Understanding and Mitigating Gender Bias in Information Retrieval Systems

Other titles in Foundations and Trends® in Information Retrieval

Mathematical Information Retrieval: Search and Question Answering
Richard Zanibbi, Behrooz Mansouri and Anurag Agarwal
ISBN: 978-1-63828-502-1

Information Discovery in E-commerce
Zhaochun Ren, Xiangnan He, Dawei Yin and Maarten de Rijke
ISBN: 978-1-63828-462-8

Fairness in Search Systems
Yi Fang, Ashudeep Singh and Zhiqiang Tao
ISBN: 978-1-63828-498-7

User Simulation for Evaluating Information Access Systems
Krisztian Balog and ChengXiang Zhai
ISBN: 978-1-63828-378-2

Multi-hop Question Answering
Vaibhav Mavi, Anubhav Jangra and Adam Jatowt
ISBN: 978-1-63828-374-4

Conversational Information Seeking
Hamed Zamani, Johanne R. Trippas, Jeff Dalton and Filip Radlinski
ISBN: 978-1-63828-200-6

Understanding and Mitigating Gender Bias in Information Retrieval Systems

Shirin Seyedsalehi
Toronto Metropolitan University

Amin Bigdeli
University of Waterloo

Negar Arabzadeh
University of Waterloo

Batool AlMousawi
University of Calgary

Zack Marshall
University of Calgary

Morteza Zihayat
Toronto Metropolitan University

Ebrahim Bagheri
University of Toronto

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

S. Seyedsalehi *et al.*. *Understanding and Mitigating Gender Bias in Information Retrieval Systems*. Foundations and Trends® in Information Retrieval, vol. 19, no. 3, pp. 191–364, 2025.

ISBN: 978-1-63828-519-9

Foundations and Trends® in Information Retrieval
Volume 19, Issue 3, 2025

Editorial Scope

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2025, Volume 19, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Understanding and Mitigating Gender Bias in Information Retrieval Systems

Shirin Seyedsalehi[1], Amin Bigdeli[2], Negar Arabzadeh[2], Batool AlMousawi[3], Zack Marshall[3], Morteza Zihayat[1] and Ebrahim Bagheri[4]

[1] *Toronto Metropolitan University, Canada*
[2] *University of Waterloo, Canada*
[3] *University of Calgary, Canada*
[4] *University of Toronto, Canada;*
ebrahim.bagheri@utoronto.ca

ABSTRACT

Gender bias is a pervasive issue that continues to influence various aspects of society, including the outcomes of information retrieval (IR) systems. As these systems become increasingly integral to accessing and navigating the vast amounts of information available today, the need to understand and mitigate gender bias within them is paramount. This monograph provides a comprehensive examination of the origins, manifestations, and consequences of gender bias in IR systems, as well as the current methodologies employed to address these biases.

Theoretical frameworks surrounding gender and its representation in artificial intelligence (AI) systems are explored, particularly focusing on how traditional gender binaries are perpetuated and reinforced through data and algorithmic

Shirin Seyedsalehi, Amin Bigdeli, Negar Arabzadeh, Batool AlMousawi, Zack Marshall, Morteza Zihayat and Ebrahim Bagheri (2025), "Understanding and Mitigating Gender Bias in Information Retrieval Systems", Foundations and Trends® in Information Retrieval: Vol. 19, No. 3, pp 191–364. DOI: 10.1561/1500000103.

processes. Metrics and methodologies used to identify and measure gender bias within IR systems are then analyzed, offering a detailed evaluation of existing approaches and their limitations.

Subsequent sections address the sources of gender bias, including biased input queries, retrieval methods, and gold standard datasets. Various data-driven and method-level debiasing strategies are presented, including techniques for debiasing neural embeddings and algorithmic approaches aimed at reducing bias in IR system outputs. The monograph concludes with a discussion of the challenges and limitations faced by current debiasing efforts and provides insights into future research directions that could lead to more equitable and inclusive IR systems.

This monograph serves as a valuable resource for researchers, practitioners, and students in the fields of information retrieval, artificial intelligence, and data science, providing the knowledge and tools needed to address gender bias and contribute to the development of fair and unbiased information systems.

1

Introduction

1.1 Information Retrieval (IR) Systems

Information Retrieval (IR) systems are fundamental to the digital era, and crucial for navigating the vast data landscape of today's world. From simple web searches to sophisticated data analytics in corporate environments, IR systems are integral to modern life and provide the tools necessary for personal and professional decision-making. IR systems do not just facilitate over 1.2 trillion searches per year on a platform like Google (Internet Live Stats, 2024) but also significantly impact various sectors such as:

- **Healthcare.** In healthcare, IR systems manage extensive patient records and research databases, enabling medical professionals to access vital information swiftly. For instance, databases like PubMed offer access to medical research, facilitating better patient care and fostering the rapid development of medical knowledge (Medicine, 2024).

- **Finance and Banking.** Financial sectors utilize IR to analyze market trends and monitor transactions. Tools provided by Bloomberg and Reuters help professionals sift through large datasets to find critical information on market developments, economic reports, and investment analytics, supporting quick and informed financial decisions (Bloomberg, 2024; Reuters, 2024).

- **Legal.** IR systems such as LexisNexis and Westlaw are indispensable in the legal arena. They allow legal professionals to efficiently search through vast quantities of legal documents, case law, and statutes, essential for case preparation, conducting due diligence, and ensuring comprehensive legal research (LexisNexis, 2024; Westlaw, 2024).
- **Academic Research.** IR systems are also crucial in academia, where platforms like Google Scholar and JSTOR enable researchers to navigate through countless scholarly articles and publications. This access supports various academic disciplines, enhancing research capabilities and fostering educational advancement (Google, 2024; ITHAKA, 2024).

Such systems have deep impacts on different aspects of society. The **economic implications** of IR systems are vast, influencing sectors from e-commerce to online advertising. They drive consumer behavior, facilitate transactions, and are instrumental in strategic business decisions, impacting billions in daily commerce. **Technological advancements** in IR have paralleled the rapid evolution of computing power and data science methodologies. Today's IR systems employ sophisticated algorithms and machine-learning techniques to improve accuracy and user experience. Furthermore, IR systems profoundly shape societal interactions and access to information, influencing education, politics, and social dynamics. In **education**, IR systems provide students and academics access to a wide array of resources, transforming how knowledge is acquired and shared. The availability of digital libraries and online courses has democratized education, making learning more accessible globally. **Politically**, IR systems play a critical role in shaping public opinion and electoral outcomes by controlling the flow of news and information. Their ability to highlight or suppress information can alter perceptions and influence decisions on a large scale. **Culturally**, IR systems facilitate the global exchange of ideas and values, promoting cross-cultural understanding and cooperation (Taksa and Flomenbaum, 2009). They have become platforms for cultural expression and identity exploration, contributing to the global cultural mosaic.

1.2 Biases and IR Systems

Information Retrieval (IR) systems, while immensely beneficial, are not immune to the influence of biases that can skew results and perpetuate societal inequalities. These biases arise from various sources including data, algorithm design, and human factors involved in the development and maintenance of such systems. Biases in IR systems can have profound implications across multiple sectors by reinforcing stereotypes and exacerbating social prejudices. Below, we explore several high-profile examples that illustrate the detrimental effects of these biases.

- **Employment and Job Recommendation Systems** One notable example involves gender bias in job recommendation algorithms. Studies have shown that certain algorithms tend to favor male candidates over equally qualified female candidates. This reflects and perpetuates existing gender disparities in job markets. For instance, a research conducted by Amazon had to scrap their AI recruiting tool because it showed bias against women. The system learned to penalize resumes that included the word "women's," as in "women's chess club captain," and it downgraded graduates of two all-women's colleges (Dastin, 2018).
- **Credit and Loan Approvals** Biases in IR systems also affect financial decisions like credit scoring and loan approvals. An investigation into Apple Card's algorithm revealed it offered higher credit limits to men than to women under similar financial circumstances. This incident sparked a broader discussion about the transparency and fairness of algorithms in financial services (Nicas, 2019).
- **Healthcare Diagnostics** In healthcare, biases in IR systems can lead to life-threatening consequences. Research has indicated that certain diagnostic algorithms prioritize the care of white

patients over equally sick patients from minority groups due to biases in the training data. For example, a widely used healthcare algorithm was found to be less likely to refer Black patients than white patients for higher-quality care, even when they were equally ill (Obermeyer *et al.*, 2019).

- **Law Enforcement and Judicial Systems** In law enforcement, predictive policing systems have come under scrutiny for perpetuating racial biases. These systems often target minority-heavy areas more aggressively, leading to a disproportionate number of arrests and convictions in these communities. Similarly, algorithms used to predict future criminal behavior for parole decisions have been criticized for being biased against people of color (Angwin *et al.*, 2016).

Tackling biases in IR systems is not only a technological imperative but also a moral obligation. Given the critical role that these systems play in shaping perceptions and decision-making processes in society, ensuring fairness, equity, and justice in digital interactions becomes paramount.

Several studies have explored bias in practical, applied industry contexts, highlighting both challenges and potential solutions. For instance, in Bogen and Rieke (2018), the authors provide recommendations to increase transparency and oversight in hiring technologies to reduce the potential harm these tools can cause. They advocate for independent audits by vendors and employers and suggest that regulators update laws to address the capabilities and risks of modern hiring technologies. The report emphasizes that without intentional intervention, these technologies could reinforce existing inequalities. Nevertheless, it argues that predictive tools also present opportunities to improve diversity if they are actively designed to address historical inequities. This balance between innovation and accountability is crucial as these technologies increasingly influence employment opportunities.

In the realm of recommendation systems, the authors in Wu *et al.* (2021) introduce FairRec, a model designed to reduce bias in news recommendations while maintaining performance levels. Traditional recommendation systems often amplify biases by capturing patterns linked to sensitive attributes like gender. FairRec mitigates this by decomposing user interests into two components: a bias-aware embedding that captures attribute-specific biases and a bias-free embedding focused on neutral interests. The model employs adversarial learning to minimize bias in the bias-free embedding and uses orthogonality regularization to keep the two embeddings distinct. Only the bias-free embedding is used in the final ranking, ensuring recommendations are independent of sensitive attributes.

Lastly, the authors in Binns *et al.* (2018) explored perceptions of justice in algorithmic decision-making. Through lab and online experiments, the study explored how various explanation styles—such as case-based, demographic, input influence, and sensitivity—affect people's sense of fairness, dignity, and accountability in scenarios like loan approvals and insurance pricing. Results indicate that people's perceptions of justice are shaped by their understanding of the decision-making process and whether they view the factors considered as appropriate. However, repeated exposure to a single explanation style led participants to focus more on scenario details than on specific explanation types. This study highlights the complexities involved in designing explanations that foster a sense of fairness and accountability in algorithmic systems, emphasizing that no single explanation style fits all needs and that users may be reluctant to assign justice or moral responsibility to machine-based decisions.

In addition, the importance of addressing biases in IR systems has been significantly recognized by the research community, prompting a vigorous response aimed at understanding and mitigating these biases. This response has been multi-faceted, focusing on various aspects of bias in IR systems—from identifying the sources of biases and understanding how they are injected into the systems, to exploring ways in which these biases are amplified and spread through societal interactions. Researchers have investigated the mechanisms through which biases are introduced into IR systems. This often originates from the data used to train algorithms, where historical inequalities or skewed data representation lead to biased decision-making processes (Barocas and Selbst, 2016; Mehrabi *et al.*, 2021). Studies have shown how machine learning algorithms can inadvertently learn and perpetuate these biases if not properly checked (Zhao *et al.*, 2017). Moreover, the research focuses on how once biases are injected, they can be intensified by the algorithms through their iterative nature. For example, feedback loops where biased outputs are used as new training data can further entrench and exacerbate these biases (Baeza-Yates, 2018). Understanding these dynamics is crucial for developing effective mitigation strategies (Friedman and Nissenbaum, 1996).

A significant portion of recent research has been devoted to developing methodologies to prevent the spread of biases. These include algorithmic fairness approaches, bias audits, and the use of fairness-enhancing interventions in the algorithmic design (Chouldechova, 2017; Holstein *et al.*, 2019). Researchers are exploring both technical solutions, such as the redesign of algorithms, and policy-based approaches, such as regulatory frameworks and transparency guidelines (Barocas *et al.*, 2020; Binns, 2018).

1.3 Section Breakdown

This work aims to contribute significantly to this ongoing discourse by providing a comprehensive overview of how biases in IR systems can be understood and addressed. Each section is dedicated to exploring a different aspect of bias in IR, from theoretical underpinnings to practical applications and case studies, thus offering a holistic view of current strategies and future directions in bias mitigation. The monograph is structured to provide a holistic approach to understanding and mitigating gender bias in Information Retrieval (IR) systems. It is composed of a series of sections that progressively investigate various dimensions of gender bias, ranging from theoretical frameworks to practical debiasing methods.

Section 2: Framing Sex, Gender, and Gender Diversity

Having outlined the biases present in information retrieval (IR) systems, we take the first step toward addressing these issues by looking at how AI systems interpret concepts like sex and gender. This next section explores how these interpretations can often reinforce social biases, helping us build a clear foundation for understanding gender bias in IR.

Section 3: Gendered Information Retrieval Systems: Metrics and Measurements

Metrics and measurements used to identify and quantify gender biases in IR systems are outlined in this section. The section discusses various approaches to assess how these systems handle fairness in algorithmic processing and result ranking.

Section 4: Understanding the Sources of Gender Bias in IR Systems

This section explores the origins of gender biases in IR systems. It analyzes how biases are integrated into algorithms through data training processes and the design of algorithms themselves. The section discusses both inadvertent and systematic insertion of biases during the development phases of IR systems.

Section 5: Data-driven Debiasing Methods

Focusing on practical approaches, this section introduces methods for data-driven bias mitigation. It covers techniques such as data augmentation, modification of training datasets, and algorithmic adjustments aimed at reducing the gender bias inherent in IR systems.

Section 6: Debiasing of Neural Embeddings

Specific techniques for debiasing neural network embeddings are covered. This section offers the details and the technical aspects of neural networks that process, providing insights into how these can be adjusted to mitigate biases.

Section 7: Method-Level Debiasing

This section extends the discussion on bias mitigation by focusing on specific methodologies that can be applied at different levels of IR system development. It includes case studies and examples where these methods have been successfully implemented.

Section 8: Challenges, Limitations, and Future Directions

The concluding section discusses the ongoing challenges in fully addressing gender bias in IR systems, the limitations of current approaches, and the potential future research directions that could lead to more comprehensive solutions.

The structure of this monograph is designed to equip researchers, practitioners, and students with a thorough understanding of the complex nature of gender biases in IR systems and provides a detailed guide on existing strategies to address these biases. Each section builds on the previous one, ensuring a comprehensive learning path for the reader.

References

Abdollahpouri, H., M. Mansoury, R. Burke, B. Mobasher, and E. Malthouse. (2021). "User-centered evaluation of popularity bias in recommender systems". In: *Proceedings of the 29th ACM conference on user modeling, adaptation and personalization*. 119–129.

Abolghasemi, A., L. Azzopardi, A. Askari, M. de Rijke, and S. Verberne. (2024). "Measuring Bias in a Ranked List Using Term-Based Representations". In: *European Conference on Information Retrieval*. Springer. 3–19.

Adam, A. (1998). "Feminist resources". In: *Artificial Knowing: Gender and the Thinking Machine*. New York: Routledge. 11–34.

Adomavicius, G. and Y. Kwon. (2011). "Improving aggregate recommendation diversity using ranking-based techniques". *IEEE Transactions on Knowledge and Data Engineering*. 24(5): 896–911.

Ainsworth, C. (2015). "Sex redefined". *Nature*. 518(7539): 288–291. DOI: 10.1038/518288a.

Albert, K. and M. Delano. (2022). "Sex trouble: Sex/gender slippage, sex confusion, and sex obsession in machine learning using electronic health records". *Patterns*. 3(8): 1–11.

Alesina, A., P. Giuliano, and N. Nunn. (2013). "On the origins of gender roles: Women and the plough". *The Quarterly Journal of Economics*. 128(2): 469–530. DOI: 10.1093/qje/qjt005.

Anderson, S. M. (2020). "Gender matters: The perceived role of gender expression in discrimination against cisgender and transgender LGBQ individuals". *Psychology of Women Quarterly*. 44(3): 323–341. DOI: 10.1177/0361684320929354.

Angwin, J., J. Larson, S. Mattu, and L. Kirchner. (2016). "Machine Bias". *ProPublica*. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Arboleda, V. A., D. E. Sandberg, and E. Vilain. (2014). "DSDs: Genetics, underlying pathologies and psychosexual differentiation". *Nature Reviews Endocrinology*. 10(10): 603–615. DOI: 10.1038/nrendo.2014.130.

De-Arteaga, M., A. Romanov, H. Wallach, J. Chayes, C. Borgs, A. Chouldechova, S. Geyik, K. Kenthapadi, and A. T. Kalai. (2019). "Bias in bios: A case study of semantic representation bias in a high-stakes setting". In: *proceedings of the Conference on Fairness, Accountability, and Transparency*. 120–128.

Ashley, F. (2023). "What Is It like to Have a Gender Identity?" *Mind*. 132(528): 1053–1073. DOI: 10.1093/mind/fzac071.

Atagi, N., N. Sethuraman, and L. B. Smith. (2009). "Conceptualizations of gender in language". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 31. URL: https://escholarship.org/uc/item/69s2385t.

Baeza-Yates, R. (2018). "Bias on the web". *Communications of the ACM*. 61(6): 54–61.

Bannister, J. J., H. Juszczak, J. D. Aponte, D. C. Katz, P. D. Knott, S. M. Weinberg, B. Hallgrímsson, N. D. Forkert, and R. Seth. (2022). "Sex differences in adult facial three-dimensional morphology: application to gender-affirming facial surgery". *Facial Plastic Surgery & Aesthetic Medicine*. 24(S2): S–24. DOI: 10.1089/fpsam.2021.0301.

Barikeri, S., A. Lauscher, I. Vulić, and G. Glavaš. (2021). "RedditBias: A real-world resource for bias evaluation and debiasing of conversational language models". *arXiv preprint arXiv:2106.03521*.

Barocas, S., M. Hardt, and A. Narayanan. (2020). *Fairness and Machine Learning: Limitations and Opportunities*. URL: https://fairmlbook.org.

Barocas, S. and A. D. Selbst. (2016). "Big Data's Disparate Impact". *California Law Review*. 104(3): 671–732. DOI: 10.15779/Z38BG31.

Basta, C., M. R. Costa-Jussa, and N. Casas. (2021). "Extensive study on the underlying gender bias in contextualized word embeddings". *Neural Computing and Applications*. 33(8): 3371–3384.

Basta, C., M. R. Costa-Jussà, and N. Casas. (2019). "Evaluating the underlying gender bias in contextualized word embeddings". *arXiv preprint arXiv:1904.08783*.

Beattie, L., D. Taber, and H. Cramer. (2022). "Challenges in translating research to practice for evaluating fairness and bias in recommendation systems". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 528–530.

Beltz, A. M., A. M. Loviska, and A. Weigard. (2021). "Daily gender expression is associated with psychological adjustment for some people, but mainly men". *Scientific Reports*. 11(1). DOI: 10.1038/s41598-021-88279-4.

Bernstein, E. S. and S. Turban. (2018). "The impact of the 'Open' workspace on human collaboration". *Philosophical Transactions of the Royal Society B: Biological Sciences*. 373(1753): 20170239. DOI: 10.1098/rstb.2017.0239.

Biega, A. J., K. P. Gummadi, and G. Weikum. (2018). "Equity of attention: Amortizing individual fairness in rankings". In: *The 41st international acm sigir conference on research & development in information retrieval*. 405–414.

Bigdeli, A. (2021). "Exploration and Mitigation of Stereotypical Gender Biases in Information Retrieval Systems". *Toronto Metropolitan University*.

Bigdeli, A., N. Arabzadeh, S. Seyedsalehi, B. Mitra, M. Zihayat, and E. Bagheri. (2023). "De-biasing Relevance Judgements for Fair Ranking". In: *Advances in Information Retrieval*. Ed. by J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo. Cham: Springer Nature Switzerland. 350–358.

Bigdeli, A., N. Arabzadeh, S. Seyedsalehi, M. Zihayat, and E. Bagheri. (2022). "A Light-Weight Strategy for Restraining Gender Biases in Neural Rankers". In: *Advances in Information Retrieval*. Ed. by M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, and V. Setty. Cham: Springer International Publishing. 47–55.

Bigdeli, A., N. Arabzadeh, S. Seyersalehi, M. Zihayat, and E. Bagheri. (2021a). "On the Orthogonality of Bias and Utility in Ad hoc Retrieval". In: *Proceedings of the 44rd International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Bigdeli, A., N. Arabzadeh, M. Zihayat, and E. Bagheri. (2021b). "Exploring Gender Biases in Information Retrieval Relevance Judgement Datasets". In: *European Conference on Information Retrieval*. Springer. 216–224.

Bingley, W. J., C. Curtis, S. Lockey, A. Bialkowski, N. Gillespie, S. A. Haslam, R. K. L. Ko, N. Steffens, J. Wiles, and P. Worthy. (2023). "Where is the human in human-centered AI? Insights from developer priorities and user experiences". *Computers in Human Behavior*. 141: 107617. DOI: 10.1016/j.chb.2022.107617.

Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy". *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*: 149–159. URL: https://dl.acm.org/doi/10.1145/3287560.3287583.

Binns, R., M. Van Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt. (2018). "'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions". In: *Proceedings of the 2018 Chi conference on human factors in computing systems*. 1–14.

Blackstone, A. (2003). "Gender roles and society". In: *Human Ecology: An Encyclopedia of Children, Families, Communities, and Environments*. Ed. by J. R. Miller, R. M. Lerner, and L. B. Schiamberg. Santa Barbara, CA: ABC-CLIO. 335–338.

Bloomberg. (2024). "Bloomberg". URL: https://www.bloomberg.com.

Bogen, M. and A. Rieke. (2018). "Help wanted: An examination of hiring algorithms, equity, and bias". *Upturn, December*. 7.

Boinodiris, P. (2024). "The importance of diversity in AI isn't opinion, it's math". URL: https://www.ibm.com/blog/why-we-need-diverse-multidisciplinary-coes-for-model-risk/.

Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov. (2017). "Enriching word vectors with subword information". *Transactions of the association for computational linguistics*. 5: 135–146.

Bolte, G., K. Jacke, K. Groth, U. Kraus, L. Dandolo, L. Fiedel, M. Debiak, M. Kolossa-Gehring, A. Schneider, and K. Palm. (2021). "Integrating sex/gender into environmental health research: Development of a conceptual framework". *International Journal of Environmental Research and Public Health*. 18(22): 12118. DOI: 10.3390/ijerph182212118.

Bolukbasi, T., K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. (2016). "Man is to computer programmer as woman is to homemaker? debiasing word embeddings". *Advances in neural information processing systems*. 29.

Borau, S., T. Otterbring, S. Laporte, and S. Fosso Wamba. (2021). "The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI". *Psychology & Marketing*. 38(7): 1052–1068. DOI: 10.1002/mar.21480.

Bordia, S. and S. R. Bowman. (2019). "Identifying and reducing gender bias in word-level language models". *arXiv preprint arXiv:1904.03035*.

Bösch, F., M. K. Angele, and I. H. Chaudry. (2018). "Gender differences in trauma, shock and sepsis". *Military Medical Research*. 5(1): 35. DOI: 10.1186/s40779-018-0182-5.

Bowman-Smart, H., J. Savulescu, M. O'Connell, and A. Sinclair. (2024). "World Athletics regulations unfairly affect female athletes with differences in sex development". *Journal of the Philosophy of Sport*. 51(1): 29–53. DOI: 10.1080/00948705.2024.2316294.

Bredella, M. A. (2017). "Sex Differences in Body Composition". In: *Sex and Gender Factors Affecting Metabolic Homeostasis, Diabetes and Obesity*. Springer International Publishing. 9–27.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.* (2020). "Language models are few-shot learners". *Advances in neural information processing systems*. 33: 1877–1901.

Bruce, V., A. M. Burton, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney. (1993). "Sex discrimination: how do we tell the difference between male and female faces?" *perception*. 22(2): 131–152. DOI: 10.1068/p220131.

Buda, M., A. Maki, and M. A. Mazurowski. (2018). "A systematic study of the class imbalance problem in convolutional neural networks". *Neural networks*. 106: 249–259.

Buolamwini, J. (2024). *Unmasking AI: My mission to protect what is human in a world of machines*. Random House.

Buslón, N., A. Cortés, S. Catuara-Solarz, D. Cirillo, and M. J. Rementeria. (2023). "Raising awareness of sex and gender bias in artificial intelligence and health". *Frontiers in Global Women's Health*. 4. DOI: 10.3389/fgwh.2023.970312.

Caira, C., L. Russo, and L. Aranda. (2023). "Artificially inequitable? AI and closing the gender gap". URL: https://oecd.ai/en/wonk/closing-the-gender-gap.

Caliskan, A., P. P. Ajay, T. Charlesworth, R. Wolfe, and M. R. Banaji. (2022). "Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics". In: *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. 156–170.

Caliskan, A., J. J. Bryson, and A. Narayanan. (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science*. 356(6334): 183–186.

Callan, J., M. Hoy, C. Yoo, and L. Zhao. (2009). "The ClueWeb09 dataset". In: *Proc. of the 32nd annual international ACM SIGIR conference on Research and development in information retrieval*. 523–524.

Cao, Y. T. and H. Daumé. (2021). "Toward gender-inclusive coreference resolution: An analysis of gender and bias throughout the machine learning lifecycle". *Computational Linguistics*. 47(3): 615–661. DOI: 10.1162/coli_a_00413.

Carpineto, C. and G. Romano. (2012). "A survey of automatic query expansion in information retrieval". *Acm Computing Surveys (CSUR)*. 44(1): 1–50.

Castets-Renard, C. and C. Lequesne. (2023). "Abortion in the age of AI: A need for safeguarding reproductive rights in the United States and the European Union". *McGill Law Journal*. 69: 1–17.

Chaloner, K. and A. Maldonado. (2019). "Measuring gender bias in word embeddings across domains and discovering new gender bias word categories". In: *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*. 25–32.

Chang, A. R. and S. M. Wildman. (2017). "Gender in/sight: Examining culture and constructions of gender". *Georgetown Journal of Gender and the Law*. 18(1): 43–80.

Chelba, C., T. Mikolov, M. Schuster, Q. Ge, T. Brants, P. Koehn, and T. Robinson. (2013). "One billion word benchmark for measuring progress in statistical language modeling". *arXiv preprint arXiv:1312.3005*.

Chen, J., H. Dong, X. Wang, F. Feng, M. Wang, and X. He. (2023a). "Bias and debias in recommender system: A survey and future directions". *ACM Transactions on Information Systems*. 41(3): 1–39.

Chen, J., X. Wang, F. Feng, and X. He. (2021). "Bias issues and solutions in recommender system: Tutorial on the recsys 2021". In: *Proceedings of the 15th ACM Conference on Recommender Systems*. 825–827.

Chen, P., L. Wu, and L. Wang. (2023b). "AI fairness in data management and analytics: A review on challenges, methodologies and applications". *Applied Sciences*. 13(18): 10258. DOI: 10.3390/app131810258.

Chen, Z. (2023). "Ethics and discrimination in artificial intelligence-enabled recruitment practices". *Humanities and Social Sciences Communications*. 10(1). DOI: 10.1057/s41599-023-02079-x.

Chmielinski, K., S. Newman, M. Taylor, J. Joseph, K. Thomas, J. Yurkofsky, and Y. Qiu. (2022). "The dataset nutrition label (2nd gen): Leveraging context to mitigate harms in artificial intelligence". *ArXiv*. abs/2201.03954.

Chouldechova, A. (2017). "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". *Big Data*. 5(2): 153–163. DOI: 10.1089/big.2016.0047.

Chowdhury, A. G., R. Sawhney, R. Shah, and D. Mahata. (2019). "# YouToo? detection of personal recollections of sexual harassment on social media". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2527–2537.

Chowdhury, S. B. R., S. Ghosh, Y. Li, J. B. Oliva, S. Srivastava, and S. Chaturvedi. (2021). "Adversarial scrubbing of demographic information for text classification". *arXiv preprint arXiv:2109.08613*.

Clark, K., M.-T. Luong, Q. V. Le, and C. D. Manning. (2020a). "Electra: Pre-training text encoders as discriminators rather than generators". *arXiv preprint arXiv:2003.10555*.

Clark, K., M.-T. Luong, Q. V. Le, and C. D. Manning. (2020b). "Electra: Pre-training text encoders as discriminators rather than generators". *arXiv preprint arXiv:2003.10555*.

Clarke, C. L., N. Craswell, I. Soboroff, A. Ashkan, E. Agichtein, and F. Díaz. (2004). "Overview of the TREC 2004 Terabyte Track". In: *TREC*. Vol. 2004. 74–85.

Clarke, C. L., N. Craswell, I. Soboroff, A. Ashkan, E. Agichtein, and F. Díaz. (2012). "The TREC 2012 Web Track". In: *TREC*. 1–12.

Clarke, C. L., F. Diaz, and N. Arabzadeh. (2023). "Preference-based offline evaluation". In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 1248–1251.

Clarke, C. L., A. Vtyurina, and M. D. Smucker. (2020). "Offline evaluation without gain". In: *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 185–192.

Clarke, C. L., A. Vtyurina, and M. D. Smucker. (2021). "Assessing top-preferences". *ACM Transactions on Information Systems (TOIS)*. 39(3): 1–21.

Craswell, N., B. Mitra, E. Yilmaz, D. Campos, and E. M. Voorhees. (2020). "Overview of the TREC 2019 deep learning track". *arXiv preprint arXiv:2003.07820*.

Crenshaw, K. (1991). "Mapping the margins: Intersectionality, identity politics, and violence against women of color". *Stanford Law Review*. 43(6): 1241. DOI: 10.2307/1229039.

Crenshaw, K. W. (2017). *On Intersectionality: Essential Writings*. No. 255. Faculty Books. URL: https://scholarship.law.columbia.edu/books/255.

Cubuk, E. D., B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. (2018). "Autoaugment: Learning augmentation policies from data". *arXiv preprint arXiv:1805.09501*.

D'Ignazio, C. and L. Klein. (2020). "What Gets Counted Counts". In: *Data Feminism*. MIT Press. 9–27.

Dai, Z., C. Xiong, J. Callan, and Z. Liu. (2018). "Convolutional neural networks for soft-matching n-grams in ad-hoc search". In: *Proceedings of the eleventh ACM international conference on web search and data mining*. 126–134.

Dastin, J. (2018). "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women". URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Dellinger, J. and S. Pell. (2024). "Bodies of evidence: The criminalization of abortion and surveillance of women in a post-Dobbs world". *Duke Journal of Constitutional Law & Public Policy*. 19(1): 1–108.

Deng, W. H., M. S. Lam, Á. A. Cabrera, D. Metaxa, M. Eslami, and K. Holstein. (2023). "Supporting user engagement in testing, auditing, and contesting AI". In: *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 556–559.

Dev, S., T. Li, J. M. Phillips, and V. Srikumar. (2020). "On measuring and mitigating biased inferences of word embeddings". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 7659–7666.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2018a). "Bert: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2018b). "Bert: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*.

Diaz, F. and D. Metzler. (2006). "Improving the estimation of relevance models using large external corpora". In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 154–161.

Diaz, F., B. Mitra, M. D. Ekstrand, A. J. Biega, and B. Carterette. (2020). "Evaluating stochastic rankings with expected exposure". In: *Proceedings of the 29th ACM international conference on information & knowledge management*. 275–284.

DiMarco, M., H. Zhao, M. Boulicault, and S. S. Richardson. (2022). "Why "sex as a biological variable" conflicts with precision medicine initiatives". *Cell Reports Medicine*. 3(4). DOI: 10.1016/j.xcrm.2022.100550.

Dinan, E., A. Fan, A. Williams, J. Urbanek, D. Kiela, and J. Weston. (2019). "Queens are powerful too: Mitigating gender bias in dialogue generation". *arXiv preprint arXiv:1911.03842*.

Doughman, J., W. Khreich, M. El Gharib, M. Wiss, and Z. Berjawi. (2021). "Gender bias in text: Origin, taxonomy, and implications". In: *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*. 34–44.

Douzas, G., F. Bacao, and F. Last. (2018). "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE". *Information sciences*. 465: 1–20.

Dubenko, E. (2022). "Across-language masculinity of oceans and femininity of guitars: Exploring grammatical gender universalities". *Frontiers in Psychology*. 13. DOI: 10.3389/fpsyg.2022.1009966.

Dwork, C., M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. (2012). "Fairness through awareness". In: *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.

Ekstrand, M. D., A. Das, R. Burke, and F. Diaz. (2021). "Fairness and Discrimination in Information Access Systems". *CoRR*. abs/2105.05779. URL: https://arxiv.org/abs/2105.05779.

Ellemers, N. (2018). "Gender stereotypes". *Annual review of psychology*. 69: 275–298.

Ellis, G. (2018). "So, what are cognitive biases?" *Cognitive biases in visualizations*: 1–10.

Eskandanian, F. and B. Mobasher. (2020). "Using stable matching to optimize the balance between accuracy and diversity in recommendation". In: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 71–79.

Fabris, A., A. Purpura, G. Silvello, and G. A. Susto. (2020). "Gender stereotype reinforcement: Measuring the gender bias conveyed by ranking algorithms". *Information Processing & Management*. 57(6): 102377.

Feast, J. (2020). "4 ways to address gender bias in AI". URL: https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai.

Fekih, S., N. Tamagnone, B. Minixhofer, R. Shrestha, X. Contla, E. Oglethorpe, and N. Rekabsaz. (2022). "Humset: Dataset of multilingual information extraction and classification for humanitarian crisis response". *arXiv preprint arXiv:2210.04573*.

Felkner, V. K., H.-C. H. Chang, E. Jang, and J. May. (2023). "Winoqueer: A community-in-the-loop benchmark for anti-lgbtq+ bias in large language models". *arXiv preprint arXiv:2306.15087*.

Fersini, E., D. Nozza, P. Rosso, *et al.* (2020). "AMI@ EVALITA2020: Automatic misogyny identification". In: *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*.

Fersini, E., P. Rosso, M. Anzovino, *et al.* (2018). "Overview of the task on automatic misogyny identification at IberEval 2018." *Ibereval@ sepln*. 2150: 214–228.

Fleisig, E. and C. Fellbaum. (2022). "Mitigating Gender Bias in Machine Translation through Adversarial Learning". *arXiv preprint arXiv:2203.10675*.

Frable, D. E. (1997). "Gender, racial, ethnic, sexual, and class identities". *Annual review of psychology*. 48(1): 139–162. DOI: 10.1146/annurev.psych.48.1.139.

Friedman, B. and H. Nissenbaum. (1996). "Bias in Computer Systems". *ACM Transactions on Information Systems*. 14(3): 330–347. DOI: 10.1145/230538.230561.

Frost, D. M. (2011). "Social stigma and its consequences for the socially stigmatized". *Social and Personality Psychology Compass*. 5(11): 824–839. DOI: 10.1111/j.1751-9004.2011.00394.x.

Garimella, A., A. Amarnath, K. Kumar, A. P. Yalla, N. Anandhavelu, N. Chhaya, and B. V. Srinivasan. (2021). "He is very intelligent, she is very beautiful? on mitigating social biases in language modelling and generation". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 4534–4545.

Gebru, T., J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford. (2021). "Datasheets for datasets". *Communications of the ACM*. 64(12): 86–92. DOI: 10.1145/3458723.

Ghabrial, M. A. (2019). ""We can shapeshift and build bridges": Bisexual women and gender diverse people of color on invisibility and embracing the borderlands". *Journal of Bisexuality*. 19(2): 169–197. DOI: 10.1080/15299716.2019.1617526.

Ghosh, B. (2018). "A diachronic perspective of Hijra identity in India". In: *Sociology of Motherhood and Beyond*. 107–119.

Gill-Peterson, J. (2024). *A Short History of Trans Misogyny*. Verso Books.

Gonen, H. and Y. Goldberg. (2019). "Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them". *arXiv preprint arXiv:1903.03862*.

González-Álvarez, J. and R. Sos-Peña. (2022). "Sex perception from facial structure: Categorization with and without skin texture and color". *Vision Research*. 201: 108127. DOI: 10.1016/j.visres.2022.108127.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. (2020). "Generative adversarial networks". *Communications of the ACM*. 63(11): 139–144.

Google. (2024). "Google Scholar". URL: https://scholar.google.com.

Guo, J., Y. Fan, Q. Ai, and W. B. Croft. (2016). "A deep relevance matching model for ad-hoc retrieval". In: *Proceedings of the 25th ACM international on conference on information and knowledge management*. 55–64.

Gurumurthy, S., R. Kiran Sarvadevabhatla, and R. Venkatesh Babu. (2017). "Deligan: Generative adversarial networks for diverse and limited data". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 166–174.

Hamidi, F., M. K. Scheuerman, and S. M. Branham. (2018). "Gender recognition or gender reductionism? The social implications of embedded gender recognition systems". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

Holstein, K., J. W. Vaughan, H. Wallach, H. D. III, and M. Dudik. (2019). "Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–16. URL: https://dl.acm.org/doi/10.1145/3290605.3300830.

Hong, L. and S. E. Page. (2004). "Groups of diverse problem solvers can outperform groups of high-ability problem solvers". *Proceedings of the National Academy of Sciences*. 101(46): 16385–16389. DOI: 10.1073/pnas.0403723101.

Hyde, J. S., M. Krajnik, and K. Skuldt-Niederberger. (1991). "Androgyny across the life span: A replication and longitudinal followup". *Developmental Psychology*. 27(3): 516–519. DOI: 10.1037/0012-1649.27.3.516.

InterACT. (2018). "InterACT statement on Intersex Terminology". URL: https://interactadvocates.org/interact-statement-on-intersex-terminology/.

Internet Live Stats. (2024). "Google Search Statistics". URL: https://www.internetlivestats.com/google-search-statistics/.

ITHAKA. (2024). "JSTOR". URL: https://www.jstor.org.

Jakiela, P. and O. Ozier. (2020). "Gendered language". *Tech. rep.* Bonn: IZA Discussion Papers, No. 13126, Institute of Labor Economics (IZA). URL: https://www.econstor.eu/bitstream/10419/216438/1/dp13126.pdf.

Jha, A. and R. Mamidi. (2017). "When does a compliment become sexist? analysis and classification of ambivalent sexism using twitter data". In: *Proceedings of the second workshop on NLP and computational social science*. 7–16.

Johnson, J. L. and R. Repta. (2012). "Sex and gender: Beyond the binaries". In: *Designing and conducting gender, sex and health research*. Ed. by J. L. Oliffe and L. Greaves. Thousand Oaks, CA: Sage. 17–37.

Jones, S. (1975). "Report on the need for and provision of an "ideal" information retrieval test collection".

Kamiran, F. and T. Calders. (2012). "Data preprocessing techniques for classification without discrimination". *Knowledge and information systems*. 33(1): 1–33. DOI: 10.1007/s10115-011-0463-8.

Kaneko, M., D. Bollegala, and N. Okazaki. (2022). "Gender bias in meta-embeddings". *arXiv preprint arXiv:2205.09867*.

Kapania, S., A. S. Taylor, and D. Wang. (2023). "A hunt for the snark: Annotator diversity in data practices". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

Karpukhin, V., B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih. (2020). "Dense passage retrieval for open-domain question answering". *arXiv preprint arXiv:2004.04906*.

Karray, F., M. Alemzadeh, J. A. Saleh, and M. N. Arab. (2008). "Human-computer interaction: Overview on state of the art". *International Journal on Smart Sensing and Intelligent Systems*. 1(1): 137–159. DOI: 10.21307/ijssis-2017-283.

Katyal, S. K. and J. Y. Jung. (2021). "The gender panopticon: AI, gender, and design justice". *UCLA Law Review*. 68: 692–785.

Keyes, O. (2018). "The misgendering machines: Trans/HCI implications of automatic gender recognition". In: *Proceedings of the ACM on human-computer interaction*. 1–22.

Khan, S. I., M. I. Hussain, S. Parveen, M. I. Bhuiyan, G. Gourab, G. F. Sarker, S. M. Arafat, and J. Sikder. (2009). "Living on the extreme margin: Social exclusion of the transgender population (hijra) in Bangladesh". *Journal of Health, Population and Nutrition*. 27(4). DOI: 10.3329/jhpn.v27i4.3388.

Klasnja, A., N. Arabzadeh, M. Mehrvarz, and E. Bagheri. (2022). "On the characteristics of ranking-based gender bias measures". In: *Proceedings of the 14th ACM Web Science Conference 2022*. 245–249.

Kleisner, K., P. Tureček, S. C. Roberts, J. Havlíček, J. V. Valentova, R. M. Akoko, J. D. Leongómez, S. Apostol, M. A. Varella, and S. A. Saribay. (2021). "How and why patterns of sexual dimorphism in human faces vary across the world". *Scientific reports*. 11(1): 5978. DOI: 10.1038/s41598-021-85402-3.

Kopeinik, S., M. Mara, L. Ratz, K. Krieg, M. Schedl, and N. Rekabsaz. (2023). "Show me a "Male Nurse"! How Gender Bias is Reflected in the Query Formulation of Search Engine Users". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

Kotek, H., R. Dockum, and D. Sun. (2023). "Gender bias and stereotypes in large language models". In: *Proceedings of The ACM Collective Intelligence Conference*. 12–24.

Krieg, K., E. Parada-Cabaleiro, G. Medicus, O. Lesota, M. Schedl, and N. Rekabsaz. (2023). "Grep-BiasIR: A Dataset for Investigating Gender Representation Bias in Information Retrieval Results". In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 444–448.

Krieg, K., E. Parada-Cabaleiro, M. Schedl, and N. Rekabsaz. (2022). "Do Perceived Gender Biases in Retrieval Results Affect Relevance Judgements?" arXiv: 2203.01731 [cs.IR].

Krieger, N. (2020). "Measures of racism, sexism, heterosexism, and gender binarism for health equity research: from structural injustice to embodied harm—an ecosocial analysis". *Annual Review of Public Health*. 41(1): 37–62. DOI: 10.1146/annurev-publhealth-040119-094017.

Kruks, S. (1992). "Gender and subjectivity: Simone de Beauvoir and contemporary feminism". *Signs: Journal of Women in Culture and Society*. 18(1): 89–110. DOI: 10.1086/494780.

Lam, M. S., M. L. Gordon, D. Metaxa, J. T. Hancock, J. A. Landay, and M. S. Bernstein. (2022). "End-user audits: A system empowering communities to lead large-scale investigations of harmful algorithmic behavior". In: *Proceedings of the ACM on Human-Computer Interaction*. 1–34.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2020). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". In: *International Conference on Learning Representations (ICLR)*. URL: https://openreview.net/forum?id=H1eA7AEtvS.

Laqueur, T. W. (1992). *Making sex: Body and gender from the Greeks to Freud*. Harvard University Press.

Lee, P. A., C. P. Houk, S. F. Ahmed, and I. A. Hughes. (2006). "Consensus statement on management of intersex disorders". *Pediatrics*. 118(2). DOI: 10.1542/peds.2006-0738.

Lemley, J., S. Bazrafkan, and P. Corcoran. (2017). "Smart augmentation learning an optimal data augmentation strategy". *Ieee Access*. 5: 5858–5869.

LexisNexis. (2024). "LexisNexis". URL: https://www.lexisnexis.com.

Li, Y., X. Wei, Z. Wang, S. Wang, P. Bhatia, X. Ma, and A. Arnold. (2022). "Debiasing Neural Retrieval via In-batch Balancing Regularization". In: *Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. Seattle, Washington: Association for Computational Linguistics. 58–66. DOI: 10.18653/v1/2022.gebnlp-1.5.

Liu, H., W. Wang, Y. Wang, H. Liu, Z. Liu, and J. Tang. (2020). "Mitigating gender bias for neural dialogue generation with adversarial learning". *arXiv preprint arXiv:2009.13028*.

Lohia, P. K., K. Natesan Ramamurthy, M. Bhide, D. Saha, K. R. Varshney, and R. Puri. (2019). "Bias mitigation post-processing for individual and group fairness". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. DOI: 10.1109/icassp.2019.8682620.

Loideain, N. N. and R. Adams. (2020). "From Alexa to Siri and the GDPR: The gendering of virtual personal assistants and the role of Data Protection Impact Assessments". *Computer Law & Security Review*. 36: 105366. DOI: 10.1016/j.clsr.2019.105366.

Manning, C. D. (2008). *Introduction to information retrieval*. Syngress Publishing,

Matthew, E. (2018). "Peters, mark neumann, mohit iyyer, matt gardner, christopher clark, kenton lee, luke zettlemoyer. deep contextualized word representations". In: *Proc. of NAACL*. Vol. 5.

Maudslay, R. H., H. Gonen, R. Cotterell, and S. Teufel. (2019). "It's all in the name: Mitigating gender bias with name-based counterfactual data substitution". *arXiv preprint arXiv:1909.00871*.

May, C., A. Wang, S. Bordia, S. R. Bowman, and R. Rudinger. (2019). "On measuring social biases in sentence encoders". *arXiv preprint arXiv:1903.10561*.

Mazzuca, C., A. M. Borghi, S. van Putten, L. Lugli, R. Nicoletti, and A. Majid. (2023). "Gender is conceptualized in different ways across cultures". *Language and Cognition*. 16(2): 353–379. DOI: 10.1017/langcog.2023.40.

McLemore, K. A. (2014). "Experiences with misgendering: Identity misclassification of transgender spectrum individuals". *Self and Identity*. 14(1): 51–74. DOI: 10.1080/15298868.2014.950691.

Medicine, U. N. L. of. (2024). "PubMed". URL: https://pubmed.ncbi.nlm.nih.gov.

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. (2021). "A Survey on Bias and Fairness in Machine Learning". *ACM Computing Surveys*. 54(6): 1–35. DOI: 10.1145/3457607.

Meinhardt-Injac, B., M. Persike, and G. Meinhardt. (2013). "Holistic face processing is induced by shape and texture". *Perception*. 42(7): 716–732. DOI: 10.1068/p7462.

Memarian, B. and T. Doleck. (2023). "Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review". *Computers and Education: Artificial Intelligence*. 5: 100152. DOI: 10.1016/j.caeai.2023.100152.

Miceli, M., T. Yang, A. A. Garcia, J. Posada, S. M. Wang, M. Pohl, and A. Hanna. (2022). "Documenting Data Production Processes: A Participatory Approach for Data Work". URL: https://arxiv.org/abs/2207.04958.

Mikołajczyk, A. and M. Grochowski. (2018). "Data augmentation for improving deep learning in image classification problem". In: *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE. 117–122.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*.

Mitra, B., F. Diaz, and N. Craswell. (2017). "Learning to match using local and distributed representations of text for web search". In: *Proceedings of the 26th international conference on world wide web*. 1291–1299.

Mort, J. A. (2023). "Fighting information termination". *Index on Censorship*. 52(1): 27–29.

Namaste, V. (2000). *Invisible lives: The erasure of transsexual and transgendered people*. University of Chicago Press.

Nass, C., J. Steuer, and E. R. Tauber. (1994). "Computers are social actors". In: *Conference Companion on Human Factors in Computing Systems - CHI '94*. DOI: 10.1145/259963.260288.

National Public Radio, N. P. R. (2024). "NPR's embedded and CBC tackle sex testing in elite sports with "tested" podcast". URL: https://www.npr.org/2024/07/12/g-s1-8943/npr-embedded-cbc-testing-in-elite-sports-tested-podcast.

Nguyen, T., M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng. (2016). "MS MARCO: A human-generated machine reading comprehension dataset". *30th Conference on Neural Information Processing Systems (NIPS 2016)*.

Nicas, J. (2019). "Apple Card Investigated After Gender Discrimination Complaints". URL: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

Niousha, R., D. Saito, H. Washizaki, and Y. Fukazawa. (2023). "Investigating the Effect of Binary Gender Preferences on Computational Thinking Skills". *Education Sciences*. 13(5): 433.

Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. (2019). "Dissecting racial bias in an algorithm used to manage the health of populations". *Science*. 366(6464): 447–453. DOI: 10.1126/science.aax2342.

Osokin, A., A. Chessel, R. E. Carazo Salas, and F. Vaggi. (2017). "GANs for biological image synthesis". In: *Proceedings of the IEEE international conference on computer vision*. 2233–2242.

Ovalle, A., A. Subramonian, A. Singh, C. Voelcker, D. J. Sutherland, D. Locatelli, E. Breznik, F. Klubicka, H. Yuan, J. Hetvi, H. Zhang, J. Shriram, K. Lehman, L. Soldaini, M. Sap, M. P. Deisenroth, M. L. Pacheco, M. Ryskina, M. Mundt, M. Agarwal, N. Mclean, P. Xu, A. Pranav, R. Korpan, R. Ray, S. Mathew, S. Arora, S. John, T. Anand, V. Agrawal, W. Agnew, Y. Long, Z. J. Wang, Z. Talat, A. Ghosh, N. Dennler, M. Noseworthy, S. Jha, E. Baylor, A. Joshi, N. Y. Bilenko, A. Mcnamara, R. Gontijo-Lopes, A. Markham, E. Dong, J. Kay, M. Saraswat, N. Vytla, and L. Stark. (2023). "Queer in AI: A case study in community-led participatory AI". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1882–1895.

Pang, L., Y. Lan, J. Guo, J. Xu, S. Wan, and X. Cheng. (2016). "Text matching as image recognition". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30. No. 1.

Pape, M., M. Miyagi, S. A. Ritz, M. Boulicault, S. S. Richardson, and D. L. Maney. (2024). "Sex contextualism in laboratory research: enhancing rigor and precision in the study of sex-related variables". *Cell*. 187(6): 1316–1326. DOI: 10.1016/j.cell.2024.02.008.

Park, S., K. Choi, H. Yu, and Y. Ko. (2023). "Never too late to learn: Regularizing gender bias in coreference resolution". In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 15–23.

Pennebaker, J. W., M. E. Francis, and R. J. Booth. (2001). "Linguistic inquiry and word count: LIWC 2001". *Mahway: Lawrence Erlbaum Associates*. 71(2001): 2001.

Pennington, J., R. Socher, and C. D. Manning. (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

Perez, L. and J. Wang. (2017). "The effectiveness of data augmentation in image classification using deep learning". *arXiv preprint arXiv:1712.04621*.

Perugia, G. and D. Lisy. (2023). “Robot’s gendering trouble: A scoping review of gendering humanoid robots and its effects on HRI”. *International Journal of Social Robotics*. 15(11): 1725–1753. DOI: 10.1007/s12369-023-01061-6.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018a). “Deep Contextualized Word Representations”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Ed. by M. Walker, H. Ji, and A. Stent. New Orleans, Louisiana: Association for Computational Linguistics. 2227–2237. DOI: 10.18653/v1/N18-1202.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018b). “Deep contextualized word representations”. *CoRR*. abs/1802.05365. URL: http://arxiv.org/abs/1802.05365.

Petreski, D. and I. C. Hashim. (2022). “Word embeddings are biased. But whose bias are they reflecting?” *AI & Society*. 38(2): 975–982. DOI: 10.1007/s00146-022-01443-w.

Pinney, C., A. Raj, A. Hanna, and M. D. Ekstrand. (2023). “Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access”. In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 269–279.

Pleiss, G., M. Raghavan, F. Wu, J. M. Kleinberg, and K. Q. Weinberger. (2017). “On fairness and calibration”. In: *The Societal Impacts of Algorithmic Decision-Making*.

Posada, J. (2023). “Platform Authority and Data Quality: Who Decides What Counts in Data Production for Artificial Intelligence”. *Tech. rep.* Berggruen Institute and Global Affairs Canada.

Prost, F., N. Thain, and T. Bolukbasi. (2019). “Debiasing embeddings for reduced gender bias in text classification”. *arXiv preprint arXiv:1908.02810*.

Qian, Y., U. Muaz, B. Zhang, and J. W. Hyun. (2019). “Reducing gender bias in word-level language models with a gender-equalizing loss function”. *arXiv preprint arXiv:1905.12801*.

Qiu, H., Z.-Y. Dou, T. Wang, A. Celikyilmaz, and N. Peng. (2023). "Gender Biases in Automatic Evaluation Metrics for Image Captioning". In: *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Qu, Y., Y. Ding, J. Liu, K. Liu, R. Ren, W. X. Zhao, D. Dong, H. Wu, and H. Wang. (2021). "RocketQA: An optimized training approach to dense passage retrieval for open-domain question answering". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 5835–5847.

Raj, A., B. Mitra, N. Craswell, and M. Ekstrand. (2023). "Patterns of gender-specializing query reformulation". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2241–2245.

Rekabsaz, N., S. Kopeinik, and M. Schedl. (2021). "Societal Biases in Retrieved Contents: Measurement Framework and Adversarial Mitigation for BERT Rankers". *arXiv preprint arXiv:2104.13640*.

Rekabsaz, N. and M. Schedl. (2020). "Do Neural Ranking Models Intensify Gender Bias?" In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2065–2068.

Reuters. (2024). "Reuters". URL: https://www.reuters.com.

Richards, J. and R. Hawley. (2011). "Sex determination: How genes determine a developmental choice". *The human genome*: 273–98. DOI: 10.1016/b978-0-08-091865-5.00008-4.

Richardson, S. S. (2022). "Sex contextualism". *Philosophy, Theory, and Practice in Biology*. 14(0). DOI: 10.3998/ptpbio.2096.

Roberts, T. K. and C. R. Fantz. (2014). "Barriers to quality health care for the transgender population". *Clinical Biochemistry*. 47(10–11): 983–987. DOI: 10.1016/j.clinbiochem.2014.02.009.

Robertson, S., H. Zaragoza, *et al.* (2009). "The probabilistic relevance framework: BM25 and beyond". *Foundations and Trends® in Information Retrieval*. 3(4): 333–389.

Rodríguez-Sánchez, F., J. Carrillo-de-Albornoz, L. Plaza, J. Gonzalo, P. Rosso, M. Comet, and T. Donoso. (2021). "Overview of exist 2021: sexism identification in social networks". *Procesamiento del Lenguaje Natural*. 67: 195–207.

Rodríguez-Sánchez, F., J. Carrillo-de-Albornoz, L. Plaza, A. Mendieta-Aragón, G. Marco-Remón, M. Makeienko, M. Plaza, J. Gonzalo, D. Spina, and P. Rosso. (2022). "Overview of exist 2022: sexism identification in social networks". *Procesamiento del Lenguaje Natural*. 69: 229–240.

Rule, N. O. (2017). "Perceptions of sexual orientation from minimal cues". *Archives of Sexual Behavior*. 46(1): 129–139. DOI: 10.1007/s10508-016-0779-2.

Russell, R. (2009). "A sex difference in facial contrast and its exaggeration by cosmetics". *Perception*. 38(8): 1211–1219. DOI: 10.1068/p6331.

Russell, R., P. Sinha, I. Biederman, and M. Nederhouser. (2006). "Is pigmentation important for face recognition? Evidence from contrast negation". *Perception*. 35(6): 749–759. DOI: 10.1068/p5490.

Samory, M., I. Sen, J. Kohne, F. Flöck, and C. Wagner. (2021). ""Call me sexist, but...": Revisiting Sexism Detection Using Psychological Scales and Adversarial Samples". In: *Proceedings of the international AAAI conference on web and social media*. Vol. 15. 573–584.

Sandfort, T. G., H. M. Bos, T.-C. Fu, D. Herbenick, and B. Dodge. (2021). "Gender expression and its correlates in a nationally representative sample of the US adult population: Findings from the National Survey of Sexual Health and Behavior". *The Journal of Sex Research*. 58(1): 51–63. DOI: 10.1080/00224499.2020.1818178.

Sanh, V., L. Debut, J. Chaumond, and T. Wolf. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". *arXiv preprint arXiv:1910.01108*.

Sapir, E. (2023). *Selected writings of Edward Sapir in language, culture and personality*. University of California Press.

Saunders, D. and B. Byrne. (2020). "Reducing gender bias in neural machine translation as a domain adaptation problem". *arXiv preprint arXiv:2004.04498*.

Saxena, S. and S. Jain. (2024). "Exploring and mitigating gender bias in book recommender systems with explicit feedback". *Journal of Intelligent Information Systems*: 1–22.

Scheuerman, M. K., M. Pape, and A. Hanna. (2021). "Auto-essentialization: Gender in automated facial analysis as extended colonial project". *Big Data & Society*. 8(2): 1–15.

Scheuerman, M. K., J. M. Paul, and J. R. Brubaker. (2019). "How computers see gender". *Proceedings of the ACM on Human-Computer Interaction*. 3(CSCW): 1–33. DOI: 10.1145/3359246.

Seyedsalehi, S., A. Bigdeli, N. Arabzadeh, B. Mitra, M. Zihayat, and E. Bagheri. (2022a). "Bias-aware Fair Neural Ranking for Addressing Stereotypical Gender Biases." In: *EDBT*. 2–435.

Seyedsalehi, S., A. Bigdeli, N. Arabzadeh, M. Zihayat, and E. Bagheri. (2022b). "Addressing gender-related performance disparities in neural rankers". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2484–2488.

Shea, M. (2023). "The sinister side to female robots?" URL: https://www.bbc.com/future/article/20230804-is-there-a-sinister-side-to-the-rise-of-female-robots.

Sheng, E., K.-W. Chang, P. Natarajan, and N. Peng. (2021). "Societal biases in language generation: Progress and challenges". *arXiv preprint arXiv:2105.04054*.

Shin, D., A. Rasul, and A. Fotiadis. (2021). "Why am I seeing this? Deconstructing algorithm literacy through the lens of users". *Internet Research*. 32(4): 1214–1234. DOI: 10.1108/intr-02-2021-0087.

Shneiderman, B. (2020). "Human-centered artificial intelligence: Reliable, safe & trustworthy". *International Journal of Human–Computer Interaction*. 36(6): 495–504. DOI: 10.1080/10447318.2020.1741118.

Shorten, C. and T. M. Khoshgoftaar. (2019). "A survey on image data augmentation for deep learning". *Journal of big data*. 6(1): 1–48.

Shrestha, S. and S. Das. (2022). "Exploring gender biases in ML and AI academic research through systematic literature review". *Frontiers in artificial intelligence*. 5: 976838.

Smith, J. J. and L. Beattie. (2022). "RecSys Fairness Metrics: Many to Use But Which One To Choose?" *arXiv preprint arXiv:2209.04011*.

Smith, J. M. (2021). "Beyond the Gender Binary in Computing Education Research". In: *Proceedings of the 17th ACM Conference on International Computing Education Research*. 444–445.

Stanovsky, G., N. A. Smith, and L. Zettlemoyer. (2019). "Evaluating gender bias in machine translation". *arXiv preprint arXiv:1906.00591*.

Steck, H. (2011). "Item popularity and recommendation accuracy". In: *Proceedings of the fifth ACM conference on Recommender systems*. 125–132.

Sun, T., J. He, X. Qiu, and X. Huang. (2022). "BERTScore is Unfair: On Social Bias in Language Model-Based Metrics for Text Generation". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Y. Goldberg, Z. Kozareva, and Y. Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 3726–3739. DOI: 10.18653/v1/2022.emnlp-main.245.

Sundararaman, D. and V. Subramanian. (2022). "Do Information Retrieval Models Exhibit Gender Bias?"

Surveillance Technology Oversight Project, Inc. (2022). "STOP: Surveillance Technology Oversight Project". URL: https://www.stopspying.org.

Svechnikov, K. and O. Söder. (2008). "Ontogeny of gonadal sex steroids". *Best Practice & Research Clinical Endocrinology & Metabolism*. 22(1): 95–106. DOI: 10.1016/j.beem.2007.09.002.

Swim, J., E. Borgida, G. Maruyama, and D. G. Myers. (1989). "Joan McKay versus John McKay: Do gender stereotypes bias evaluations?" *Psychological Bulletin*. 105(3): 409.

Tadiri, C. P., V. Raparelli, M. Abrahamowicz, A. Kautzy-Willer, K. Kublickiene, M.-T. Herrero, C. M. Norris, L. Pilote, and G. Consortium. (2021). "Methods for prospectively incorporating gender into health sciences research". *Journal of Clinical Epidemiology*. 129: 191–197.

Taksa, I. and J. M. Flomenbaum. (2009). "An integrated framework for research on cross-cultural information retrieval". In: *2009 Sixth International Conference on Information Technology: New Generations*. IEEE. 1367–1372.

Tang, K., W. Zhou, J. Zhang, A. Liu, G. Deng, S. Li, P. Qi, W. Zhang, T. Zhang, and N. Yu. (2024). "GenderCARE: A Comprehensive Framework for Assessing and Reducing Gender Bias in Large Language Models". *arXiv preprint arXiv:2408.12494*.

Tannenbaum, C., R. P. Ellis, F. Eyssel, J. Zou, and L. Schiebinger. (2019). "Sex and gender analysis improves science and engineering". *Nature*. 575(7781): 137–146. DOI: 10.1038/s41586-019-1657-6.

Te'eni, D., J. Carey, and P. Zhang. (2007). *Human-Computer Interaction: Developing Effective Organizational Information Systems*. John Wiley & Sons, Inc.

Tolmeijer, S., N. Zierau, A. Janson, J. S. Wahdatehagh, J. M. Leimeister, and A. Bernstein. (2021). "Female by default? – Exploring the effect of voice assistant gender and pitch on trait and trust attribution". In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3411763.3451623.

UNESCO, I. (2024). "Challenging systematic prejudices: An investigation into gender bias in large language models". URL: https://unesdoc.unesco.org/ark:/48223/pf0000388971.

United Nations University. (2019). "Taking Stock: Data Evidence on Gender Equality in Digital Access, Skills, and Leadership". URL: https://collections.unu.edu/eserv/UNU:7350/EQUALS-Research-Report-2019.pdf.

Urbanek, J., A. Fan, S. Karamcheti, S. Jain, S. Humeau, E. Dinan, T. Rocktäschel, D. Kiela, A. Szlam, and J. Weston. (2019). "Learning to speak and act in a fantasy text adventure game". *arXiv preprint arXiv:1903.03094*.

Voorhees, E. M. (2004). "Overview of the TREC 2004 robust retrieval track". In: *TREC*. Vol. 2004. No. 5. 3.

Wang, J., Y. Liu, and X. E. Wang. (2021). "Are gender-neutral queries really gender-neutral? mitigating gender bias in image search". *arXiv preprint arXiv:2109.05433*.

Wang, Y. and M. Kosinski. (2018). "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images." *Journal of personality and social psychology*. 114(2): 246.

Waseem, Z. and D. Hovy. (2016). "Hateful symbols or hateful people? predictive features for hate speech detection on twitter". In: *Proceedings of the NAACL student research workshop*. 88–93.

Weischedel, R., S. Pradhan, L. Ramshaw, J. Kaufman, M. Franchini, M. ElBachouti, N. Xue, M. Palmer, J. D. Hwang, C. Bonial, *et al.* (2012). "OntoNotes Release 5.0". URL: https://catalog.ldc.upenn.edu/LDC2013T19.

Westlaw. (2024). "Westlaw". URL: https://legal.thomsonreuters.com/en/westlaw.

Whorf, B. L. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

Wu, C., F. Wu, X. Wang, Y. Huang, and X. Xie. (2021). "Fairness-aware news recommendation with decomposed adversarial learning". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. No. 5. 4462–4469.

Xiong, C., Z. Dai, J. Callan, Z. Liu, and R. Power. (2017). "End-to-end neural ad-hoc ranking with kernel pooling". In: *Proceedings of the 40th International ACM SIGIR conference on research and development in information retrieval*. 55–64.

Yang, J., A. A. Soltan, D. W. Eyre, Y. Yang, and D. A. Clifton. (2023). "An adversarial training framework for mitigating algorithmic biases in clinical machine learning". *NPJ Digital Medicine*. 6(1): 55.

Yi, X., E. Walia, and P. Babyn. (2019). "Generative adversarial network in medical imaging: A review". *Medical image analysis*. 58: 101552.

Yin, H., B. Cui, J. Li, J. Yao, and C. Chen. (2012). "Challenging the long tail recommendation". *arXiv preprint arXiv:1205.6700*.

Young, E., J. Wajcman, and L. Sprejer. (2023). "Mind the gender gap: Inequalities in the emergent professions of artificial intelligence (AI) and data science". *New Technology, Work and Employment*. 38(3): 391–414. DOI: 10.1111/ntwe.12278.

Zehlike, M., K. Yang, and J. Stoyanovich. (2022). "Fairness in ranking, part i: Score-based ranking". *ACM Computing Surveys*. 55(6): 1–36.

Zemel, R. S., L. Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. (2013). "Learning fair representations". In: *International Conference on Machine Learning*.

Zerveas, G., N. Rekabsaz, D. Cohen, and C. Eickhoff. (2021). "CODER: An efficient framework for improving retrieval through COntextual Document Embedding Reranking". *arXiv preprint arXiv:2112.08766*.

Zhang, D. T., S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. Manyika, J. Niebles, M. Sellitto, Y. Shoham, J. Clark, and R. Perrault. (2021). "The AI Index 2021 Annual Report". *ArXiv*. abs/2103.06312.

Zhang, G. and S. Ananiadou. (2022). "Examining and mitigating gender bias in text emotion detection task". *Neurocomputing*. 493: 422–434.

Zhao, J., S. Mukherjee, S. Hosseini, K.-W. Chang, and A. H. Awadallah. (2020). "Gender bias in multilingual embeddings and cross-lingual transfer". *arXiv preprint arXiv:2005.00699*.

Zhao, J., T. Wang, M. Yatskar, R. Cotterell, V. Ordonez, and K.-W. Chang. (2019). "Gender bias in contextualized word embeddings". *arXiv preprint arXiv:1904.03310*.

Zhao, J., T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang. (2017). "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2979–2989. URL: https://aclanthology.org/D17-1323/.

Zhao, J., T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang. (2018a). "Gender bias in coreference resolution: Evaluation and debiasing methods". *arXiv preprint arXiv:1804.06876*.

Zhao, J., Y. Zhou, Z. Li, W. Wang, and K.-W. Chang. (2018b). "Learning Gender-Neutral Word Embeddings". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Ed. by E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii. Brussels, Belgium: Association for Computational Linguistics. 4847–4853. DOI: 10.18653/v1/D18-1521.

Zhao, J., Y. Zhou, Z. Li, W. Wang, and K.-W. Chang. (2018c). "Learning gender-neutral word embeddings". *arXiv preprint arXiv:1809.01496*.

Zimman, L. (2014). "The discursive construction of sex". In: *Queer Excursions: Retheorizing Binaries in Language, Gender, and Sexuality*. Oxford University Press. 13–34.

Zou, J. Y., D. J. Hsu, D. C. Parkes, and R. P. Adams. (2013). "Contrastive learning using spectral methods". *Advances in Neural Information Processing Systems*. 26.