# Limitations and potentials of ChatGPT-4 in gender expression: A future research outlook from a feminist perspective

iD Yunning Wang[1*]
[1]Department in Anthropology, Seoul National University, Seoul, South Korea; dearmydorothy@gmail.com (Y.W.).

**Abstract:** This paper explores the limitations and potentials of ChatGPT in gender expression, with a particular focus on its impact on gender bias and stereotypes. As a generative model relying on vast training data, ChatGPT often reflects societal gender biases and stereotypical representations of traditional gender roles in its content generation. This phenomenon may lead ChatGPT to unintentionally reproduce or even reinforce these biases, further solidifying associations of "male dominance" and "female subordination," thereby limiting the diversity of gender role expressions. Adopting a feminist perspective, this study reviews core feminist principles and analyzes the potential of ChatGPT-4 and its image generation model, DALL-E, to express female subjectivity, promote diverse representations, and challenge stereotypes. Additionally, this paper reveals the technical and ethical challenges posed by data bias and algorithmic prejudice in ChatGPT, highlighting their implications for gender-equitable expression. To achieve more inclusive and diverse gender representation, this study proposes future directions for improvement, aiming to advance ChatGPT's development from a feminist perspective through diversified and balanced data handling and optimized generative model structures.
**Keywords:** ChatGPT, Feminism, Gender roles, Gender bias, Stereotypes.

## 1. Introduction

With the rapid advancement of artificial intelligence technology, Generative AI has become a core driver of content creation across various domains. Leveraging deep learning and transformer-based models like DALL-E, Generative AI can produce diverse forms of content, including text, images, audio, and video. Its efficiency and versatility show broad application potential in fields such as journalism, art, and entertainment. However, while Generative AI brings about a creative revolution, it also raises extensive discussions about gender bias and stereotypes. Since generative content is often trained on historical data that inevitably contains societal biases and traditional gender views, the outputs of Generative AI may unintentionally reinforce and intensify fixed gender roles and stereotypes.

Feminism has long advocated for social equality by challenging gender stereotypes, supporting diverse expressions in areas like art, literature, and film. As Generative AI sees growing use in content creation, a feminist perspective has gradually entered technological discussions to address the negative impact of potential gender bias in generated content. Traditional gender roles often position men as symbols of authority, rationality, and strength, while portraying women as nurturing, emotionally expressive, and supportive. These ingrained gender notions are pervasive in society and may inadvertently be perpetuated and solidified during technology design and data training processes. If Generative AI adopts these biases uncritically, it risks reinforcing stereotypes, such as the belief that men are better suited for STEM fields while women are more appropriate for caregiving roles, subtly reproducing societal views within content creation.

In this context, feminism provides a unique theoretical lens for Generative AI, advocating for gender equality and diverse representation to challenge entrenched traditional gender roles. Feminist theory posits that the potential of Generative AI lies not only in enhancing content generation

efficiency but also in supporting the broader social goal of gender equality through technological innovation. However, the current technical limitations, data biases, and gender prejudices within Generative AI present significant challenges to feminist expression. Thus, the purpose of this study is to theoretically explore the potential and limitations of Generative AI in feminist expression, with a focus on the issue of gender bias in image generation. Images, as the most immediate medium for conveying gender concepts, can vividly reflect societal views on gender, providing valuable material for academic analysis. Accordingly, this study will use ChatGPT-4 and DALL-E as research subjects to conduct an in-depth analysis of how gender bias is perpetuated and reinforced in generated content, solidifying specific gender roles. It will also explore data improvements and model optimizations to address these issues, aiming to advance Generative AI toward a more inclusive and diverse direction.

## 2. Theoretical Framework of ChatGPT and Feminism
### 2.1. Basic Concepts of ChatGPT-4 and DALL-E
ChatGPT-4 and DALL-E are two generative AI models developed by OpenAI that excel in the fields of text and image generation, respectively, and can be used in tandem in certain applications to accomplish more complex generative tasks. ChatGPT-4 specializes in understanding and generating natural language, while DALL-E can transform textual descriptions into images, showcasing significant potential for multi-modal content generation.

### 2.1.1. Basic Concept of ChatGPT-4
ChatGPT-4 is an advanced natural language generation model based on the Transformer architecture, trained on a large corpus of text data, and has an impressive capacity for language understanding and generation. ChatGPT-4 uses a decoder structure to generate text word by word in an autoregressive manner, enabling it to produce contextually relevant content with a natural flow of meaning.

ChatGPT-4 demonstrates exceptional performance in language generation tasks such as text-based conversations, content creation, summarization, and translation. Although ChatGPT-4 itself cannot directly generate images, it can produce descriptive text for images and collaborate with image generation models like DALL-E. By providing detailed textual descriptions, ChatGPT-4 can guide DALL-E in generating images that meet user requirements. However, due to the inherent biases in the training data, ChatGPT-4's content may inadvertently reflect existing societal views, particularly on sensitive topics like gender and race, potentially amplifying these biases.

### 2.1.2. Basic Concept of DALL-E
DALL-E is an image generation model developed by OpenAI specifically designed to convert textual descriptions into images, based on the Transformer architecture. DALL-E can understand detailed natural language descriptions and generate contextually appropriate images. It has been trained on a vast dataset of image-text pairs, enabling it to create highly detailed and diverse images based on complex text inputs. Through multi-modal data embedding, DALL-E translates textual features into visual features, which are then used to generate images.

The DALL-E model can be considered an extension of ChatGPT-4. DALL-E can produce images in various styles and themes, with applications in creative design, advertising, educational content development, and more. Although DALL-E is capable of generating highly creative content, it may also reflect the biases present in its training data, particularly regarding gender, occupations, and social roles. When generating content based on text provided by ChatGPT-4, these biases may compound, unintentionally reinforcing stereotypes.

### 2.2. Overview and Extension of Feminism
Feminism, also referred to as Womanism, primarily seeks to critique societal structures to uncover the roots of gender inequality. At its core, feminist theory aims to understand and analyze the nature of gender inequality, with a particular focus on issues related to sexual politics, power

dynamics, and sexuality. Broadly speaking, feminism represents a movement and theory advocating for women's equal participation in social, political, cultural, and economic spheres. In a narrower sense, it is a philosophy and social movement centered on advancing women's rights, aiming to eliminate gender discrimination and promote gender equality.

In this context, Feminist Aesthetics emerged as an extension of feminist theory within the realms of art and aesthetics. Feminist aesthetics challenges male-dominated standards of beauty in traditional aesthetics, questioning the gender biases embedded in artistic creation and evaluation. Traditional aesthetics often uphold male perspectives as authoritative and superior, relegating female aesthetic expressions to a subordinate status, thus limiting the expression of women's subjectivity and creativity. Feminist aesthetics, by emphasizing female self-expression, subjectivity, and diverse representation, seeks to redefine gender roles within the aesthetic field (Li, N. 2023).

Feminist aesthetics is not only an artistic assertion of gender equality but also a critique of social structures and cultural norms. It emphasizes the social responsibility of art, aiming to reveal gender inequality and challenge gender stereotypes through artistic expression, thereby promoting social progress. With the rise of feminist movements in the 20th century, feminist aesthetics has gradually become an independent aesthetic ideology, permeating fields such as literature, painting, and film, and turning art into a medium for expressing and disseminating gender equality.

### 2.3. Intersection of ChatGPT-4 and Feminism

The relationship between ChatGPT-4 and feminism is evident in the presence of gender bias within generated content and the strategies to address it. Since ChatGPT-4 relies on large-scale data for training, this data often contains gender stereotypes, such as "scientists are male" or "nurses are female." These biases are learned during model training and may unintentionally appear in generated content, thereby exacerbating gender bias. Feminism critiques this phenomenon, arguing that generative AI may reinforce traditional gender role assignments, particularly displaying gender preferences in areas such as professions and emotional characteristics. In response, feminism advocates for integrating diverse perspectives into AI training data to ensure balanced gender representation across roles. Including diverse representations, such as female scientists and male nurses, can help models reflect a more inclusive expression of gender. Furthermore, feminism emphasizes the formation of diverse development teams in AI design to reduce potential biases during development.

Drawing on Deleuze and Guattari's theory of "minor literature," we can further understand how to technically achieve these feminist-inspired diversities. The "deterritorialization" characteristic in "minor literature" describes a way of expressing marginalized voices through the deconstruction and reconstruction of language. This concept can be implemented in generative AI through optimized algorithms and data filtering. For example, during data processing, AI can filter for more diverse corpora and images that include minority gender roles to balance and deconstruct traditional gender norms in the data. By incorporating balanced weighting strategies in model generation algorithms, the generation probabilities of different genders and roles can be dynamically adjusted, reducing the replication of mainstream gender biases and promoting inclusivity in generated content.

Additionally, "deterritorialization" can be achieved by enhancing the model's ability to recognize gender diversity, so that even if certain gender roles are underrepresented in the data, the model can still assign importance to them through a "minoritizing" approach. This allows AI outputs to present a broader range of gender roles. By combining feminist values with the theory of "minor literature," generative AI can better deterritorialize gender expression, moving beyond traditional gender symbols and roles and advancing toward the goal of gender-equal representation.

Based on Chat GPT 4o "Please help generate XXX career related pictures" Instruction



**Figure 1.**

## 3. Gender Bias Issues in ChatGPT-4

*3.1. Occupational Stereotypes in ChatGPT-4*

In the process of ChatGPT-4 content generation, gender bias has become an increasingly prominent issue. DALL-E models rely on extensive pre-existing data for training, which is typically sourced from the internet and public resources containing diverse gender, occupational, and cultural elements. However, because the training data often embodies societal gender stereotypes, DALL-E models tend to replicate these stereotypes when generating occupation-related images, resulting in gendered biases associated with certain professions. For example, *Figure 1* illustrates distributions such as "doctor as male," "nurse as female," and "secretary as female." These are reflections of societal perceptions embedded in the data that DALL-E models have "learned," rather than intentional design choices.

Influenced by traditional norms, society often holds specific expectations regarding gender roles in occupations, and these expectations are similarly reflected in ChatGPT-4-generated content. Women are typically expected to fill more "caring" roles, such as nursing, teaching, or secretarial positions, whereas men are perceived as more suitable for technical, authoritative, or leadership roles, such as engineering, entrepreneurship, or medicine. These views are prevalent in training data, causing ChatGPT-4 to "automatically" assign corresponding genders to specific roles, reinforcing the notion of "What Women Should Be" and "What Men Should Be."

These biases are not merely technical issues in content generation but reflect deeply entrenched gender biases within the workplace and society. In career development, women often face limitations due to traditional views, such as encountering glass ceilings or gender discrimination in fields like high-tech, engineering, or management. Meanwhile, men are often subject to higher expectations and greater pressures in various professions; they are frequently expected to shoulder more responsibility, serve as primary economic providers, or achieve significant career milestones. This double standard is further embedded through ChatGPT-4 content generation, reinforcing subconscious notions of "Women Suited for Caregiving Roles" and "Men Suited for Leadership and Technical Roles."

Moreover, under broader societal and cultural influence, men are often defaulted as the primary figures in many fields—a perception deeply rooted in social structures and cultural norms. This underlying gender bias subtly affects the outputs of DALL-E generated content. When ChatGPT-4 is trained on large sets of societal data, which frequently embody such stereotypes, DALL-E models naturally develop a tendency to prioritize male representations in content generation.

Based on Chat GPT 4o "Please help me generate a picture of the XXX teacher" Instruction



| Chemistry Teacher | Biology Teacher | Science Teacher | Art Teacher | History Teacher | Political Teacher |

*The above includes three teachers each in Science and Liberal Arts.*

**Figure 2.**

This phenomenon is clearly illustrated in Figure 2, where six different teaching professions—whether in the humanities or sciences—are defaulted to male representations when no specific gender is specified. Although society often associates "women as better suited for the humanities and men for the sciences," ChatGPT-4 still defaults to male representations across all teaching roles, highlighting the deep-seated societal bias of "male as the default gender." This notion of "male as the standard" not only limits diversity in generated content but also perpetuates gender stereotypes to some extent. It reflects how society often views men as the norm, with women positioned as peripheral or secondary. In both technical and educational fields, men are typically seen as authoritative and capable figures, while women are comparatively marginalized. This mindset has permeated ChatGPT-4's training data, leading models to default to male representations in professional contexts. In other words, ChatGPT-4 is not actively choosing male imagery but rather reproducing societal norms embedded in the data it learns from, illustrating how technology can inadvertently reflect and reinforce gender biases.

Based on the Chat GPT 4o "Please think of a picture of Female/Male Elements and draw it." Instruction



| Picture 1 | Picture 2 | Picture 3 | Picture 4 | Picture 5 | Picture 6 |

**Figure 3.**

### 3.2. The Reinforcement of Gender Symbols in ChatGPT-4

In Figure 3 presents six images illustrating the stereotypical reinforcement of gender symbols in ChatGPT-4-generated content. Female figures are depicted as gentle and delicate, often accompanied by floral and natural elements, while male figures are linked with grand landscapes like mountains, embodying resilience and strength. This symbolic contrast not only highlights the issue of fixed gender symbols in ChatGPT-4-generated content but also reflects the binary gender understanding rooted in Western traditional aesthetics. Through an analysis of three sets of images (picture 1 & 2, picture 3 & 4, picture 5 & 6), we can better understand the aesthetic differences, detail expressions, and metaphorical imagery associated with gender symbols.

In the first set (picture 1 and picture 2), the female figure is placed amidst a cluster of flowers, with soft tones creating a serene atmosphere, emphasizing aesthetic symbols traditionally associated with "delicacy." In contrast, the male figure is shown against a background of mountains and rocks, conveying strength and confidence. This contrast between flowers and mountains metaphorically reinforces traditional aesthetic assumptions about "delicate" femininity and "strong" masculinity. Flowers symbolize beauty, fragility, and dependency in women, while mountains suggest stability and resilience in men, perpetuating biases that overlook the potential for women to embody strength

and wisdom.

In the second set (picture 3 and picture 4), the female figure faces the sunlight, surrounded by flowers, projecting a harmonious and pure image. Conversely, the male figure is shown in a contemplative pose amid mountainous surroundings, embodying a sense of struggle against nature. This pairing further emphasizes the depiction of women as "nature dependents" and men as "nature conquerors." Such biases restrict female representations to frames of sensitivity and grace, while men are granted attributes of power and exploration. Western aesthetic perspectives reinforce the definition of women as inherently delicate, positioning men as the embodiment of rationality and strength, thereby solidifying gender role divisions.

In the third set (picture 5 and picture 6), the female profile merges with soft botanical patterns, suggesting aesthetics of gentleness and emotional sensitivity, while the male figure is surrounded by cosmic elements, like stars and the universe, symbolizing rationality and ambitious pursuits. This visual contrast further reveals the traditional aesthetic association of women with softness and receptivity and men with power and exploratory imagery. This binary opposition not only confines women to symbols of emotional beauty but also reinforces the dominance of men in fields of rationality, logic, and strength.

This fixation on gender symbols reflects the long-standing suppression of female creative expression within Western aesthetics. Women's creations have often been regarded as "decorative" or "secondary," while mainstream aesthetic criticism emphasizes rationality and objectivity, overlooking women's subjective expressions. With the rise of feminist aesthetics in the 20th century, artists sought to challenge this binary limitation through critique and reinvention, calling for a redefinition of women's roles in art and granting them space to express strength, independence, and intellect. From a feminist aesthetic perspective, female representations should not be confined solely to gentleness and sensitivity but should embody the diversity of women's identities and potential.

### 3.3. The Entrenchment of Gender Roles in ChatGPT-4

In the ChatGPT-4 generated series titled "A Successful Woman's Life (Figure 4)", a fixed gender role assignment is clearly evident: regardless of the level of achievement women attain professionally, they are ultimately depicted within the roles of motherhood and childbearing. This default association with reproductive duties has become an implicit benchmark for a "successful life" for women, suggesting that even accomplished women are still bound by societal expectations concerning family and caregiving responsibilities. This bias not only reflects society's entrenched views on female roles but also exposes the gender stereotypes embedded within ChatGPT-4 training data and generation models.



Based on Chat GPT 4o "Please draw a successful woman's life in chronological order" instruction - Random

Here is the series of illustrations depicting the life journey of a successful woman in chronological order. Let me know if you'd like any adjustments.

Here is the series of illustrations depicting the life stages of a successful woman in chronological order. Let me know if you need any adjustments.

Figure 4.

In Figure 4, the narrative of a woman's life progression—from student to career professional, and ultimately to mother—inevitably incorporates themes of family and child-rearing, with motherhood presented as a "completion" element in the final stages. No matter how much progress a woman

makes in her career, her ultimate role is framed within family and caregiving contexts. This narrative logic not only overlooks the many women who choose to remain child-free or focus solely on career development but also implies an implicit expectation that successful women must balance both career and motherhood. This underlying gender bias reinforces the notion that "motherhood is the ultimate fulfillment of female identity," subtly positioning women's professional advancement as a binary balancing act between "work" and "reproduction."

From a feminist perspective, this narrative undermines female agency. Feminism advocates for breaking fixed gender roles, promoting gender equality, and enabling diverse self-expression. However, in generating content, ChatGPT-4 often unintentionally amplifies these entrenched gender role images by relying on existing data patterns and societal expectations for women. As a result, the content portrays women as inextricably bound to the societal responsibility of "motherhood." This phenomenon not only limits diversity in content creation but also subtly reinforces gender role biases, leading audiences to subconsciously associate women's success with family and motherhood.
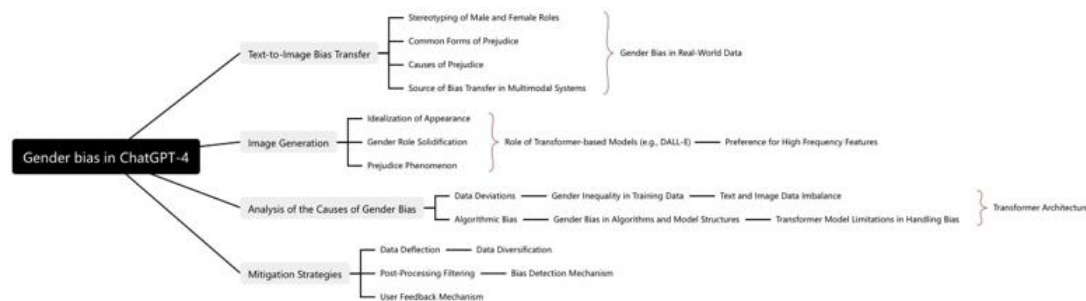


**Figure 5.**

### 3.4. Analysis of Causes of Gender Bias in ChatGPT-4

The causes of gender bias in ChatGPT-4 (Figure 5) can be further attributed to data bias and algorithmic bias.

First, data bias encompasses the gender inequality present within training data. This includes disparities in the representation of men and women in certain professions, which leads to a tendency in generated content to favor specific gendered images. For example, the differing representation of men or women in particular occupations within the data results in a bias towards certain gender portrayals.

Second, algorithmic bias refers to gender biases embedded in the design of models and structures. For instance, Transformer architectures may give more weight to high-frequency gender characteristics, which can inadvertently fix biases within generated content. This bias is not only a result of data distribution discrepancies but also stems from limitations within the algorithm's structure, further reinforcing gender biases in ChatGPT-4 generated content.

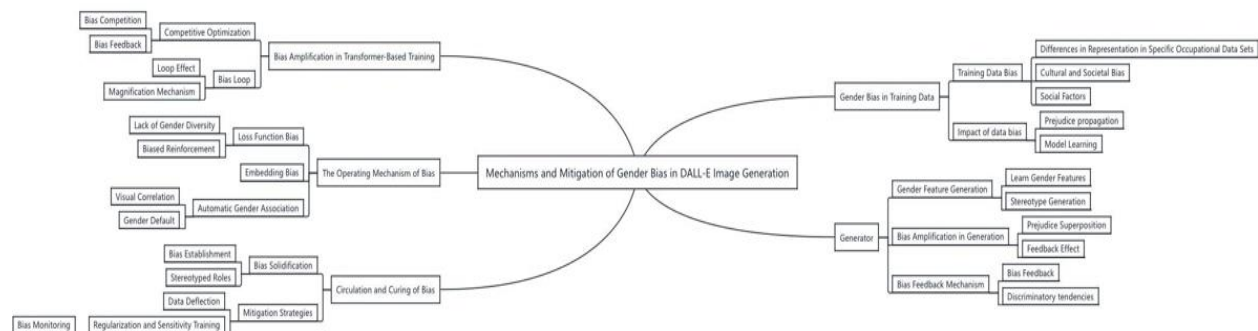In this analysis, we will focus specifically on image generation.



**Figure 6.**

In image generation (Figure 6), DALL-E models are prone to idealizing appearances and reinforcing fixed gender roles. The root cause of this phenomenon lies in the biased design of DALL-E. DALL-E models tend to generate content based on high-frequency features within training data, which often include gender biases. As a result, generated images frequently reinforce prevalent societal gender stereotypes. The fixation of gender roles causes the depiction of male and female images in generated content to become overly standardized, hindering diverse expressions.

Using "3.1 Occupational Stereotypes in ChatGPT-4" as a case study, we can conduct an in-depth analysis of the causes of gender bias in DALL-E. This example helps to clarify why DALL-E exhibits gender bias when generating content. In examining the images of different professions generated by DALL-E, such as STEM teachers, humanities teachers, surgeons, nurses, and engineers, a pattern of fixed gender distribution emerges. For instance, professions like doctors, engineers, and science teachers are often depicted as male, whereas roles such as nurses and preschool teachers tend to appear as female.

This phenomenon reflects how DALL-E models, during training, absorb substantial bias information that mirrors societal role divisions, resulting in a lack of gender diversity in generated content. The attention mechanism in DALL-E, particularly within multi-head attention weights, tends to recognize and amplify high-frequency gendered features, such as "male scientist" or "female nurse." This tendency leads to automatic gender associations in generated content, further embedding societal gender stereotypes within the output.At the same time, From the perspective of "minor literature," generative AI models like ChatGPT and DALL-E exhibit "minority" characteristics within mainstream data frameworks. According to Deleuze and Guattari's concept of "minor literature," marginalized groups are often assigned fixed identity labels within dominant cultures, a phenomenon that is particularly evident in gender expressions generated by AI. Due to the substantial presence of gender stereotypes in training data, such as "doctors are male" or "nurses are female," AI models tend to reproduce these mainstream biases in generated content, thus reinforcing existing gender role assignments. This stereotyped expression reflects the "minority" dilemma—minority gender roles and diverse representations struggle to gain adequate visibility within mainstream data.

To overcome this dilemma, generative AI can adopt the "deterritorialization" strategy inspired by "minor literature," reducing the impact of traditional biases through adjustments in technical frameworks. For instance, during model training, "minoritized" datasets could be introduced by increasing data on marginalized gender roles, such as female scientists or male nurses, to disrupt fixed gender roles. Additionally, balanced weighting could be incorporated into algorithm design, granting equal generation probabilities for various gender roles and thereby reducing the recurrence of biases. Through these technical approaches, generative AI can achieve "deterritorialization" within mainstream data frameworks, moving beyond the binary distribution of traditional gender roles. This would foster a more inclusive and diverse expression of gender, laying the groundwork for gender equality in AI-generated content.



Figure 7.

# 4. Challenges of ChatGPT-4 in Feminist Applications

**"Ethical and Technical Constraints"** primarily highlights the bias issues in ChatGPT-4 generated content, particularly the gender stereotypes arising from data biases, aesthetic tendencies, and limitations in algorithmic design. These biases are perpetuated or even intensified within generated content, deviating from ChatGPT-4's initial goal of promoting gender diversity and equal

representation.

On the other hand, **"Cultural and Social Barriers"** reflect the social resistance encountered in ChatGPT-4 applications, especially in contexts where the acceptance of feminist principles and diverse gender expressions varies significantly across cultures and societies. This cultural and societal limitation hinders the practical implementation of technical and ethical improvements in ChatGPT-4, further entrenching gender biases.

The interplay between these two factors (Figure 7) creates a "bias-resistance-reinforcement" loop, where biases are continually amplified. This not only increases the difficulty for ChatGPT-4 to overcome gender bias in feminist expressions but also makes society's acceptance of diverse female representations more conservative. Breaking this cycle in ChatGPT-4 applications requires more than isolated improvements in technology or culture; it necessitates interdisciplinary and cross-cultural solutions to gradually diminish and ultimately eliminate the cycle of bias.

### 4.1. Ethical and Technical Constraints

ChatGPT-4 faces complex ethical and technical challenges in generating feminist content, particularly concerning data bias and the balance between aesthetics and ethics. First, data bias is difficult to eliminate entirely. The content generation process in ChatGPT-4 systems relies heavily on vast amounts of historical data, which often contain embedded gender biases and stereotypes. Although efforts are made to filter and clean training data, biases can still persist in subtle forms, re-emerging in generated content and making it challenging for ChatGPT-4 to avoid perpetuating gender stereotypes.

Additionally, achieving a balance between aesthetics and ethics in content generation presents a significant challenge. ChatGPT-4 content often aims for visual appeal; however, the over-emphasis on beautifying and idealizing female representations can unintentionally reinforce gendered features, overlooking the authenticity and diversity of female subjectivity. Addressing this issue requires integrating ethical and aesthetic considerations within the generation algorithms and objective functions. Yet, this is technically challenging, as balancing these aspects is difficult to quantify and optimize, especially in highly automated generative systems.

To gain deeper insights into the mechanisms of bias formation in ChatGPT-4 generated content, we can further analyze this from an algorithmic and mathematical perspective. In DALL-E models, bias arises not only from the selection and distribution of training data but also from the specific computations within the algorithms. Taking the attention mechanism in DALL-E as an example, the weight assignments in the formula prioritize high-frequency features (such as certain gender characteristics), and these weight distributions significantly impact the resulting data.

### 4.2. Cultural and Social Barriers

In the global application of ChatGPT-4, cultural and social barriers pose significant challenges, particularly in terms of the embedded social biases within generated content and the varying acceptance of feminist ideas across different cultures. Although ChatGPT-4 has shown innovation in generating diverse content forms such as text and images, the inherent social biases in training data limit the model's applicability across global cultural contexts. These biases not only solidify gender roles on a technical level but also inadvertently reflect traditional gender views and stereotypes, such as automatically associating women with "weakness," "decorative" or "nurturing" roles, while men are typically portrayed as symbols of "strength," "leadership," or "rationality." When these biases are globally disseminated, they risk reinforcing gender norms in various regions, complicating efforts to achieve gender equality.

These cultural biases stem from the complex social contexts within ChatGPT-4's training data, much of which is sourced from publicly available internet resources that encompass traditional views on gender roles across regions. In some cultures, these views are more deeply entrenched and align closely with existing gender expectations. For instance, in certain regions, women are primarily expected to fulfill family and caregiving roles, while men are associated with professional achievements. Repeated exposure to such gender roles in ChatGPT-4 generated content not only

reinforces gender division but may also subtly shape local audiences' subconscious acceptance of these roles, further entrenching gender bias. This is particularly pronounced in highly patriarchal societies such as Saudi Arabia, Nigeria, India, Iran, Yemen, and Russia, where gender inequality is not only a societal norm but also institutionally embedded in laws and cultural traditions.

Notably, countries like Saudi Arabia, a leader in AI, have made significant advancements in the global AI sector. Saudi Arabia ranked 14th in the 2024 Global AI Index by Tortoise Intelligence, leading Arabnations with a 17-spot rise from previous rankings, highlighting its rapid progress in AI development. However, despite this technological success, deeply ingrained cultural rejection of gender equality and persistent inequality in societal structures remain major challenges. This paradox reveals that even in technologically advanced nations, the promotion of feminist ideas and gender-diverse expression continues to face significant obstacles (Liu, Y. 2024).

Furthermore, the acceptance of feminist ideas within ChatGPT-4 generated content varies widely across global cultures. In some cultures, feminist principles are more readily embraced, allowing generated content to freely display female subjectivity, equitable gender role distribution, and diverse representations of women. In others, however, feminist ideas are met with resistance or viewed as challenges to established social norms. Thus, when ChatGPT-4 content incorporates feminist perspectives or critiques of traditional gender concepts, it may provoke cultural conflicts, hindering the global reach of such content. This disparity in cultural acceptance not only creates obstacles for the worldwide application of ChatGPT-4 but also intensifies the challenges in achieving its goals of genderequality and diverse representation of women.

## 5. Future Improvement Strategies

In the application of ChatGPT-4 technology, reducing gender bias and enhancing content diversity and inclusivity require a multifaceted approach. To achieve fairer and more diverse content representation, the following improvement strategies are proposed from three perspectives: data optimization, algorithm refinement, and public education. These strategies aim to advance ChatGPT-4 in promoting gender equality and diverse expression.

### 5.1. Data Optimization and Debiasing Strategies

Gender bias in DALL-E models often stems from the homogeneity and stereotypes within training data. Therefore, data optimization and debiasing strategies form the foundation for addressing this issue. First, it is essential to establish more diverse and inclusive training datasets. By incorporating datasets that reflect varied cultures, professions, and life roles, we can ensure a balanced representation of gender roles within the training data. This approach helps to mitigate the replication of traditional stereotypes and promotes a broader, more equitable portrayal of gender in ChatGPT-4 generated content.

Assuming that gender bias is reflected through sample proportions, a bias measurement formula can be defined as follows:

$$\text{Bias}_{\text{DALL-E}} = \frac{1}{N} \sum_{i=1}^{N} |p_{\text{male}}(i) - p_{\text{female}}(i)|$$

Let $N$ represent the number of different scenarios or categories in the generated content, and let $p_{male(i)}$ and $p_{female(i)}$. denote the proportion of male and female characters generated in the i-th scenario, respectively.

$N$: The number of scenarios or categories in the generated content, such as different professions, activities, or environments.

$p_{male(i)}$ and $p_{female(i)}$: The proportion of male and female characters generated in the ii-th scenario. For example, if DALL-E generates 100 images of doctors, with 70 of them featuring male characters, then $p_{male(doctor)}=0.7$.

$|p_{male(i)} - p_{female(i)}|$: The absolute difference between the male and female proportions in

the ii-th scenario, representing the degree of gender imbalance in that scenario.

### 5.1.1. Specific Methods for Reducing Bias

*Increasing Dataset Diversity:* Introduce more instances of women in traditionally "male-dominated" occupations (e.g., engineers, scientists) and men in traditionally "female-dominated" occupations (e.g., nurses, teachers). This approach helps balance gender ratios and prevents the generated outputs from favoring specific gender representations, thus reducing the bias index in generated content.

*Balancing Gender Ratios:* Ensure that the gender ratio within specific roles or occupations in the training dataset is as balanced as possible. For example, if 90% of the data labeled "scientist" consists of male instances, incorporating more data for female scientists can create a more equal distribution, thereby lowering the absolute difference $| male(i) - female(i) |$ in the bias formula.

*Data Resampling:* When it is difficult to gather new data, resampling (increasing or reducing certain sample categories) can help achieve balance. For male-dominated occupations, this might mean reducing the proportion of male samples or increasing the frequency of female samples. Additionally, resampling can adjust the frequency of stereotypes, such as "women as weak" or "men as strong," helping to prevent the reinforcement of gender symbols in generated content.

When $\rho_{male} \approx \rho_{female}$, the bias index approaches zero. This indicates that the closer the gender ratio is to being balanced in specific occupations, the smaller the data bias, which, in turn, reduces the manifestation of gender bias in the generated content.

### 5.1.2. Example Data on Gender Distribution in Professions

For a particular profession (e.g., "scientist"), if the current data shows a male proportion of $\rho_{male}=0.85$ and a female proportion of $\rho_{female}=0.15$, we can calculate the bias as follows:

$$\text{Bias}_{\text{DALL-E (scientist)}} = |p_{\text{male}} - p_{\text{female}}| = |0.85 - 0.15| = 0.70$$

This high bias value indicates a significant skew toward male representation, suggesting that gender stereotypes for this profession may be amplified in the generated content.

**Table 1.**

| Profession | $\rho_{male}$ | $\rho_{female}$ | $\text{Bias}_{\text{DALL-E}}$ |
|---|---|---|---|
| Scientist | 0.85 | 0.15 | 0.7 |
| Engineer | 0.80 | 0.20 | 0.6 |
| Nurse | 0.20 | 0.80 | 0.6 |
| Teacher | 0.30 | 0.70 | 0.4 |
| Doctor | 0.90 | 0.10 | 0.8 |

In such a distribution Table 1 a high bias value indicates a severe gender imbalance within the profession, suggesting that the gender ratio is heavily skewed. This imbalance often results in reinforced gender stereotypes in generated content.

### 5.1.3. Balanced Example Data

Through dataset diversification and balancing, we can adjust the "scientist" profession data to have a male proportion of $\rho_{male}=0.5$ and a female proportion of $\rho_{female}=0.5$.

Now, the bias value is close to zero:

$$\text{Bias}_{\text{DALL-E}} = 0$$

To gain a clearer understanding, we can introduce hypothetical data(Table 1) to demonstrate how balancing gender ratios reduces bias, highlighting the importance of diversity and inclusivity in optimizing datasets.

By diversifying and balancing the dataset, ChatGPT-4 can significantly reduce preferences for

certain genders or gender-specific traits, preventing the rigid representation of female and male roles in generated content.This adjustment also enhances the diversity of generated content, allowing a broader portrayal of gender roles and identities that transcends traditional gender symbols. Furthermore, ChatGPT-4 can more accurately reflect the varied role distribution in modern society, showcasing both women in non-traditional professions and men in supportive and caregiving roles, aligning with values of gender equality and inclusivity.

*5.1.4. To Eliminate Gender Bias in Future Chatgpt-4 Research, the Following Technical Methods Can Be Implemented*

1. Increase Data Samples for Minority Genders：For professions skewed toward male representation (e.g., scientists or engineers), increase the sample size of female data to gradually bring the female proportion closer to 50%.

Similarly, for professions skewed toward female representation (e.g., nurses or teachers), increase male samples to balance the gender ratio. For example, if the current distribution for the "scientist" profession is $\rho$ male=0.85 and $\rho$ female=0.15, add more female scientist data samples until both proportions approach 50%.

2. Resample Data：Balance proportions through downsampling or upsampling. For professions with an excess of male samples, reduce the frequency of male samples (downsampling) and increase female samples (upsampling) to achieve a more balanced distribution. If male data predominates for scientists, reduce the usage frequency of male samples and increase female samples to reach equilibrium.

3. Generate Synthetic Data：When sufficient data is unavailable, use techniques compatible with

DALL-E to create synthetic data samples for the underrepresented gender, helping balance gender ratios. For instance, generate additional virtual female scientist images using DALL-E to achieve a balanced gender proportion.

4. Adjust Weights：In algorithms, assign higher weights to data from the minority gender, increasing its influence during content generation to achieve gender balance.

*5.2. Public Education and Social Advocacy*

The application of ChatGPT-4 technology is not solely a technical issue; it also involves public understanding and acceptance of gender equality. Thus, public education and social advocacy are essential for promoting the fair application of ChatGPT-4. Raising awareness among AI developers about gender bias issues can serve as an effective preventative measure. By strengthening gender equality education within AI development teams, developers can be encouraged to proactively avoid gender bias in data selection and algorithm design. Additionally, internal review mechanisms can be established to regularly evaluate the fairness of generated content, reducing bias at its technical source (Wang, Y. 2024).

Moreover, advocating for responsible ChatGPT-4 applications and spreading awareness of gender equality play crucial roles in shaping public understanding and usage of this technology. Promoting gender equality through workshops, educational events, and similar initiatives can help the public and institutions recognize the risks of gender bias and increase awareness of its impact in ChatGPT-4. This approach not only fosters a responsible technology usage culture at the societal level but also creates a supportive environment for the ongoing development of ChatGPT-4 in advancing gender-equal expression (Wang, Y. 2024).

## 6. Conclusion

This study explored the potential and challenges of ChatGPT-4 in feminist expression, specifically examining how this technology can promote feminist ideals in the context of gender equality and diverse representation. As one of the most widely used tools in AI-generated content, ChatGPT-4 wields significant influence. Leveraging advanced techniques like deep learning and

DALL-E model, it generates rich content across various formats—including text, images, and video—offering new avenues for expressing gender equality. However, ChatGPT-4 also faces substantial technical and ethical challenges in practical application, particularly related to data bias, algorithmic design, and varying levels of cultural acceptance. While ChatGPT-4 has considerable potential for feminist expression, hidden gender biases within its generated content often persist and may even intensify, undermining diverse gender portrayals and creating a "vicious cycle" of gender bias.

ChatGPT-4 holds significant potential for promoting diversity and inclusivity. By balancing gender ratios within datasets, optimizing algorithmic structures, and incorporating diverse training data, ChatGPT-4 can support the representation of women, non-traditional gender roles, and varied gender portrayals, establishing a technical foundation for reducing gender bias and challenging stereotypes (Wang, X. 2024). Additionally, ChatGPT-4 can contribute to reshaping public perceptions of gender through the automatic generation of art, film scenes, and advertising images, encouraging diverse expressions of gender roles and advancing feminist ideals within mainstream culture. This capability highlights ChatGPT-4's unique potential for fostering gender-equal representation. However, if the model heavily relies on high-frequency gender patterns from its training data, such biases are not only difficult to eliminate but may also be reinforced within generated content. This perpetuation of gender bias strengthens traditional gender narratives, creating a "vicious cycle" that contradicts feminist goals of diversity and inclusivity.

The challenges ChatGPT-4 faces in reducing bias and achieving gender diversity are also considerable. Firstly, the generated content frequently reflects social biases embedded within the training data, allowing these biases to persist in technology and, in some cases, reinforcing gender role entrenchment rather than diminishing it. Furthermore, algorithmic biases within the model's design—such as an overreliance on high-frequency gender features during content generation—lead ChatGPT-4 to reproduce specific gender distributions in professions or roles. This dependence exacerbates the spread of gender bias in generated content, making it challenging for ChatGPT-4 to break the fixation on traditional gender roles. Additionally, varying levels of acceptance of feminist concepts across cultures present significant challenges, as ChatGPT-4 may struggle to reflect gender diversity in certain cultural contexts and could even encounter resistance due to cultural conflicts, further limiting its potential to promote gender-equal representation.

To disrupt this cycle of bias, future research must propose innovative solutions to address these challenges. Firstly, debiasing strategies can be enhanced by using advanced tools to screen and cleanse gender bias from training data, thereby reducing the likelihood of gender stereotypes manifesting in generated content. Secondly, improvements in algorithmic design can incorporate diversity considerations directly within content generation processes, dynamically adjusting models to reduce reliance on high-frequency gender patterns. Additionally, fostering public education and raising gender equality awareness among AI developers can help integrate technology with social responsibility, facilitating a more holistic development of ChatGPT-4 in promoting gender equality and diverse expression.

Looking ahead, ChatGPT-4 presents vast opportunities for research in feminist expression. Future studies could explore interdisciplinary collaborations to analyze gender representation in ChatGPT-4-generated content from anthropological and sociological perspectives, fostering greater cultural sensitivity within algorithms and designing more inclusive systems. By merging technological advancements with humanistic insights, ChatGPT-4 can effectively address gender bias issues, promote gender-equal representation, and support the mainstreaming and advancement of feminist ideals in popular culture.

much-needed comfort and motivation during this study. Thank you all for being an integral part of my life and for making this achievement possible.

## Copyright:

## References

[1]     Wu, M. (2024, November 11). "China Origin · Henan Story: AIGC Digital Human Heritage Plan":
[2]     Exploring the International Expression of Chinese Culture. China News Publishing Radio and Television News, (008).
[3]     Wang, X. (2024). Collaboration and Reshaping: Communication Strategies for Chinese Traditional Painting and Calligraphy in the AIGC Era. Art Communication Research, 2024(06), 21-30.
[4]     Zheng, H. (2024). Trustworthy Artificial Intelligence: Governance Focus of AIGC under the Risks of Technological Ethics. Science and Technology Progress and Policy, 1-11. Retrieved from http://kns.cnki.net/kcms/detail/42.1224.G3.20241101.1042.002.html
[5]     Yi, X. Q., Qin, T. S., & He, J. F. (2024). High-Precision Path Planning and Experimental Research of Intelligent Drawing Robot Based on AIGC. China-Arab Science and Technology Forum (Bilingual), 2024(11), 98-102.
[6]     Xu, C. (2024, November 1). Exploration and Effectiveness of AIGC Technology in the Construction and Teaching Practice of Art Courses in Colleges and Universities. Dahe Fine Arts News, (012).
[7]     Qin, S. Y., & Li, X. Y. (2024). From ChatGPT to Sora: Reshaping Production Processes and Addressing Trust Crises in the AIGC Transformation of the Film Industry. Audiovisual, 2024(11), 3-8. https://doi.org/10.19395/j.cnki.1674-246x.2024.11.001
[8]     Liu, Y. (2024). Research on the Application of Generative Artificial Intelligence Technology in News Communication. News and Cultural Development, 2024(20), 178-180.
[9]     Fan, D. Q. (2024, October 26). Exploring AIGC Empowerment to Drive Industrial Innovation and Breakthroughs. Dongguan Daily, (A03).
[10]    Liu, X. H., & Li, J. Y. (2024). Integration Paths and Development Trends of the Media Industry under AIGC. Media Forum, 7(20), 9-12.
[11]    He, Y. F., Yuan, H. F., Bo, Z. M., & Yang, X. B. (2024). Cross-Domain Hybrid Pattern Innovation Design Based on AIGC—A Case Study of Huangmei Embroidery. Fashion Designer, 2024(11), 90-95. https://doi.org/10.20100/j.cnki.cn11-4548/ts.2024.11.006
[12]    Li, N. (2023). Li Na: A Western Feminist Aesthetic Perspective. ARTLINKART. Retrieved from https://www.artlinkart.com
[13]    Wang, Y. (2024). The public and private dimensions of Chinese women's bodies—From reproductive policies to bodily autonomy. Humanities and Social Science Research, 7(3),
[14]    12. https://doi.org/10.30560/hssr.v7n3p12
[15]    Wang, Y. (2024). Women's dilemma and social reflection in film narrative: A case study of Kim Ji-young, Born 1982. Journal of Politics and Law, 17(4), 38. https://doi.org/10.5539/jpl.v17n4p38