

**Title: Gendered Media Narratives: A Computational Analysis of Media  
and Public Discourse on Male and Female Music Celebrities**

**Candidate Number: 202146996**

**Seminar Group Number: 5**

**Word Count: 2998 words**

## **Literature Review**

### ***Gendered Media Representation***

The disparity in how successful men and women are represented in the media has been extensively documented across politics, sports, and entertainment (Yingnan, 2023). Despite the growing success of women across industries, they continue to receive disproportionately less media coverage than men (Shor, 2019). This gap persists even after accounting for occupational differences such as fame and achievements, as well as media-related inequalities like newsmaker biases and unequal access to reporters (Wasburn, 2011). When women do receive coverage, studies show that it disproportionately emphasizes their personal lives, physical appearance, and relationships rather than their professional accomplishments (Ridgeway, 2013). In sports, media coverage remains heavily skewed toward men, with between 90% and 95% of articles focusing on male athletes or men's sports (Cooky, 2013). A similar pattern emerges in politics, where female candidates receive more coverage on topics like family welfare and public safety, whereas their male counterparts are more frequently questioned on economic and national security issues (Goodyear, 2013). These findings indicate that media representation is not merely a reflection of public interest but is shaped by persistent gender biases that influence the quantity of coverage.

Media narratives around male and female celebrities differ not only in volume but also in tone and content. Studies on gender portrayal in the entertainment industry have found that female musicians are more likely to be sexualized and objectified than their male counterparts (Panuzzo, 2014). Furthermore, the media often justifies and even glorifies the self-destructive behaviours of male artists while portraying women with similar behaviours as problematic or in need of intervention (Castagnetti, 2013). This disparity aligns with broader gender stereotypes in media, which associate men with ambition, strength, and independence, while women are often framed as emotional, passive, and reliant on relationships for their success (Eagly & Karau, 2002).

### ***Role Congruity Theory: Explaining Gendered Media Bias***

Role Congruity Theory posits that when a social group's perceived traits conflict with the expectations of a particular role, they are judged more harshly (Eagly & Karau, 2002). This framework helps explain why successful women in male-dominated fields receive different media treatment than their male counterparts. Women are often perceived as violating traditional gender norms when they exhibit ambition or authority, leading to negative evaluations in the media (Carmen, 2024). Journalists may unconsciously adhere to these stereotypes, resulting in more critical or personal coverage of female celebrities (Ritter, 2004). Additionally, studies suggest that non-stereotypical behaviour is exaggerated in media portrayals, further reinforcing gender biases (Ramasubramaniam, 2014). Supporting this,

research highlights that female role models portrayed as highly successful in male-dominated fields can trigger upward comparison threats, especially when their success is perceived as unattainable. Such portrayals often lead to self-deflation rather than self-inspiration among women, ultimately reducing interest in counter-stereotypical roles and careers (Luong et al., 2020)

### ***The Media Depiction-Public Perception Feedback Loop***

This undeniable gender reporting disparity reflects the complex interplay between societal expectations and media influence. On one hand, media coverage plays a crucial role in shaping public attitudes toward celebrities, often reinforcing existing gender stereotypes (Opplinger & Brian, 2007). Exposure to stereotypical portrayals of women in the media can influence audiences to adopt similar attitudes, further entrenching gender norms. On the other hand, the advent of social media has introduced a more interactive dynamic, where audience reactions and online discourse can shape media narratives in return (Moscovici, 2020). This also includes users being able to differentiate popular opinions (upvoted) from less popular ones. The extent to which public discourse challenges or reinforces media portrayals remains relatively unexplored, particularly in digital spaces like Reddit where users actively engage with news content.

### ***Gendered Reporting in Musical Journalism***

This study focuses on music journalism as its media domain, as it presents a unique intersection of fact-driven reporting and opinion-based critique, offering both explicit and implicit insights into gender representation in popular culture. Despite long-standing perceptions of the music industry as a “boy’s club” (UK Parliament, 2023) and a predominantly ‘masculine domain’ (Kruse, 2022)—particularly in genres like rock, metal, hip-hop, and country—female musicians have seen unprecedented success in recent years. In 2023, nearly half (48.5%) of the Top 10 streamed songs in the UK were by female artists, reflecting their growing influence and visibility (The Guardian, 2024). This increasing success makes female musicians a subject of greater public discourse, intensifying media coverage and scrutiny. Moreover, musical artists often foster strong parasocial relationships with their audiences, with fans closely associating aspects of their self-identity with their idols (Wang, 2023). As a result, the way female musicians are framed in media narratives carries heightened significance, potentially shaping public perceptions more directly than in other industries (Lieb, 2018). Additionally, given that music journalists frequently emphasize female artists’ youth, appearance, and personal lives over their professional achievements (Johnston, 2012), understanding how these biases impact public discourse is crucial to evaluating the broader implications of media influence on gender perception and cultural norms.

### ***Computational Approaches to Gender Representation Analysis***

Traditional content analysis has been the dominant method for studying gender representation in media. However, manual coding is time-intensive and limits the scale of analysis (Baker, 2014). Recent advancements in computational methods, such as natural language processing and machine learning, allow for large-scale analysis of text data, providing a more comprehensive understanding of media biases. Unsupervised techniques like topic modelling can identify emergent themes in media discourse, revealing patterns of gendered language use (Zotos, 2014). Given the dynamic nature of media narratives, these methods offer a valuable approach to studying how gendered portrayals evolve over time and across different media platforms.

### ***Relevance and Contributions of this Study***

Despite extensive research on gender representation in media, few studies have leveraged computational methods to analyse these issues at scale. Most existing work focuses on text classification methods to quantify themes into static categories, case-study approaches or qualitative insights, limiting generalizability (Shor, 2019). Additionally, prior research almost always assumes a one-way influence of media on public perception, overlooking the evolving two-way interaction facilitated by social media. This study aims to fill these gaps by employing large-scale text analysis to examine gendered narratives in music journalism and how these are received and reflected back in online discussion platforms like Reddit. This research also employs unsupervised methods to identify patterns in large scale data and dynamically create categories, staying relevant among fast swooping social changes (Zotos, 2014). By integrating topic modelling with discourse analysis, this research provides a more nuanced understanding of the interplay between media portrayal and public discourse.

### **Research Questions**

How do media portrayals of successful female and male celebrities in the music industry differ?

- Are there differences in the language, sentiment, and framing of media articles about female artists as compared to their male counterparts?
- To what extent do public discussions on Reddit reflect, adopt, or resist the linguistic patterns and narratives presented in media coverage of music celebrities?
- Does media framing shape public perception, or do public attitudes influence media portrayals?

### **Research Design**

This research will undertake an inductive bottom-up computational approach, as it seeks to uncover patterns and relationships towards analysing gender bias in the media and its relation to public perception of female celebrities in the music industry. Using topic modelling inductively via

unsupervised learning methods allows for the theory to emerge inductively from patterns within the corpus, eliminating biases enforced by the researcher's prior conceptualizations and remaining relevant to the given social context (Krippendorff, 2004). Additionally, since this study aims to understand the directionality between media portrayals and their influence on public perceptions, this approach does not impose prior conceptualisations, given the complexity and evolving nature of social media.

This study also employs a mixed methods approach by integrating quantitative text analysis methods such as topic modelling, sentiment analysis, and linguistic framing with qualitative discourse analysis in the form of manually coding user discussions on Reddit and interpreting obtained themes in depth. While the quantitative text analysis provides a systematic, high-level, generalisable comparison of how male and female artists are portrayed in the media and whether narrative frames differ by gender, the qualitative component of this research provides greater granularity into these findings by further evaluating public interaction with media. This combination allows for both a computational analysis of media sentiment and an exploration of public discourse, enabling a comparison between the two.

Reddit has been chosen as a platform to analyse public discourse since the site provides a format in which information is voted on by registered users rather than curated and organized by a centralized media organization. This has created the widely held perception of Reddit as a news source that is truly democratic in its mediation (Zhu, 2013), making it ideal to scrape unfiltered opinions. Media sources will be scraped only from publicly available newspapers, and the limits of the Reddit API will be adhered to with clearly established endpoints and transparency in data sourcing.

It is also important to consider that gender bias is inherently embedded in multiple parts of text analysis algorithms, including the training data, resources, pretrained models (e.g., word embeddings), and algorithms themselves (Zhao et al., 2018). These biases could amplify existing gender stereotypes within the data for example, word embeddings like BERT may associate terms such as "successful" more strongly with male celebrities and "beautiful" with female celebrities, which could skew the results. To mitigate these risks, this study will implement debiasing techniques, such as fine-tuning pretrained models on gender-balanced corpuses based on the research by (Kaneko, 2019). Additionally, findings from computational analysis will be cross validated with qualitative manual coding of a subset of data to ensure accuracy and reduce over-reliance on potentially biased models.

This study accounts for the researcher's role in biasing results by critically examining how researcher gender, positionality and assumptions may influence data selection, interpretation, and analysis. Efforts have been made to minimize these biases through transparent methodological choices and the adoption of computational methods such as topic modelling, which are less directly influenced by the researcher's preconceptions.

## **Data Collection Strategies**

The selection sample strategy for this study focuses on identifying the most successful figures in the music industry currently. To minimise the confounding effects of structural inequalities, data will be collected on men and women in equivalent roles with comparable levels of achievement. To achieve an unbiased sample, 100 top male and female musical artists will be chosen based on their inclusion in the 'Billboard Top 100 artists of 2024'. This ranking is compiled through collating metrics on album sales (physical and digital), airplay and streaming numbers (Billboard, 2023), ensuring that the selected individuals are not only currently relevant but also equitably comparable across genders.

Established US and UK music journalism publications will be selected based on their popularity with readers, their focus on listeners rather than musicians, and their representation of diverse genres, audiences, journalists, ownership, and artists, drawing from previous studies in the field (Whipple, 2021) (**Appendix 1**). News articles from these major media outlets mentioning the selected celebrities will then be gathered through web scraping. Additionally, an auto-scraper will be developed to input collected headlines and articles into Reddit, identifying and extracting discussion threads where available. The Reddit API will then be used to retrieve posts and comments discussing these articles and headlines. Media sentiment will be operationalised by aggregating polarity scores and word frequencies, while gender framing and public reactions will be examined to assess the narratives shaping celebrity portrayals.

The articles collected will cover four years of newspaper coverage between 2021 and 2024. This date range will be selected as these years have seen the re-advent of high music sales and greater streaming revenue driven by a surge in female artists (IFPI, 2024). This boom in the music industry presents an opportunity to study the improved representation of women in media and its effect on a greater interest in female artists.

## **Data Analysis Strategies**

Before large-scale data can be used for effective analysis and training AI algorithms, it needs to be pre-processed into measurable components. Thus, the obtained corpuses of text will be pre-processed by stemming, lemmatizing, and tokenizing for easier analysis by helping map multiple words to a common root word.

To compare the portrayal of male and female artists in media coverage, LDA (Latent Dirichlet Allocation), a type of topic modelling will be carried out as it automatically identifies topics that best

describes the content of a text (Kozlowski, 2022). This technique models each article as a distribution over topics, and each topic as a distribution over words, computing the probability distribution for each topic on each article. Eventually, each topic can be represented by a list of its most probable words. The topic distribution will be compared across gender groups using word distributions and topic prevalence scores. This will help identify whether coverage of women disproportionately focuses on their personal lives, relationships, or controversies while coverage of men emphasizes achievements and professional success. The topic distribution will be compared across gender groups using word distributions and topic prevalence scores. Expected results can be found in **Appendix 2**.

To further assess sentiment differences, pre-trained models such as BERT-based classifiers will be used to determine whether articles frame male and female artists in a positive, neutral, or negative light since they are generally the best at capturing contextual nuances and providing detailed sentiment classifications (Vonitsanos, 2020). This deep learning model will process words in relation to surrounding text for contextual understanding and use this to systematically analyse sentiment trends. By aggregating sentiment scores, this study will assess whether women receive more negative sentiment in media coverage compared to men. Additionally, dictionary-based methods will be employed using a curated set of gendered descriptors (e.g., “pretty,” “scandalous” for women vs. “genius,” “legendary” for men) to quantify linguistic framing differences.

To explore whether public discourse reinforces or challenges media narratives, sentiment analysis will be conducted to compare public sentiment on Reddit with media sentiment determining whether public discourse mirrors media negativity towards women. Stance detection models trained on labelled stance detection datasets will be used to categorize Reddit comments into agreeing, disagreeing, or neutral stances towards media portrayals, providing insight into whether public sentiment aligns with or contests media narratives. These models predict an author’s viewpoint on a specific issue based on linguistic patterns within their text and have been commonly used in fake news detection (Sobhani, 2017). Moreover, semantic similarity analysis using BERT-based embeddings and cosine similarity metrics will be applied to measure the linguistic resemblance between media articles and Reddit discussions. This technique quantifies textual similarity by representing texts as vector embeddings and computing their cosine similarity score, where high similarity suggests that public discourse mirrors media narratives, while low similarity indicates reinterpretation or rejection of the viewpoint.

To determine whether media sentiment influences public sentiment or vice versa, Granger causality tests will be conducted. The Granger causality test will assess whether the yearly frequency of prejudice-denoting words in media coverage can predict their future occurrence in public discourse, or vice versa (Rozado, 2023). This will be achieved using time-series regression models, which will

establish whether shifts in sentiment within media articles precede changes in public sentiment or whether public discussions influence subsequent media portrayals. The results will provide insights into the directionality of influence between media narratives and public perception.

### **Potential Impact and Relevance:**

This study's theoretical contribution lies in extending research on media bias through computational text analysis techniques, providing a large-scale, data driven approach to understanding gender bias in the field of music journalism. Research indicates that exposure to stereotypical, objectifying, and sexualizing media representations of women can strengthen gender stereotypes, sexism, and detrimental effects on well-being (Santonniccolo, 2023). By examining how media narratives construct gendered portrayals of artists, this study can, therefore, have big consequences for youth mental health.

Studies have shown that when journalists actively monitor their sources and adopt inclusive reporting practices, achieving gender balance in media coverage becomes a tangible goal (Mamasodikova, 2022). Therefore, a more nuanced understanding of these disparities can support journalists and policy makers in their efforts to achieve gender parity in their sources. Additionally, unexpected findings—such as public discourse contradicting media narratives—could suggest that audiences actively resist gender biases in media, highlighting potential shifts in societal attitudes, which can be expanded on in further research.

### **Limitations and Future Research:**

One of the key challenges in this research is the use of topic modelling as an unsupervised machine learning technique, as model validation is challenging (Mohseni, 2021). Since topic assignment relies on human judgment, it is inherently subjective and prone to lower accuracy. Additionally, the number of topics extracted is determined subjectively, which can affect the interpretability and validity of results. Future research could incorporate log-likelihood measures and topic coherence metrics to improve model evaluation and enhance robustness (Hosseiny, 2023). Similarly, sentiment analysis presents its own set of challenges as machine-based techniques struggle to detect nuanced linguistic features such as sarcasm, irony, and contextual meaning, which are essential in the context of social media data, like Reddit threads (Sharma, 2024).

The study also faces data limitations related to its reliance on Reddit as a source of public discourse. Reddit is known to be predominantly left leaning, which may limit its ability to accurately



reflect broader public opinion. Moreover, artist-specific subreddits tend to be dominated by fan communities, which may discourage critical dialogue and bias discussions toward positive sentiment (Betteridge, 2016). Moreover, Reddit is a male-dominated platform, and prior research suggests that it fosters a community culture shaped by hegemonic masculinity, where male experiences are universalized while women's perspectives are often ignored or stigmatized (Dewey, 2015). As a result, insights derived from Reddit may not be fully generalizable to the wider public. Future research could diversify data sources by analysing Instagram comments, Twitter discussions, or public opinion surveys to capture a broader and more balanced spectrum of perspectives.

Despite these limitations, this methodology represents a promising tool for capturing gender bias in the music industry. Future studies could enhance its intersectionality by incorporating race, ethnicity, or geographical variations in media representations. Additionally, longitudinal sentiment analysis can be carried out to monitor shifts in public discourse over time, particularly in response to significant news scandals or award shows. Expanding the research scope to other industries, such as film, sports, or politics, could provide further insights into how media portrayal biases extend beyond the music industry.

## **References:**

- Asr, F. T., Mazraeh, M., Lopes, A., Gautam, V., Gonzales, J., Rao, P., & Taboada, M. (2021). The gender gap tracker: Using natural language processing to measure gender bias in media. *PloS one*, 16(1), e0245533.
- Baker, P. (2014). Using corpora to analyze gender. *Using Corpora to Analyze Gender*, 1-224.
- Brugnoli, E., Simone, R., & Delmastro, M. (2024). Combining Natural Language Processing and Statistical Methods to Assess Gender Gaps in the Mediated Personalization of Politics. *Social Science Computer Review*.  
<https://doi.org/10.1177/08944393241269097>
- Calasanti, A., & Gerrits, B. (2021). “You’re not my nanny!” Responses to racialized women leaders during COVID-19. *Politics, Groups, and Identities*, 1–18.  
<https://doi.org/10.1080/21565503.2021.1942084>
- Castagnetti, Carolina, and Luisa Rosti. "Unfair tournaments: gender stereotyping and wage discrimination among Italian graduates." *Gender & Society* 27.5 (2013): 630-658.
- Cooky, C., Messner, M. A., & Hextrum, R. H. (2013). Women play sport, but not on TV: A longitudinal study of televised news media. *Communication & Sport*, 1(3), 203-230.
- Cousineau, L. S. (2021). Masculinity in two reddit men’s communities: Feminism is foul, privilege is precarious, and being beta is bad.
- del Carmen Triana, M., Song, R., Um, C. T., & Huang, L. (2024). Stereotypical perception in management: A review and expansion of role congruity theory. *Journal of Management*, 50(1), 188-215.
- Deligianni, A., & Horne, Z. (2023). Analysing hate speech towards women on Reddit.
- Betteridge, J. G. (2016). Oppression in the Reddit Hivemind: Tracing Patterns of Misogyny in Electronic Public Spaces.
- Goodyear-Grant, E. (2013). Gendered news: Media coverage and electoral politics in Canada. UBC Press.

- Dolan, K. (2009). The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior*, 32(1), 69–88. <https://doi.org/10.1007/s11109-009-9090-4>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3), 573.
- Elsesser, K. M., & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations*, 64(12), 1555–1578.
- Farrell, T., Fernandez, M., Novotny, J., & Alani, H. (2019). Exploring Misogyny across the Manosphere in Reddit. *Proceedings of the 10th ACM Conference on Web Science - WebSci '19*, 87–96. <https://doi.org/10.1145/3292522.3326045>
- Furl, K. (2022). Denigrating Women, Venerating “Chad”: Ingroup and Outgroup Evaluations among Male Supremacists on Reddit. *Social Psychology Quarterly*, 85(3), 279–299. <https://doi.org/10.1177/01902725221090907>
- Hopkins, D. J., Kim, E., & Kim, S. (2017). Does newspaper coverage influence or reflect public perceptions of the economy? *Research & Politics*, 4(4), 205316801773790. <https://doi.org/10.1177/2053168017737900>
- Hosseiny Marani, A., & Baumer, E. P. (2023). A review of stability in topic modeling: Metrics for assessing and techniques for improving stability. *ACM Computing Surveys*, 56(5), 1-32.
- Jacobi, C., Van Atteveldt, W., & Welbers, K. (2018). Quantitative analysis of large amounts of journalistic texts using topic modelling. In *Rethinking Research Methods in an Age of Digital Journalism* (pp. 89-106). Routledge.

- Jaworska, S., & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society*, 23(4), 401–431. <https://doi.org/10.1177/0957926512441113>
- Johnston M (2012) How not to write about female musicians: A handy guide. The Village Voice, 10 February. Available at: <https://www.villagevoice.com/2012/02/10/how-not-to-writeabout-female-musicians-a-handy-guide/> (accessed 27 October 2016).
- Kaneko, M., & Bollegala, D. (2019). Gender-preserving debiasing for pre-trained word embeddings. *arXiv preprint arXiv:1906.00742*.
- Kozlowski, D., Lozano, G., Felcher, C. M., Gonzalez, F., & Altszyler, E. (2022). Large-scale computational content analysis on magazines targeting men and women: the case of Argentina 2008-2018. *Feminist Media Studies*, 23(5), 1–19. <https://doi.org/10.1080/14680777.2022.2047090>
- Krippendorff, K. (2004). Measuring the reliability of qualitative text analysis data. *Quality and quantity*, 38, 787-800.
- Kruse H (2002) Abandoning the absolute: Transcendence and gender in popular music discourse. In: Jones S (ed.) Pop Music and the Press. Philadelphia, PA: Temple University Press, pp. 134–155.
- Leavy, S. (2020). Uncovering gender bias in media coverage of politicians with machine learning. *arXiv preprint arXiv:2005.07734*.
- Lieb, K. (2018). *Gender, branding, and the modern music industry: The social construction of female popular music stars*. Routledge.
- Luong, K. T., Knobloch-Westerwick, S., & Niewiesk, S. (2019). Superstars within reach: The role of perceived attainability and role congruity in media role models on women's social comparisons. *Communication Monographs*, 1–21. <https://doi.org/10.1080/03637751.2019.1622143>

- Mamasodikova, S., & Ergasheva, G. (2022, April). THE GENDER GAP TRACKER: USING LANGUAGE PROCESSING TO GENDER BIAS IN MEDIA. In *E Conference Zone* (pp. 217-220).
- Mapes J (2016) Gwen Stefani: This is what the truth feels like. Pitchfork.com, 21 March. Available at: <https://pitchfork.com/reviews/albums/21647-this-is-what-the-truth-feels-like/> (accessed 15 October 2018).
- Mohseni, S., Block, J. E., & Ragan, E. (2021, April). Quantitative evaluation of machine learning explanations: A human-grounded benchmark. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (pp. 22-31).
- Moscovici, S. (2002). De ce o pledoarie pentru studierea reprezentărilor sociale în psihologie?. *Psihologia socială*, (09), 6-16.
- Oppliger PA and Brian A (2007) Effects of gender stereotyping on socialization. In: Preiss RW, Gayle BM, Burrell N, et al. (eds) *Mass Media Effects Research: Advances through MetaAnalysis*. New York: Routledge, pp.199–214.
- Päkkilä, E. (2017). " *Fallible, sentimental and ruled by contradictory impulses*": representations of female pop artists in *The Guardian's* online news articles
- Panuzzo B (2014) Covering performers, discovering femininities: US hip-hop journalism and female artists. *IASPM Journal* 4(2): 43–54.
- Ramasubramanian, S., & Murphy, C. J. (2014). Experimental studies of media stereotyping effects. In *Laboratory experiments in the social sciences* (pp. 385-402). Academic Press.
- Rao, P., & Taboada, M. (2021). Gender bias in the news: A scalable topic modelling and visualization framework. *Frontiers in Artificial Intelligence*, 4, 664737.
- Ridgeway, C. L., & Kricheli-Katz, T. (2013). Intersecting cultural beliefs in social relations: Gender, race, and class binds and freedoms. *Gender & Society*, 27(3), 294-318.

- Ritter, B. A., & Yoder, J. D. (2004). Gender differences in leader emergence persist even for dominant women: An updated confirmation of role congruity theory. *Psychology of Women Quarterly*, 28(3), 187-193.
- Roels, G., & Su, X. (2014). Optimal design of social comparison effects: Setting reference groups and reference points. *Management Science*, 60(3), 606-627.
- Rozado, D., Al-Gharbi, M., & Halberstadt, J. (2023). Prevalence of prejudice-denoting words in news media discourse: A chronological analysis. *Social Science computer review*, 41(1), 99-122.
- Rudman, L., & Kilianski, S. (2010). Personality and Social Psychology Bulletin Implicit and Explicit Attitudes Toward Female Authority On behalf of: Society for Personality and Social Psychology can be found at: Personality and Social Psychology Bulletin Additional services and information for. *Pers Soc Psychol Bull*, 26(11).  
<https://doi.org/10.1177/0146167200263001>
- Santoniccolo, F., Trombetta, T., Paradiso, M. N., & Rollè, L. (2023). Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization. *International journal of environmental research and public health*, 20(10), 5770.
- Sharma, N. A., Ali, A. S., & Kabir, M. A. (2024). A review of sentiment analysis: tasks, applications, and deep learning techniques. *International journal of data science and analytics*, 1-38.
- Shor, E., Van De Rijt, A., & Fotouhi, B. (2019). A large-scale test of gender bias in the media. *Sociological Science*, 6, 526-550.
- Shor, E., van de Rijt, A., & Fotouhi, B. (2019). A Large-Scale Test of Gender Bias in the Media. *Sociological Science*, 6, 526–550. <https://doi.org/10.15195/v6.a20>

- Shor, E., van de Rijt, A., Miltsov, A., Kulkarni, V., & Skiena, S. (2015). A Paper Ceiling: Explaining the Persistent Underrepresentation of Women in Printed News. *American Sociological Review*, 80(5), 960–984. <https://doi.org/10.1177/0003122415596999>
- Smirke, R. (2024, March 21). *IFPI Global Report 2024: Music Revenues Climb 10% to \$28.6 Billion*. Billboard. <https://www.billboard.com/business/business-news/ifpi-global-report-2024-music-business-revenue-market-share-1235637873/>
- Sobhani, P. (2017). *Stance detection and analysis in social media* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Thorpe, V. (2024, February 4). “*They’re really breaking through*”: how women put the music business in a spin. The Guardian; The Guardian. <https://theguardian.com/music/2024/feb/04/theyre-really-breaking-through-how-women-put-the-music-business-in-a-spin>
- Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, 27(4), 401–422. <https://doi.org/10.1177/0957926516634546>
- Van der Pas, D. J., & Aaldering, L. (2020). Gender differences in political media coverage: A meta-analysis. *Journal of Communication*, 70(1), 114-143.
- Vonitsanos, G., Kanavos, A., & Mylonas, P. (n.d.). *Social Media Insights into Climate Change: Sentiment Analysis Using VADER and RoBERTa*. <http://www.image.ece.ntua.gr/papers/1138.pdf>
- Vu, H. T., Lee, T.-T., Duong, H. T., & Barnett, B. (2017). Gendering Leadership in Vietnamese Media: A Role Congruity Study on News Content and Journalists’ Perception of Female and Male Leaders. *Journalism & Mass Communication Quarterly*, 95(3), 565–587. <https://doi.org/10.1177/1077699017714224>

- Wang, S., Kim, M. S., & Kim, S. (2023). The Effects of Idol Fans' Interactions in Online Fan Community Platforms on Their Well-Being and Loyalty.
- Wasburn, P. C., & Wasburn, M. H. (2011). Media coverage of women in politics: The curious case of Sarah Palin. *Media, Culture & Society*, 33(7), 1027-1041.
- Whipple, K., & Coleman, R. (2021). Facing the music: Stereotyping of and by women in US music journalism. *Journalism*, 23(10), 146488492110287.  
<https://doi.org/10.1177/14648849211028770>
- Whipple, K., & Coleman, R. (2022). Facing the music: Stereotyping of and by women in US music journalism. *Journalism*, 23(10), 2060-2078.
- Women and Equalities Committee. (2023). *Misogyny in music - Women and Equalities Committee*. Parliament.uk.  
<https://publications.parliament.uk/pa/cm5804/cmselect/cmwomeq/129/report.html>
- Yingnan, S., & Guoliang, W. (2023). Gender Disparities in Sports Media Coverage: Phenomenon, Analysis, and Transformation. *Frontiers in Sport Research*, 5(9).
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In *Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings 33* (pp. 338-349). Springer Berlin Heidelberg.
- Zhu, J., & Su, Y. (2022). On the Gender Order of Sports: Based on the Embodiment, Representation, and Expression of the Body. *Zhejiang Sports Science*, 44(05), 20-25.
- Zotos, Y. C., & Tsihla, E. (2014). Female stereotypes in print advertising: a retrospective analysis. *Procedia-social and behavioral sciences*, 148, 446-454.



## Appendix:

### Appendix 1: List of Music News Publishers (based on study conducted by Whipple, 2021):

1. Billboard magazine, founded in 1894: Considered one of the most reputable sources of music industry news (Sisario, 2014), it is known for regularly publishing charts that track the songs and albums most played, purchased and downloaded in the US.
2. SPIN magazine, founded in 1985 as an alternative to more establishment music outlets and as a way for readers to learn about music options.
3. MTV News, owned by media conglomerate Viacom and created in 1993 after the music video television channel, which was established in 1981.
4. Pitchfork.com, a digital native publication created in 1995 to focus on independent music, now includes a large variety of genres. It is one of the most popular independent music sites (Itzkoff, 2006).
5. Rolling Stone magazine, founded in 1967, had a circulation of 1.4 million in 2016 (Alliance for Allied Media, 2017). Its annual 'Hot Issue' and '500 Greatest' lists have become classics of the subfield.
6. Alternative Press magazine, AltPress.com, was founded in 1985 to cover punk rock, indie and alternative music, and was included because of its focus on genres that tend to receive less attention.
7. Complex Magazine: focusing on hip-hop and culture
8. New York Times: mainstream news with well-regarded Arts section, receiving two Pulitzer Prizes for music criticism.
9. The Guardian: Founded in 1821, The Guardian is renowned for its in-depth journalism and cultural coverage, including critically acclaimed music reporting and reviews.
10. BBC: Established in 1922, the BBC is a globally trusted source for news and culture, with extensive coverage of music and its societal impact.
11. CNN: Launched in 1980, CNN provides comprehensive coverage of entertainment news, including music industry trends and artist features.

### Appendix 2: Expected Differences in Topic Proportions by Gender Categories

