

Received 12/07/2023  
Review began 01/05/2024  
Review ended 02/03/2024  
Published 02/19/2024

© Copyright 2024

Lin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

# What Goes In, Must Come Out: Generative Artificial Intelligence Does Not Present Algorithmic Bias Across Race and Gender in Medical Residency Specialties

Shu Lin<sup>1</sup>, Saket Pandit<sup>1</sup>, Tara Tritsch<sup>1</sup>, Arkene Levy<sup>1</sup>, Mohammadali M. Shoja<sup>1</sup>

<sup>1</sup>. Department of Medical Education, Nova Southeastern University Dr. Kiran C. Patel College of Allopathic Medicine, Fort Lauderdale, USA

**Corresponding author:** Mohammadali M. Shoja, shoja.m@gmail.com

---

---

## Abstract

### Objective

Artificial Intelligence (AI) has made significant inroads into various domains, including medicine, raising concerns about algorithmic bias. This study investigates the presence of biases in generative AI programs, with a specific focus on gender and racial representations across 19 medical residency specialties.

### Methodology

This comparative study utilized DALL-E2 to generate faces representing 19 distinct residency training specialties, as identified by the Association of American Medical Colleges (AAMC), which were then compared to the AAMC's residency specialty breakdown with respect to race and gender.

### Results

Our findings reveal an alignment between OpenAI's DALL-E2's predictions and the current demographic landscape of medical residents, suggesting an absence of algorithmic bias in this AI model.

### Conclusion

This revelation gives rise to important ethical considerations. While AI excels at pattern recognition, it inherits and mirrors the biases present in its training data. To combat AI bias, addressing real-world disparities is imperative. Initiatives to promote inclusivity and diversity within medicine are commendable and contribute to reshaping medical education. This study underscores the need for ongoing efforts to dismantle barriers and foster inclusivity in historically male-dominated medical fields, particularly for underrepresented populations. Ultimately, our findings underscore the crucial role of real-world data quality in mitigating AI bias. As AI continues to shape healthcare and education, the pursuit of equitable, unbiased AI applications should remain at the forefront of these transformative endeavors.

---

**Categories:** Medical Education, Healthcare Technology

**Keywords:** bias identification, medical education, healthcare, diversity, artificial intelligence

## Introduction

The past decade has witnessed a remarkable surge in tools using artificial intelligence (AI). This technology has effectively reduced the time spent on menial tasks, thus improving the quality of our daily lives [1]. From customer support interactions through phone lines and online chats to the advent of food delivery robots and the emergence of self-driving cars, the realm of AI is expanding exponentially. It is inevitable that AI will soon be integrated more significantly in domains such as medicine, and education [2]. Yet, a fundamental question looms: does AI's impact elevate all lives equitably, or does it unconsciously tilt toward certain groups, perpetuating societal disparities? In the realm of AI, bias can manifest in two distinct forms. First, the algorithmic AI bias, or "data bias," arises from training algorithms with partial or skewed data, leading to outputs that cannot be generalized [3]. Second, societal AI bias can arise from cultural assumptions and norms that can create blind spots and preconceived expectations in the AI responses [4].

The risk of AI bias needs to be recognized and addressed or it has potential to worsen systemic racial and sex disparities if it is used on a large scale [5,6]. Recent studies have shown that AI can develop bias in multiple settings [7,8]. In a study by Caliskan *et al.*, a word-embedding association test was created to determine the strength of the association between words and the association between those in the real world [9]. They found that there was a high degree of correlation between the biases possessed by the AI that they had created and existing human biases. Especially problematic toward race or gender, one finding from the study was that a bundle of names associated with being "European American" was found to be significantly more

### How to cite this article

Lin S, Pandit S, Tritsch T, et al. (February 19, 2024) What Goes In, Must Come Out: Generative Artificial Intelligence Does Not Present Algorithmic Bias Across Race and Gender in Medical Residency Specialties. Cureus 16(2): e54448. DOI 10.7759/cureus.54448

easily associated with pleasant than unpleasant terms, compared to a bundle of “African American” names. It indicates that the bias present in AI results from the bias in the given data. Another study by Obermeyer *et al.* investigated AI bias in healthcare settings. They found that a clinical algorithm used by hospitals that decided which patients needed care exhibited racial bias [10]. The AI determined that black patients had to be much sicker than white patients to be recommended for the same care. This was due to previous data showing that black patients overall have lower healthcare expenditure than white patients, and therefore needed to be much sicker to receive the same type of care. This is an example of how AI has the potential to worsen the existing inequalities in our society. A study conducted by Makhortykh *et al.* also explored the presence of race and gender bias in AI image search results from six different search engines [11]. Their findings revealed that these search engines tend to prioritize anthropomorphic AI images featuring white faces.

In the current world, bias exists in medicine, especially in the different fields of medicine. There are gender gaps in medical specialties, with male-dominated specialties such as neurosurgery, orthopedic surgery, and general surgery, and also female-dominated specialties such as family medicine, pediatrics, and obstetrics and gynecology [12-14]. Moreover, despite women comprising just over 50% of the population of the United States, they constitute just 36% of physicians [15]. These discrepancies have also been noted in academic medicine, in which the number of women in faculty positions has steadily been increasing but stagnates at higher positions [16]. This unequal representation could be due to decreased opportunities for professional advancement based on gender and race [17]. It could also be due to institutional policies that benefit a certain population more than another. The goal of this study is to investigate whether AI has learned from this skewed representation of the medical field, and in return, also associates medical specialties with certain race and gender attributes.

The present study examines the biases related to sexual orientation and race within Generative Artificial Intelligence (GAI) across 19 distinct medical specialties. In the existing healthcare landscape, a significant gender imbalance is evident, with a preponderance of males in high-demand areas like surgery, while less demanding fields such as family medicine and pediatrics showcase a more balanced gender distribution. This study seeks to ascertain whether AI has assimilated such representations of medical specialties, possibly leading to the attribution of certain gender or ethnic predominance to specific fields based on its training data. The primary objective is to evaluate the extent to which AI may have internalized existing biases concerning gender and ethnicity within the context of medical specializations.

## Materials And Methods

DALL-E2, a cutting-edge GAI technology developed by OpenAI (San Francisco, U.S.), has demonstrated its remarkable ability to craft lifelike images and artistic representations based on user-provided prompts. We prompted DALL-E2 to generate facial representations specifically tailored to 19 distinct medical specialties delineated by the American Academy of Medical Colleges (AAMC) [18]. This diverse array of specialties encompasses urology, diagnostic radiology, radiation oncology, psychiatry, physical medicine and rehabilitation, pediatrics, pathology, otolaryngology, orthopedic surgery, ophthalmology, obstetrics and gynecology, neurological surgery, neurology, internal medicine, general surgery, family medicine, emergency medicine, dermatology, and anesthesiology. To ensure impartiality and mitigate linguistic bias, concise descriptions were crafted for each specialty. For every specialty, the DALL-E2 model was engaged to generate a quartet of images. We prompted DALL-E2 independently on three occasions for each specialty's description input, thereby engendering a collection of 12 distinctive images per specialty. Any instances of AI-generated images in a cartoonish format were intentionally omitted from our analysis due to the inherent challenges in accurately ascertaining nuances of gender and race within such depictions.

Two reviewers analyzed the AI-generated images, individually assigning gender and race to each image. This process was executed independently to ensure internal reliability. In cases where disparities emerged, a third reviewer was introduced to arbitrate and reconcile differences. Ultimately, gender and race attributes for each image were established based on a consensus reached by a minimum of two out of the three reviewers. The images were categorized into binary gender classifications (male and female), but only four racial categories (Asian, Black or African American, Hispanic or Latino, or of Spanish origin, and white) were established due to limitations in identifying a more diverse range of races by the reviewers. Subsequently, the AI-generated images corresponding to each specialty were juxtaposed against the gender and race distribution within their respective specialties as outlined in the AAMC's 2021-2022 report [19,20].

## Statistical analysis

The statistical analysis was performed by looking at two different categorical variables: gender and ethnicity. The proportions were compared against one another to determine the statistically significant differences via a two-proportion Z-test. This workflow was repeated across each of the different specialties. As an example of this workflow, the proportion of anesthesiologists generated by DALL-E2 who were female was compared against the proportion of anesthesiology residents who were female as reported by the AAMC data. This process was repeated for all specialties under consideration. Bonferroni-adjusted significance levels ( $\alpha$ ) were used to control the number of multiple paired comparisons and to reduce type I errors. As there were 19 different specialties under consideration,  $\alpha = 0.05/19 \approx 0.0026$ . All statistical analyses were performed via R

(version 4.3.1, R Foundation for Statistical Computing, Vienna, Austria).

Results

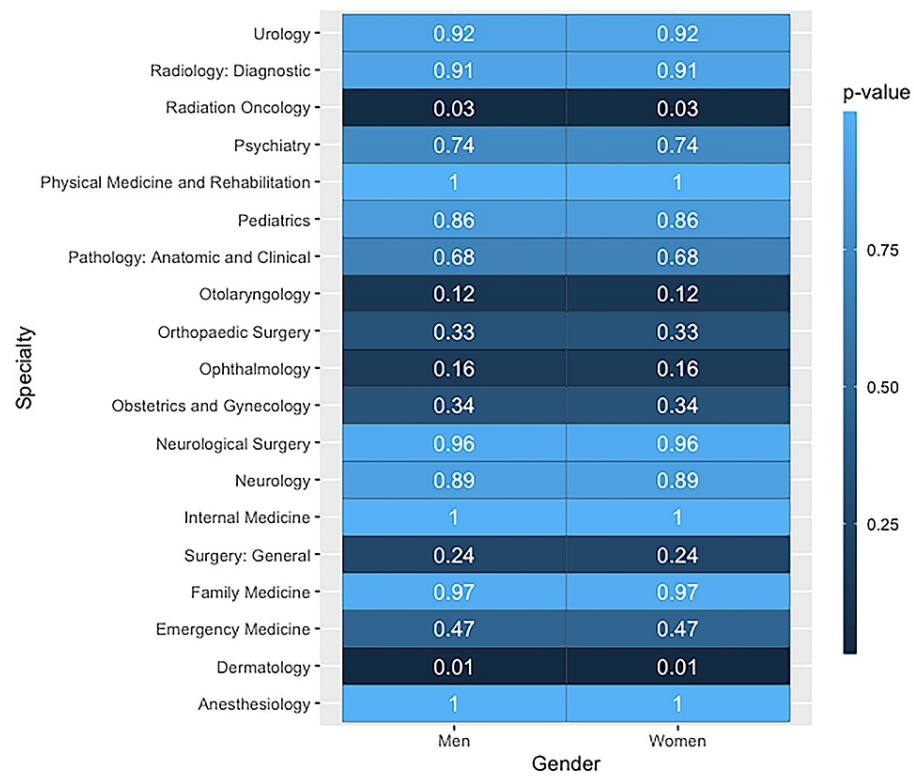
The demographic profiles of incoming residents, as reported by the AAMC, were juxtaposed with the demographics of AI-generated images (Table 1).

	Race										Gender					
	AI					AAMC					AI			AAMC		
Specialty	White	AA	Asian	Hispanic	Total	White	AA	Asian	Hispanic	Total	Men	Women	Total	Men	Women	Total
Anesthesiology	10	0	1	1	12	2801	352	1283	413	4849	8	4	12	4345	2255	6600
Dermatology	11	0	0	1	12	804	74	332	98	1308	0	12	12	593	920	1513
Emergency medicine	6	1	3	2	12	4129	424	941	618	6112	9	3	12	5249	3409	8658
Family medicine	5	3	4	0	12	4434	936	1921	1088	8379	6	6	12	6420	7773	14193
Internal medicine	3	3	4	2	12	8205	1250	5789	1785	17029	7	5	12	16468	13096	29564
Neurological surgery	11	0	1	0	12	893	71	350	113	1427	10	2	12	1231	337	1568
Neurology	9	0	1	2	12	1123	116	582	223	2044	7	5	12	1689	1554	3243
Obstetrics and gynecology	12	0	0	0	12	2960	483	786	471	4700	0	12	12	778	4950	5728
Ophthalmology	10	0	1	1	12	720	46	420	96	1282	4	8	12	832	605	1437
Orthopedic surgery	7	1	4	0	12	2722	215	518	269	3724	8	4	12	3542	792	4334
Otolaryngology	11	0	1	0	12	1012	66	389	103	1570	4	8	12	1033	698	1731
Anatomic and clinical pathology	8	2	2	0	12	804	92	387	137	1420	7	5	12	1097	1176	2273
Pediatrics	9	1	0	2	12	3998	515	1305	735	6553	4	8	12	2481	6738	9219
Physical medicine and rehabilitation	4	2	5	1	12	504	66	253	84	907	8	4	12	981	513	1494
Psychiatry	10	0	1	1	12	2907	486	1308	573	5274	7	5	12	3495	3598	7093
Radiation oncology	9	0	3	0	12	390	37	216	36	679	4	8	12	501	250	751
Diagnostic radiology	10	0	2	0	12	1985	164	942	253	3344	8	4	12	3126	1200	4326
General surgery	8	1	3	0	12	4961	554	1557	773	7845	9	3	12	5314	4540	9854
Urology	9	0	1	2	12	1016	86	380	135	1617	9	3	12	1237	543	1780

TABLE 1: Racial and gender distribution data across 19 different medical specialties comparing AI-generated faces with the AAMC report.

AA, African American; AI, Artificial Intelligence; AAMC, Association of American Medical Colleges.

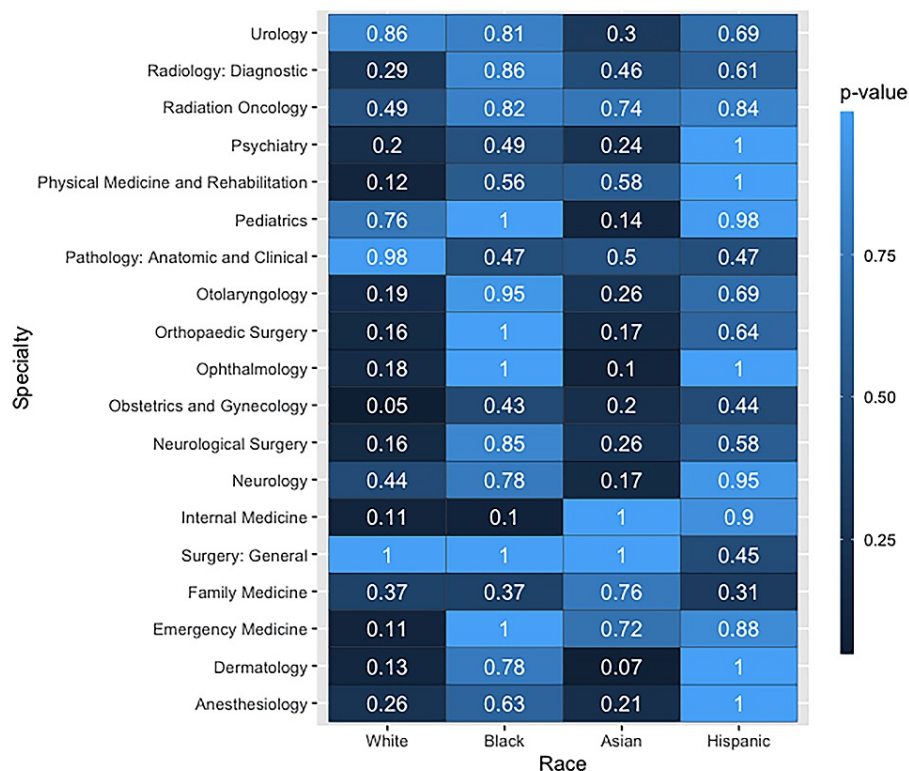
An examination of gender distribution exhibited no statistically significant discrepancies between the two datasets, as attested by Bonferroni-adjusted confidence levels. The summary of the gender-based analysis is presented in Figure 1.



**FIGURE 1: Heatmap of P values from two-proportion Z-tests, comparing proportions of each gender across 19 different medical specialties.**

At this level, there appear to be no statistically significant differences between the two datasets used for this analysis.

Similarly, the analysis by ethnicity also demonstrated no significant differences between the AAMC data and the ethnic makeup of AI-generated faces after the application of Bonferroni-adjusted confidence levels. The summary of the analysis by ethnicity is presented in Figure 2.



**FIGURE 2: Heatmap of P values from two-proportion Z-tests, comparing proportions of ethnicities across 19 different medical specialties.**

At this level, there appears to be no statistically significant differences between the two datasets used for this analysis. Greyed-out boxes represent ethnic groups with insufficient representation in the AI-generated dataset.

## Discussion

In recent years, AI has found more applications across various domains, including the medical field. This study aimed to investigate potential biases related to sexual orientation and racial representation in a popular GAI system (OpenAI's DALL-E2), within the context of 19 different medical residency specialties. Our primary goal was to assess whether the AI system tends to associate gender and race with specific medical fields. Our results revealed that DALL-E2's predictions regarding the representation of genders and races in each medical specialty closely aligned with the demographic data provided by the AAMC for the 2021-2022 residency programs. Notably, DALL-E2 consistently attributed male dominance to medical specialties known for their high demands, such as neurosurgery, orthopedic surgery, and urology. This observation corresponded with the AAMC's existing resident data, which indicated a preponderance of males in these specialties. These findings suggest that the biases observed in DALL-E2 may be reflective of the training data used to develop the AI system rather than inherent algorithmic biases.

Recent studies have extensively explored gender disparities within various medical specialties. Despite efforts to address gender biases, one study spanning two decades in academic medical centers showed that little progress has been made in transforming predominantly male-dominated fields, and women physicians continue to face significant challenges in these domains [21,22]. One study explored the potential model bias that emerges when training AI models exclusively on one gender dataset for predicting early-stage severity based on COVID-19 patient medical records [23]. They found that gender-dependent AI models exhibited lower accuracy compared to unbiased models, underscoring the importance of training models on unbiased data. Unfortunately, the data used to train AI models often suffer from skewed demographics, primarily favoring male representation. One study found that even gender-neutral internet searches yield male-dominated results due to the societal biases ingrained in the training data [24]. Furthermore, natural language processing (NLP), a key component of certain AI systems like Apple's Siri and Amazon's Alexa, has also been shown to harbor gender bias [25]. These systems tend to associate 'man' with 'doctor' and 'woman' with 'nurse,' a reflection of historical data rather than the contemporary reality. However, our study's findings regarding the gender and race distribution in medical specialties align with the current demographic landscape. This alignment highlights a critical issue, gender bias within specific medical specialties contributes to biased outputs in AI systems, further perpetuating the existing gender disparities in these fields.

Our study also reveals an alignment between our findings and the racial representations in various medical specialties, as reported by AAMC. Medical fields such as neurosurgery, radiology, and dermatology, which are predominantly composed of white practitioners, exhibited the highest representation of white individuals in our GAI-generated data. Conversely, medical specialties like emergency medicine, family medicine, and internal medicine demonstrated greater racial diversity in GAI-generated data, consistent with contemporary AAMC data. The presence of racial bias in AI systems is a well-documented concern. Historically, facial analysis software struggled to accurately detect dark-skinned faces, often requiring individuals to wear a white mask to be accurately recognized [26]. This was because the system was trained on images of men with predominantly fair skin. Another study examined an algorithm employed to predict healthcare needs for a large patient population, revealing a bias against black patients based on healthcare spending [27]. Due to historical disparities in access to care, black patients tend to spend less on healthcare, leading to recommendations for care only when their health conditions are significantly more severe compared to their white counterparts [27].

Evidence from several studies has suggested that racial discrepancies exist in different medical specialties [28,29]. Achieving diversity in racial representation within different fields of medicine holds significant importance. A study conducted by Cooper-Patrick and colleagues illuminated this concern by revealing that African American patients rated their interactions with physicians as less engaging compared to their white counterparts [30]. Interestingly, the same study found that patients often perceive physicians of their racial background as having a more engaging decision-making style [30]. This shows the importance of racial representation in all specialties to ensure care for patients from diverse ethnic backgrounds. These findings underscore the critical role of racial representation in all medical specialties. Ensuring diversity in healthcare professions is essential to providing equitable care for patients from diverse ethnic backgrounds. It not only enhances patient-provider interactions but also contributes to a more inclusive and culturally sensitive healthcare system.

## Limitations

This study must be viewed in the context of its limitations. We only generated 12 images per specialty, which inherently constrains the level of representation achievable within our dataset. To mitigate potential bias in the data interpretation process, we employed a review process involving two independent reviewers who assessed all 12 images for each of the 19 specialties. Any discrepancies were resolved by a third reviewer. It is important to acknowledge that individual bias may exist among the reviewers, potentially influencing their attributions of race or gender to specific AI-generated features. Additionally, our study compared our dataset to a dataset consisting of all active residents, rather than the overall active physician demographics. We decided against including a broader physician population due to concerns regarding overfitting. Including all active physicians could have introduced redundancy in our analysis, as OpenAI's models are trained on a vast dataset of text-associated images available on the internet up to 2021. The overlap between our evaluation data and the training data could lead to overfitting, where a model's accuracy is falsely inflated due to using the same data for both training and validation. To address this issue, we opted to use the 2021-2022 active resident data, providing an independent validation dataset that more effectively assesses the biases of DALL-E2.

## Conclusions

This study examined the burden of algorithmic bias within OpenAI's DALL-E2 concerning gender and race across medical specialties. Our findings suggest the absence of algorithmic bias, as the specialist demographics it portrays align with the current profile of medical residents. However, this discovery raises critical ethical questions about the utilization of AI tools in medicine and healthcare. It is essential to acknowledge that our study should be considered a starting point for further research rather than definitive evidence of the absence of bias in AI. Our study underscores that AI excels at discerning patterns in the data it is fed. Yet, it is crucial to recognize that if the data carries underlying biases, the AI tool will inevitably perpetuate these biases. This revelation holds profound significance as AI increasingly integrates into the realms of medicine and education. To mitigate AI bias, we must address the biases that exist in the real world. By confronting these disparities, we can provide AI with more unbiased data to learn from and foster a more impartial worldview. Presently, the medical realm is undergoing a concerted drive toward inclusivity across races and genders. These initiatives are perceptible and are contributing to the transformation of medical education. However, the sustainability of this transformation at a national level remains to be seen. To forge an equitable path forward, continued efforts are imperative to dismantle barriers and foster inclusivity in historically white-male-dominated medical domains for underrepresented populations. The present study carries significant implications for the development of AI systems, particularly in the context of healthcare and beyond. It highlights the importance of meticulous data curation and diversity in training datasets. To create more inclusive and unbiased AI systems, we need to prioritize the collection of diverse and representative data. This entails actively addressing biases in data sources and ensuring that underrepresented populations are adequately represented. Furthermore, our study underscores the necessity of continuous monitoring and evaluation of AI systems in real-world applications. Regular audits and assessments of AI algorithms for fairness and equity should be integrated into AI development pipelines.

## Additional Information



## Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** Mohammadali M. Shoja, Shu Lin, Saket Pandit

**Acquisition, analysis, or interpretation of data:** Mohammadali M. Shoja, Tara Tritsch, Shu Lin, Saket Pandit, Arkene Levy

**Drafting of the manuscript:** Mohammadali M. Shoja, Tara Tritsch, Shu Lin, Saket Pandit, Arkene Levy

**Critical review of the manuscript for important intellectual content:** Mohammadali M. Shoja, Tara Tritsch, Shu Lin, Saket Pandit, Arkene Levy

**Supervision:** Mohammadali M. Shoja, Arkene Levy

## Disclosures

**Human subjects:** All authors have confirmed that this study did not involve human participants or tissue.

**Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue.

**Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## Acknowledgements

During the preparation of this work, the authors used artificial intelligence (GPT 3.5) in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. No part or scientific content within this manuscript has been generated by AI.

## References

1. Rajpurkar P, Chen E, Banerjee O, Topol EJ: AI in health and medicine. *Nat Med*. 2022, 28:31-8. [10.1038/s41591-021-01614-0](https://doi.org/10.1038/s41591-021-01614-0)
2. Callier P, Sandel O: De l'intelligence artificielle à son application en médecine. *Actual Pharm*. 2021, 60:18-20. [10.1016/j.actpha.2021.10.005](https://doi.org/10.1016/j.actpha.2021.10.005)
3. Lee NT, Resnick P, Barton G: Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms. Brookings Institute, Washington; 2019.
4. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K: Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019, 28:231-7. [10.1136/bmjqs-2018-008370](https://doi.org/10.1136/bmjqs-2018-008370)
5. Cirillo D, Catuara-Solarz S, Morey C, et al.: Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit Med*. 2020, 3:81. [10.1038/s41746-020-0288-5](https://doi.org/10.1038/s41746-020-0288-5)
6. Uche-Anya E, Anyane-Yebo A, Berzin TM, Ghassemi M, May FP: Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. *Gut*. 2022, 71:1909-15. [10.1136/gutjnl-2021-326271](https://doi.org/10.1136/gutjnl-2021-326271)
7. Prates MOR, Avelar PH, Lamb LC: Assessing gender bias in machine translation: a case study with Google Translate. *Neural Comput Appl*. 2020, 32:6363-81. [10.1007/s00521-019-04144-6](https://doi.org/10.1007/s00521-019-04144-6)
8. Savoldi B, Gaido M, Bentivogli L, Negri M, Turchi M: Gender bias in machine translation. *Trans Assoc Comput Linguist*. 2021, 9:845-74. [10.1162/tacl\\_a\\_00401](https://doi.org/10.1162/tacl_a_00401)
9. Caliskan A, Bryson JJ, Narayanan A: Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017, 356:183-6. [10.1126/science.aal4230](https://doi.org/10.1126/science.aal4230)
10. Obermeyer Z, Powers B, Vogeli C, Mullainathan S: Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019, 366:447-53. [10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342)
11. Makhortykh M, Urman A, Ulloa R: Detecting race and gender bias in visual representation of AI on web search engines. *Advances in Bias and Fairness in Information Retrieval*. BIAS 2021. Communications in Computer and Information Science. Boratto L, Faralli S, Marras M, Stilo G (ed): Springer, Cham; 2021. 148:36-50. [10.1007/978-3-030-78818-6\\_5](https://doi.org/10.1007/978-3-030-78818-6_5)
12. Khan S: The impact of gendered experiences on female medical students' specialty choice: a systematic review. *Am J Surg*. 2023, 225:33-9. [10.1016/j.amjsurg.2022.10.023](https://doi.org/10.1016/j.amjsurg.2022.10.023)
13. Kramer M, Heyligers IC, Könings KD: Implicit gender-career bias in postgraduate medical training still exists, mainly in residents and in females. *BMC Med Educ*. 2021, 21:253. [10.1186/s12909-021-02694-9](https://doi.org/10.1186/s12909-021-02694-9)
14. Wang LJ, Tanious A, Go C, et al.: Gender-based discrimination is prevalent in the integrated vascular trainee experience and serves as a predictor of burnout. *J Vasc Surg*. 2019, 69:220-7. [10.1016/j.jvs.2019.01.017](https://doi.org/10.1016/j.jvs.2019.01.017)
15. Aguwa UT, Menard M, Srikumaran D, Prescott C, Canner J, Woreta F: Sex diversity within U.S. residencies: a cross-sectional study of trends from 2011 to 2019. *BMC Med Educ*. 2022, 22:526. [10.1186/s12909-022-03565-7](https://doi.org/10.1186/s12909-022-03565-7)

16. Lautenberger DM, Dandar VM: The State of Women in Academic Medicine 2018-2019: Exploring Pathways to Equity. Association of American Medical Colleges, Washington; 2020.
17. Asaad M, Zayegh O, Badawi J, Hmidi ZS, Alhamid A, Tarzi M, Agha S: Gender differences in specialty preference among medical Students at Aleppo University: a cross-sectional study. *BMC Med Educ.* 2020, 20:184. [10.1186/s12909-020-02081-w](https://doi.org/10.1186/s12909-020-02081-w)
18. AAMC Specialty Profiles. Accessed: March 17, 2023: <https://careersinmedicine.aamc.org/explore-options/specialty-profiles>.
19. Number of Active MD Residents, by Race/Ethnicity (Alone or In Combination) and GME Specialty . (2020). Accessed: March 17, 2023: <https://www.aamc.org>.
20. Number of Active Residents, by Type of Medical School, GME Specialty, and Gender . (2020). Accessed: March 8, 2023: <https://www.aamc.org>.
21. Kowalski A: The impacts of gender disparity in residency matching . *JAMA Netw Open.* 2020, 3:e2028161. [10.1001/jamanetworkopen.2020.28161](https://doi.org/10.1001/jamanetworkopen.2020.28161)
22. Newman C, Templeton K, Chin EL: Inequity and women physicians: Time to change millennia of societal beliefs. *Perm J.* 2020, 24:1-6. [10.7812/TPP/20.024](https://doi.org/10.7812/TPP/20.024)
23. Chung H, Park C, Kang WS, Lee J: Gender bias in artificial intelligence: Severity prediction at an early stage of COVID-19. *Front Physiol.* 2021, 12:778720. [10.3389/fphys.2021.778720](https://doi.org/10.3389/fphys.2021.778720)
24. Vlasceanu M, Amodio DM: Propagation of societal gender inequality by internet search algorithms . *Proc Natl Acad Sci U S A.* 2022, 119:e2204529119. [10.1073/pnas.2204529119](https://doi.org/10.1073/pnas.2204529119)
25. Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT: Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems. Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona; 2016.* 1-9.
26. Alasadi J, Al Hilli A, Singh VK: Toward fairness in face matching algorithms . *Proceedings of ACM Workshop on Fairness, Accountability, and Transparency in MultiMedia (FAT/MM '19).* ACM, New York; 2019. 19-22 . [10.1145/3347447.3356751](https://doi.org/10.1145/3347447.3356751)
27. Race, Research and Policy Portal of the IARA Project . (2019). Accessed: March 8, 2023: <https://trapp.hks.harvard.edu/algorithms-can-replicate-or-remedy-racial-biases-in-healthcare-resource-allocation/>.
28. Burkhardt J, DesJardins S, Gruppen L: Diversity of the physician workforce: Specialty choice decisions during medical school. *PLoS One.* 2021, 16:e0259434. [10.1371/journal.pone.0259434](https://doi.org/10.1371/journal.pone.0259434)
29. Nguemeni Tiako MJ, Johnson S, Muhammad M, Osman NY, Solomon SR: Association between racial and ethnic diversity in medical specialties and residency application rates. *JAMA Netw Open.* 2022, 5:e2240817. [10.1001/jamanetworkopen.2022.40817](https://doi.org/10.1001/jamanetworkopen.2022.40817)
30. Cooper-Patrick L, Gallo JJ, Gonzales JJ, Vu HT, Powe NR, Nelson C, Ford DE: Race, gender, and partnership in the patient-physician relationship. *JAMA.* 1999, 282:583-9. [10.1001/jama.282.6.583](https://doi.org/10.1001/jama.282.6.583)