*Corresponding Author's Email: suni@umh.es
Proceedings of the Global Conference on Women's Studies
Vol. 4, Issue. 1, 2025, pp. 71-94
DOI: https://doi.org/10.33422/womensconf.v4i1.1390

ISSN: 2783-7777 online

Diamond Scientific Publishing

# Interdisciplinary Strategies to Mitigate Gender Bias in AI-Generated Imagery

**Maria Asunción Vicente* and César Fernández**

Miguel Hernandez University, Elche, Spain

## Abstract

This work reports preliminary findings from Project LENA, an interdisciplinary investigation of gender in images produced by artificial intelligence (AI) platforms. Employing a mixed-methods framework that combines experimental prompt engineering, quantitative-qualitative visual analysis, and critical evaluation, we examine how textual prompts, model architectures (e.g. DALL·E, Stable Diffusion, Grok), and training datasets collectively shape stereotyped gender representations. Initial results show that inclusive prompting, diversified training corpora, and heightened user awareness substantially reduce biased outputs, yielding more equitable imagery. Building on these insights, we propose practical guidelines for the ethical deployment of generative AI, with special attention to educational settings. Project LENA underscores the importance of critical digital literacy, algorithmic transparency, and an intersectional feminist perspective in fostering an inclusive digital culture that empowers educators and learners to engage reflectively and transformatively with emerging visual-generation technologies. We conclude by outlining future research that integrates human-centered design principles with policy frameworks.

**Keywords:** ethics, fairness, images, machine learning, technology

## 1 Introduction

The rapid advancement of generative artificial intelligence (GAI) has transformed visual content production, with platforms such as DALL·E, Stable Diffusion, Midjourney and Grok democratizing access to creative tools (Bommasani et al., 2021; Vincent, 2022). However,

these systems also reproduce entrenched gender and racial stereotypes, as documented in recent studies (García-Ull & Melero-Lázaro, 2023; Gorska & Jemielniak, 2023; Kalluri, 2024; Sun et al., 2024). For example, prompts associated with professions such as *"nurse"* frequently generate female figures, while *"engineer"* often produces male representations—even when no gender is specified. These outcomes illustrate how biases embedded in training datasets, algorithmic architectures, and cultural assumptions are perpetuated in AI-generated imagery.

### 1.1 Background

Some studies suggest that GAI may partially mitigate long-standing biases found in traditional image repositories. For instance, Freixà et al. (2025) compared stock photography platforms (e.g., Shutterstock, Getty Images) with AI-based repositories (e.g., Lexica, Adobe Stock). Their findings revealed that traditional databases overrepresented women, whereas AI-generated images achieved a more balanced gender distribution. In this context, the authors argue that generative systems have the potential to reduce certain historical imbalances in visual representation.

Nevertheless, the broader body of literature highlights the tendency of generative models to reproduce or even intensify stereotypes**.** Girrbach et al. (2025)**,** through a large-scale analysis, demonstrated that text-to-image systems reinforced traditional roles, depicting women primarily in caregiving contexts and men in technical or financial domains. Similarly, Sun et al. (2023) found that DALL·E 2 systematically underrepresented women in male-dominated occupations while portraying them disproportionately as young, smiling, and submissive. Expanding on this**,** Zhou et al. (2024) examined multiple platforms (Midjourney, Stable Diffusion, and DALL·E 2) and observed consistent patterns of gender and racial bias, with women often depicted as younger and more expressive, whereas men were shown as older and more serious. Early contributions such as Bianchi et al. (2023) further confirmed that generative systems amplify stereotypes even when prompts are neutral, and that technical mitigation strategies (e.g., inverse prompts, filtering) are insufficient to eliminate the problem.

Overall, the evidence paints a complex and somewhat contradictory picture. While some generative platforms may introduce a degree of balance in gender representation (Freixà et al., 2025), the majority of empirical studies indicate that these systems largely perpetuate cultural stereotypes. The persistence of these biases is attributable to several factors: the biased nature of training datasets**,** design choices that privilege certain cultural representations, and the lack of diversity within development teams**.**

### 1.2 Project LENA

Project LENA, funded by the Spanish Ministry of Equality, seeks to investigate these issues and propose strategies to mitigate gender bias in AI-generated imagery. By combining technical experimentation with critical perspectives from feminist research, the project aims to empower educators, learners, and policymakers to engage more ethically with AI technologies.

The acronym LENA was deliberately chosen as a symbolic tribute to *Lena Söderberg*. The so-called "Lena" image originated from a cropped photograph of the Swedish model published in *Playboy* magazine in November 1972, later scanned and widely adopted in the field of

digital image processing. Over the decades, this picture became one of the most frequently used benchmarks for testing and comparing algorithms in computer vision, appearing in research papers, textbooks, and conferences (The Lenna story, n.d.; Thompson, 2019). However, the long-standing reliance on this image also highlights the lack of diversity and persistence of gender bias in technology**.** Recent critiques, including the "Losing Lena" campaign, have argued that continued use of this image perpetuates exclusionary practices in the tech community (Women Love Tech, 2020). Figure 1 includes both the historical 'Lena' test image and a recent photograph of Lena Söderberg, reproduced in accordance with the ethical guidelines promoted by the 'Losing Lena' campaign and used solely for critical and educational purposes.

For this reason, we deliberately adopted the name *LENA* for our project: both to acknowledge this historical controversy and to reframe its meaning. By doing so, we aim to transform Lena Söderberg's symbolic role—from a reminder of bias to an emblem of equality and diversity in the digital age.

*Figure 1. Lena Söderberg, a Swedish model whose image—originally taken from Playboy magazine—became the most widely used test image in the history of digital image processing. Left: examples of algorithmic filtering applied to the "Lena" image. Right: a more recent photograph of Söderberg during an interview, where she expressed her fatigue with the continued use of her image. Source: Women Love Tech. (2020, November 21). Losing Lena – Why removing one image will end tech's original sin. Women Love Tech. https://womenlovetech.com/losing-lena-why-we-need-to-remove-one-image-and-end-techs-original-sin/*

Within this context, the present study is situated as part of project LENA. The project adopts a multi-phase mixed-methods design that combines technical experimentation with critical feminist analysis. This article reports exploratory findings from the initial phases of the project, focusing on illustrative case studies, prompt-based experiments, and preliminary analytical categories. While later phases will incorporate more extensive quantitative metrics and validation procedures, the present contribution seeks to establish a structured methodological and conceptual framework, and to offer early empirical insights relevant for both educational practice and policy-oriented discussions.

# 2 Materials and Methods

Project LENA is designed as an interdisciplinary, multi-phase research initiative aimed at systematically identifying, analyzing, and mitigating gender bias in AI-generated imagery. The project adopts a mixed-methods exploratory–descriptive design, combining technical experimentation in text-to-image systems with qualitative and descriptive quantitative analysis informed by feminist and critical digital studies. The methodological framework has been formally defined at the project level and is being implemented progressively across successive phases.

## 2.1 Methodological Framework and Project Phases

Project LENA is structured around a five-phase methodological framework designed to progressively identify, analyze, and mitigate gender bias in AI-generated imagery. This phased approach ensures methodological transparency, replicability, and scalability, while allowing exploratory findings to inform subsequent stages of evaluation and validation.

***Phase 1. Prompt design.***

This initial phase involved a systematic review of the literature on gender bias in artificial intelligence and visual representation. Based on this review, a set of textual prompts was developed with two complementary objectives: (a) to elicit potential gender bias in AI-generated imagery through neutral or minimally specified descriptions, and (b) to design control prompts incorporating bias-aware and inclusive specifications. These prompts constitute the empirical entry point of the study and were iteratively refined during early experimentation.

***Phase 2. Image analysis.***

In this phase, images were systematically generated across multiple widely used text-to-image platforms. The resulting outputs were examined to identify recurring visual patterns and stereotypical representations associated with gender. This phase focused on detecting stable tendencies across repeated generations rather than isolated outputs, providing the empirical basis for the identification of preliminary analytical categories.

***Phase 3. Evaluation protocol.***

Building on insights from the initial phases, Phase 3 is dedicated to the development of a standardized assessment tool for measuring gender bias in AI-generated imagery. This includes the operationalization of analytical dimensions, the construction of bias rubrics, and the definition of clear criteria to support replicability and methodological rigor. At the time of writing, this phase is under active development.

***Phase 4. Blind evaluation.***

Phase 4 involves the application of the evaluation protocol through blind assessment procedures. AI-generated images are compared with real-world visual references and evaluated by a diverse panel of human assessors to reduce subjective bias. This phase

incorporates interrater reliability analysis and constitutes the primary validation stage of the project.

***Phase 5. Dissemination and transfer.***

The final phase focuses on the dissemination of findings and the transfer of knowledge to key stakeholders. This includes the publication of scientific outputs, the development of educational resources, and engagement with educators, policymakers, and technology developers to promote ethical, bias-aware use of generative AI systems.

The present article reports findings derived from Phases 1 and 2**,** while explicitly situating them within the broader multi-phase framework of Project LENA. Phases 3 to 5 are described to ensure transparency regarding the project's overall design and future methodological trajectory.

## 2.2 Study Design

In this exploratory phase, the LENA corpus was constructed by generating images from a controlled set of 60 unique prompts, derived from 20 base scenarios covering professional and social roles, each instantiated in three variants (neutral, inclusive, and deliberately biased). These prompts were systematically executed across four widely used text-to-image platforms: DALL·E 3 (via ChatGPT), Grok/Aurora-Flux, Stable Diffusion-based tools such as Lexica.art, and Gemini/Nano Banana Pro.

## 2.3 Sampling Strategy and Image Generation

Given the exploratory nature of the initial phases, the study employs a purposeful sampling strategy. Prompts were designed based on prior literature on gender bias in AI-generated imagery and were iteratively refined to include both neutral descriptions and bias-aware formulations. Sampling focused on:

- **Occupational roles** (e.g., teacher, scientist, construction worker);
- **Gender-related descriptors** (e.g., woman, man);
- **Comparative prompt structures**, contrasting neutral prompts with detailed, inclusive prompts explicitly requesting the avoidance of gender bias and hypersexualization.

## 2.4 Analytical Categories and Coding Schema

Drawing on both the literature and the conceptual framework of Project LENA, a preliminary *coding schema* was developed to guide visual analysis. Images were examined according to four primary analytical dimensions:

1. **Role representation**: association of genders with specific professional or domestic roles;
2. **Physical appearance**: age, body type, attractiveness, and conformity to dominant beauty standards;

3. **Sexualization**: presence of provocative poses, clothing, or framing unrelated to the prompt context;
4. **Emotional expression and agency**: portrayal of authority, passivity, assertiveness, or dependence.

These categories function as analytical lenses rather than exhaustive or mutually exclusive codes, allowing images to be interpreted across multiple dimensions. The schema was developed during Phase 1 and will be further refined and operationalized in subsequent project phases.

### 2.5 Evaluation Procedure and Reliability

In the exploratory phases reported here, image evaluation was conducted through structured qualitative assessment by members of the interdisciplinary research team. At this stage, formal interrater reliability metrics were not calculated, as the primary objective was to identify salient patterns and inform the development of a more robust evaluation protocol.

To address this limitation, Project LENA has defined a subsequent evaluation framework (Phase 3) that includes the development of explicit bias assessment rubrics aligned with the analytical categories, blind evaluation procedures involving multiple independent evaluators and calculation of interrater agreement to assess reliability. In these later phases, inter-rater reliability will be quantified using standard agreement statistics (e.g., Cohen's/Fleiss' kappa and intraclass correlation coefficients) applied to independently coded subsets of images.

These procedures are currently being implemented and will be reported in future publications.

### 2.6 Prompt Engineering and Platform Configuration

Prompt engineering constituted a central methodological component of the study. Two primary levels of prompt specification were compared:

- **Neutral prompts**, containing minimal descriptive detail.
- **Bias-aware prompts**, incorporating explicit information about gender, age, appearance, and professional context, as well as instructions to avoid stereotypical or sexualized representations.

Where possible, default platform settings were used to reflect typical user interactions. The study did not aim to optimize platform-specific parameters, but rather to observe how different systems respond to comparable prompt structures under standard conditions.

### 2.7. Future Phases

To ensure transparency and replicability, the overall methodological design of Project LENA includes additional phases beyond those reported in this article. Phase 3 and 4 focus on the systematic validation of the analytical framework through blind assessment and quantitative reliability measures, while Phase 5 addresses dissemination, educational transfer, and policy-oriented evaluation.

# 3 Results

This section presents the preliminary empirical findings from the initial phases of Project LENA. These results illustrate how gender bias emerges across different generative AI platforms and case studies, highlighting both recurring patterns of stereotypical representation and the potential of methodological strategies to mitigate such biases.

## 3.1 Revealing Stereotypes Using Lexica.art

Exploratory searches conducted using Lexica.art (a tool built on Stable Diffusion that indexes millions of AI-generated images) revealed stable and recurrent patterns of gender bias when prompts referred to occupations or gendered terms without explicit specification. Across repeated queries and multiple generated images, stereotypical representations emerged consistently, indicating that these outputs were not isolated anomalies but rather systematic tendencies of the generative model.

When prompted with occupational terms, such as *"housekeeper"* or *"construction worker"*, the platform overwhelmingly produced gendered representations aligned with traditional role stereotypes. Domestic and caregiving roles were predominantly associated with women, whereas technical or manual labor roles were depicted almost exclusively by men (Figures 2 and 3). Similarly, searches for *"scientist"* revealed marked differences in representational style: male figures were typically portrayed as engaged in professional activity and framed as authoritative, while female figures appeared younger, more stylized, and less frequently depicted in active scientific contexts (Figure 4).

Gender-based prompts further reinforced these asymmetries. Searches for *"woman"* yielded images strongly associated with beauty, fashion, and aestheticized presentation, whereas *"man"* produced representations emphasizing professionalism, authority, and seriousness (Figures 5 and 6). Taken together, these results suggest that gender bias in AI-generated imagery operates not only through explicit occupational stereotypes but also through broader cultural norms governing appearance, agency, and emotional expression.

Based on repeated observations across prompts and outputs, four recurrent categories of gender bias were identified and used to structure subsequent analysis: *role stereotypes, physical appearance norms, sexualization, and emotional expression*. Table 1 summarizes these categories, providing representative examples and outlining their broader social implications.

*Table 1.* Main types of gender bias identified in AI-generated images (Lexica.art examples)

| Bias type | Typical example | Implications |
|---|---|---|

| | | |
|---|---|---|
| **Role stereotypes** | *"Housekeeper"* → smiling women in aprons; *"Construction worker"* → men in hard hats | Reinforces traditional divisions: women in domestic roles, men in professional/technical work |
| **Physical appearance** | Female *"scientists"* depicted as younger, stylized, and conventionally attractive | Promotes unrealistic beauty standards; reduces credibility of female professionals |
| **Sexualization** | Women shown in provocative poses even without such prompts | Normalizes objectification and hyper-sexualization of women |
| **Emotional expression** | Men appear strong, serious, and authoritative; women appear passive or dependent | Associates strength with masculinity and dependence with femininity |

These findings indicate that AI-generated imagery systematically reproduces restrictive gender norms, potentially reinforcing unequal representations in educational, professional, and cultural contexts.

*Figure 2. Search results for "housekeeper" in Lexica.art. Source: Retrieved August 5, 2025, from https://lexica.art/?q=housekeeper*

*Figure 3. Search results for "construction worker" in Lexica.art. Source: Retrieved August 5, 2025, from https://lexica.art/?q=construction+worker*

*Figure 4. Search results for "scientist" in Lexica.art. Source: Retrieved August 5, 2025, from https://lexica.art/?q=scientist*

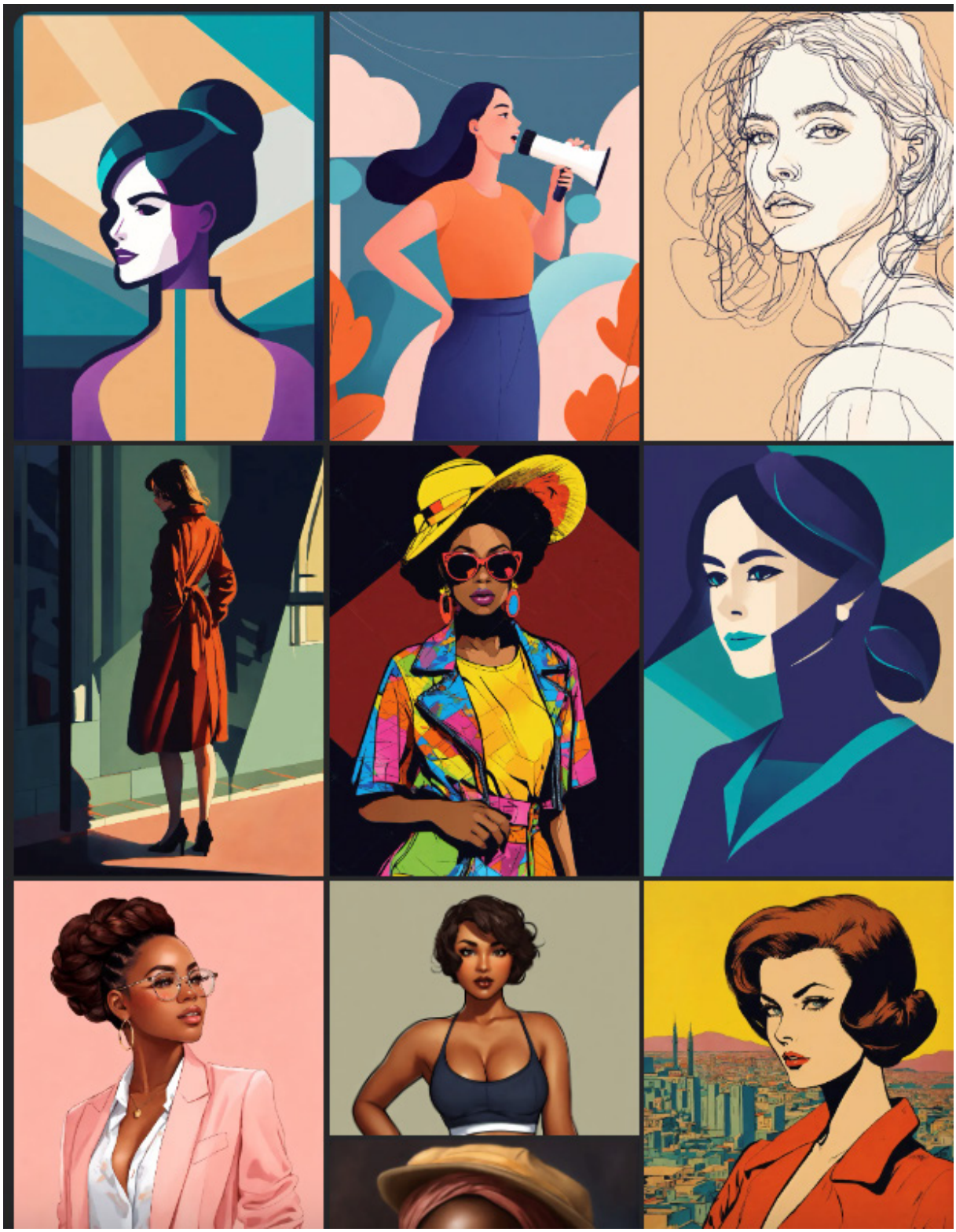

*Figure 5. Search results for "woman" in Lexica.art. Source: Retrieved August 5, 2025, from https://lexica.art/?q=woman*

*Figure 6. Search results for "man" in Lexica.art. Source: Retrieved August 5, 2025, from https://lexica.art/?q=man*

### 3.2 Descriptive Frequency Patterns Across Bias Categories

To move beyond purely illustrative examples, the observed images were examined descriptively to identify relative frequencies of bias categories across prompts and platforms. While this phase did not aim to produce inferential statistics, repeated image generation allowed for the identification of dominant representational tendencies. Across occupational prompts:

- **Role stereotypes** were the most frequently observed form of bias, particularly in domestic and technical professions.
- **Physical appearance norms** disproportionately affected female representations, which were consistently younger and more aligned with conventional beauty standards.
- **Sexualization** occurred primarily in images depicting women, even in prompts unrelated to appearance or intimacy.
- **Emotional expression and agency** differed markedly by gender, with men more often portrayed as authoritative and women as passive or dependent.

These patterns appeared consistently across platforms, although their intensity varied depending on model architecture and training data exposure. Importantly, the recurrence of these categories across different prompts and systems suggests that gender bias is structural rather than incidental**.**

### 3.3 Prompt Engineering as a Bias Mitigation Strategy

To explore how prompt formulation influences bias in generative AI, we conducted a simple comparative experiment using two levels of instruction (Table 2).

The first test used a simple prompt**:** *"Create an image of a math teacher standing next to a chalkboard in a classroom."* Figures 7, 8, and 9 show the results obtained across different GAI platforms. In Figure 7**,** Copilot (DALL·E 3) generated only male teachers after four iterations. Figure 8 presents outputs from Grok, and Figure 9 from Gemini 2.5. In all cases, the systems initially produced male teachers, reflecting a prominent role-based stereotype. Although the prompt itself was neutral, the outputs diverged in terms of age and appearance but were consistent in one respect: the implicit assumption that mathematics teachers are men.

In contrast, when we applied a detailed prompt—adding specifications about gender, age, physical appearance, clothing, and explicitly requesting the avoidance of hypersexualization—the results were more balanced and consistent across platforms. The outputs uniformly depicted female teachers aligned with the description provided. Figure 10 illustrates these results, produced using the following prompt *"Create an image of a young female math teacher, around 25 years old, standing next to a chalkboard in a classroom, explaining the Pythagorean theorem. She is holding a book in one hand and a piece of chalk in the other. She has curly red hair and wears black-rimmed glasses. Ensure the image avoids gender bias and hypersexualization. She is dressed in light blue."*

*Table 2. Experimental image generation using two levels of prompt instruction.*

| **Simple Prompt** | **Detailed Prompt** ***(adding specific details about gender, appearance, and explicitly requesting to avoid bias produced more consistent results)*** |
|---|---|
| *"Create an image of a math teacher standing next to a chalkboard in a classroom"* | *"Create an image of a young female math teacher, around 25 years old, standing next to a chalkboard in a classroom, explaining the Pythagorean theorem. She is holding a book in one hand and a piece of chalk in the other. She has curly red hair and wears black-rimmed glasses. Ensure the image avoids gender bias and hypersexualization. She is dressed in light blue".* |

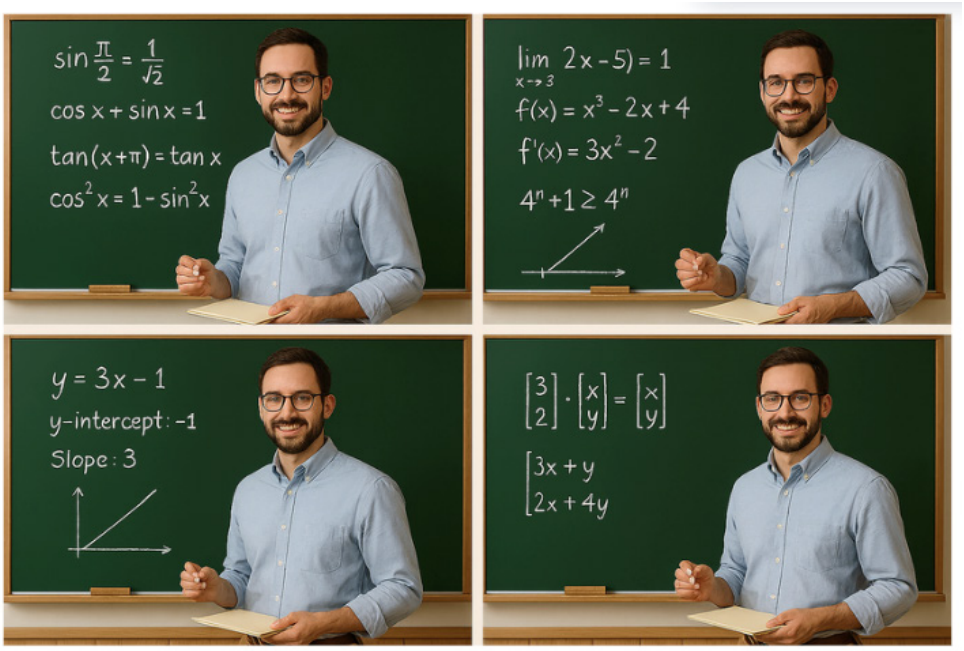


*Figure 7. Examples of AI-generated images produced by the Copilot platform (DALL·E 3).*

*Source: Authors' own elaboration*

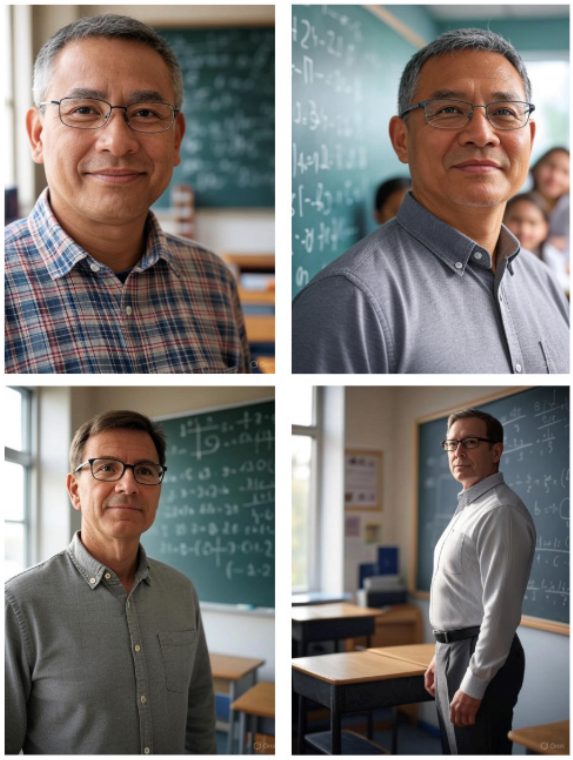

*Figure 8. Examples of AI-generated images produced by GROK. Source: Authors' own elaboration*

*Figure 9. Examples of AI-generated images produced by Gemini. Source: Authors' own elaboration*

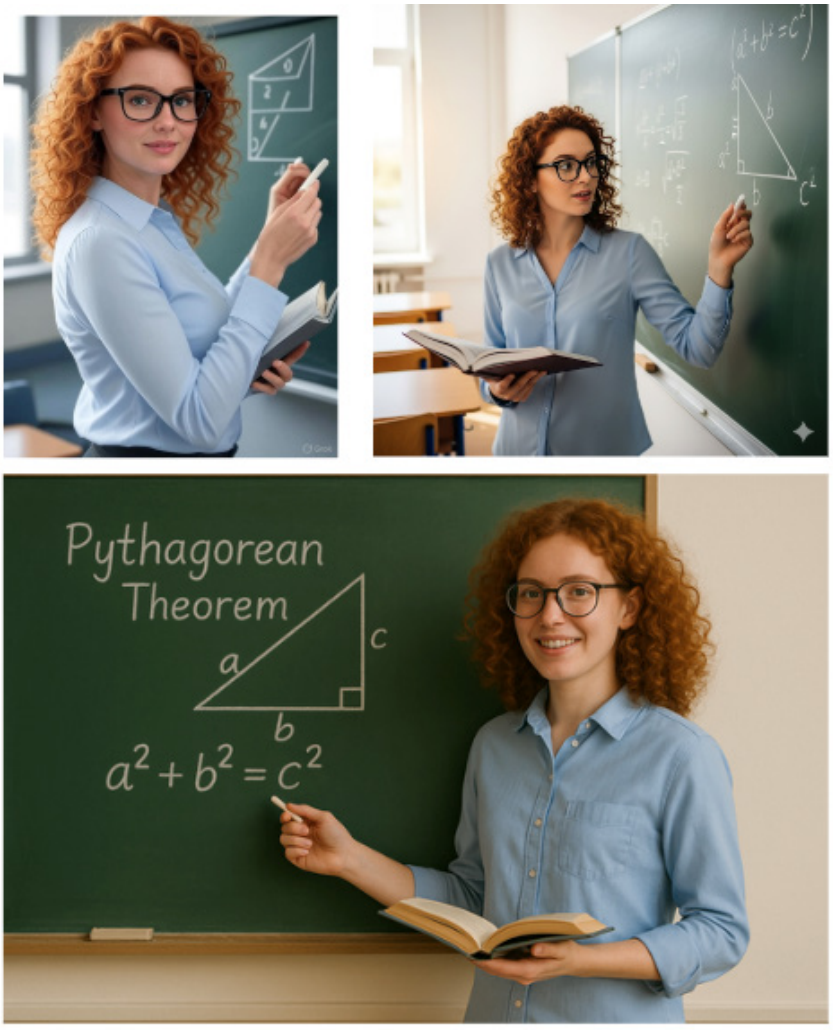


*Figure 10. Examples of AI-generated images created by three different platforms using the same detailed prompt. Top row: GROK (left) and Gemini (right). Bottom: DALL·E 3. Source: Authors' own elaboration*

This example, centered on the professional role of a mathematics teacher, illustrates how gender bias can appear even in prompts that seem neutral. Yet, such biases are not confined to education; comparable patterns have been documented across a range of professions and contexts, from healthcare to engineering, where AI-generated outputs often reproduce entrenched stereotypes (Gorska & Jemielniak, 2023; Sun et al., 2024). These observations reinforce the need for ethical and bias-aware prompting practices, not only as a technical safeguard but also as a pedagogical and cultural strategy.

To extend this analysis beyond the educational domain, we next examined representations in a traditionally male-dominated occupation: bricklayers at a construction site. By default, most GAI platforms generated male bricklayers, reinforcing conventional role stereotypes. However, when provided with a detailed prompt, the systems were able to produce a wider variety of outputs, representing individuals of different genders, ages, and skin tones. Figure 11 illustrate this sequence, showing the contrast between default outputs and those generated through more inclusive prompting.

Create an image of a bricklayer laying bricks at a construction site.

Create an image of a female bricklayer laying bricks at a construction site.

Create an image of an older woman (about 60 years old) bricklayer laying bricks at a construction site.

Create an image of an older woman (about 60 years old) bricklayer laying bricks at a construction site. The woman is slender and muscular. She is wearing a bubblegum pink sleeveless t-shirt and has a heart tattoo on her right shoulder.

*Figure 11. AI-generated depictions of bricklayers at a construction site, produced with the Grok 4 platform. Source: Authors' own elaboration*

### 3.4 Limits of Prompt-based Bias Mitigation

While prompt engineering proved effective in altering individual outputs, the results also highlight its inherent limitations as a standalone strategy. Inclusive prompting requires users to anticipate potential biases and actively counteract them, thereby shifting part of the responsibility for fairness onto individuals rather than system designers.

Moreover, the need for highly detailed prompts to obtain balanced representations underscores the persistence of biased defaults within generative models. Without explicit intervention, systems tend to reproduce dominant cultural stereotypes embedded in their training data. As such, prompt engineering should be understood as a *mitigating practice*, not a definitive solution.

These findings reinforce the importance of addressing gender bias at multiple levels, including dataset composition, model design, platform governance, and user education, an approach that aligns with the broader objectives of Project LENA.

### 3.5 Practical Recommendations for Ethical Use of GAI Platforms

As part of the early outcomes of the LENA Project, a set of interdisciplinary recommendations has been developed to guide the ethical and bias-aware use of generative AI platforms. These recommendations are informed by the project's empirical findings and by broader debates at the intersection of computer science, feminist research, and digital ethics. Their purpose is to provide users—researchers, educators, designers, and the wider public—with strategies to mitigate gender bias and foster more inclusive representations.

#### 3.5.1. Technical Practices

First, from a technical perspective, intentional prompt design remains a crucial entry point. Users should avoid neutral prompts that implicitly reproduce stereotypes and instead employ bias-aware formulations that include diverse descriptors of gender, age, ethnicity, and professional roles. Small linguistic adjustments can significantly shift outcomes, while iterative testing of prompt variations allows users to identify and reduce biased outputs.

#### 3.5.2. Educational Practices

Second, an educational dimension is essential. Promoting critical digital literacy ensures that users do not approach generative systems as neutral tools but rather as technologies shaped by cultural assumptions. Training programs for students, teachers, and professionals should emphasize the capacity of AI to reproduce inequalities, while also demonstrating strategies for mitigation through reflective use.

#### 3.5.3. Cultural and Ethical Practices

Third, these practices must be embedded in a broader ethical and cultural framework. From an intersectional perspective, it is insufficient to address gender bias in isolation: systems must also be examined for racial, age-related, and body-normativity biases, which often converge in AI-generated imagery. By adopting feminist and human-centered approaches, users and developers alike can foreground diversity and challenge exclusionary norms that have historically dominated technological design.

### 3.5.4. Policy and Institutional Practices

Finally, at the policy and institutional level, recommendations extend beyond user practices to structural reforms. Developers and organizations should commit to diversifying training datasets, making algorithmic decision-making more transparent, and submitting systems to regular external audits. Regulatory bodies, in turn, must establish clear guidelines for accountability and provide oversight mechanisms that ensure generative AI is deployed in ways that promote equity and fairness.

Taken together, these recommendations underscore that mitigating gender bias in AI-generated imagery cannot be reduced to a purely technical adjustment. Table 3 summarizes the multi-level recommendations proposed by Project LENA for mitigating gender bias in generative AI systems.

*Table 3. Multi-level recommendations for mitigating gender bias in generative AI*

| Dimension | Focus | Example recommendations |
|---|---|---|
| **Technical** | How users configure and operate GAI systems | Use bias-aware prompts that explicitly specify diverse genders, ages, ethnicities and roles; iteratively test prompt variations and monitor outputs; activate safety filters, content-moderation options and appropriate model settings to reduce stereotypical results. |
| **Educational** | How GAI is integrated into teaching and training | Embed critical digital literacy in curricula so learners understand GAI as socio-technical systems; use AI-generated images as classroom material to identify stereotypes and discuss their origins; train students and professionals to experiment with mitigation strategies such as inclusive prompting. |
| **Cultural and ethical** | How norms and values shape visual representations | Apply an intersectional lens that considers gender, race, age and body-normativity together; encourage designers and educators to question default representations and deliberately seek non-stereotypical, diverse visual narratives; frame GAI use within broader discussions on power, inequality and representation |
| **Policy and institutional** | How organizations and regulators govern GAI | Require developers to diversify training datasets, document data curation and model design, and submit systems to periodic external audits; establish institutional and regulatory frameworks that allocate responsibility for bias mitigation to platforms and governing bodies, rather than placing the burden solely on individual users. |

## 4 Discussion

The findings presented in this study provide further evidence that gender bias remains a structural feature of AI-generated imagery. Across multiple platforms and prompt types, recurrent patterns emerged in the representation of professional roles, physical appearance, sexualization, and emotional expression. These patterns were observed consistently across repeated generations, suggesting that they are not isolated anomalies but rather reflect systemic

biases embedded in training datasets, model architectures, and optimization objectives**.** This observation aligns with recent large-scale audits of text-to-image systems, which document persistent gendered and racialized representations across models such as Midjourney, Stable Diffusion, and DALL·E (Bianchi et al., 2023; Sun et al., 2024; Zhou et al., 2024).

By situating these results within the framework of Project LENA, this study contributes to ongoing scholarly debates on algorithmic bias by combining technical experimentation with critical feminist analysis. From this perspective, generative AI systems are not neutral tools but socio-technical assemblages that reproduce historically situated power relations (Crawford, 2021; D'Ignazio & Klein, 2020). The exploratory nature of the findings does not diminish their relevance; on the contrary, early-stage diagnostic approaches are essential for identifying bias mechanisms before they become further normalized through large-scale deployment and everyday use of generative technologies.

The comparative analysis of prompt formulations demonstrates that prompt engineering can partially mitigate biased outputs, enabling more diverse and inclusive representations when sufficient contextual detail is provided. Similar effects have been reported in prior studies showing that explicit constraints can redirect generative outputs away from stereotypical defaults (Gorska & Jemielniak, 2023; Sun et al., 2023). However, the need for highly detailed prompts to achieve equitable representations also reveals the persistence of biased default assumptions within generative models. In this sense, inclusive prompting operates as a compensatory strategy, capable of mitigating representational bias at the output level but insufficient as a structural remedy.

Importantly, these results caution against framing bias mitigation as a task that rests primarily on users. While educators and practitioners can employ bias-aware prompting as a pedagogical and professional tool, responsibility for equitable representation must also be assumed at the level of platform design, dataset curation, and governance**.** Recent policy-oriented scholarship emphasizes that without institutional accountability and regulatory oversight, bias mitigation strategies risk shifting responsibility from system designers to end users, thereby reproducing existing inequalities (Kalluri, 2024; Floridi et al., 2018). Differences observed across platforms in this study—particularly between systems with stronger filtering mechanisms and more "open" models—further highlight the role of governance choices in shaping representational outcomes.

From an educational perspective, the findings underscore the importance of critical digital literacy. Generative AI systems should not be presented as objective or neutral technologies, but as tools shaped by cultural assumptions, economic incentives, and historical power structures. Integrating reflective and critical practices into educational contexts can help learners recognize biased outputs, question default representations, and develop more ethical and responsible forms of engagement with AI technologies (D'Ignazio & Klein, 2020; Crawford, 2021). In this sense, education becomes not only a site of AI adoption, but also a key arena for resistance, critique, and transformation.

# 5 Conclusions

This study has examined preliminary evidence of gender bias in AI-generated imagery through an interdisciplinary, mixed-methods approach embedded within Project LENA. The results indicate that generative AI platforms consistently reproduce gendered stereotypes across occupational and gender-based prompts, reinforcing traditional norms related to roles, appearance, and agency. At the same time, the study shows that methodological interventions, particularly bias-aware prompt engineering, can influence representational outcomes and reduce stereotypical defaults. However, such interventions are insufficient when applied in isolation. Effective mitigation of gender bias in AI-generated imagery requires multi-layered strategies that integrate technical adjustments, critical awareness, and institutional accountability.

The contribution of this article lies in establishing a structured exploratory framework for analyzing gender bias in generative imagery, while articulating clear directions for methodological refinement and empirical expansion. By explicitly situating the findings within an ongoing, publicly funded research project, the study clarifies both its current scope and its future trajectory.

Future phases of Project LENA will extend this work by incorporating systematic quantitative metrics, blind evaluation procedures, and interrater reliability analyses. In addition, forthcoming research will address intersectional dimensions of bias (including race, age, and body norms) and will contribute to policy-oriented discussions on the ethical governance of generative AI.

In conclusion, addressing gender bias in AI-generated imagery is not merely a technical challenge but a cultural and ethical imperative. Only through interdisciplinary collaboration and sustained critical inquiry can generative AI technologies evolve in ways that promote fairness, inclusivity, and social responsibility.

**Acknowledgment**

The LENA Research Team is composed of Rosario Carmona Paredes, Irene Carrillo Murcia, Ángela Coves Soler, Miguel Onofre Martínez Rach, José Joaquín Mira Solves, and Victoria Soto Sanz (Miguel Hernández University); Eva Gil Hernández and Daniel García Torres (FISABIO); and Almudena Arroyo Rodríguez and María Calderón Fernández (Fundación San Juan de Dios, Comillas Pontifical University).

This research was supported by the Institute for Women of the Spanish Ministry of Equality under the 2024 Call for Feminist Research Projects, within the project *"Investigación sobre sesgos de género en la generación de imágenes por Inteligencia Artificial (LENA), Exp. 19-3-ID24."*

**Declaration of AI use**

Parts of this manuscript benefited from the assistance of generative AI tools (ChatGPT and Grok), which were employed to improve clarity, refine style, and translate draft sections from Spanish to English. The authors reviewed, verified, and edited all AI-generated text, ensuring that the final content accurately reflects the intended meaning and original scholarly contributions. Responsibility for the ideas, interpretations, and conclusions remains entirely with the authors.

## References

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 1534–1545. https://doi.org/10.1145/3593013.3594095

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). *On the opportunities and risks of foundation models*. *Journal of Machine Learning Research, 22*(1), 1–199. https://www.jmlr.org/papers/v22/21-0896.html

Freixà, P., Redondo-Arolas, M., Codina, L., & Lopezosa, C. (2025). IA, fotografía de stock y bancos de imágenes: sesgos de género y estereotipos. Hipertext.net, 30, 41–78. https://doi.org/10.31009/hipertext.net.2025.i30.05

García-Ull, F. J., & Melero-Lázaro, M. (2023). Gender stereotypes in AI-generated images. *El Profesional de la Información, 32*(5), e320507. https://doi.org/10.3145/epi.2023.sep.05

Girrbach, S., Leiser, F., & Weller, A. (2025). A large-scale analysis of gender biases in text-to-image generation. arXiv. https://arxiv.org/abs/2503.23398

Gorska, A. M., & Jemielniak, D. (2023). The invisible women: Uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies, 23*(8), 4370–4375. https://doi.org/10.1080/14680777.2023.2263659

Kalluri, P. R. (2024). AI image generators often give racist and sexist results: Can they be fixed? *Nature, 627*, 722–725. https://doi.org/10.1038/d41586-024-00599-8

Sun, L., Spruit, S., Langer, M., Dolezal, J., & Heine, C. (2023). *Stereotype amplification and bias in DALL·E 2 generated images*. arXiv. https://doi.org/10.48550/arXiv.2305.10566

Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2024). Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI. *Journal of Computer-Mediated Communication, 29*(1). https://doi.org/10.1093/jcmc/zmae003

The Lenna story. (n.d.). *Lenna.org*. Retrieved August 28, 2025, from http://www.lenna.org

Thompson, C. (2019, June 17). *Finding Lena, the patron saint of JPEGs*. *Wired*. https://www.wired.com/story/finding-lena-the-patron-saint-of-jpegs/

Vincent, J. (2022). AI art is here and the world is already different. *Nature*, 610, 628–630. https://doi.org/10.1038/d41586-022-02818-9

Women Love Tech. (2020, November 21). *Losing Lena – Why removing one image will end tech's original sin*. *Women Love Tech*. https://womenlovetech.com/losing-lena-why-we-need-to-remove-one-image-and-end-techs-original-sin/

Zhou, Z., Zhang, M., Chen, X., & Wang, Y. (2024). *Measuring bias in text-to-image generative models: An analysis of Midjourney, Stable Diffusion, and DALL·E 2*. arXiv. https://doi.org/10.48550/arXiv.2403.02726

**Appendix A. Software References**

- Google DeepMind. (n.d.). *Gemini*. DeepMind. Retrieved August 28, 2025, from https://deepmind.google/technologies/gemini
- Grok. (n.d.). *Grok AI*. xAI. Retrieved August 28, 2025, from https://x.ai
- Lexica. (n.d.). *Lexica.art*. Lexica. Retrieved August 28, 2025, from https://lexica.art
- Microsoft. (n.d.). *Copilot*. Microsoft. Retrieved August 28, 2025, from https://copilot.microsoft.com
- OpenAI. (n.d.). *DALL·E*. OpenAI. Retrieved August 28, 2025, from https://openai.com/dall-e
- Stability AI. (n.d.). *Stable Diffusion*. Stability AI. Retrieved August 28, 2025, from https://stability.ai