

Article

Gender Bias in Text-to-Image Generative Artificial Intelligence When Representing Cardiologists

Geoffrey Currie ^{1,2,*} , Christina Chandra ³ and Hosen Kiat ^{4,5,6}

¹ School of Dentistry and Medical Sciences, Charles Sturt University, Wagga Wagga, NSW 2678, Australia

² Department of Radiology, Baylor College of Medicine, Houston, TX 77030, USA

³ Faculty of Science, School of Psychology, UNSW, Sydney NSW 2052, Australia

⁴ Cardiac Health Institute, Sydney, NSW 2121, Australia

⁵ College of Health and Medicine, Australian National University, Canberra, ACT 2601, Australia

⁶ Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW 2109, Australia

* Correspondence: gcurrie@csu.edu.au; Tel.: +61-02-69332822

Abstract: Introduction: While the global medical graduate and student population is approximately 50% female, only 13–15% of cardiologists and 20–27% of training fellows in cardiology are female. The potentially transformative use of text-to-image generative artificial intelligence (AI) could improve promotions and professional perceptions. In particular, DALL-E 3 offers a useful tool for promotion and education, but it could reinforce gender and ethnicity biases. Method: Responding to pre-specified prompts, DALL-E 3 via GPT-4 generated a series of individual and group images of cardiologists. Overall, 44 images were produced, including 32 images that contained individual characters and 12 group images that contained between 7 and 17 characters. All images were independently analysed by three reviewers for the characters' apparent genders, ages, and skin tones. Results: Among all images combined, 86% (N = 123) of cardiologists were depicted as male. A light skin tone was observed in 93% (N = 133) of cardiologists. The gender distribution was not statistically different from that of actual Australian workforce data ($p = 0.7342$), but this represents a DALL-E 3 gender bias and the under-representation of females in the cardiology workforce. Conclusions: Gender bias associated with text-to-image generative AI when using DALL-E 3 among cardiologists limits its usefulness for promotion and education in addressing the workforce gender disparities.

Keywords: generative artificial intelligence; diversity; inclusivity; cardiology; bias



Citation: Currie, G.; Chandra, C.; Kiat, H. Gender Bias in Text-to-Image Generative Artificial Intelligence When Representing Cardiologists. *Information* **2024**, *15*, 594. <https://doi.org/10.3390/info15100594>

Academic Editors: Danilo Avola and Heming Jia

Received: 5 August 2024

Revised: 14 September 2024

Accepted: 25 September 2024

Published: 30 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are gender and ethnicity biases across medicine generally [1,2] and across the cardiology specialisation specifically [3–5]. Only 13% of cardiologists in the US are female, despite approximately 50% of both the population and medical students being female [3,5,6]. This is worse for interventional cardiology, with only 4.3% being female [3,5,6]. These figures are mirrored in Australia, where 15% of cardiologists and 4.8% of interventional cardiologists are female, respectively [7]. While female medical graduates represent 50% of the new medical workforce in the US, UK, and Australia, among training fellows, only 20–27% are female for general cardiology and only 4–10% for interventional cardiology [3,5–8]. White male stereotyping and masculine cultures can dissuade females from a career in cardiology, and the lack of gender diversity can undermine health outcomes for female cardiac patients [1,4,7,9]. For example, women with acute myocardial infarction have worse survival outcomes if their cardiologist is male [9]. With respect to ethnicity, in the US, ethnic minority groups make up just 8% of cardiologists [5]. Under-represented gender and ethnic groups in cardiology are less likely to receive guideline-based quality care, while cardiologists from those under-represented groups are more likely to provide evidence-based care to other minorities [4]. Furthermore, diversity in the cardiology workforce can reduce cardiovascular health disparities in the community [3,4].

There are a number of contributing factors to gender disparities among cardiologists. The perception that a career in cardiology is associated with a poor work–life balance has been widely reported among prospective trainees [6,8]. Women tend to put a higher value on the work–life balance than males, which means the perception of a poor work–life balance in cardiology exacerbates the gender disparity [10]. There is also a perception of a lack of diversity in cardiology, a value again more heavily considered by female trainees than their male counterparts [3,6,10]. With gender and ethnic disparities comes the perception of discrimination in cardiology [6]. In particular, interventional cardiology is portrayed as an “old boys club” that lacks female role models, creates a poor work–life balance, and involves high radiation doses, all of which discourage women from pursuing a career in interventional cardiology [7,11]. In the US, female cardiologists are less likely to be married than their male colleagues (74% compared to 89%) and less likely to have children than the men (72% compared to 86%) [12]. This vicious cycle requires an aggressive intervention that includes increasing the visibility of women in cardiology [13]. Artificial intelligence (AI) may be a valuable tool to use in that context to positively influence perceptions, change the workforce, and help develop educational interventions that aid efforts to craft a more positive professional reputation.

Artificial Intelligence in Cardiology

While there is a rich history of AI in cardiology, generative AI has only recently been utilised. This includes generative adversarial networks (GANs) [14], generative text-to-text applications associated with large language models (LLMs) like ChatGPT or Gemini, and diffusion models for text-to-image creation like Dall-E or Midjourney. Generative AI could be used for digital marketing in medicine, answering patient questions, or translating and summarising medical information for patients. In particular, LLMs could enhance patient care, improve communication and education, improve medical education and training, enhance administrative operations, help identify adverse events, augment risk prediction, streamline the cardiology workup, and provide medical diagnosis and decision support [15,16]. Nonetheless, their accuracy, appropriateness, and privacy remain concerns [15,16]. There are a number of key issues to consider [17]:

- Lack of transparency of algorithms.
- Variable accuracy and potential errors, fabrication, and hallucination.
- Lack of algorithm responsibility for errors or accountability, which raises a question regarding liability.
- How the lack of a detailed patient history or medical knowledge limits the accuracy and appropriateness of applications.
- Privacy and confidentiality of the data.
- Variability in the success with text prompts and a reliance on users’ patience or skills for prompt engineering.

Despite the limitations of generative AI, reports suggest it could add USD 2.6–4.4 trillion annually to the economy, with healthcare making up USD 150–260 billion of bonus [18].

Text-to-image generative AI could transform patient and medical education [19]. There are a broad variety of potential applications, such as generating promotional or marketing material, producing educational or information posters, producing anatomical and pathological images for medical education, and interpreting imported images for patient or medical purposes. Certainly, text-to-image generative AI promises to capitalise on the power of images in making up visual-rich learning aids, but there is variability in both the general capability and specifically in medically related images across the various applications available (Figures 1 and 2). Previously, DALL-E 3 was used to produce images that should have been accurate, educationally valuable depictions of congenital heart disease, but an evaluation revealed that 81% were anatomically incorrect, 85% were incorrectly annotated, and 78% were not usable in education [19]. Similarly, Midjourney production of medical art is limited by algorithm bans on medical terms, a shallow pool of stock medical images, and a variable quality of outputs for medical images [20]. In another

investigation, DALL-E 3 was used to produce a variety of cardiology images including a 12-lead ECG representing specific conditions, training images for ECG interpretation, and information posters for performing CPR [21]. DALL-E 3 was proposed to have an improved image generation and prompt interpretation capability over DALL-E 2. Nonetheless, DALL-E 3 produced outputs that were recognisable as ECGs but were not fit for purpose, instead resembling a malfunctioning ECG more than something authentic [21]. The ECG strips created also did not reflect the conditions prompted, although DALL-E 3 managed to represent the very basic information adequately. Meanwhile, the CPR posters produced were satisfactory for education but not fit for the purpose of their design because key information was not featured prominently [21].



Figure 1. In response to the prompt “create a realistic image of a typical interventional cardiologist”, the following images were produced by DALL-E 3 (top left), Firefly 2 (top right), Stable Diffusion 2.1 (bottom left), and Midjourney 5.2 (bottom right). Readers can evaluate the professional representativeness and accuracy, image quality, image-to-text alignment, and visual reasoning of this suite of images in the context of the cardiology profession.

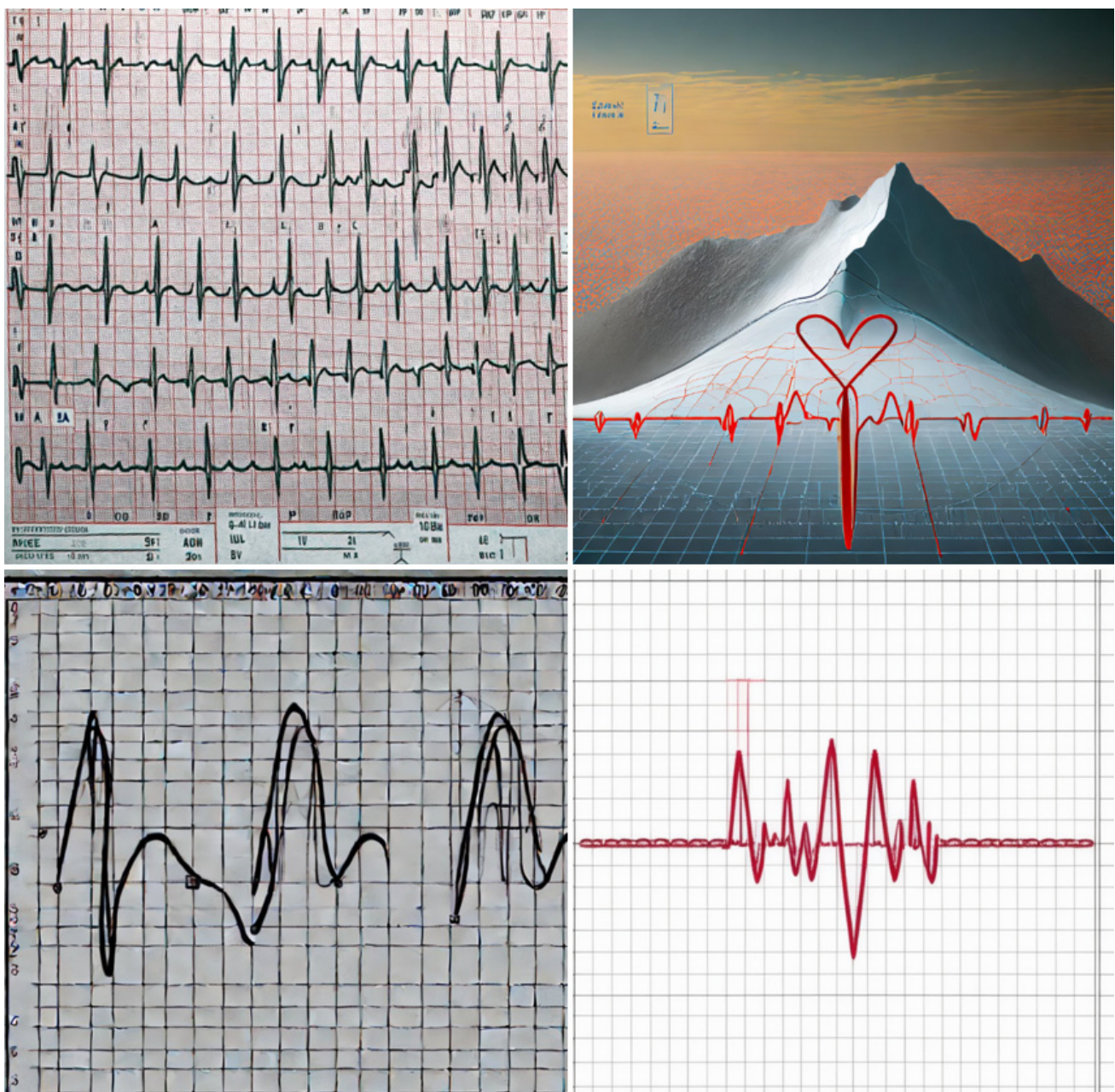


Figure 2. In response to the prompt “create an image of the printout for a single-lead ECG trace that represents sinus tachycardia”, the following images were produced by DALL-E 3 (**top left**), Firefly 2 (**top right**), Stable Diffusion 2.1 (**bottom left**), and Midjourney 5.2 (**bottom right**). Readers can evaluate the accuracy, image quality, image-to-text alignment, and visual reasoning of this suite of images.

Strategies for creating an inclusive and diverse cardiology workforce will benefit from educational and promotional material that reflects inclusivity. Social media and generative AI play an increasingly important role in this space. Generative AI (e.g., Adobe Firefly, Midjourney, and DALL-E) extends a low-cost, rapid, and accessible opportunity to craft high-quality images that could be used by cardiologists, professional organisations, other medical and health colleagues, advocacy groups, and patients themselves. DALL-E 3 is the latest (at the time of writing) text-to-image extension for ChatGPT’s subscription version, GPT-4o. DALL-E 3 can be used, according to ChatGPT, to create educational material and

produce artistic interpretations that could be used to represent cardiology and cardiologists in materials (print or virtual) for public awareness, patient information, and recruitment.

Despite the potential benefits of generative AI, these diffusion models (e.g., DALL-E 3) have the potential to reinforce or amplify gender and ethnicity biases. Misrepresentation or lack of diversity and inclusivity in produced images can perpetuate stereotypes and negatively shape external perceptions of the profession. Furthermore, reinforcing biases and stereotypes can exacerbate discrimination, threaten trust, undermine cultural safety, and influence recruitment, further constructing barriers to inclusivity and diversity. Generative adversarial network (GAN) and diffusion model images are synthetic or fake [22,23], so it remains possible, and indeed probable, that the algorithm will produce images that are less than representative of the cardiology workforce (human resources).

The potential for reinforcement or amplification of biases and stereotypes has recently been highlighted by a number of studies. In an investigation of generative AI representations of surgeons, Midjourney and Stable Diffusion amplified gender and ethnicity biases among surgeons, while DALL-E 2 produced a representative distribution for gender and ethnicity [24]. Midjourney and Stable Diffusion depicted 98% of surgeons as male and white [24]. In a similar study among surgeons, DALL-E 2 depicted 71.9% of surgeons as male and 50% with a light skin tone, while Midjourney produced 87.5% and 100% of images in these ways, respectively [25]. In a similar evaluation among ophthalmologists, DALL-E 2 produced images that were 77.5% male and 75% light skin tone [26]. Nonetheless, these three studies used prompts calling for individual images of either surgeons or ophthalmologists, which increased the likelihood of bias in any individual image. The generation of images containing groups of surgeons or ophthalmologists would produce a more realistic insight into the depth of these biases. Nonetheless, individual images remain an authentic prompt, as this may reflect a large proportion of user applications of generative AI text-to-image applications.

2. Methods

In March 2024, DALL-E 3 was utilised via GPT-4 to generate a series of individual and group images of cardiologists. Cardiologists were considered in four groupings based on specialisation: general cardiologist, interventional cardiologist, nuclear cardiologist, and cardiothoracic surgeon. For each of the four specialist roles, eight iterations of images were generated, to produce thirty-two individual images, using the following prompt:

- Create an image of a typical

Subsequently, three iterations of group images were generated, to produce 12 images, using the following prompts:

- Create an image of a group of 10 cardiologists socialising at a coffee shop.
- Create an image of a group of 10 cardiologists socialising at a bar.
- Create an image of a group of 10 cardiologists on an expert panel at a professional conference.
- Create an image of a group of five cardiologists discussing a difficult case.

Collectively, 44 images of cardiologists were produced for evaluation, of which 32 were individual characters and the remaining 12 were comprised of multiple (5 to 17) characters. A total of 143 characters were evaluated. This reflected an 80% statistical power at the 95% confidence level for detection of a 10% difference in proportion from the known population statistics. All images were independently evaluated by two reviewers (GC, HK) for apparent gender (male, female, unclear), age (<35, 35–55, >55, unclear), and ethnicity (Caucasian, non-Caucasian, unsure). Skin tone was evaluated using the Massey–Martin NIS Skin Color Scale (SCS) as 1–3 (light), 4–6 (mid), or 7–10 (dark) [27]. Discrepancies in responses were resolved by third-reviewer (CC) consensus. All data were AI generated and analysed by the authors; thus, ethical approval was not required.

The statistical significance among nominal data was calculated using Chi-Square analysis with the Pearson Chi-Square (X^2) test for a normal distribution and the Likelihood

Ratio Chi-Square (G^2) test for a non-normal distribution. A p -value less than 0.05 was considered significant.

3. Results

Inter-observer agreement was 100% for gender, 60.8% for age, and 91.6% for skin tone between observer 1 and 2. Discrepancies associated with age were exclusively associated with one observer indicating unsure while the other used an age bracket, rather than differences in age groupings themselves. All discrepancies were resolved by consensus.

Among the images of individual cardiologists ($N = 32$), DALL-E 3 generated 100% as male ($N = 32$) and 90.6% as Caucasian ($N = 29$) (Figure 3). Light skin tones were evident in 96.9% ($N = 31$). The age in 28.1% was under 35 years ($N = 9$), with 65.6% aged 35 to 55 years ($N = 21$) and 6.3% older than 55 years ($N = 2$). The data are summarised in Table 1. Among the different professional roles in cardiology, all over-represented males (100%). A general cardiologist had a statistically lower age representation than nuclear cardiologists, interventional cardiologists, and cardiothoracic surgeons ($p < 0.0001$). There was no statistically significant difference across the four cardiology roles for ethnicity ($p = 0.1908$) or skin tone ($p = 0.4119$).

Table 1. Percentage distribution of various characteristics of the cardiology workforce when using DALL-E 3 for individual images of various classifications of cardiologists.

Characteristic	Cardiologist	Interventional Cardiologist	Percentage (N) Nuclear Cardiologist	Cardiothoracic Surgeon	Total
Gender					
Male	100% (8)	100% (8)	100% (8)	100% (8)	100% (32)
Female	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Ethnicity					
Caucasian	87.5% (7)	75% (6)	100% (8)	100% (8)	90.6% (29)
Non-Caucasian	12.5% (1)	0% (0) *	0% (0)	0% (0)	3.1% (1) *
Skin Tone					
Light	87.5% (7)	100% (8)	100% (8)	100% (8)	96.9% (31)
Mid	12.5% (1)	0% (0)	0% (0)	0% (0)	3.1% (1)
Dark	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Age					
<35	87.5% (7)	0% (0)	25% (2)	0% (0)	28.1% (9)
35–55	12.5% (1)	100% (9)	50% (4)	100% (9)	65.6% (21)
>55	0% (0)	0% (0)	25% (2)	0% (0)	6.3% (2)

* Discrepancies from 100% reflect an unsure classification.

Among the images of groups of cardiologists ($N = 111$), DALL-E 3 generated 82.0% ($N = 91$) as male and 92.0% with a light skin tone ($N = 102$) (Figure 4). The 95% confidence interval of the gender distribution ranged from 12.0% to 26.2% for female cardiologists and, therefore, the distribution did not vary in a statistically significant way from that of the actual gender statistics for cardiologists ($p = 0.3854$). The age for 8.1% was under 35 years ($N = 9$), with 71.2% aged 35 to 55 years ($N = 79$) and 20.7% older than 55 years ($N = 23$). The data are summarised in Table 2.



Figure 3. Representative examples of DALL-E 3-generated images of “a cardiologist” (**top left**), “an interventional cardiologist” (**top right**), “a nuclear cardiologist” (**bottom left**), and “a cardiothoracic surgeon” (**bottom right**). There are some obvious errors like wearing two watches, an arm where a head should be, and the absence of a head covering during surgery, but a more detailed critique will be left to the readers.



Figure 4. Representative examples of DALL-E 3-generated images of “a group of 10 cardiologists socialising at a coffee shop” (**top left**), “a group of 10 cardiologists discussing a difficult case” (**bottom left**), “a group of 10 cardiologists socialising at a bar” (**top right**), and “a group of 10 cardiologists on an expert panel at a professional conference” (**bottom right**).

Table 2. Percentage distribution of various characteristics of the cardiology workforce when using DALL-E 3.

Characteristic	Individual Data	Percentage (N) Group Data	Collective Data
Gender			
Male	100% (32)	82.0% (91)	86.0% (123)
Female	0% (0)	18.0% (20)	14.0% (20)
Skin Tone			
Light	96.9% (31)	92.0% (102)	93.0% (133)
Mid	3.1% (1)	7.2% (8)	6.3% (9)
Dark	0% (0)	0.9% (1)	0.7% (1)
Age			
<35	28.1% (9)	8.1% (9)	12.6% (18)
35–55	65.6% (21)	71.2% (79)	69.9% (100)
>55	6.3% (2)	20.7% (23)	17.5% (25)

Collectively for individual and group images of cardiologists (N = 143), DALL-E 3 generated 86.0% (N = 123) as male and 14.0% as female (N = 20), and 93.0% had a light skin tone (N = 133). The 95% confidence interval for the gender distribution was 9.2% to 20.6% female, which did not vary significantly from actual cardiologists’ data ($p = 0.7342$). The age for 12.6% was under 35 years (N = 18), with 69.9% aged 35 to 55 years (N = 100) and 17.5% older than 55 years (N = 25). The data are summarised in Table 2.

4. Discussion

For the collective DALL-E 3-depicted gender distribution among cardiologists, there was no statistically significant difference between the percentage of females depicted as cardiologists (14%) and the US, UK, and Australian workforce data (13–15%). This was also true for the depicted (7%) versus actual ethnicity in the US (8%). Previous research among surgeons and ophthalmologists indicated that DALL-E 2 produced a gender and ethnicity bias [24–26], but in this study, no such bias has been demonstrated for cardiologists, despite representations of males being higher than those reported for surgeons (71.9%) and ophthalmologists (77.5%). This reflects, in part, the gender bias inherent in the current cardiology workforce. Nonetheless, among individual images, a clear gender bias toward males (100%) was shown, which was statistically higher ($p = 0.0017$) than the previously reported gender bias of DALL-E 2. In contrast to previous studies, this investigation included group images of cardiologists, to evaluate biases more rigorously. In isolation, the group images reflected improved diversity; however, with 82% and 92% for male and light skin tone, respectively, there remained a significant bias. The collective (group and individual) data revealed a statistically higher proportion of men as cardiologists (86%) compared to the worst case in the previous studies (77.5%) ($p = 0.0148$). While this might suggest that DALL-E 3 deliberately represented cardiologists according to known workforce data, there is no convincing evidence that this is the case. It is most probably a coincidence based on the confluence of general DALL-E 3 gender bias and the under-representation of females in the cardiology workforce.

The ongoing portrayal of cardiology as lacking gender and ethnic diversity is a contributing factor to the slow progress being made toward striking a gender balance. Certainly, the DALL-E 3 representation of cardiologists perpetuates that perception, which can potentially harm the profession and reinforce biases. Yet, generative AI could provide a powerful tool for improving the visibility of diverse groups in cardiology. A lack of visibility of women or specific cultural groups in cardiology undermines cultural safety for the potential future workforce and for cardiac patients. Visual media produced by generative AI, in the absence of new training data that are highly curated to eliminate historical and institutionalised biases, require careful prompt engineering to craft images more suitable for professional use. The attractiveness of generative AI is ease of use, accessibility and rapid outcomes; all of which need to be deprioritised to use multiple prompt iterations and careful quality assurance to produce an appropriate image for professional use. This requires explicit user-generated prompts relating to gender and ethnicity and aiming to reflect inclusivity (Figure 5). This would be appropriate for producing educational, recruitment or promotion material. Unfortunately, use outside the profession is more likely to rely on the immediate gratification provided by less complicated use. Compounding this representation of cardiology as being white male dominated for users outside cardiology is the more probable prompts using individual images of cardiologist. As reflected in these results, individual images of cardiologists amplify the inherent historical and institutionalised biases.



Figure 5. Overcoming bias with prompt engineering by specifying in a DALL-E 3 prompt “a realistic image of a female cardiologist performing a treadmill stress test on an obese male patient” (**top left**) and “a female cardiologist monitoring the images on a computer while a patient has a cardiac PET scan” (**bottom left**). Anomalies in the generated images threaten professionalism because of the attempt at photo-realism. For graphical messaging, it may offer benefits in avoiding criticism of inaccuracies in photo-realistic images if we instead create caricatures. The original prompts for the images on the left were reloaded into DALL-E 3 with the additional prompt to “convert to a Disney-style image”. While less real, inaccuracies can be forgiven.

4.1. Summary of DALL-E 3 from ChatGPT

User Prompt:

There is a perception among graduating doctors that a career as a cardiologist has poor diversity and a poor work–life balance. This makes a career as a cardiologist less attractive

to women and ethnic minorities, who value a work–life balance and diversity more than men. How can DALL-E help address this issue?

ChatGPT (GPT-4):

Addressing the perception of cardiology as a field with poor diversity and little work–life balance requires a multifaceted approach. While DALL-E creates images based on text descriptions, it can play a role in visual storytelling and education. Here is how it might help:

- **Visual Campaigns:** DALL-E could be used to create visual representations of diversity in cardiology, showing a variety of people in different roles and settings, to help challenge stereotypes.
- **Educational Materials:** DALL-E could be used to develop educational content that can be used in presentations or on social media to highlight the importance of a work–life balance, showcase successful programs, and promote initiatives supporting diversity.
- **Role Models:** DALL-E could be used to generate illustrations featuring diverse cardiologists in a variety of work and life scenarios, possibly paired with real stories of cardiologists who manage a healthy work–life balance.
- **Highlighting Initiatives:** DALL-E could be used to generate images showcasing initiatives, programs, and events that promote diversity and a work–life balance in cardiology, which can help attract a broader range of medical graduates to the field.

For an effective strategy, such visual materials should be used as part of broader initiatives, such as policy changes, mentorship programs, and flexible work schedules that directly address the concerns around a work–life balance and diversity.

User Note:

These four bullet points were evaluated using DALL-E 3 via GPT-4, and the resultant images are available as Supplementary Information.

User Prompt:

Are there risks of using DALL-E and worsening gender or ethnicity biases?

ChatGPT (GPT-4):

Yes, using an AI system like DALL-E could potentially worsen gender or ethnicity biases, especially if it is not used thoughtfully. Here are a few risks to consider:

- **Reinforcing Stereotypes:** If DALL-E generates images based on biased prompts or existing stereotypes, it can reinforce those biases. For instance, if it creates images of cardiologists that predominantly depict one gender or ethnicity, it may perpetuate the idea that cardiology is not inclusive.
- **Lack of Representation:** If the AI is not provided with diverse and inclusive prompts, it might produce images that lack representation of minority groups.
- **Cultural Sensitivity:** The AI might not fully understand cultural nuances, which can lead to the creation of images that are culturally insensitive or inappropriate.
- **Bias in Training Data:** The AI's training data may contain historical biases, which can be reflected in the images it generates.

To minimise these risks, it is important to use carefully crafted prompts that encourage diversity and avoid stereotypes. Additionally, reviewing and vetting the AI-generated images for biases before use is critical. It is also beneficial to have diverse teams working on AI development and implementation, to ensure a variety of perspectives are considered.

User Prompt:

How does DALL-E compare to other text-to-image generative AI like Firefly, Mid-journey, and Stable Diffusion when it comes to gender and ethnicity biases in created images?

ChatGPT (GPT-4):

When discussing the potential for AI systems like DALL-E, Midjourney, and Stable Diffusion to exacerbate or mitigate gender and ethnicity biases, several key factors are typically considered:

- **Training Data:** All these models learn from vast datasets sourced from the internet, which inevitably contain biases present in human culture. The model's output reflects these biases, so if the training data over-represent certain demographics, the AI might generate images that do so as well.
- **Bias Mitigation Efforts:** Different AI platforms may employ various strategies to address and mitigate biases. These could include curating training datasets, adjusting algorithms, or setting guidelines for usage.
- **User Input:** The way users phrase prompts can also influence the output and potentially perpetuate stereotypes. Tools might offer guidance on crafting neutral prompts to help reduce bias.
- **Filtering and Moderation:** Some systems may have more stringent content moderation to prevent the generation of biased or harmful images.
- **Transparency and Updates:** The organisations behind these models may differ in how transparent they are about their models' limitations and biases and how actively they work to update and improve their models.
- **Research and Improvement:** Ongoing research into bias in AI and machine learning is vital. Models that are regularly updated with the latest research findings might handle bias more effectively.

As of my last update in April 2023, DALL-E has become known for its approach to mitigating biases by imposing content restrictions and refining its training data. However, the specifics of these mechanisms compared to other systems like Firefly, Midjourney, and Stable Diffusion vary, and the effectiveness of each system's bias mitigation strategies will best be determined through direct comparison and testing, which is a rapidly developing area of AI research. Each new version or iteration of these models potentially comes with improvements or changes in how biases are addressed.

4.2. Limitations

It was thought possible that DALL-E 3 may respond to the characteristics of the individual user and bias outcomes; however, neither GPT-4 nor DALL-E 3 has input from or connection to either the user's computer or user account. While a range of individual and group images were produced, these did not represent the very broad range of possible prompts for representations of cardiologists or the cardiology profession. The results were, however, indicative of potential biases. Similarly, the number of iterations for each prompt was finite. It is possible that a larger number of iterations could have skewed the results in either of two directions with respect to the biases.

4.3. Future Directions

The results reflect the performance of generative AI at the time of writing. The algorithms and technology are advancing rapidly and capabilities are expected to significantly improve over a short period of time. The applications investigated throughout this article are likely to be within the scope of generative AI in the near future. To that end, the future of text-to-image generative AI is likely to have a significant footprint in the education space, for both professionals and patients. Re-evaluation of the capabilities of text-to-image generative AI with respect to the purported benefits where expectations are not met, which were outlined in the introduction, will be required at regular intervals (e.g., bi-annually) as AI technology and algorithms are enhanced. The horizon promises to be transformational.

5. Conclusions

Text-to-image generative AI using DALL-E 3 suffers from inherent gender and ethnic biases that reinforce cardiology as being a white, male domain. Despite this, both the gender and ethnic biases among cardiologists approximate the under-representation of females

and ethnic minorities in cardiologists' ranks. This is most probably incidental rather than the deliberate alignment of images with workforce data. In either case, the representation of women in cardiology does not reflect the progress in addressing gender disparity reflected in higher proportions of female trainee cardiologists. The assimilation of generative AI tools into patient education, promotion, and marketing across cardiology demands mitigation strategies to reduce the deleterious effects of amplifying historical and institutionalised biases. These results highlight the need for the development of appropriate use criteria and minimum standards of quality and professionalism for the use of text-to-image generative AI in cardiology. The insights gleaned are useful for guiding users utilising text-to-image generative AI for professional purposes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/info15100594/s1>.

Author Contributions: Conceptualization, G.C.; Methodology, G.C.; Formal analysis, G.C., C.C. and H.K.; Investigation, G.C.; Resources, G.C.; Data curation, G.C.; Writing—original draft, G.C., C.C. and H.K.; Writing—review & editing, G.C., C.C. and H.K.; Project administration, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: There are no conflicts of interests or funding declarations to be made.

References

1. Crews, D.C.; Collins, C.A.; Cooper, L.A. Distinguishing workforce diversity from health equity efforts in medicine. *JAMA Health Forum* **2021**, *2*, e214820. [CrossRef] [PubMed]
2. Lee, R. How should medical school admissions drive health care workforce diversity? *AMA J. Ethics* **2021**, *23*, E912–E918.
3. Zafar, M.D.B.; Jamil, Y.; Bilal, M.; Rathi, S.; Anwer, A. Impact of racial, ethnic and gender disparities in Cardiology. *Curr. Probl. Cardiol.* **2023**, *48*, 101725. [CrossRef]
4. Crowley, A.L.; Damp, J.; Sulistio, M.S.; Berlacher, K.; Polk, D.M.; Hong, R.A.; Capers, Q., IV. Perceptions on diversity in cardiology: A survey of cardiology fellowship training program directors. *J. Am. Heart Assoc.* **2020**, *9*, e017196. [CrossRef]
5. Mehta, L.S.; Fisher, K.; Rzeszut, A.K.; Lipner, R.; Mitchell, S.; Dill, M.; Douglas, P.S. Current demographic status of cardiologists in the United States. *JAMA Cardiol.* **2019**, *4*, 1029–1033. [CrossRef]
6. Capdeville, M. Gender disparities in cardiovascular fellowship training among 3 specialties from 2007 to 2017. *J. Cardiothorac. Vasc. Anesth.* **2019**, *33*, 604–620. [CrossRef]
7. Burgess, S.; Shaw, E.; Ellenberger, K.; Thomas, L.; Grines, C.; Zaman, S. Women in medicine: Addressing the gender gap in interventional cardiology. *J. Am. Coll. Cardiol.* **2018**, *72*, 2663–2667. [CrossRef]
8. Burgess, S.; Shaw, E.; Zaman, S. Women in cardiology: Underwhelming rate of change. *Circulation* **2019**, *139*, 1001–1002. [CrossRef] [PubMed]
9. Greenwood, B.N.; Carnahan, S.; Huang, L. Patient-physician gender concordance and increased mortality among female heart attack patients. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 8569–8574. Available online: <https://pubmed.ncbi.nlm.nih.gov/30082406/> (accessed on 3 October 2022). [CrossRef] [PubMed]
10. Douglas, P.S.; Rzeszut, A.K.; Merz, C.N.B.; Duvernoy, C.S.; Lewis, S.J.; Walsh, M.N.; Gillam, L. Career preferences and perceptions of cardiology among US internal medicine trainees: Factors influencing cardiology career choice. *JAMA Cardiol.* **2018**, *3*, 682–691. [CrossRef] [PubMed]
11. Yong, C.M.; Abnoui, F.; Rzeszut, A.K.; Douglas, P.S.; Harrington, R.A.; Mehran, R. Society for Cardiovascular Angiography and Interventions Women in Innovations (SCAI WIN). Sex differences in the pursuit of interventional cardiology as a subspecialty among cardiovascular fellows-in-training. *Cardiovasc. Interv.* **2019**, *12*, 219–228.
12. Lewis, S.J.; Mehta, L.; Douglas, P.; Gulati, M.; Limacher, M.; Poppas, A.; Walsh, M.; Bohannon, K.; Rzeszut, A.; Duvernoy, C. The professional life of cardiologists: Insights from the third American College of Cardiology professional life survey. *J. Am. Coll. Cardiol.* **2016**, *67*, 1928. [CrossRef]
13. Lundberg, G.; Tamirisa, K.; Le, E.; Wood, M.; York, M.; Singh, T. Addressing Gender Equity in Cardiology. *Am. J. Med.* **2020**, *133*, 1113–1115. [CrossRef]
14. Karatzia, L.; Aung, N.; Aksentijevic, D. Artificial intelligence in cardiology: Hope for the future and power for the present. *Front. Cardiovasc. Med.* **2022**, *9*, 945726. [CrossRef] [PubMed] [PubMed Central]
15. Boonstra, M.J.; Weissenbacher, D.; Moore, J.H.; Gonzalez-Hernandez, G.; Asselbergs, F.W. Artificial intelligence: Revolutionizing cardiology with large language models. *Eur. Heart J.* **2024**, *45*, 332–345. [CrossRef]

16. Gala, D.; Makaryus, A.N. The utility of language models in cardiology: A narrative review of the benefits and concerns of ChatGPT-4. *Int. J. Environ. Res. Public Health* **2023**, *20*, 6438. [\[CrossRef\]](#)
17. Ooi, K.B.; Tan GW, H.; Al-Emran, M.; Al-Sharafi, M.A.; Capatina, A.; Chakraborty, A.; Wong, L.W. The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *J. Comput. Inf. Syst.* **2023**, 1–32. [\[CrossRef\]](#)
18. Chui, M.; Hazan, E.; Roberts, R.; Singla, A.; Smaje, K. The Economic Potential of Generative AI. 2023. Available online: <https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/the%20economic%20potential%20of%20generative%20ai%20the%20next%20productivity%20frontier/the-economic-potential-of-generative-ai-the-next-productivity-frontier.pdf> (accessed on 5 August 2024).
19. Temsah, M.H.; Alhuzaimi, A.N.; Almansour, M.; Aljamaan, F.; Alhasan, K.; Batarfi, M.A.; Nazer, R. Art or Artifact: Evaluating the Accuracy, Appeal, and Educational Value of AI-Generated Imagery in DALL· E 3 for Illustrating Congenital Heart Diseases. *J. Med. Syst.* **2024**, *48*, 54. [\[CrossRef\]](#)
20. Mazzoli, C.A.; Semeraro, F.; Gamberini, L. Enhancing Cardiac Arrest Education: Exploring the potential use of MidJourney. *Resuscitation* **2023**, *189*, 109893. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zhu, L.; Mou, W.; Wu, K.; Zhang, J.; Luo, P. Can DALL-E 3 Reliably Generate 12-Lead ECGs and Teaching Illustrations? *Cureus* **2024**, *16*, e52748. [\[CrossRef\]](#)
22. Currie, G.; Rohren, E. Radiation dosimetry, artificial intelligence and digital twins: Old dog, new tricks. *Semin. Nucl. Med.* **2022**, *52*, 498–503. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Currie, G. The emerging role of artificial intelligence and digital twins in pre-clinical molecular imaging. *Nucl. Med. Biol.* **2023**, *120–121*, 108337. [\[CrossRef\]](#)
24. Ali, R.; Tang, O.Y.; Connolly, I.D.; Abdulrazeq, H.F.; Mirza, F.N.; Lim, R.K.; Johnston, B.R.; Groff, M.W.; Williamson, T.; Svokos, K.; et al. Demographic representation in 3 leading artificial intelligence text-to-image generators. *JAMA Surg.* **2024**, *159*, 87–95. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Cevik, J.; Lim, B.; Seth, I.; Sofiadellis, F.; Ross, R.J.; Cuomo, R.; Rozen, W.M. Assessment of the bias of artificial intelligence generated images and large language models on their depiction of a surgeon. *ANZ J. Surg.* **2024**, *94*, 287–294. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Choudhry, H.S.; Toor, U.; Sanchez, A.J.; Mian, S.I. Perception of race and sex diversity in ophthalmology by artificial intelligence: A DALL E-2 study. *Clin. Ophthalmol.* **2023**, *17*, 2889–2899. [\[CrossRef\]](#)
27. Massey, D.S.; Martin, J.A. *The NIS Skin Color Scale*; Princeton University Press: Princeton, NJ, USA, 2003.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.