional media. Participants experienced this duality, when its outputs mirrored narratives from conventional media outlets. From a technical point of view, ChatGPT is trained on a large and diverse corpus of online textual data (Brown et al., 2020; Kalyan 2023; Dodge et al. 2022), which includes a variety of sources that carry their own set of biases. The model absorbs the biases present in the training data that are the product of other mediums and sources of information. In other words, the presence of political biases in the results of artificial intelligence can be seen as a reflection of the broader media environment from which it learns.

Chadwick's (2017) concept of hybridity emphasizes complexity, interdependence, and transition (p. 10). It encapsulates the role of ChatGPT within the modern media landscape, where it functions as both a "new" communicative AI introducing interactive, generative capacities (Hancock et al., 2020, p. 89) and a medium that reflects the influence of existing media forms and reproducing already existing content.

This hybridity positions ChatGPT as an integral component of the media environment, blending elements of both old and new media albeit in a new mode of presumably dialogic interaction, of prompt-and-response. Participants, especially those engaging with ChatGPT for the first time, encountered a new media experience where they not only received information about the Russian-Ukrainian war but also engaged with the tool by attempting to "teach" it or express emotions in reaction to the tool's responses. However, the nature of the responses often felt familiar, as if derived from older media forms, demonstrating the "not only but also" aspect of hybridity that Chadwick (2017) describes (p. 4). This duality highlights that while ChatGPT is an entirely new medium representing new media, it also remediates "old" media by carrying forward issues and biases inherent to traditional media. That also aligns with studies that indicate the presence of political and other biases in GPT models (Afgiansyah, 2023; Rozado, 2023; Hartmann et al., 2023). As a result, this reduces the trust in communicative AI as a reliable source of information putting into question the notion of media neutrality.

Thus, as a part of the hybrid media system, ChatGPT, influenced by the interdependence between old and new media logics, transmits information that users may perceive as biased. This issue becomes particularly sensitive in the context of geopolitical topics, such as the Russian-Ukrainian war, where communicative AI has the potential to inadvertently contribute to the spread of misinformation.

Given the results, it is important to emphasize the importance of media literacy in the context of communicative AI such as ChatGPT. As AI tools become increasingly integrated into the media landscape, the question arises of how to train users to critically evaluate AI-generated information. This can help reduce the risks of misinformation that arise in a hybrid media system where old and new media intersect. Additionally, this study focus on Ukrainian women in the diaspora emphasizes how emotional factors influence the use of AI communication tools. Their emotional reactions may be enhanced by their refugee status, living abroad, and the constant need to follow the news of the war. This suggests that trust in communication AI is deeply contextual and is shaped by specific circumstances and emotional needs of users.

Conclusion

The results indicated that several factors influenced users' evaluation of ChatGPT's responses in the context of the Russian-Ukrainian war, including biases, lack of source transparency, accuracy issues such as hallucinations and data limitations, terminological concerns like the use of "conflict" instead of "war", or the presence of Russianisms in Ukrainian texts. These issues negatively impacted users' trust in ChatGPT, suggesting that the tool may not meet expectations for reliable and unbiased information, especially in sensitive geopolitical contexts.

The level of trust was significantly influenced by both *informational* and *emotional* factors, deeply tied to the participants' backgrounds as Ukrainians and the context of their conversations. The limitations of ChatGPT's training data, primarily available only until January 2022, often resulted in outdated or incorrect information, particularly concerning the full-scale invasion of Ukraine in February 2022. This led to confusion and frustration among participants, further diminishing their trust in the tool.

Language errors, such as the use of Russianisms and politically sensitive terminology, also eroded the perceived reliability of ChatGPT. Participants reacted strongly to terms like “conflict” instead of “war”, and their attempts to correct ChatGPT had limited success, highlighting the emotional impact of these language issues. The lack of transparency regarding the sources of information further fueled skepticism, as participants were frustrated by ChatGPT’s inability to provide verifiable sources.

In conclusion, this study underscores the intricate dynamics of trust in communicative AI, particularly in sensitive geopolitical contexts such as the Russian-Ukrainian war. The findings suggest that even though AI systems like ChatGPT are likely to become increasingly (and imperceptibly) integrated into our communication networks, their ability to act as reliable and trustworthy sources of information is deeply challenged by complexity of the hybrid media system in which they operate. Thus, the trust users place in communicative AI systems is fragile, easily undermined by inconsistencies, biases, and the system’s failure to meet human expectations of the tool’s competence. These insights call for ongoing research into the ethical and practical implications of communicative AI in media systems, particularly in ensuring that these systems can be trusted as partners in communication rather than just tools that replicate or assemble information.

In addition to empirical results, this study also offers a methodological approach to how communicative AI can be studied qualitatively through a combination of traditional ethnographic techniques, such as interviews, and real-time interactions akin to prototype testing in software studies. This approach uncovers layers of user interaction and perception that might be missed by conventional methods like content analysis. By treating ChatGPT as both a tool and an active participant, the method views the entire experience as a media practice rather than merely evaluating media effects.

Furthermore, the interview sessions were not only a data collection method but also a meaningful experience for the participants. Participants with refugee status and personal experiences of war appreciated the opportunity to express their own opinions on information related to deeply painful events. This underscores the significance of creating spaces for dialogue in sensitive contexts like the Russian-Ukrainian war.

References

Afgiansyah, A. (2023). Artificial Intelligence Neutrality: Framing Analysis of GPT Powered-Bing Chat and Google Bard. *Jurnal Riset Komunikasi, 6*(2). https://doi.org/10.38194/jurkom.v6i2.908

Bang, Y., et al. (2023). A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *arXiv*. https://doi.org/10.48550/arXiv.2302.04023

Bolin, G., & Ståhlberg, P. (2023). *Managing Meaning in Ukraine: Information, Communication, and Narration since the Euromaidan Revolution*. The MIT Press. https://directory.doabooks.org/handle/20.500.12854/111596

Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

Chadwick, A. (2017). *The Hybrid Media System: Politics and Power*. Oxford University Press. https://doi.org/10.1093/oso/9780190696726.001.0001

Ghosh, S., & Caliskan, A. (2023). ChatGPT Perpetuates Gender Bias in Machine Translation. *arXiv*. https://doi.org/10.48550/arXiv.2305.10510

Guzman, A. L. (2018). *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves*. New York, NY: Peter Lang Publishing Inc.

Hepp, A., et al. (2023). ChatGPT, LaMDA, and the Hype Around Communicative AI. *Human-Machine Communication, 6*(1). https://doi.org/10.30658/hmc.6.4

Hughes, A. (2023). ChatGPT: Everything You Need to Know About OpenAI's GPT-4 Tool. *Science Focus*. https://www.sciencefocus.com/future-technology/gpt-3

Kaplan, A. D., et al. (2023). Trust in Artificial Intelligence: Meta-Analytic Findings. *Human Factors, 65*(2), 337–359. https://doi.org/10.1177/00187208211013988

Kalyan, K. S. (2023). A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4. *arXiv*. https://doi.org/10.48550/arXiv.2310.12321

Lee, S. K., & Sun, J. (2023). Testing a Theoretical Model of Trust in Human-Machine Communication. *Behaviour & Information Technology, 42*(16), 2754–2767. https://doi.org/10.1080/0144929X.2022.2145998

Luhmann, N. (1979). *Trust and Power*. Chichester, Toronto: Wiley.

Magda, Y. (2016). *Hybrid War*. CreateSpace Independent Publishing Platform.

Narayanan, A. (2023, March 20). ChatGPT is a Bullshit Generator. But It Can Still Be Amazingly Useful. *AI Snake Oil*. https://www.aisnakeoil.com/p/chatgpt-is-a-bullshit-generator-but

Pantti, M. (Ed.). (2016). *Media and the Ukraine Crisis: Hybrid Media Practices and Narratives of Conflict*. New York, NY: Peter Lang Publishing Inc.

Petriv, O. (2023, March 23). ChatGPT Is a Russian Propaganda "Consumer." *Centre for Democracy and Rule of Law*. https://cedem.org.ua/en/analytics/chatgpt-propaganda

Rozado, D. (2023). The Political Biases of ChatGPT. *Social Sciences, 12*(3). https://doi.org/10.3390/socsci12030148

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics, 26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Shevchuk, A. O. (2021). Russianisms (Surzhyk) as One of the Main Problems of Media Texts. In *Polit. Challenges of Science Today*. Available at: https://er.nau.edu.ua/handle/NAU/50070

Sundar, S. S., & Liao, M. (2023). Calling BS on ChatGPT. *Journalism & Communication Monographs, 25*(2), 165–180. https://doi.org/10.1177/15226379231167135

Williamson, B., et al. (2023). Re-Examining AI, Automation, and Datafication in Education. *Learning, Media and Technology, 48*(1), 1–5. https://doi.org/10.1080/17439884.2023.2167830

The Love to Hate Taylor: Dislike and Toxic Practices in Anti-fandom Culture Surrounding Taylor Swift

> I can't stand her but when I say so, turns out I'm "misogynistic". Let's get something clear, I don't hate her for being a woman, I hate her for being the classic rich privileged person who has never encountered a problem in her life and won't do anything if poorer people are suffering (butijustkeeponla, 2024)

Lisa Kinnunen

The digital age has ushered in a captivating phenomenon: the rise of the anti-fan. While these passionate critics existed long before the introduction of social media, the contemporary media landscape has provided them with an unparalleled platform to challenge the traditional dynamics of fandom. Consider the Reddit comment above made by an "anti-Swifter", a vocal group dedicated to despising pop star Taylor Swift (butijustkeeponla, 2024). This opening quote offers a glimpse into the multifaceted world of anti-fandom, revealing the complex web of motivations and behaviours that drive their online presence and opposition.

The evolution of fandom in the digital era has intensified fan-celebrity interactions and given rise to "stan culture". Eminem's 2000 song *Stan* highlighted obsessive fandom's darker aspects, depicting a fan driven to despair by unmet expectations, while platforms like Perez Hilton's infamous gossip blog blurred public-private boundaries and fostered fan entitlement (Adegbuyi, 2021b). Today, "stanning" encompasses various forms of engagement, from fan fiction to tracking behaviours, creating perceived connections that often blur the lines between parasocial and real relationships (Adegbuyi, 2021b). In contrast, anti-fans have emerged on social media, critiquing celebrity culture and questioning the authenticity of celebrity images and toxic parasocial relationships (Adegbuyi, 2021a; Graham, 2020). These communities provide unique spaces for real-time discussions and amplify anti-fan perspectives, while algorithms increase the visibility of their content, reshaping media discourse (Adegbuyi, 2021a; Chin & Huang, 2023; Dowell, 2024).

A thriving fandom combines admiration with constructive criticism, differentiating critics from anti-fans, who actively oppose media figures (Dowell, 2024). In contrast to non-fans, who engage passively, anti-fans are deeply involved, vocalising dissent against celebrities while refusing to adhere to traditional fan norms (Gray, 2003). While negativity can appear in both fandom and anti-fandom, it does not only equate to anti-fandom, as negativity can be present in fandom as well (Dowell, 2024). Toxicity signifies a deliberate hostility aimed at harming others, and this study addresses the gap in fan studies by exploring how anti-fans escalate negativity into toxicity. Historically, fan studies have emphasised positive emotions like love and admiration, neglecting the equally intense feelings of dislike and hatred (Click, 2019). Anti-fans disrupt celebrity adoration narratives by expressing intense dislike and critical interpretations of celebrities' actions, as highlighted by Graham (2020), who notes that "haters are just fanboys with the sign switched". This perspective emphasises the need to consider both admiration and dissent to fully understand fandom dynamics.

This study delves deeper through an examination of previous research in the field, as well as employing a case study for the analysis of this study. While fan studies have primarily examined the positive aspects of fandom, the realm of anti-fandom, fuelled by burning negativity and toxicity, offers a unique and critical perspective. To illuminate the inner workings of these groups, this study examines the unexplored realm of anti-fandom, specifically targeting global pop star Taylor Swift's anti-fan movement on social media. Swift's broad popularity and cultural impact makes her an ideal subject for studying anti-fan behaviours and the dynamics of online dissent.

The Taylor Swift Phenomenon

Taylor Swift's career, marked by record-breaking achievements like her *Eras Tour* and "Taylor's Versions" re-releases, cements her as a major cultural icon. However, her success has sparked intense anti-fan responses, especially around her advocacy for feminism and public relationships (Irfan & Atukunda, 2024; Stark, 2024). Despite under constant scrutiny, Swift was named Time Magazine's Person of the Year in 2023, recognising her ground-breaking contributions to music, activism and social discourse (Lansky, 2023). Her impact extends to realms like American football, where her appearances with NFL player Travis Kelce sparked frustration among NFL fans (Stark, 2024). Her potential political influence is underscored by the 2024 American presidential election, where both camps see her endorsement as pivotal. Her dominance in these spaces generates both admiration and animosity, illustrating the complex dynamics of celebrity perception (Helmore, 2024; Lansky, 2023).

Swift's loyal fanbase, known as "Swiftes", fiercely defends her, creating a positive atmosphere at her concerts and cultivating a culture of kindness (Hedbom, 2024). However, this unwavering support also leads to challenges in fostering balanced critiques of her work and persona (Yadav, 2023). The Swifties' dedication contrasts sharply with the sentiments of anti-fans, who view Swift's crafted public image with scepticism. This positivity may unconsciously fuel anti-fandom, as those rejecting Swift's wholesome image or fanbase kindness may respond with intensified criticism and negativity. Swift's case thus offers a unique lens on the complexities of fan and anti-fan culture, where extreme admiration and opposition coexist and influence public discourse.

Aim and Research Questions

This research aims to shed light on the understudied phenomenon of anti-fandom culture in the digital age, focusing on how anti-fan communities targeting media figures manifest and operate on social media. By examining online interactions and the dynamics within these communities, the study seeks to uncover aspects of negativity and toxicity in anti-fandom practices and explore the role of parasocial interactions and relationships within anti-fandom culture. Ultimately, this research aims to provide a nuanced understanding of anti-fandom culture, offering insights valuable to scholars and those interested in the complexities of online fan culture and celebrity interactions. To examine this, three main research questions has been constructed:

1. How are anti-fans voicing their negativity and (eventual) toxicity? How do they communicate with fellow anti-fans?
2. How are anti-fans using social media platforms to do this? What practices and framings are they focusing on?
3. How do anti-fans demonstrate parasocial tendencies in their interactions on social media?

Previous Research and Theoretical Framework

This review summarises previous research, offering a brief historical overview of key areas of this study: fan culture studies, fan culture on social media, anti-fan culture, parasocial interactions and relationships, as well as media practice and social media logic. This overview serves as the foundation upon which the theoretical framework and the subsequent analysis of this study are built.

Fan Culture Studies

To gain a comprehensive understanding of fan culture, it is essential to explore the foundational works of early scholars in the field. These contributions illuminate the origins and evolution of fan culture, providing historical context. Initially, fandom studies primarily focused on film, television and popular music, often portraying fans as obsessive and irrational, overlooking the diverse nature of fan communities.

Jenkins (1992), in his influential book *Textual Poachers*, argues against this stereotype, emphasising the complexity and variety within fandom. Fiske (1992) further develops this discourse in *The Cultural Economy of Fandom*, suggesting that fan culture serves as a response to gaps left by mainstream culture. He argues that fandom offers social status and self-esteem through cultural capital and positions popular culture as a reflection of the people's tastes rather than an imposition (Fiske, 1992). Fiske asserts that audiences are active participants rather than passive consumers, engaging critically with media texts and creating their own interpretations through fan art and fan fiction (Fiske, 1989/2011).

The fandom surrounding *Star Trek* is often cited as foundational in media fandom studies. Bacon-Smith (1992) argues that this franchise laid the groundwork for understanding media fandom. Also, Jenkins (1992) critiques the oversimplified portrayal of fans as socially immature. He highlights that different fan groups share experiences but also exhibit differences based on their cultural standing and interests (Jenkins, 1992). Similarly, Fiske (1986) introduces the idea of polysemy in popular television, suggesting that multiple interpretations allow viewers from diverse backgrounds to find personal meaning within shows, reflecting their unique social positions. Both Jenkins and Fiske recognise that fandom often emerges as a response to power imbalances between consumers and producers. Jenkins (1992) describes fandom as a platform for consumer activism, where fans engage with producers to express opinions and influence media narratives. This collective engagement fosters a sense of identity and community among fans.

While Jenkins and Fiske's work significantly advances our understanding of fan engagement, it does not fully capture the complexities of contemporary fandom, particularly in the digital age. To fully comprehend today's fan culture, research must extend beyond traditional media and investigate the dynamic landscape of online fandom.

Fan Culture on Social Media

Social media has transformed fan culture, enabling global connections and fostering diverse, inclusive communities. These platforms have reshaped how fans interact, share content and collaborate, shaping the landscape of fan culture in the digital era. Fans now actively engage in content creation and discussion, exemplifying the concept of participatory culture, a term Jenkins (1992) initially introduced in *Textual Poachers*. He framed fans as a creative community that repurposes mass entertainment through activities like fan fiction, fan art and video edits. Later, Jenkins et al. (2016) revisited this concept, highlighting its evolution within a dynamic digital environment enabled by online platforms and digital media. It involves everyday digital interactions that promote diversity and democracy, allowing individuals and groups to express themselves freely (Jenkins et al., 2016). Key characteristics include minimal barriers to expression, strong support for sharing, and informal mentorship within communities. While participatory culture can foster connections, it may not always yield positive societal outcomes. Ultimately, participation encompasses cultural engagement beyond mere activity, distinguishing it from interactivity by emphasising shared practices and collective experiences (Jenkins et al., 2016).

Fans contribute to media convergence by engaging with diverse content across platforms. Jenkins (2008) explores how media convergence represents the interconnectedness of platforms, industries and audiences, where content flows across various media channels and fans pursue integrated media experiences. He argues that convergence is not just technological but a cultural shift where consumers actively participate in shaping, interpreting and circulating media content, making participatory culture integral to today's media landscape.

Over the past two decades, the convergence of media and digitalisation has transformed how media is produced and consumed, empowering fans through platforms that enable greater engagement and influence. Galuszka (2015) discusses how digital tools allow fans to collectively impact producers, while Jenkins (2006) describes online fan communities as expansive, self-organised groups that actively create, debate and share interpretations of cultural artefacts, unconstrained by geographical boundaries. This shift has brought fandom from niche subcultures into the mainstream, although it also introduces challenges, such as a generational divide between older fans and newcomers who navigate today's fast-paced digital fandom spaces (Jenkins, 2006).

Anti-fan Culture

Fan studies often overlook dislike and hatred among audiences. Scholars like Gray (2003) highlight the importance of studying anti-fans, individuals who actively dislike media content. Gray's study on viewers of *The Simpsons* laid the foundation for anti-fandom studies, emphasising the need to understand anti-fans' perspectives within fan culture (Gray, 2003; Click, 2019). Anti-fans, while opposite in sentiment to fans, engage deeply, often deriving enjoyment from shared dislike, forming communities that echo fan structures (Click, 2019).

Additionally, Gray (2019) further explored anti-fandom through hatewatching, where viewers continue watching disliked content, often for cultural relevance or hope for improvement. Gilbert (2019) notes that hatewatching fosters communal engagement, framing it as a collective identity performance within anti-fandom (Gilbert, 2019). This communal dislike is often social, creating commentary on taste and quality while reinforcing shared attitudes toward media.

Anti-fandom has gained visibility on social media, amplifying both fan and anti-fan voices. Click (2019) argues that platforms like Twitter host "snarky" and critical engagements alongside positive ones, showing how convergence culture allows both fandom and anti-fandom to coexist online. Recuero (2024) connects anti-fandom to online toxicity, suggesting that such discourses can polarise communities, foster hate speech and build echo chambers, while Gray (2019) highlights its potential for rivalry and hostility between fan bases. Furthermore, Gray and other scholars link anti-fandom to Fiske's (1986) concept of polysemy, suggesting that anti-fandom provides valuable insights into how digital interactions shape modern media landscapes.

Additionally, Click (2019) advocates for more research on the emotional dynamics of dislike in digital media, noting the impact of such negative engagements on individuals (Click, 2019). Building on fan and anti-fan scholarship, this study will next explore parasocial relationships, which are central to both fandom and anti-fandom, particularly in celebrity culture.

Parasocial Interactions and Relationships

In her article, Jenson (1992) suggests that intense fandom often functions as psychological compensation for modern life's perceived emptiness. This aligns with Horton and Wohl's (1956) concept of parasocial interactions, where individuals develop feelings of closeness with media figures through one-sided emotional investments. Jenson (1992) highlights that a fragmented society can lead to a fragmented self, suggesting that fandom may fill identity voids left by societal alienation. Building on Horton and Wohl, Hartmann (2008, 2016) distinguishes between short-term parasocial interactions and enduring parasocial relationships. While interactions give an illusion of engagement, relationships are formed through deeper emotional ties, which can be either positive or negative, thus broadening the concept's application.

In the digital age, parasocial dynamics have evolved significantly, with Stever and Lawson (2013) highlighting how social media, particularly platforms like Twitter, enhances perceived direct interactions and fosters parasocial bonds. Yuksel and Labrecque (2016) emphasise that online spaces facilitate both positive and negative parasocial relationships, though most studies have focused on brand-related interactions rather than personal connections. Hartmann (2016) notes that these connections can range from admiration to aversion, creating complex social dynamics in digital environments. Mardon et al. (2023) further explore how social media fosters negative parasocial relationships, where shared dislike of media figures can drive anti-fan communities. Despite facilitating both positive and negative bonds, the existing research does not fully address the impact of anti-fan communities, a gap this study seeks to fill by exploring the complexities of negative parasocial relationships.

While Horton & Wohl's foundational work emphasised the illusion of intimacy, recent research has refined these concepts, distinguishing between parasocial interactions, brief exchanges during media use, and parasocial relationships, which are long-term bonds resembling friendships (Schramm & Hartmann, 2008). Furthermore, the term "parasocial processing" expands this idea to include all user responses toward online personas, even without perceived interaction (Schramm & Hartmann, 2008). For this study, the term PSI will encompass this broader definition. Schramm and Hartmann (2008) propose a Two-Level Model of *PSI-Process Scales*, which categorises *automatic cognitive*, *affective* and *behavioural* responses triggered by media personas. This model suggests that such responses are immediate and involuntary, varying in intensity and significance, which is particularly relevant for anti-fandom studies. The three response categories include: (1) *perceptual-cognitive responses*, where individuals form impressions of the persona; (2) *affective responses*, capturing the range of emotions felt towards the persona; and (3) *behavioural responses*, encompassing user reactions such as non-verbal cues or desires for interaction (Schramm & Hartmann, 2008). Social media platforms foster anti-fan communities, allowing users to connect over negative sentiments towards media figures through one-sided communication, creating a sense of belonging while maintaining anonymity. By applying the PSI framework, this study explores how anti-fan behaviour manifests as negative responses to media figures or content.

Media Practices

Couldry (2012) introduces the concept of media practice, emphasising how individuals integrate media into their daily lives, moving beyond passive consumption to engage actively with media. This approach encourages researchers to consider what people do, say and think in relation to media, thereby revealing the social dynamics at play (Couldry, 2012). By shifting focus from media texts to the broader range of media practices, Couldry's framework captures the diverse interactions individuals have with media, facilitating a deeper understanding of how these practices shape and are shaped by the media landscape (Couldry, 2012).

Within Couldry's framework, two key categories are applied to this study: *presencing* and *commentary*. Presencing refers to the methods individuals use to maintain a public presence through media, which involves constructing an identity that transcends mere personal communication (Couldry, 2012). It emphasises the importance of media platforms and skills in facilitating self-expression while also raising questions about the effort required to sustain an online presence. Commentary, on the other hand, addresses the need for signalling amidst an abundance of content, as individuals now share and comment on materials globally. This practice has significant implications for media economics and sociology, influencing how commentary is organised and understood in the digital landscape (Couldry, 2012). Overall, these concepts provide valuable insights into how social dynamics evolve within the media environment and shape individual experiences and interactions.

Media Logic and Social Media Logic

The rise of online fan culture is driven by the interplay of media logic and social media logic. Media logic, a concept established by Altheide and Snow in 1979, refers to how media formats shape the presentation and interpretation of information. It encompasses the tools and rules that guide communication and cultural evolution across various platforms. Furthermore, media logic refers to how events, actions and performances are shaped by the governing technologies, media and formats of communication, which become institutionalised and guide social interactions and cultural shifts (Altheide, 2016).

While initially linked to traditional mass media like television, media logic has evolved with digital platforms such as social media, smartphones and the internet, adapting to new mediums and their unique characteristics. This progression highlights that media logic is not linear or medium-specific but a dynamic conceptual model of how mediation shapes societal change (Altheide, 2016).

The concept of social media logic includes four key elements, although this study will primarily focus on three of them: *programmability*, *popularity*, and *connectivity* (van Dijck & Poell, 2013). Programmability highlights how platforms influence user-generated content while allowing users to impact information flow. Popularity illustrates the evolving dynamics of fame and content visibility on social media compared to traditional media. Lastly, connectivity emphasises the mutual relationships among users, platforms and advertisers, creating an interactive environment where users actively shape content (van Dijck & Poell, 2013). Together, these elements reveal the complexities of online engagement and the transformative nature of social media on culture.

Methodology

Given the focus on anti-fandom culture and its online manifestation, the study combines a quantitative sampling method with qualitative analysis of some representative posts. This aligns with Jensen (2002), who notes that qualitative research involves sampling elements within cultural contexts to understand concepts like citizenship or reality. Similarly, Brennen (2017) argues that researchers analyse discourse to understand how communities are formed, how meaning is constructed, and how social realities are shaped. Goffman's (1974) framing analysis explores how people use "frames" to make sense of social interactions. He defines a frame as a construct influenced by contextual principles that shape how individuals perceive events and engage in situations (Goffman, 1974; Entman 1993; Van Gorp, 2007; Persson, 2018). Snow et al. (1986) introduces frame alignment processes, including *frame bridging*, as a mechanism to connect individuals' interpretations to collective goals, fostering a shared identity within social movements (Snow et al., 1986; van Dijk, 2023).

Moreover, this research uses a case study approach to analyse the dynamics of anti-fandom culture within an online community focused on Taylor Swift. While case studies provide valuable insights into the case itself, they also allow for a broader understanding of social patterns by identifying recurring themes and providing detailed descriptions. They focus on observing phenomena within their natural contexts and exploring their connections to wider social structures and themes (Jensen, 2002).

Materials

This study examines anti-fandom in a long-standing Facebook group dedicated to expressing dislike toward Taylor Swift, chosen for its global reach and diverse user base (Vitak, 2016), which facilitates the exploration of online phenomena like anti-fandom. Established in 2007 and with nearly 1,900 members, the group serves as a structured community where anti-fans express shared dislike and build collective identities.

The study employed a systematic approach, utilising a data-scraping tool to collect and preserve the chronological order of all posts (*N=624*) from the Facebook group, which created a timeline of community interactions. Initial data analysis revealed distinct frames of meaning, leading to a two-step sampling process (Jensen, 2002). First, key events were identified across different time frames to select representative materials. A deeper analysis of these samples uncovered patterns and connections, enhancing understanding of group dynamics and the evolution of shared sentiments. Due to periods of heightened activity, a subset of (*n=413*) posts from the total was selected for analysis, specifically from time frames with a high volume of comments, providing a rich basis for examining engagement levels and discussion content.

The study analysed data from the Facebook group spanning its entire existence from 2007 up to April 1, 2024, providing insights into its evolution. While not focused on direct period comparisons, the historical context highlighted changes within the community. The analysis concentrated on two main periods: a peak activity period (years 2023–2024) with a large sample of posts (*n=311*) and comments that reflected current dynamics, and an earlier active period (years 2009–2011) with a smaller sample (*n=102*) that offered valuable historical context. From these selections, the analysis refined its focus to (*n=18*) posts from the earlier period and (*n=39*) from the peak period, totalling (*n=57*) posts. This purposeful sampling prioritised posts with high comment activity and multimedia content to facilitate a deeper exploration of the group's dynamics.

Data Analysis

The analysis process involved collecting all selected Facebook posts to conduct a comprehensive framing analysis, capturing the full context of each post, including comments, reactions and visual elements. This holistic approach was essential for accurately identifying the frames that anti-fans use to express their dislike, supported by a systematic coding scheme for the (*n=57*) posts. The study linked established knowledge on parasocial interactions, anti-fandom culture and online toxicity to reveal the complex connections between behaviour, community building, and the formation of online anti-fan identities. The research utilised a coding scheme (see *table 1*) grounded in studies of anti-fandom and participatory culture, focusing on *anti-fan behaviour, toxicity* and *community dynamics*. By employing the concept of frame bridging, the coding scheme demonstrated how anti-fans connect various ideas to create a shared frame of reference around their dislike, which is crucial for mobilising participation and fostering a sense of identity within their community.

Frame Package	Sub-Category	References	Description	Examples
Anti-fan behaviour				
	Antipathy	Gray (2003)	Expressing strong dislike or hatred towards the artist	Using insults, minimising the artist's achievements, celebrating their perceived failures.
	Hatewatching	Gilbert (2019), Gray (2019)	Engaging with disliked media content with the intention of finding fault.	Listening, watching or reading about the artist to point out flaws and find something to critique.
	Snarky Expressions	Click (2019)	Using sarcasm, jokes or mean comments about the media object.	Counter positive comments with sarcasm, using mocking emojis or writing ironic or backhanded compliments.
Community				
	Collective Discursive Practices	Jenkins (2006, 2008), Jenkins et al. (2016)	Engaging in shared discussions and debates that solidify the anti-fan narrative.	Engaging in dedicated forums, sharing content (articles, blog posts, etc.) online, and fuelling discussions meant to tear down their target.
	Shared Identity Construction	Jenkins (2006, 2008), Jenkins et al. (2016)	Developing and reinforcing a sense of belonging and shared	Expressing solidarity and sharing personal experiences with other members through supportive

			purpose within the community.	comments or reactions.
	Creation/sharing of (anti)fan art	Jenkins (2006, 2008), Jenkins et al. (2016)	Creation/sharing of critical and/or satirical artistic representations.	Crafting memes that mock the artist's work or persona. Creating image edits that satirise the person or thing's appearance or actions.
Toxicity				
	Polarisation	Recuero (2024), Gray (2019)	Framing discussion about the media object/figure as an opposition between anti-fans and fans.	Fosters an "us vs. them" divide, portraying themselves as morally superior and demonising the opposing side.
	Hate speech	Recuero (2024)	Using language that attacks or demeans the artist or their fans based on personal characteristics.	Attacks ranging from derogatory and/or misogynistic slurs and body shaming to threats of violence and sexual harassment, targeting both the artist and their fanbase.
	Aggressive Behaviour	Recuero (2024)	Engaging in actions intended to intimidate, harass or silence opposing viewpoints.	Sending threatening messages and coordinate online attacks. Engaging in cyberbullying behaviours with the intent to humiliate or harm others.

Table 1. Coding scheme with framing packages.

Ethical Considerations

To ensure participant safety and anonymity, this research will not disclose identifying information about individual group members, despite Facebook itself may not enforce strict anonymity, as some members are using their real names in the Facebook group. Given the sensitive nature of the research on negativity and toxicity in anti-fan identities, revealing names could expose participants to harassment or judgement. Consequently, all examples and quotes in the results and analysis sections will be anonymised, prioritising ethical protection while still enabling valuable insights from the collected data.

Results and Analysis

This section analyses the data from the Facebook group to reveal the characteristics of anti-fandom culture, exploring themes and narratives that illustrate how anti-fans express themselves, interact and construct their collective identity. Using the key frame packages from the coding scheme, the findings connects to existing research and the theoretical framework. The analysis will focus on interconnected themes, including *the anti-fan behaviour, the anti-fan practices* and *the anti-fan community*, and lastly, the theme of *parasocial interactions and relationships* will be explored.

The Anti-fan Behaviour

The data showed that a strong current of antipathy toward Taylor Swift emerged as a defining theme within the anti-fan community, shaping their identity, communication style and collective sense of purpose. This intense dislike went beyond individual complaints, creating a shared narrative that criticised Swift's perceived flaws, her music and her public persona. Many anti-fans echoed sentiments similar to Gray's (2003) definition of anti-fans as those who view a figure or work as "inane, stupid, morally bankrupt and/or aesthetic drivel" (Gray, 2003, p. 70). For example, one member compared listening to Swift's music to the harsh sound of "bricks in a blender", underscoring a dismissive view of her talent. Another remarked, "I don't think she sucks. I know she sucks", reflecting an uncompromising stance that extends beyond opinion into certainty. Collectively, these comments demonstrate how antipathy functions not only as individual expression but as a unifying force, creating a community bound by shared critiques and collective disapproval.

Sarcasm and snark dominated the anti-fan community's comments, as members used sharp humour to belittle Swift, aligning with Click's (2019) observations. Awards were dismissed as undeserved, lyrics twisted to reveal supposed arrogance, and her personal life became a target for mean-spirited jokes, with one example of a comment being: "I'm just happy she finally wrote a song admitting SHE is the problem!". Anti-fans also strategically employed emojis to amplify sarcasm; laughing-crying emojis often accompanied such remarks to underscore contempt, while vomiting emojis vividly expressed antipathy. This use of emojis not only enhances the emotional impact of their comments but also serves as a visual signifier of their negative sentiments toward Swift.

While the anti-fan community's intense antipathy toward Swift might suggest baseless negativity, their critiques often delve into substantive issues like racism, LGBTQ+ rights and environmental responsibility. Some members express these criticisms with a blend of snark and pointed observations, demonstrating a desire to hold Swift accountable for her influence and messaging. For example, a member asserts that Swift has "the ability to change from a republican, to a democrat to a liberal", suggesting she adapts her views to gain popularity. The member further critiques her support for LGBTQ+ rights while simultaneously supporting conflicting groups, stating that Swift "plays the victim so well" and exhibits an "attention-seeking mentality of a 13-year-old". This comment underscores a deeper disdain for Swift's public persona, as it reflects a belief that, unlike other celebrities who "go to sports and hides", she loves being in the spotlight, portraying herself in ways that align with their critiques.

In response to Swift being named Time Magazine's 2023 Person of the Year, many members expressed strong disdain for the title, blending their antipathy with genuine critiques. They not only questioned Swift's worthiness but also criticised the magazine's choice, arguing it overlooked true heroes like war fighters and journalists. This reflects Gray's (2003) insights into anti-fandom, suggesting that anti-fans often engage in substantive critiques rather than mere hatred. However, alongside this capacity for thoughtful criticism, a troubling aspect emerged, as some comments revealed intense loathing expressed through degrading, misogynistic, and even violent language, highlighting an undercurrent of toxicity within the group. These toxic comments, though not the majority, were unmistakably present, highlighting an unsettling undercurrent in some interactions, with one example being this comment:

> someone should snip Taylor's Achilles tendon with a pair of scissors and she'll never be able to walk again nor tour or go to the studio again. stupid lifeless bitch

The anti-fan community's negativity illustrates how anti-fan behaviour can drift into toxicity, aligning with Recuero's (2024) findings on online hate speech and aggression. Recuero argues that social media fosters environments for negativity, facilitating cyberbullying and trolling, as seen in this anti-fan group's harmful discourse, which amplifies polarisation and can lead to violent interactions. Additionally, this behaviour reflects Couldry's (2012) analysis of media practices, where high engagement does not necessarily lead to positive interactions; here, anti-fans actively spread negativity, using platforms to fuel toxic discourse and cyberbullying. Viewed through Jenkins et al.'s (2016) lens of participation, the anti-fan community's active engagement exemplifies how shared practices and culture can drive collective negativity on social media.

In this context, anti-fans engage in a shared culture of negativity and hostility towards Swift and her supporters, actively contributing to this environment through hateful content and cyberbullying, which reinforces their negative sentiments. This behaviour reflects Jenkins et al.'s (2016) distinction between mere activity and meaningful participation, as such practices can have harmful social consequences. Although platforms for participatory culture encourage user expression, they also enable the spread of hostility, highlighting the need for critical awareness and possible interventions to address online toxicity. Notably, aggression within this community often appears in subtle, symbolic forms, such as indirect threats disguised as snarky remarks (e.g., "let's convince her she is [the almighty] … that she can jump into a volcano unscathed!"). Such hidden hostility, though seemingly playful, fosters a shared sense of belonging rooted in negativity and symbolic violence, emphasising the importance of analysing subtext in online interactions, where hostility is often masked by humour or harmless language.

The anti-fan community's reliance on sarcasm and coded language reflects van Dijck & Poell's (2013) concept of social media logic, where algorithms amplify negativity. Moreover, the analysis revealed the creation and distribution of elaborate conspiracy theories and misinformation targeting Swift, encompassing false narratives and harmful speculation about her character and intentions. This behaviour, initially unexpected, was incorporated into the coding scheme and extended beyond mere criticism, encompassing false narratives about Swift's achievements. Members accused her of "buying" awards and spread misleading stories about her entry into the music industry, as illustrated by comments questioning how she gained accolades and disparaging the integrity of media outlets. In essence, this phenomenon with its snarky comments and indirect aggression creates echo chambers that reinforce confirmation bias, leading individuals to accept false information that aligns with their beliefs while dismissing fact-checking efforts, ultimately harming the public's ability to discern truth from fiction and contributing to wider online polarisation.

The Anti-fan Practices

Now having examined anti-fan behaviour on social media, this study now delves into specific practices that anti-fans engage in. The study builds upon Gray's (2019) concept of *hatewatching*, which describes viewers actively engaging with media they dislike. While Gray's studies focused mostly on television shows and movies, the concept can also be applied to anti-fans of celebrities and other forms of media. Expanding on this concept, the study introduces *hate-consuming*, a broader term that encompasses various forms of media consumption beyond just visual content. Anti-fans actively seek out and consume content related to their disliked figure, engaging in activities such as hate-listening to music, hate-watching interviews, and hate-reading articles. This behaviour illustrates a unique engagement driven by negativity, as anti-fans participate in online discussions, share information and contribute to discourse, even if it elicits negative emotions. Their pursuit of content to fuel their disapproval emphasises the multifaceted nature of anti-fandom culture. The concept of hate-consuming connects to participatory culture, as highlighted by Jenkins (1992), showcasing that their engagement, while negative, is far from passive. This aligns with Jenkins et al.'s (2016) assertion about the active roles individuals play in cultural contexts. Hate-consuming signifies an active contribution to the anti-fandom culture, reflecting Gilbert's (2019) observation that hate-watching has become a prevalent form of media consumption. Ultimately, through their engagement, anti-fans become active participants in the broader cultural landscape, despite their negativity.

Additionally, the study observed a significant shift in anti-fan activity during the second data collection period, with an explosion of snarky memes targeting Swift's personality, appearance and career. This contrasts with the first period, where such creative expressions were rare. These memes, as exemplified by *Figure 1*, serve as a tool for playful mockery and bonding within the anti-fan community. This aligns with Jenkins' (1992) concept of participatory culture and Couldry's (2012) media practices theory, where individuals actively participate in shaping their cultural experience. The anti-fan community uses memes to go beyond passive consumption and actively create and share content that reinforces their shared antipathy towards Swift. By recontextualising her image and actions, they contribute to a narrative that fuels their dislike.

Figure 1. Example of snarky meme shared by anti-fans.

The research highlights the potential dangers of media practices, particularly the use of memes to spread negativity. While snarky memes dominate, a smaller but concerning trend of toxic memes targeting Swift with misogynistic and sexist content emerged. Examples like those in *Figure 2* and *Figure 3* showcase the potential for online harassment and the escalation of negativity within the anti-fan community, where they share memes that likens Swift to a "blow-up doll" and calls her "stupid", as well as calls her "trashy" with a sexual undertone. These examples showcase the anti-fans' willingness to share and encourage misogynistic views based on harassing Swift's appearance in a sexist way. This aligns with Couldry's (2012) concern about the ease with which nega-

Figure 2. Example of a toxic meme comparing Swift to a blow-up.

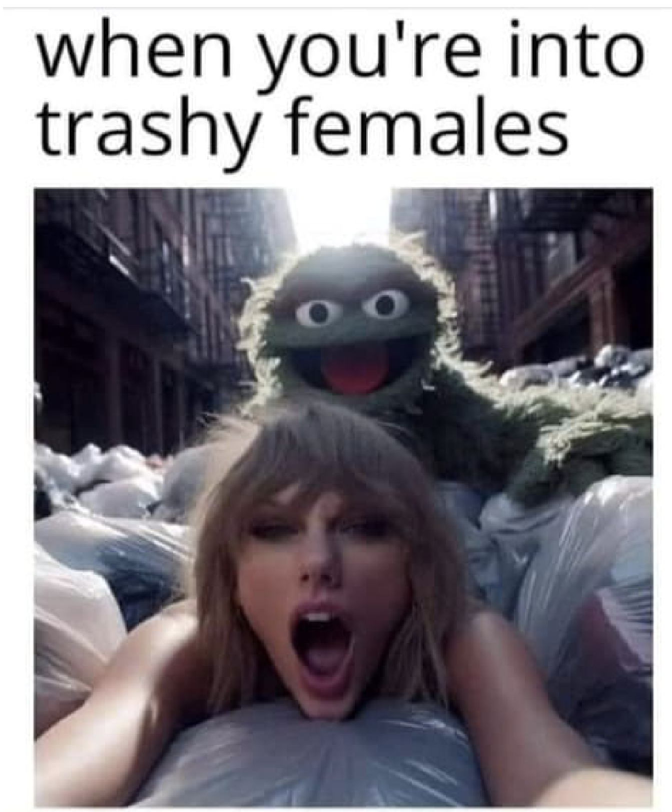

Figure 3. Example of a toxic meme constructed to show Swift in a degrading, sexual manner.

tivity can spread online, as meme-makers actively contribute to shaping negative perceptions of their targets.

Anti-fan communities weaponise memes as tools for degradation and harassment, embodying Gray's (2019) concept of hatewatching and this study's proposed idea of hate-consumption, highlighting the risks of online anonymity and unchecked negativity. These practices, examined through Jenkins' (1992) and Couldry's (2012) frameworks, emphasise the need for critical media engagement, as memes help maintain anti-fans' public presence and reinforce shared disdain for Taylor Swift. Platforms like Facebook facilitate this through social media logic (van Dijck & Poell, 2013), yet the lack of moderation can intensify toxicity, underscoring the importance of media literacy to distinguish between constructive and harmful discourse. However, it is crucial to acknowledge that Facebook might not represent the full extent of online negativity, as other platforms with less moderation could harbour even more extreme anti-fan communities.

The Anti-fan Community

The analysis of anti-fan behaviour surrounding Taylor Swift highlights the community's focus on antipathy toward her career and persona. While earlier discussions centred on expressions of negativity like memes and conspiracy theories, this section shifts to explore the shared identity within the anti-fan community. The data analysis reveals a significant evolution in interaction patterns within the Facebook group, moving from minimal engagement in 2009-2011 to a surge in discussions and collaborative content creation by 2023-2024. Members began to experience a sense of belonging and shared purpose, as evident in supportive comments on posts that mocked Swift. Many expressed gratitude for finding a community where they could openly share their disdain for her, emphasising their relief and validation in connecting with like-minded individuals. Some even noted a growing trend in anti-fan sentiment, suggesting an increase in searches for negative commentary about Swift, further reinforcing their views.

One member of the anti-fan community expressed confusion and frustration over Taylor Swift's popularity, questioning; "how could this person be the greatest in all the world, what the h*** is going on?". This feeling of isolation was alleviated by finding a Facebook group that provided a sense of belonging and validation for their scepticism about Swift's success. Another member, despite uncertainty about how they found the group, continued to engage, indicating that it fulfilled a need for connection with others who share their views. However, the group's antipathy extends beyond Swift, it also includes her fans and supporters, fostering a hostile environment for anyone expressing positive opinions about her (see examples below). This aligns with Gray's (2019) idea that anti-fandom should be understood in terms of rivalries, where the negativity directed at Swift is intensified against her fans. The aggressive and toxic language used by the anti-fans reveals a deeper loathing that seeks to silence any positive discourse about Swift rather than engage in healthy competition:

> For me, her fans play a big part into why I despise her. They are delusional and think they know her personally because they've seen her in concert and follow her socials. They are obsessive and it's pathetic.
>
> You can tell Swifties are a zombie cult because they attack everyone with a brain.

This behaviour contributes to broader online polarisation, as discussed by Recuero (2024), who argues that negativity fosters toxicity, creating filter bubbles and echo chambers where anti-fans primarily interact with like-minded individuals, amplifying their negative views toward Swift and her supporters. Recuero also highlights how algorithms exacerbate an "us vs. them" mentality, limiting exposure to opposing perspectives. The antipathy directed at Swift's fans exemplifies affective polarisation, where strong negative feelings toward differing views lead to distrust and hostility that can impact real-world interactions. There were instances where "Swifties" entered the anti-fan Facebook group disrupted this negativity and challenged the echo chamber effect.

While some interactions remained snarky, others escalated into aggression and violence. Anti-fans perceived the presence of Swifties as an attack, further entrenching their hateful opinions and creating a hostile environment that hinders constructive dialogue. Moreover, the anti-fan community often views itself as victims of censorship, which cultivates conspiracy theories and narratives of victimhood. This pervasive negativity fosters a hostile environment that discourages diverse viewpoints, contrary to Jenkins et al.'s (2016) vision of participatory culture promoting constructive dialogue. Disagreements and internal conflicts arise, revealing the blurred line between shared antipathy and acceptable negativity, and highlighting ethical dilemmas surrounding derogatory tactics against a public figure, even as some members advocate for "girl power".

This study reveals anti-fandom to be a complex ecosystem that includes not just haters, but also critics, trolls and potential lurkers who may disapprove of the negativity. This diversity underscores how anti-fans utilise tactics like frame bridging, connecting disparate ideas to establish a common narrative of dislike (Snow et al., 1986). The anti-fan community uses e.g. hate speech, cyberbullying and anti-fan art to solidify group identity and foster a sense of belonging. These strategies bridge individual grievances into a collective frame, motivating participation and reinforcing member bonds through a shared sense of purpose and negativity.

Parasocial Interactions and Relationships

The Facebook group studied thrives on a complex interplay of parasocial interactions and shared opposition to a celebrity. Unlike traditional fans who form positive parasocial relationships (Horton & Wohl, 1956), anti-fans develop connections based on dislike. This highlights how social media can foster unexpected connections, even if it is through negativity. This study will employ the PSI-Process model (Schramm & Hartmann, 2008) to frame and understand anti-fan behaviour within the Facebook group. The model's three core response categories (*cognitive*, *affective* and *behavioural*) will be used to analyse qualitative data related to anti-fandom. Key findings will be mapped onto the model to illustrate how anti-fans frame their interactions and relationships within the community (see *table 2)*. The analysis will target the most relevant aspects of the PSI-Process model related to anti-fandom, acknowledging that users may not experience all response categories in every interaction (Schramm & Hartmann, 2008).

The PSI-Process model emphasises how anti-fans engage in negativity-focused scrutiny, magnifying flaws and contributing to hate-consuming behaviour. Unlike traditional fans, who often focus on positive aspects, anti-fans engage in negativity-focused scrutiny. They actively seek out and magnify perceived flaws, tailoring their content to align with the community's collective negativity. This behaviour contributes to the formation of echo chambers, where negative information is amplified and opposing viewpoints are suppressed. Anti-fans interpret future developments negatively, dismissing positive news and fostering a climate of distrust. Unlike fans who build positive parasocial relationships, anti-fans deliberately distance themselves from the media figure, focusing on flaws and undermining their public image. Their communities thrive on shared disdain, reinforcing their identity and solidarity.

Furthermore, the PSI-Process model examines the emotional responses of anti-fans toward media figures, emphasising their intense dislike, which results in critical behaviours such as insults and harassment. This antipathy fosters in-group sympathy and solidarity among anti-fans, contrasting with the empathy fans feel for celebrities. Instead, anti-fans exhibit counter-empathy, bonding over shared negative experiences. Their para-verbal behaviour includes aggressive communication directed at the celebrity, while they use playful, informal language within their community to reinforce their shared identity. Anti-fans feel driven to "speak out" against the celebrity by sharing information, debunking rumours, and engaging in discussions that promote collective negativity. While these actions strengthen their community, they

Response	**Process (Schramm & Hartmann, 2008)**	**Item example (Schramm & Hartmann, 2008)**	**In relation to media figure**	**In relation to fellow anti-fans**
Cognitive				
	Attention allocation	“I carefully followed the behaviour of PERSONA”	Hate-consuming	Shared identity construction
	Comprehension of persona’s action and situation	“I hardly thought about why PERSONA did certain things s/he did” (inverted)	Echo chambers	Shared identity construction
	Evaluations of persona and persona’s actions	“I became aware of aspects of PERSONA that I really liked or disliked”	Shown antipathy by hate-consuming	Shown sympathy by shared identity
	Anticipatory observation	“I kept asking myself how things would evolve around PERSONA”	Shown antipathy by conspiracy	–
	Construction of relations between persona and self	“Occasionally, I wondered if PERSONA was similar to me or not”	–	Shared identity construction
Affective				
	Sympathy/antipathy	“Sometimes I really loved PERSONA for what s/he did”	Antipathy	Sympathy
	Empathy/counter empathy	“If PERSONA felt bad, I felt bad as well; if PERSONA	Counter empathy in the form of antipathy	Empathy

		felt good, I felt good as well”		
Behavioural				
	(Para-)verbal behaviour	“Occasionally, I said something to PERSONA on impulse”	Snarky expressions through aggressive behaviour	Snarky expressions through shared identity and fellowship
	Behavioural intentions	“Sometimes I felt like speaking out on PERSONA”	Collective discursive practices through antipathy	Collective discursive practices through shared identity

Table 2. PSI-Process model based on Schramm & Hartmann (2008), with key findings from framing analysis.

can also lead to toxic practices, highlighting the complex dynamics of anti-fan behaviour.

In essence, the analysis of parasocial interactions within the anti-fan community on social media reveals a notable inversion of traditional models proposed by Horton & Wohl (1956). While conventional parasocial interactions emphasise positive connections between fans and media figures, anti-fans exhibit negative parasocial relationships characterised by strong antipathy toward celebrities, as exemplified by Taylor Swift in this study. These one-sided interactions focus on criticism rather than admiration, showcasing how social media facilitates communities based on shared negative sentiments. This dynamic foster collective identities among anti-fans, uniting them through their loathing of Swift, which transforms individual grievances into a shared sentiment. Regular group activities, such as sharing memes and engaging in discussions, validate and amplify their negative feelings, reinforcing a cohesive group dynamic. Ultimately, this interconnectedness driven by mutual antipathy highlights the need to expand the understanding of parasocial interactions to include diverse emotional engagements, even those rooted in negativity and toxicity.

Conclusion

This study aimed to enhance the understanding of online fan culture by examining anti-fan communities, particularly how negativity circulates within these groups. It reveals that anti-fans use social media to construct and amplify toxic sentiments, forming a unique form of parasocial interaction characterised by negativity. This shows how anti-fans have developed a distorted intimacy with the media figure and fellow community members. Furthermore, the study introduces the concept of *hate-consuming*, where anti-fans actively engage with and consume content about the media figure they dislike, often using humour and memes to express their discontent, although some of these take on e.g. a misogynistic tone, highlighting the potential dangers of online negativity as it can devolve into harassment. The findings highlight the dangers of negativity in online fan culture, demonstrating how echo chambers amplify toxic sentiments and polarise communities. The research stresses the importance of understanding the complex relationships between community identity, shared disdain and parasocial interactions in the digital age. As anti-fan groups evolve, they pose significant challenges for researchers, particularly in distinguishing genuine sentiments from performative behaviours fuelled by the need for social acceptance. Overall, this study enriches the dialogue around anti-fandom, shedding light on the intricate dynamics that define contemporary online interactions.

Reference List

Adegbuyi, F. (2021a, August 31). *The Anti-Fan Phenomenon*. Retrieved May 7 2024, from Every: http://every.to/cybernaut/the-anti-fan-phenomenon

Adegbuyi, F. (2021b, July 14). *The Blurred Lines of Parasocial Relationships*. Retrieved May 8 2024, from Every: https://every.to/cybernaut/

the-blurred-lines-of-parasocial-relationships

Altheide, D. L. (2016). Media Logic. In G. Mazzoleni, *The International Encyclopedia of Political Communication.* Wiley.

Bacon-Smith, C. (1992). *Enterprising Women: Television fandom and the creation of popular myth.* University of Pennsylvania Press.

butijustkeeponla. (2024, March 19). I cant stand her but when I say so, turns out im "misogynistic". Let's get something clear, I dont hate [Comment on Reddit post "*Why I Hate Taylor Swift*"]. Reddit. https://www.reddit.com/r/TaylorSwiftHateClub/comments/14k2wyj/comment/kvl0p2y/

Brennen, B. (2017). *Qualitative research methods for media studies.* New York: Routledge.

Chin, C.-Y., & Huang, W.-Y. (2023). Discovering fans and anti-fans among social media users based on their emotional reactions and comments. *Journal of Information Science*, pp. 1–13.

Click, M. (2019). *Anti-Fandom: Dislike and Hate in the Digital Age.* (M. Click, Ed.) New York: New York University Press.

Couldry, N. (2012). *Media, Society, World: Social Theory and Digital Media Practice.* Polity Press.

Dowell, M. (2024, February 7). *In Every Fandom, 'Anti-Fans' Dominate the Algorithm*. Retrieved May 7 2024, from Medium: https://megdowell.medium.com/in-every-fandom-anti-fans-dominate-the-algorithm-2cc93fe81ab7

Entman, R. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication, 43*(4), pp. 51-58.

Fiske, J. (1986). Television: Polysemy and popularity. *Critival Studies in Mass Communication, 3*(4), pp. 391-408.

Fiske, J. (1989/2011). *Understanding Popular Culture.* Routledge.

Fiske, J. (1992). The Cultural Economy of Fandom. In L. A. Lewis, *The Adoring Audience: Fandom culture and popular media* (pp. 30–49). London & New York: Routledge.

Galuszka, P. (2015). New Economy of Fandom. *Popular Music and Society, 38*(1), pp. 25–43.

Gilbert, A. (2019). Hatewatch with Me: Anti-fandom as Social Performance. In M. Click, *Anti-Fandom: Dislike and Hate in the Digital Age* (pp. 62-80). New York: New York University Press.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience.* Boston: Northeastern University Press.

Graham, P. (2020, January). *Haters*. Retrieved March 8 2024, from Paul Graham: https://paulgraham.com/fh.html

Gray, J. (2003). New Audiences, New Textualities: Anti-fans and non-fans. *International Journal of Cultural Studies, 6*(1), pp. 64–81.

Gray, J. (2019). How Do I Dislike Thee? Let Me Count The Ways. In M. Click, *Anti-Fandom: Dislike and Hate in the Digital Age* (pp. 25–41). New York: New York Univeristy Press.

Hartmann, T. (2008). Parasocial interactions and paracommunication with new media characters. In E. A. Konjin, S. Utz, M. Tanis, & S. B. Barnes, *Mediated interpersonal communication.* Routledge.

Hartmann, T. (2016). Parasocial Interaction, Parasocial Relationships, and Well-Being. In L. Reinecke, & M. B. Oliver, *The Routledge handbook of media use and well-being: International perspectives on theory and research on positive media effects.* Routledge.

Hedbom, P. (2024, May 19). *Taylor har väldigt snälla fans*. Retrieved May 23 2024, from Aftonbladet: https://www.aftonbladet.se/nojesbladet/a/dROKmj/taylor-har-valdigt-snalla-fans

Helmore, E. (2024, February 3). *'Being mean will only rally her fans': Taylor Swift is winning whether she backs Biden or thumps Trump*. Retrieved May 8 2024, from The Guardian: https://www.theguardian.com/music/2024/feb/03/taylor-swift-biden-trump-presidential-race

Horton, D., & Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry, 19*(3), pp. 215-229.

Irfan, A., & Atukunda, C. (2024, March 5). *I don't want to be a Taylor Swift fan anymore*. Retrieved May 8 2024, from Shado Magazine: https://shado-mag.com/opinion/i-dont-want-to-be-a-taylor-swift-fan-anymore/

Jenkins, H. (1992). *Textual Poachers: Television Fans and Participatory Culture.* New York: Routledge.

Jenkins, H. (2006). *Fans, bloggers and gamers: Exploring participatory culture.* New York & London: New York University Press.

Jenkins, H. (2008). *Convergence Culture: Where old and new media collide.* New York: New York University Press.

Jenkins, H., Ito, M., & boyd, d. (2016). *Participatory Culture in a Networked Era: A Conversation on Youth, Learning and Politics.* Polity Press.

Jensen, K. B. (2002). *A Handbook of Media and Communication Research: Qualitative and quantitative methodologies.* (K. B. Jensen, Ed.) London & New York: Routledge.

Jenson, J. (1992). Fandom as Pathology: The Consequences of Characterization. In L. A. Lewis, *The Adoring audience: fan culture and popular media.* London & New York: Routledge.

Lansky, S. (2023, December 6). *Person of The Year 2023: Taylor Swift*. Retrieved May 8 2024, from TIME Magazine: https://time.com/6342806/person-of-the-year-2023-taylor-swift/

Mardon, R., Cocker, H., & Daunt, K. (2023). When parasocial relationships turn sour: social media influencers, eroded and exploitative intimacies, and anti-fan communities. *Journal of Marketing Management, 39*(11-12), pp. 1132–1162.

Persson, A. (2018). *Framing Social Interaction: Continuities and Cracks in Goffman's Frame Analysis.* London & New York: Routledge.

Recuero, R. (2024). The Platformization of Violence: Toward a Concept of Discursive Toxicity on Social Media. *Social Media + Society, 10*(1).

Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales: A new measure to assess the intensity and breadth pf parasocial processes. *Communications, 33*(4), pp. 385–401.

Snow, D. A., Rochford, E. B., Worden, S. K., & Benford, R. D. (1986). Frame Alignment Processes, Micromobilization, and Movement Participation. *American Sociological Review, 51*(4), pp. 464–481.

Stark, D. (2024, February 7). *The Taylor Swift hate has gone too far*. Retrieved May 8 2024, from The Daily Campus: https://dailycampus.com/2024/02/07/the-taylor-swift-hate-has-gone-too-far/

Stever, G. S., & Lawson, K. (2013). Twitter as a way or celebrities to communicate with fans: Implications for the study of parasocial interaction. *North American Journal of Psychology, 15*(2), pp. 339–354.

van Dijck, J., & Poell, T. (2013). Understanding Social Media Logic. *Media and Communication, 1*(1), pp. 2–14.

van Dijk, T. A. (2023). Analyzing frame analysis: A critical review of framing studies in social movement research. *Discourse Studies, 25*(2), pp. 153–178.

Van Gorp, B. (2007). The Constructionist Apporach to Framing: Bringing Culture Back In. *Journal of Communication, 57*, pp. 60–78.

Vitak, J. (2016). Facebook as a research tool in the social and computer sciences. In L. Sloan, & A. Quan-Haase, *The SAGE Handbook of Social Media Research Methods* (pp. 627-642). SAGE Publications.

Yadav, N. (2023, October 31). *Is Taylor Swift Impossible to Critique?* Retrieved May 8 2024, from The Swaddle: https://www.theswaddle.com/is-taylor-swift-impossible-to-critique

Yuksel, M., & Labrecque, L. (2016). "Digital buddies": parasocial interactions in social media. *Journal of Research in Interactive Marketing, 10*(4).

Authors

Juliana Frias Lisboa has over 17 years of experience in media and technology, including a decade at a global tech company. Having lived in Brazil, Italy, Singapore, and now Sweden, she has led digital transformation and social impact initiatives at all scales, from global to local. With a deep passion for equity and education, she continues to explore the evolving intersection of digitalization and social change.

Lisa Kinnunen holds a Bachelor's degree in International Marketing from Mälardalen University as well as a Master's degree in Media, Communication and Cultural Analysis from Södertörn University. Her interests focus on pop culture, particularly its presence on social media, as well as the dynamics of social media interactions, engagement, and content creation.

Nataliia Murashko holds a Master of Arts in Media, Communication, and Cultural Analysis from Södertörn University, Sweden. She is an interpreter and translator, with an interest in media, technology, and diasporic communities. Currently pursuing a PhD at Södertörn University.

This volume contains adaptions of three excellent master theses that were written and defended within the international master's programme media, communication, and cultural analysis at Södertörn University in 2024. Running since 2009, the programme has more than 100 alumni who are employed in the media, academia and education. In 2020, the programme coordinator together with the programme and the department councils, chose to distinguish the best theses in a printed volume. This is the fifth in the series.

The contributions in this volume cover three different topics: how gender influences the representation of influential people in AI; how Ukrainian women from the European diaspora interact with ChatGPT about the Russian–Ukrainian war; and the anti-fandom culture surrounding Taylor Swift on social media. Although stretching across three such different topics, the chapters share an interest in on how media can be understood in relation to gender. In two of the chapters, gendering is explicitly made visible as an ongoing process, while in the study on women interacting with ChatGPT it is done indirectly. However, common for the three chapters is that they fruitfully re-center our attention on how gender frames our everyday digital practices and discourses.

SÖDERTÖRN UNIVERSITY | STOCKHOLM
sh.se