ChatGPT

Relevant Literature on Feminist AI Literacies and Diversity-Reflective Prompting (2023–2025)

1. **Shah, S. S. (2025).** ***Gender bias in artificial intelligence: Empowering women through digital literacy*****. Premier Journal of Artificial Intelligence, 1, Article 1000088.** DOI: 10.70389/PJAI.1000088
   Summary: This narrative review examines how systemic gender biases are embedded in AI systems across domains (e.g. hiring, healthcare, finance) and explores digital literacy as a tool to combat these biases [1]. Key findings indicate that biases arise from underrepresentation of women in AI development, biased training data, and algorithmic design choices [1]. Digital literacy programs for women are highlighted as a promising intervention: they raise critical awareness of AI bias, encourage women's participation in AI careers, and foster women-led AI projects [1]. The review concludes that improving feminist digital competencies can be transformative for AI equity, emphasizing inclusive AI design, gender-responsive education, and sustained efforts to mitigate bias [2].
   Quality: Medium. Published in a new peer-reviewed journal, this article underwent external peer review [3]. The **Premier Journal of Artificial Intelligence** is a relatively new outlet (volume 1), so its impact and reputation are not yet established. The methodology is a **narrative literature review** with a systematic search and thematic analysis, lending it a solid scholarly foundation. The work is comprehensive (91 references) and clearly relevant to the topic. However, as a very recent publication (Jan 2025) it has no citation record yet, and the journal's impact factor is unknown, tempering its perceived influence. Overall, the content is well-structured and pertinent, but the source's newness and limited track record suggest a moderate confidence level in its authority.

2. **Fraile-Rojas, B., De-Pablos-Heredero, C., & Méndez-Suárez, M. (2025).** ***Female perspectives on algorithmic bias: Implications for AI researchers and practitioners*****. Management Decision.* (Advance online publication).** DOI: 10.1108/MD-04-2024-0884
   Summary: This study uses NLP and machine learning to analyze 172,041 tweets (over 1 year) from female users discussing gender inequality in AI [4]. It identifies prominent themes, notably *the future of AI technologies* and *women's active role in ensuring gender-balanced systems* [5]. The findings show that algorithmic bias directly affects women's experiences, prompting them to engage in online discourse about injustices [5]. Women in these digital dialogues often lead constructive conversations (frequently linking with gender/race empowerment groups) and even create entrepreneurial solutions when faced with bias [6]. The authors note that *feminist critical thought is indispensable* for developing balanced AI systems [7]. They call for an **intersectional, collaborative approach** to AI design and policy, as female voices demand technologies that are fair and inclusive of all genders and races [8]. This underscores how feminist digital literacies – manifested in social media activism and knowledge-sharing – can make AI biases visible and push for their reduction.
   Quality: High. This article appears in **Management Decision**, a well-established peer-reviewed journal (Scopus Q1; Impact Factor ~5.1 [9]). The study's methodology is **robust**, combining social media mining, sentiment analysis, and clustering to derive insights from a large dataset, which strengthens the reliability of its conclusions. The journal's strong reputation and the article's comprehensive reference list (91 sources) indicate thorough scholarship. Although only recently published (Jan 2025) with minimal citations so far (1 noted), its placement in a high-

impact journal and clear empirical approach suggest significant influence. The text is well-written, and the topic is directly relevant to bias in AI, making this source a credible and valuable contribution.

3. **Jääskeläinen, P., Sharma, N. K., Pallett, H., & Åsberg, C. (2025). *Intersectional analysis of visual generative AI: The case of Stable Diffusion*. AI & Society.* Advance online publication. DOI: 10.1007/s00146-025-02207-y**
   Summary: **This open-access paper provides a feminist intersectional critique of a popular generative AI system (Stable Diffusion) through qualitative visual analysis of 180 AI-generated images [10] [11]. The authors deliberately prompted images representing various intersecting social categories (e.g. wealth/poverty, citizen/immigrant) and examined how power systems like racism, sexism, heteronormativity, ableism, colonialism, and capitalism are reflected and amplified in the outputs [12]. They found that Stable Diffusion's default outputs frequently** perpetuate harmful stereotypes **and assume a default subject position of "white, able-bodied, masculine-presenting" [11]. The imagery tended toward Eurocentric and North American cultural aesthetics, revealing implicit biases rooted in the training data and institutional context of the tool [13]. Crucially, the paper argues that generative AI is *not culturally neutral* but mirrors and reinforces societal power imbalances [13]. To counter this, the authors advocate a** social justice-oriented approach **to AI: first, by** acknowledging and rendering visible **these cultural-aesthetic biases in AI outputs, and second, by engaging in "reparative" strategies to symbolically and materially mend the injustices inflicted on marginalized groups [14]. This intersectional lens illustrates how feminist AI literacy (critical awareness of how AI reproduces structural inequalities) can unveil hidden biases and inform interventions to reduce them.**
   Quality: High. **Published in** AI & Society **(a peer-reviewed Springer journal focusing on social impacts of AI), this article has strong scholarly credentials. The journal is reputable in the technology and society domain (established track record, Scopus Q2, and known editorial standards). The study's methodology is a** rigorous qualitative analysis**, drawing on feminist STS and critical theory to interpret AI outputs—appropriate for uncovering nuanced cultural biases. The text is richly theorized and well-supported by references to both technical and feminist literature. While very recent (March 2025) with no citation data yet, the work's open access availability and timely subject matter likely enhance its reach. The clarity of findings and direct relevance to intersectional bias in AI signal a significant contribution. Overall, the combination of a credible venue, novel insights, and methodological depth justifies a high quality assessment.

4. **Skilton, R., & Cardinal, A. (2024). *Inclusive prompt engineering: A methodology for hacking biased AI image generation*. In** Proceedings of the 42nd ACM International Conference on Design of Communication (SIGDOC '24) **(pp. 76–80). ACM.** DOI: 10.1145/3641237.3691655
   Summary: This conference paper introduces "inclusive prompt engineering" as a strategy to probe and mitigate biases in generative AI image systems. The authors developed a methodology to "hack" biased image generation by systematically modifying prompts and providing users with tools to generate more diverse outputs. For example, they suggest using AI to offer a list of **alternative descriptors** that users can incorporate into their prompts in order to steer image generation away from default stereotypes [15]. In user studies, they observed that when participants encountered stereotypical or undesired outputs, they tried adding negative qualifiers (e.g. *"without X"*) to prompts – but current models often failed to obey these negations, revealing a gap between user intent and the model's behavior [16]. This indicates that naive prompting alone cannot always overcome ingrained biases. The paper's findings underscore the need for improved prompt interfaces and workflows that actively promote inclusive representation in AI outputs. By empowering users with diversity-reflective prompting

techniques (such as suggested attributes to include or avoid), **feminist AI literacy** practices can help make biases visible (by deliberately eliciting and examining biased outputs) and then reduce those biases (by creatively rephrasing or expanding prompts to counteract them).
Quality: Medium. This work was **peer-reviewed** and presented at an ACM conference (SIGDOC 2024), ensuring a baseline of scholarly quality. Conferences in human-computer interaction and design of communication, while not as highly ranked as top-tier CS conferences, still vet submissions rigorously; the acceptance of this paper suggests its methods and insights were found credible. The authors' approach is innovative, though the page length (5 pages) indicates it's a concise contribution. The conference proceedings do not have an impact factor, but ACM digital library listings show the paper has already attracted several citations (7 by early 2025), suggesting it's influencing subsequent research. Methodologically, the combination of **user-centered experimentation and interface design proposals** is sound for exploring prompt bias. The text is clear and relevant, though less expansive than a journal article. Given the novelty of the topic and the solid peer-review, this source is of good quality, albeit with the inherent limitations of a short conference paper.

5. **Djeffal, C. (2025).** ***Reflexive prompt engineering: A framework for responsible prompt engineering and AI interaction design*****. In** Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) **(pp. 1757–1768). ACM.** DOI: 10.1145/3715275.3732118
Summary: This FAccT conference paper proposes "Reflexive Prompt Engineering" as a comprehensive framework to embed ethical and inclusive principles directly into the way we craft prompts for generative AI. Djeffal argues that prompt engineering should go beyond optimizing functionality and actively incorporate **fairness, accountability, transparency, and societal values** [17]. The framework consists of five interconnected components – prompt design, system selection, system configuration, performance evaluation, and prompt management – each considered from a perspective of social responsibility [18]. Practically, this means that those developing or using AI systems can modify and manage prompts in order to **prevent discriminatory outcomes and promote inclusive representation** in AI outputs [19]. For example, prompts might be formulated or adjusted to explicitly counteract biases (e.g. asking the model to include diverse genders or ethnicities in a scenario) and include validation steps to ensure accessibility [19]. The paper highlights that achieving this requires balancing technical precision with ethical consciousness [20]. Ultimately, Djeffal positions responsible prompt engineering as *"an essential component of AI literacy"*, bridging the gap between AI development and deployment by empowering stakeholders to align AI behavior with human rights and diversity values during the prompt-crafting process [21]. This forward-looking approach illustrates how **diversity-reflective prompting**, grounded in feminist and intersectional awareness, can systematically reduce bias in AI systems by design.
Quality: High. This paper was peer-reviewed and accepted at **ACM FAccT 2025**, one of the premier international conferences on AI fairness and accountability. FAccT has a rigorous selection process (low acceptance rate) and a strong reputation (CORE A rank), indicating that the work underwent thorough scrutiny by experts. The content is largely conceptual but draws on empirical observations and case studies, lending it credibility. As a single-author piece, it synthesizes interdisciplinary insights from computer science, ethics, and law, reflecting methodological robustness in constructing the framework. The publication is very recent (presented in mid-2025), so citation data is not yet available; however, being part of FAccT proceedings suggests it will be influential in the responsible AI community. The writing is of high quality—clear and well-argued—and the topic is directly relevant to bias mitigation via prompt literacy. Given the venue's prestige and the comprehensive nature of the framework, this source is of high quality.

6. **Shin, P. W., Ahn, J. J., Yin, W., Sampson, J., & Narayanan, V. (2024). *Can prompt modifiers control bias? A comparative analysis of text-to-image generative models*. arXiv preprint arXiv: 2406.05602.** https://arxiv.org/abs/2406.05602
   Summary: This preprint study investigates whether adding explicit **prompt modifiers** can reduce societal biases in text-to-image generative AI models. The authors evaluated three leading models (Stable Diffusion, DALL·E 3, and Adobe Firefly) by comparing their output images for baseline prompts versus bias-mitigating prompts (e.g. adding descriptors like race or gender) [22]. The analysis revealed that all models exhibited notable biases – for instance, a simple prompt "monk" yielded predominantly Asian male images across models, whereas Firefly's outputs were more gender-balanced [22] [23]. When modifiers were introduced (e.g. "monk who is Black"), some models still defaulted to stereotypes (Stable Diffusion and DALL·E often continued showing Asian men despite the prompt) [24]. **Prompt engineering showed potential to adjust bias**, but results were inconsistent: certain biases persisted or the effectiveness depended on the phrasing/order of modifiers [25]. The study highlights the *challenges and limits* of diversity-reflective prompting – while it can expose hidden model biases and sometimes nudge outputs towards inclusivity, it is not a comprehensive fix [22] [25]. The authors call for the development of standardized metrics and more complex strategies to control AI biases [26]. In sum, this work demonstrates how *feminist AI literacy in practice (through creative prompt experiments)* can make biases visible, and also shows that such prompting must be combined with broader ethical AI development efforts for truly bias-resistant systems [22].
   Quality: Medium. As an **arXiv preprint**, this paper has not yet undergone formal peer review, which introduces some uncertainty about its rigor. Nevertheless, the research appears methodologically sound: it involves a comparative experiment with multiple models and both qualitative and quantitative bias evaluations, supported by an NSF grant [27]. The authors are affiliated with a reputable institution (Pennsylvania State University) and the work is detailed (the full manuscript spans ~20 pages with extensive data and a proposed taxonomy for bias sensitivity). The preprint has garnered attention in the community for its timely topic, but its citation count and impact are not fully established (it serves as a foundation for future research rather than a validated end product). The text quality is high – well-structured and clearly reporting findings. Relevance to the topic is direct, as it provides empirical evidence on prompt-based bias mitigation. In summary, while the content is strong, the lack of peer-review means this source should be viewed with moderate caution; it offers valuable insights but awaits validation through formal publication (which would likely further solidify its credibility).

---

[1] [2] [3] Gender Bias in Artificial Intelligence: Empowering Women Through Digital Literacy - Premier Science
https://premierscience.com/pjai-24-524/

[4] [5] [6] [7] [8] [9] Female perspectives on algorithmic bias: implications for AI researchers and practitioners | CoLab
https://colab.ws/articles/10.1108%2Fmd-04-2024-0884

[10] [11] [12] [13] [14] Intersectional analysis of visual generative AI: the case of stable diffusion | AI & SOCIETY
https://link.springer.com/article/10.1007/s00146-025-02207-y

[15] [16] [17] [18] [19] [20] [21] Inclusive Prompt Engineering: A Methodology for Hacking Biased AI Image Generation
https://www.researchgate.net/publication/385325948_Inclusive_Prompt_Engineering_A_Methodology_for_Hacking_Biased_AI_Image_Generation

22 23 24 25 26 27 Can Prompt Modifiers Control Bias? A Comparative Analysis of Text-to-Image Generative Models

https://arxiv.org/html/2406.05602v1