GEDIS - Gender Diversity in Information Science: Challenges in Higher Education
Project Reference: 2024-1-ES01-KA220-HED-000246558
https://ub.edu/GEDIS

Co-funded by the European Union

„Mind the Gap: Gender Data and AI Bias“

Gender Data Gap: Evidence Brief –

Supporting Evidence for OER

Co-funded by the European Union. The opinions and views expressed are solely those of the author(s) and do not necessarily reflect those of the European Union or the Spanish Service for the Internationalisation of Education (SEPIE). Neither the European Union nor the granting authority can be held responsible for them.

GEDIS - Gender Diversity in Information Science: Challenges in Higher Education
Project Reference: 2024-1-ES01-KA220-HED-000246558
https://ub.edu/GEDIS

Co-funded by the European Union

GEDIS

Gender Diversity in Information Science:

Challenges in Higher Education

Barcelona, 05/09/2025

Citation: Bosshammer, Svetlana and Daniela Varbanova. (2025). *Mind the Gap: Gender Data and AI Bias.* DOI: 10.5281/zenodo.17164102

Co-funded by the European Union

Executive Summary

The Gender Data Gap — systematic absence, under-representation, or distortion of data on women and girls — perpetuates a male default in AI, policy, and practice. This brief synthesizes 13 empirical cases (2020–2025) across four bias types — representative, algorithmic, cultural, intersectional — and outlines key drivers (historically male-centric data, digital divide, methodological limits on intersectionality, underinvestment). It consolidates actionable guidance: developers should diversify datasets, integrate fairness mechanisms, and monitor bias; organisations should require pre/post-deployment audits, establish AI ethics committees, and hire diversely; regulators should classify HR/finance as high-risk, mandate gender impact statements, and enforce algorithmic transparency.

Table of contents

1. Definition & Framework

The Gender Data Gap is the systematic absence, under-representation, or distortion of data about women and girls across how information is collected, represented, analyzed, and used. Because the evidence base that informs organisational and technological decisions is often skewed toward (white) men — missing, incomplete, or lower-quality for women — the resulting models, policies, and products are calibrated to male bodies, preferences, and life paths and fail to capture women's experiences, needs, and contributions.[1,2]

Gender statistics are defined by data that are: (1) collected and presented by sex as primary classification, (2) reflect gender issues, (3) based on concepts that adequately capture diversity of women and men's lives, and (4) use collection methods that account for stereotypes and cultural factors that may induce gender bias.[3]

2. Relevance & Impact

The Gender Data Gap skews decisions across society and management science by normalizing a male default rather than a population-representative view. It has real safety and health costs, because male-oriented data and training materials correlate with worse outcomes for women. It also slows workplace equality, as policies and measurement scales calibrated on male patterns disadvantage women and reproduce leadership gaps. This gap undermines claims of "neutral" management theory, since

[1] Sperber et al., "Gender Data Gap and Management Science," 2–8.

[2] PARIS21 and UN Women, Gender Data Outlook 2024.

[3] United Nations Statistics Division, Integrating a Gender Perspective into Statistics.

canonical constructs are often normed on male behaviour, shifting attention to "fix-the-women" rather than structural change. The AI wave amplifies these problems: systems trained on male-skewed or unrepresentative traces learn proxies for gender and automate discrimination at scale — making closure of the gap urgent.[4]

Availability does not guarantee use. Although the supply of gender data has expanded across many sectors, uptake remains concentrated in a few established areas, notably violence against women and unpaid care. Many other domains — including those linked to AI — still see limited application.

To broaden use, awareness and access must improve. Dissemination should be purposeful, linking producers and users to stimulate wider uptake and reveal the needs of new user groups. Clear, plain-language presentation of gender data is essential to reach the general public and stakeholders with lower data literacy.

„An investment in gender data is ultimately an investment in the lives of women, girls, boys and men".[5]

3. Methodology

This evidence brief employed a systematic approach to identify documented cases of gender bias in AI systems. A comprehensive literature search was conducted across major academic databases (Web of Science, Scopus, PubMed, Google Scholar) covering publications from 2018 to 2025, using targeted keywords combining "artificial intelligence," "algorithmic bias," and "gender discrimination." From an initial corpus of over 30 relevant studies, 13 representative cases were selected based on empirical evidence of measurable gender disparities in AI outcomes. Key limitations

[4] Sperber et al., " Gender Data Gap and Management Science"

[5] PARIS21 and UN Women, Gender Data Outlook 2024

include focus on English-language sources and compressed project timeframes that limited systematic review protocols.

4. Causes of the Gender Data Gap

Based on academic research, the Gender Data Gap stems from multiple interconnected factors:

- **Historical male-centric data collection practices** — Systematic prioritization of men's experiences, bodies, and life patterns as the default norm in research design and data gathering[6]
- **Digital divide and technological exclusion** — Women's limited access to smartphones, internet, and digital platforms reducing their representation in increasingly important digital data sources[7]
- **Methodological challenges in capturing intersectionality** — Difficulty in designing research frameworks that adequately represent the diversity of women's experiences across race, class, disability, and other identity markers[8].
- **Institutional underinvestment in gender-specific research** — Systematic underfunding of studies focused on women's experiences and gender-disaggregated analysis[9].
- **Biases in research and policy frameworks** — Male-dominated academic and policy institutions perpetuating research priorities that reflect masculine perspectives and concerns[10].

[6] Sperber et al., "Gender Data Gap and Its Impact on Management Science — Reflections from a European Perspective."

[7] Musizvingoza, "Bridging the Gender Data Gap: Harnessing Synthetic Data for Inclusive AI," UNU Macau (blog).

[8] Buvinic et al., Mapping Gender Data Gaps.

[9] Car Criado Perez, Invisible Women: Exposing Data Bias in a World Designed for Men.

[10] Sperber et al., "Gender Data Gap and Its Impact on Management Science — Reflections from a European Perspective."

5. Illustrative Cases

We have categorized the identified cases into four types of bias—representative, algorithmic, cultural, and intersectional—where bias refers to systematic skew in data, models, or design that disadvantages certain groups.

5.1. Representative bias

Occurs when datasets under-or over-represent specific demographic groups, leading to poorer model performance for those groups.

Case 1: Commercial face-analysis misclassifies darker-skinned women up to 34.7% vs 0.8% for lighter-skinned men. Face recognition errors hit darker-skinned women hardest, increasing false matches in policing and border control.

An intersectional audit of three commercial gender classification systems (Microsoft, IBM, Face++) revealed severe algorithmic bias. The study found that darker-skinned women were misclassified at rates up to 34.7% compared to just 0.8% for lighter-skinned men. Using the newly created Pilot Parliaments Benchmark dataset balanced by gender and skin type, researchers exposed that existing datasets were overwhelmingly composed of lighter-skinned subjects (79.6-86.2%). Despite comprising only 21.3% of the dataset, darker-skinned females accounted for 61-72% of all classification errors, with maximum subgroup disparities reaching 34.4% between best and worst classified groups[11].

[11] Buolamwini and Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in Proceedings of the 1st Conference on Fairness, Accountability, and Transparency, 77–91.

Case 2: DALL·E 2 underrepresents women (38% vs 62% men) and shows women smiling ~2.2× more. In female-dominated jobs women are more often pictured with downward head pitch (subordination cue).

An audit of DALL·E 2 (15300 images across 153 US occupations) finds both representational and presentational gender bias: women appear in only 38. 4% of occupational images (vs 46. 4% in Google Images), with underrepresentation in male-dominated fields and overrepresentation in female-dominated roles; DALL·E 2 matches census gender gaps in 88 occupations (Google: 18), indicating amplification. Presentationally, women are 2. 19× more likely to be shown smiling, and in female-dominated jobs more often depicted with downward head pitch (a subordination cue). These effects exceed those in Google Images, suggesting DALL·E 2 not only reproduces but amplifies occupational gender stereotypes — through both who is shown and how they are portrayed.[12]

Case 3: AI STEM images often show 75–100% men, reinforcing "STEM = male."

A UNDP Serbia policy analysis of AI image generators finds systematic underrepresentation of women in STEM: AI-generated STEM images show men in 75–100% of visuals, despite women comprising ~28–40% of STEM graduates globally. These systems reproduce — and often amplify — inequalities: prompts for "engineer/mathematician/scientist" predominantly yield male figures, reinforcing "STEM = male." Such visuals risk self-fulfilling effects by shaping aspirations and professional identity, potentially lowering

[12] Sun et al., "Smiling Women Pitching Down: Auditing Representational and Presentational Gender Biases in Image-Generative AI."

women's recruitment and retention in STEM. UNDP warns that without deliberate human oversight, AI can worsen existing disparities.[13]

Case 4: Across generators, professionals are shown as men 76% vs women 8%; only 7% of doctors depicted are women.

A nine-generator audit shows strong gender bias in professional depictions: with neutral prompts across law, medicine, engineering, and research, men appear in 76% of images and women in 8%; excluding ambiguous/non-human figures, 91% of identifiable professionals are coded male. In medicine, only 7% of doctors shown are women, despite near-parity in reality. The pattern holds across tools, with one exception (HotPotAI: 67% women). These results indicate that AI image generators perpetuate occupational gender stereotypes and may shape career perceptions as such imagery spreads.[14]

5.2. Algorithmic bias:

Results from model objectives, loss functions, or decision thresholds that optimize for overall accuracy without ensuring fairness across all groups.

Case 5: Amazon's experimental hiring AI down-ranked CVs with "women's" terms after learning from male-dominated histories.

A review of AI hiring (49 studies + 10 interviews) shows systematic bias across gender, race, skin tone, and personality, rooted in male-skewed historical data and designers' feature/target choices. Illustrative case: Amazon's experimental recruiter (ended in 2018), trained on a decade of male-dominated CVs, learned to penalize "women's" terms and down-rank female candidates. Mitigations span representative training

[13] United Nations Development Programme Serbia, "Reproducing Inequality: How AI Image Generators Show Biases Against Women in STEM," UNDP Serbia Blog.

[14] Gorska and Jemielniak, "The Invisible Women: Uncovering Gender Bias in AI-Generated Images of Professionals," 4370–75.

data, model transparency and audits, and organisational governance with external oversight (e.g., GDPR).[15].

Case 6: Music recommenders are more accurate for men — up to 6.3 percentage points better.

Using a large music-listening dataset (~2 billion events), recommenders were up to 6.3 percentage points more accurate for men; women made up ~21% of users. The highest-accuracy models showed the biggest gaps, and standard debiasing cut them only slightly — revealing a persistent tension between accuracy and fairness.[16].

Case 7: "Gender-blind" credit scoring approves fewer and smaller loans for women. Despite better repayment by women, profit-optimized models favor larger loans and tilt access toward men.

A qualitative study of 25 fintech executives and data scientists shows "gender-blind" ML credit scoring approves fewer and smaller loans for women. Across markets, women make up ~20–35% of borrowers despite better repayment; models infer gender via proxies (income, job formality, smartphone use). A digital divide shrinks the female applicant pool (women 8% less likely to own phones; 20% less likely to use mobile internet). Profit-optimized objectives (e.g., lifetime value) favor larger loans and tilt access toward men — embedding "encoded gender norms" and creating an "objective algorithm paradox[17].

Case 8: AI heart-disease tools trained on male-heavy data miss symptoms in women, delaying diagnosis.

[15] Chen, "Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices."

[16] Melchiorre et al., "Investigating Gender Fairness of Recommendation Algorithms in the Music Domain."

[17] Smith, "Mindsets and Management: AI and Gender (In)Equitable Access to Finance," manuscript accepted for ACM FAccT 2025.

This study revealed systematic gender discrimination in GPT-4's coronary artery disease risk assessment when psychiatric comorbidities were introduced. Despite identical clinical presentations, nearly half of women experienced reduced risk assessments when psychiatric conditions were added to their cases — a bias not observed equally in men. This suggests GPT-4 systematically undervalues women's chest pain symptoms as psychological rather than cardiac, echoing documented clinical biases where women are perceived as more "emotional." The findings are particularly concerning given that no scientific evidence supports differential CAD risk between genders for psychiatric conditions, revealing the algorithmic bias as medically unfounded and highlighting urgent needs for bias detection systems in AI healthcare applications.[18]

5.3. Cultural bias

Arises from embedding social stereotypes and historical norms into system design, such as choice of names, voices, or role assumptions.

Case 9: Llama 2 assigns women domestic roles 4× more often than men; men get higher-status jobs.

A UNESCO-commissioned audit of LLMs (GPT-2, GPT-3. 5, Llama 2) finds systemic gender bias: female names cluster with domestic terms (“home, ” “family, ” “children”) while male names link to professional ones (“business, ” “executive, ” “career”); in story generation Llama 2 places women in domestic roles ~4× more than men and assigns men higher-status jobs; intersectionally, 20% of stories about Zulu women depict them as domestic servants/cooks/housekeepers; open-source models show the strongest bias, with ~20% of outputs containing explicitly sexist content.[19].

[18] Achtari et al., “Gender Bias in AI’s Perception of Cardiovascular Risk.”

[19] UNESCO and IRCAI, “Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models.”

Case 10: Default female personas in voice assistants reinforce "women as helpers" stereotypes.

A comprehensive analysis of AI voice assistants and chatbots reveals systematic adoption of feminine personas through female names, voices, and caregiving behaviours across major systems including Siri, Alexa, Cortana, and ChatGPT. Companies justify these choices by appealing to user preferences for "warm, trustworthy" voices, while feminist scholars warn this reinforces stereotypes of women as "submissive and overly accommodating aides. " The research demonstrates that this "feminization" is an intentional, historically grounded design phenomenon with material social consequences — voice-perception studies confirm users attribute nurturing qualities to female voices and authority to male voices, organizing human-AI interaction along gendered lines. When AI assistants act as helpful young women, they legitimate the idea that women's primary role is to please and serve others, perpetuating the "assistant = woman" equation as generative AI becomes increasingly pervasive in homes, cars, and devices.[20]

5.4. Intersectional bias

Emerges when multiple factors (e.g., gender plus race, dialect, or age) combine to amplify discrimination beyond individual dimensions.

Case 11: YouTube auto-captions make more errors for women and perform worst on Scottish English.

An experimental audit of YouTube's automatic captioning system revealed systematic biases against both women and speakers of certain dialects. The study analyzed 80 native English speakers (8 women and 8 men from each of 5 regions:

[20] Dave and Kushwah, "The Feminization of Generative AI: Gender Delegation Through Names and Personas," TechRxiv preprint.

California, Georgia, New England, New Zealand, and Scotland) using "accent tag challenge" videos.

The results demonstrated significant gender bias, with higher word-error rates for women than men across all dialect groups. No dialect showed reliably better accuracy for women. Scottish English speakers experienced the worst performance overall, while the combination of being both female and speaking certain dialects created compounded disadvantages[21].

Case 12: Gender-neutral STEM ads reached men ~20% more; women 25–34 were 40% less likely to see them.

Beyond individual user bias, research on advertising algorithms demonstrates systemic discrimination in content delivery. A field study (Facebook's advertising platform, gender-neutral STEM career advertisements across 191 countries) found that men received approximately 20% more STEM job advertisement impressions despite gender-neutral targeting, with women aged 25-34 being 40% less likely to see such opportunities due to algorithmic cost-optimization prioritizing cheaper demographics[22].

Case 13: Book recommenders shift exposure toward male authors.

Similarly, book recommendation algorithms systematically shift author exposure patterns based on gender, with different models producing varying degrees of bias toward male authors.[23]

[21] Tatman, "Gender and Dialect Bias in YouTube's Automatic Captions," in Proceedings of the First Workshop on Ethics in Natural Language Processing.

[22] Lambrecht and Tucker, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," 2966–2981.

[23] Ekstrand and Kluver, "Exploring Author Gender in Book Rating and Recommendation," 377–420.

6. Best Practices and Recommendations

This section presents best practices for mitigating gender bias in AI — practices that society must demand from companies, developers, and policymakers. We share these recommendations with students, faculty, and librarians because the education system shapes the future decision-makers who will implement and uphold these standards.

6.1. Recommendations for AI Developers

- **Ensure dataset diversity**[24]

Curate your datasets to include balanced representations of different genders, age groups, skin tones, and cultural backgrounds. This approach promotes equitable model performance across all demographic segments.

Integrate algorithmic fairness mechanisms[25, 26, 27]

Incorporate post-processing techniques — such as score normalization or adversarial debiasing — that adjust model outputs to achieve parity between male and female outcomes without necessitating full model retraining.

- **Implement continuous bias monitoring**[28]

Establish ongoing evaluation pipelines to track performance disparities across demographic groups. Regularly audit model predictions to detect and address any degradation in accuracy for underrepresented populations.

[24] Buolamwini and Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in Proceedings of the 1st Conference on Fairness, Accountability, and Transparency, 77–91.

[25] Terhörst et al., "Post-Comparison Mitigation of Demographic Bias in Face Recognition Using Fair Score Normalization," arXiv preprint.

[26] Dhar et al., "PASS: Protected Attribute Suppression System for Mitigating Bias in Face Recognition," arXiv preprint.

[27] Thakur et al., "Language Models Get a Gender Makeover: Mitigating Gender Bias with Few-Shot Data Interventions," arXiv preprint.

[28] Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

6.2. Recommendations for Organizations

- **Mandate pre- and post-deployment audits**[29]

Conduct systematic technical and procedural evaluations before and after AI system deployment to identify and mitigate emergent biases. Ethical guidelines advocate for regular multidisciplinary audits encompassing data, algorithms, and user impacts.

- **Establish AI ethics committees**[30]

Form dedicated interdisciplinary bodies to oversee AI governance, integrating technical, legal, and social expertise. Such committees ensure ongoing accountability and alignment with organisational values and ethical standards.

- **Invest in diverse hiring practices**[31,32,33]

Prioritize recruitment of professionals from varied demographic and disciplinary backgrounds to enhance team perspectives and reduce blind spots in AI development.

6.3. Recommendations for Regulators

- **Classify HR and financial services as high-risk sectors**[34, 35, 36, 37]

[29] Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

[30] Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

[31] Boinodiris, "The Importance of AI Diversity: Driving Trustworthy AI," IBM Consulting.

[32] Bradford, "Why Diversity in AI Makes Better AI for All: The Case for Inclusivity and Innovation," SHRM.

[33] Suighi and Rachel, "Breaking the Echo Chamber: Why Diversity Is Crucial for AI's Future," AIM Research Council.

[34] European Union, "AI Act," Shaping Europe's Digital Future.

[35] Crisanto et al., "Regulating AI in the Financial Sector: Recent Developments and Main Challenges," Bank for International Settlements (FSI Insights).

[36] van der Merwe and Veldsman, "AI Risk Management for HR: 3 Key Risks To Manage & HR Actions To Take," AIHR.

[37] Soleimani et al., "Reducing AI Bias in Recruitment and Selection: An Integrative Grounded Approach," The International Journal of Human Resource Management.

Due to their direct impact on employment, credit access, and economic opportunities, AI applications in hiring and lending warrant stricter oversight and mandatory bias mitigation requirements.

- **Require gender impact statements**[38]

Mandate that AI developers and deployers publish assessments detailing potential gender-related harms, data gaps, and mitigation plans. These statements, grounded in human-rights frameworks, enable transparency and accountability throughout the AI lifecycle.

- **Enforce algorithmic transparency standards**[39]

Establish legal requirements for disclosing AI decision-making logic, key variables, and performance metrics. Providing explainable outputs improves stakeholders' ability to detect and counteract bias in critical contexts.

[38] Jobin, Ienca, and Vayena, "The Global Landscape of AI Ethics Guidelines."

[39] Hou, Tseng, and Yuan, "Is This AI Sexist? The Effects of a Biased AI's Anthropomorphic Appearance and Explainability on Users' Bias Perceptions and Trust," International Journal of Information Management.

Acknowledgements — Use of AI tools

We used an AI assistant (ChatGPT, GPT-5 Thinking; Sept 2025) for drafting support (summarization, copy-editing, clarity). We also used Perplexity for literature discovery. All sources cited in this brief were located in peer-reviewed or official venues and independently verified by the authors. Generative outputs were not treated as evidence. No personal or confidential information was entered into these tools.

References

Sperber, Sonja, Susanne Täuber, Corinne Post, and Cordula Barzantny. 2023. "Gender Data Gap and its Impact on Management Science — Reflections from a European Perspective." *European Management Journal* 41, no. 12-8. https://doi.org/10.1016/j.emj.2022.11.006.

PARIS21 and UN Women. *Gender Data Outlook 2024: Unlocking Capacity, Driving Change*. Paris: PARIS21, 2024.

United Nations Statistics Division. *Integrating a Gender Perspective into Statistics*. New York: United Nations, 2016. Accessed August 23, 2025. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/gender/Integrating-a-Gender-Perspective-into-Statistics-E.pdf.

Ronald, Musizvingoza. "Bridging the Gender Data Gap: Harnessing Synthetic Data for Inclusive AI." *UNU Macau* (blog). August 2, 2024. https://unu.edu/macau/blog-post/bridging-gender-data-gap-harnessing-synthetic-data-inclusive-ai.

Buvinic, Mayra, Rebecca Furst-Nichols, and Gayatri Koolwal. *Mapping Gender Data Gaps*. Washington, DC: Data2X, March 2014. http://data2x.org/wp-content/uploads/2019/05/Data2X_MappingGenderDataGaps_FullReport.pdf.

Criado Perez, Caroline. Invisible Women: Exposing Data Bias in a World Designed for Men. London: Chatto & Windus, 2019.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency*. Proceedings of Machine Learning Research 81, 2018. https://proceedings.mlr.press/v81/buolamwini18a.html.

Chen, Zhisheng. "Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices." *Humanities and Social Sciences Communications* 10 (2023): 567. https://doi.org/10.1057/s41599-023-02079-x.

Tatman, Rachael. "Gender and Dialect Bias in YouTube's Automatic Captions." In *Proceedings of the First Workshop on Ethics in Natural Language Processing*. Valencia, Spain: Association for Computational Linguistics, 2017. https://aclanthology.org/W17-1606/

UNESCO and IRCAI. "Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models." Paris: United Nations Educational, Scientific and Cultural Organization; Ljubljana: International Research Centre on Artificial Intelligence under the auspices of UNESCO, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000388971

Sun, Luhang, Mian Wei, Yibing Sun, Yoo Ji Suh, Liwei Shen, and Sijia Yang. "Smiling Women Pitching Down: Auditing Representational and Presentational Gender Biases in Image-Generative AI." *Journal of Computer-Mediated Communication* 29, no. 1 (2024): zmad045. https://doi.org/10.1093/jcmc/zmad045.

United Nations Development Programme Serbia. "Reproducing Inequality: How AI Image Generators Show Biases Against Women in STEM." Policy analysis. UNDP Serbia Blog. Belgrade: UNDP Serbia, 2024. https://www.undp.org/serbia/blog/reproducing-inequality-how-ai-image-generators-show-biases-against-women-stem.

Gorska, Anna M., and Dariusz Jemielniak. "The Invisible Women: Uncovering Gender Bias in AI-Generated Images of Professionals." *Feminist Media Studies* 23, no. 8 (2023): 4370–4375. https://doi.org/10.1080/14680777.2023.2263659.

Melchiorre, Alessandro Benedetto, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. "Investigating Gender Fairness of Recommendation Algorithms in the Music Domain." *Information Processing & Management* 58, no. 5 (2021): 102666. https://doi.org/10.1016/j.ipm.2021.102666.

Lambrecht, Anja, and Catherine Tucker. "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads." *Management Science* 65, no. 7 (2019): 2966–2981. https://doi.org/10.1287/mnsc.2018.3093.

Ekstrand, Michael D., and Daniel Kluver. "Exploring Author Gender in Book Rating and Recommendation." *User Modeling and User-Adapted Interaction* 31, no. 3 (2021): 377–420. https://doi.org/10.1007/s11257-020-09284-2.

Melchiorre, Alessandro B., Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. "Investigating Gender Fairness of Recommendation Algorithms in the Music Domain." *Information Processing & Management* 58, no. 5 (2021): 102666. https://doi.org/10.1016/j.ipm.2021.102666.

Smith, Genevieve. "Mindsets and Management: AI and Gender (In)Equitable Access to Finance." Manuscript accepted for presentation at ACM FAccT 2025. Updated July 17, 2025. https://arxiv.org/pdf/2504.07312.

Achtari, Margaux, Adil Salihu, Olivier Muller, Emmanuel Abbé, Carole Clair, Joëlle Schwarz, and Stephane Fournier. "Gender Bias in AI's Perception of Cardiovascular Risk." *Journal of Medical Internet Research* 26 (2024): e54242. https://doi.org/10.2196/54242.

Dave, Nigam, and Shrishti Kushwah. "The Feminization of Generative AI: Gender Delegation Through Names and Personas." TechRxiv preprint. June 15, 2025. https://www.techrxiv.org/users/928336/articles/1299795-the-feminization-of-generative-ai-gender-delegation-through-names-and-personas.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research 81:77–91. 2018. https://proceedings.mlr.press/v81/buolamwini18a.html.

Terhörst, Philipp, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. "Post-Comparison Mitigation of Demographic Bias in Face Recognition Using Fair Score Normalization." Preprint, arXiv, February 6, 2020. https://doi.org/10.48550/arXiv.2002.03592.

Dhar, Prithviraj, Joshua Gleason, Aniket Roy, Carlos D. Castillo, and Rama Chellappa. "PASS: Protected Attribute Suppression System for Mitigating Bias in Face Recognition." Preprint, arXiv, August 9, 2021. https://doi.org/10.48550/arXiv.2108.03764

Thakur, Himanshu, Atishay Jain, Praneetha Vaddamanu, Paul Pu Liang, and Louis-Philippe Morency. "Language Models Get a Gender Makeover: Mitigating Gender Bias with Few-Shot Data Interventions." Preprint, arXiv, June 7, 2023. https://doi.org/10.48550/arXiv.2306.04597.

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, no. 2 (September 2019). https://doi.org/10.1038/s42256-019-0088-2.

Boinodiris, Phaedra. "The Importance of AI Diversity: Driving Trustworthy AI." IBM Consulting. Accessed August 23, 2025. https://www.ibm.com/think/insights/ai-diversity-importanceAuthor.

Bradford, Nichol. "Why Diversity in AI Makes Better AI for All: The Case for Inclusivity and Innovation." SHRM. October 7, 2024. https://www.shrm.org/topics-tools/flagships/ai-hi/why-diversity-in-ai-makes-better-ai-for-all--the-case-for-inclus.

Suighi, Giorgio, and Rachel. "Breaking the Echo Chamber: Why Diversity Is Crucial for AI's Future." AIM Research Council. January 8, 2025. https://council.aimresearch.co/breaking-the-echo-chamber-why-diversity-is-crucial-for-ais-future/.

Hou, Tsung-Yu, Yu-Chia Tseng, and Chien Wen (Tina) Yuan. "Is This AI Sexist? The Effects of a Biased AI's Anthropomorphic Appearance and Explainability on Users 'Bias Perceptions and Trust." *International Journal of Information Management* 76 (2024). https://doi.org/10.1016/j.ijinfomgt.2024.102775.

European Union. "AI Act." Shaping Europe's Digital Future. July 31, 2025. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

Crisanto, Juan Carlos, Cris Benson Leuterio, Jermy Prenio, and Jeffery Yong. "Regulating AI in the Financial Sector: Recent Developments and Main Challenges." FSI Insights on policy implementation, no. 63. December 2024. Bank for International Settlements. https://www.bis.org/fsi/publ/insights63.pdf.

van der Merwe, Marna, and Dieter Veldsman. "AI Risk Management for HR: 3 Key Risks To Manage & HR Actions To Take." AIHR. Accessed August 23, 2025. https://www.aihr.com/leading-hr/ai-risk-management/.

Soleimani, Mohammad, Ali Intezari, Julia Arrowsmith, Daniel J. Pauleen, and Nesrin Taskin. "Reducing AI Bias in Recruitment and Selection: An Integrative Grounded Approach." *The International Journal of Human Resource Management* (2025). https://doi.org/10.1080/09585192.2025.2480617.