**Title**: Who does the fairness in health AI community represent?

**Authors:**

Philippines: 2
Turkey: 3
Canada: 2
Brazil: 1
US: 9
Germany: 1
Denmark: 2
Peru: 1
Lebanon: 1

Ethnicity:
White: 3 (10)
Asian: 5 (8)
Black: 3 (3)
Hispanic: 0 (1)

Gender:
Male:12
Female: 10

Isabelle Rose I. Alberto (Philippines)
iialberto@up.edu.ph
ORCID: 0000-0002-7206-4770
University of the Philippines College of Medicine, Manila, Philippines

Nicole Rose I. Alberto (Philippines)
nialberto@up.edu.ph
ORCID: 0000-0001-9166-8134
University of the Philippines College of Medicine, Manila, Philippines

Yuksel Altinel (Turkey)
dryukselaltinel@gmail.com
ORCID 0000-0003-0113-4839
Bagcilar Research and Training Hospital, General Surgery Department, University of Health Sciences, Istanbul Turkiye

Sarah Blacker (Canada White)
sblacker@yorku.ca
ORCID 0000-0002-6146-8972
Department of Social Science
York University
Toronto, Ontario, Canada

William Warr Binotti (Brazil)
wbinotti@tuftsmedicalcenter.org
ORCID 0000-0001-6761-8807
New England Eye Center, Tufts Medical Center, Boston, MA, USA
Department of Medicine, Carney Hospital, Boston, MA, USA

Leo Anthony Celi (US Asian)
lceli@mit.edu

ORCID 0000-0001-6712-6626
Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA , USA
Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Tiffany Chua (US Asian)
tlchua@dons.usfca.edu
University of San Francisco, San Francisco, CA, USA

Amelia Fiske (Germany)
a.fiske@tum.de
ORCID 0000-0001-7207-6897
Institute for History and Ethics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany

Molly Griffin (US White)
mgriffi8@bidmc.harvard.edu
ORCID: 0000-0002-1615-2645
Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

Gulce Karaca (Turkey)
gkaraca@mgh.harvard.edu
ORCID 0000-0001-5211-0680
Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

Nkiruka Mokolo (US Black)
nmokolo21@email.mmc.edu
ORCID 0000-0002-6991-1828
Meharry Medical College School of Medicine Nashville, TN, USA

David Kojo N. Naawu (US Black)
dnathan22@email.mmc.edu
ORCID: 0000-0002-0470-107X
Meharry Medical College School of Medicine, Nashville, TN, USA

Jonathan Patscheider (Denmark), Corresponding Author
jpatscheider@truststamp.net
ORCID 0000-0003-4413-8573
Trust Stamp Denmark, Copenhagen, Denmark

Anton Petushkov (US White)
antonpet@umich.edu
ORCID 0000-0003-3661-236X
University of Michigan, Ann Arbor, MI, USA

Justin Quion (US Asian), Corresponding Author
justinmicq@gmail.com
ORCID 0009-0009-6844-3047

Charles Senteio (US Black)
charles.senteio@rutgers.edu
ORCID: 0000-0002-0254-3127
Department of Library and Information Science
Rutgers University School of Communication and Information, New Brunswick, NJ, USA

Simon Taisbak (Denmark)
ORCID 0009-0005-6264-2118
Inviso by Devoteam
Simon.taisbak@devoteam.com

İsmail Tırnova (Turkey)
tirnova77@gmail.com
ORCID: 0000-0003-4488-1607
Department of General Surgery
Başkent University School of Medicine, Istanbul, Turkey

Harumi Tokashiki (Peru)
harumi.tokashiki@tufts.edu
ORCID 0000-0002-7307-1884
Department of Medicine, Carney Hospital, Boston, MA, USA

Adrian Velasquez (US Asian)
adrian@sleepbetterclinic.com
ORCID 0000-0003-2700-467X
Warren Alpert School of Medicine at Brown University, Providence, RI, USA
Department of Medicine, Carney Hospital, Boston, MA, USA

Antonio Yaghy (Lebanon)
ayaghy@tuftsmedicalcenter.org
ORCID 0000-0002-5054-495X
New England Eye Center, Boston, MA, USA

Keagan Yap (Canada Asian)
keaganyap@college.harvard.edu
ORCID 0009-0002-4486-4516
Harvard College, Cambridge, MA, USA

The authors contributed equally to the manuscript and are listed alphabetically.

## ABSTRACT

OBJECTIVE: Artificial intelligence (AI) and machine learning are central components of today's medical environment. The fairness of AI, i.e. the ability of AI to be free from bias, has repeatedly come into question. This study investigates the diversity of the members of academia whose scholarship poses questions about the fairness of AI.

METHODS: The articles that combine the topics of fairness, artificial intelligence, and medicine were selected from Pubmed, Google Scholar, and Embase using keywords. Eligibility and data extraction from the articles were done manually and cross-checked by another author for accuracy. 375 articles were selected for further analysis, cleaned, and organized in Microsoft Excel; spatial diagrams were generated

using Public Tableau. Additional graphs were generated using Matplotlib and Seaborn. The linear and logistic regressions were analyzed using Python.

RESULTS: We identified 375 eligible publications, including research and review articles concerning AI and fairness in healthcare. When looking at the demographics of all authors, out of 1984, 794 were female, and 1190 were male. Out of 375 first authors, 155 (41.33%) were female, and 220 (58.67%) were male. For last authors 110 (31.16%) were female, and 243 (68.84%) were male. In regards to ethnicity, 234 (62.40%) of the first authors were white, 103 (27.47%) were Asian, 24 (6.40%) were black, and 14 (3.73%) were Hispanic. For the last authors, 234 (66.29%) were white, 96 (27.20%) were Asian, 12 (3.40%) were black, and 11 (3.11%) were Hispanic. Most authors were from the USA, Canada, and the United Kingdom. The trend continued for the first and last authors of the articles. When looking at the general distribution, 1631 (82.2%) were based in high-income countries, 209 (10.5 %) were based in upper-middle-income countries, 135 (6.8%) were based in lower-middle-income countries, and 9 (0.5 %) were based in low-income countries.

CONCLUSIONS: Analysis of the bibliographic data revealed that there is an overrepresentation of white authors and male authors, especially in the roles of first and last author. The more male authors a paper had the more likely they were to be cited. Additionally, analysis showed that papers whose authors are based in higher-income countries were more likely to be cited more often and published in higher impact journals. These findings highlight the lack of diversity among the authors in the AI fairness community whose work gains the largest readership, potentially compromising the very impartiality that the AI fairness community is working towards.

KEYWORDS: artificial intelligence, fairness, health equity, healthcare disparity

## INTRODUCTION

The fields of medicine and technology are undeniably intertwined; progress in one field often drives innovation in the other. It is no surprise that artificial intelligence (AI) is making headlines with its promise to inform or even automate clinical decision-making. However, the greatest impact this innovation has is not the language models trained on billions of parameters nor the generative models that create images from text prompts. Rather, complex bias exists within the data and it takes form in various ways, ranging from outcomes that are inconsistent across demographics, to subconsciously tainted tests and treatment decisions, and to influencing local clinical practice patterns in the form of institutional bias. [1]

Fairness is considered a critical element of trustworthy AI [2]. The fairness in health AI movement gained increasing momentum in the United States following an article by ProPublica, an independent nonprofit news organization focusing on accountability, justice, and safety, revealed that a software used by judicial courts across the US was discriminating against Black and Hispanic prisoners during parole hearings [3]. Definitions of fairness and metrics to evaluate fairness in medical algorithms subsequently mushroomed in the medical literature. Yet criteria of fairness also extend beyond the algorithms themselves to the people involved in their creation as flawed data alone does not account for the bias found within algorithms. It is also important to consider the bias woven into the very fabric of the algorithm itself that reflects the human assumptions of those who created it. This is a problem of representation and exclusion, and of epistemic narrowing that can lead to the perpetuation of structural inequities. Through her concept of “strong objectivity,” the Science and Technology Studies scholar Sandra Harding has shown that the exclusion of marginalized authors, including People of Color, scholars based in low-income countries, and women, among others, from research and publishing is not only unjust, but also diminishes the scientific knowledge produced. In order to attain a stronger version of scientific objectivity, and to create a science that can work towards equity and justice, Harding argues that we need to fortify that science by increasing the diversity of academic authorship as much as possible [4]. Which leads to the crux of the matter: how diverse is the fairness of the AI community proposing these definitions and metrics of fairness? It is important to bring attention to the diversity, or lack thereof, within this community to help prevent future propagation of bias and promote equity in health care.

## METHODS

### Search Strategy

In order to analyze the AI fairness community in healthcare, an in-depth descriptive study was done measuring aspects related to the publications in the field with a bibliometric review. The AI field was defined with terms that included machine learning, deep learning, convolution neural network, and natural language processing. Fairness overlapped in terms of health equity and health disparities.
The combination of searches from PubMed, Google Scholar, and Embase yielded results that spanned 31 years (1991-2022). The collected data were manually curated to secure the field of interest.

The search was conducted with the help of librarian Paul Bain, Ph.D., MLIS, from Harvard Medical School's Countway Library [Appendix].

### Eligibility of articles

Studies were considered eligible for inclusion if they met the following criteria: (i) Does the paper discuss machine learning fairness? (ii) Is the paper related to healthcare?, and (iii) does it discuss clinical applications?

If the three questions were affirmative, then the article was eligible. If a paper's eligibility was still uncertain, it was cross-checked by another author.

### Data Items

The bibliometric data obtained from Embase, PubMed, and Google Scholar provided the authors' first and last names, gender, race, article title, abstract, keywords, and URL. The articles were manually vetted to obtain enough bibliometric data and to ensure a thorough mapping and measurement of academic

trends. For each eligible study, the following data were extracted: type of article - opinion or research, which country the paper originated from, the journal it was published in, publication year, number of times each article was cited, whether or not funding was provided, and the name of the funding organization if provided. Additionally, the originating countries were classified based on the World bank classification for income to the following: low-income (<1.045 USD per year), lower-middle-income (1.046-4.095 USD per year), upper-middle income (4.096-12.695 USD per year), and high income (>12.695 USD per year) (cf. GDP per capita (current US$) | Data (worldbank.org)) [5].

**Approach to identifying each author's nationality, race, and sex (Joanthan to add few lines about method in cases of determining ethnicity and gender)**

To ensure consistency in our dataset and perform statistical analysis, we used pre-defined groups provided by search platforms to classify the gender and race/ethnicity of authors. Race and ethnicity was classified as White, Asian, Black, Hispanic, or 'none', by the search platforms, while Gender was recorded as male, female or none.

When collecting data on the author's gender, race, and ethnicity, the study relied on a variety of sources, including self-identification in terms of ethnicity and race, and the author's chosen pronouns. If this information was not available, information found on web pages and articles, and details related to the authors' affiliations or memberships in social or support groups, was used to determine gender, ethnicity and race. If this information was still unclear, the authors' gender, race, and ethnicity were determined based on photographs found on multiple websites including, university websites, private web pages, YouTube, and social media platforms such as LinkedIn. To maintain the accuracy and validity of the data, the study cross-checked every authors' gender, race, and ethnicity, against multiple sources of information, and in addition, each article and its inherent information were verified by another author of the paper to ensure the validity, consistency, and accuracy in the data collection process. When collecting data on the authors' countries of origin, the study went back as far as possible on the authors' past,

reviewing information available on LinkedIn or faculty and research web pages. If the authors did not disclose their home country, the study considered the country of their furthest educational background.

Lastly, the income level of the author's country of origin (cf. GDP per capita (current US$) | data (worldbank.org)) [3] their affiliated institution, and whether it is a minority-serving institution (cf. MSI List 2021.pdf (rutgers.edu), and their highest academic degree obtained (MD, Ph.D., etc.) [6] were investigated. To confirm the statistical certainty of the paper, the accuracy of the author's review of race, ethnicity, gender, country of origin, income level, etc. for each article's datapoint was verified multiple times by other authors involved in the study.

The approach used to identify race, ethnicity, and gender has its limitations. When analyzing the bibliometric data, the collected information on the author’s race, ethnicity, and gender was found to be unreliable and inconsistent, which reflects the inherent complexity of the topic at hand. Not all of the websites used as sources in this study allowed for authors to state their own identity. As a result, not all information was equally accessible via web searches. Authors who identified as multiracial, nonbinary, or other situations where the data was unclear, were not included, as the pre-defined groups provided by search platforms did not account for this. Moreover, the social constructs that can vary significantly depending on their socio-political context, gender, ethnicity and race, cannot and should not be directly determined from a picture. This means that in some cases, the information found may not completely reflect the author’s preferred identities, which highlights some of the methodological challenges and the complexities of this kind of intersectional demographic data gathering, and the difficulty of analyzing data regarding race, ethnicity, and gender on a large scale with quantitative methods

Nonetheless, this approach was chosen, because such determinations are difficult to make without engaging with a more in-depth survey of all authors in order to accurately record their preferred race, ethnicity, and gender.

Research on diversity requires a high level of reflexivity, including reflecting on one's own positionality in relation to matters of fairness in research. As such, we would like to situate ourselves in relation to this scholarship. Among the authors on this paper, 12 identify as male and 10 identify as female and 0 identify as non-binary. In terms of ethnicity, 10 authors identify as White, 8 identify as Asian, 3, identify as Black and 1 identifies as Hispanic. The authors come from the following countries of origin: USA (9), Turkey (2), Philippines (2), Canada(2), Denmark (2), Germany (1), Brazil(1), Lebanon(1), Peru (1). The idea for this research was inspired by conversations we have had with others in the field on matters of race, gender, and representation within AI and academia, and our own experiences of relative privileges working within this system.


**Figure 1.** Distribution of first and Last author sex - AI Fairness community in healthcare

**Statistical Analysis**

Regression analyses were performed to evaluate multiple factors that influence the number of citations and the presence of funding during the study. Only papers identified as research papers were included in this analysis as opinion pieces generally have little direct funding nor are widely cited beyond the community.

**RESULTS**

Bibliographic data was directly obtained from Embase, which yielded 242 articles. Data from the PubMed API yielded a total of 875 articles. Finally, another 497 articles from Google Scholar were found using the package PyPaperBot 1.2.2 [7] in Python 3.9.12. In total,1614 articles potentially related to AI fairness were collected.

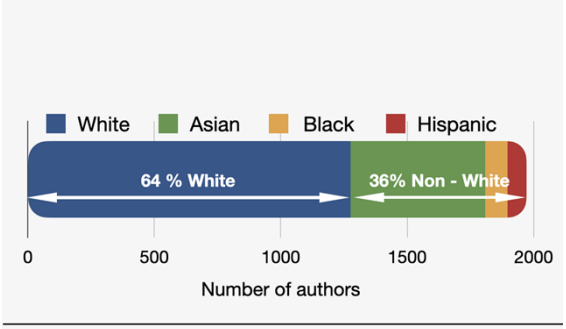


**Figure 2.** Distribution of First, Last and all Authors race - AI Fairness community in healthcare

Data was cleaned in Microsoft Excel. Spatial diagrams were generated using Public Tableau. Additional graphs were generated using Matplotlib 3.5.1 [8] and Seaborn 0.11.2 [9]. The linear and logistic regressions were analyzed using the Python package statsmodels 0.13.2, and the t-tests were analyzed using the Python package scipy 1.7.3.

As scope of the analysis was to map the gender and ethnic representation in the community of AI fairness within healthcare, scientific and non-scientific articles were screened for eligibility. The authors used manual vetting to narrow down the list of articles by reading the abstracts or full texts when the abstracts did not provide enough information. Out of the 1614 articles initially found through the search, 375 (23%) were determined to be eligible for further analysis, with a total number of authors of 1984.

**Distribution of each author's ethnicity and gender**

The results showed that, overall, 794 (40.0 %) of the authors were female, and 1190 (60%) were male (Table 1; Appendix). When looking specifically at the first and last authors, 155 (41.3 %) of the first authors were female, and 110 (32.2 %) of the last authors were female (Figure 1).

When the author's race distribution was analyzed, the study categorized the race of 1966 authors out of the total of 1984 authors in our curated database. The study found that the majority of the authors were

White (1270; 64.0%), followed by Asian (533; 26.9%), Black (89; 4.5%), and Hispanic (74; 3.7%) (Figure 2).

When dividing the authors into two groups, whites and non-whites, the study found that among the first authors, 234 (62.4%) were white and 141 (37.6%) were non-white. Among the last authors, 251 (66.9 %) were white, and 124 (33.1 %) were non-white (Table 1).

**Distribution of each nationality**

When looking at the country of origin, it was clear that most articles were from the USA. The total author nationality distribution showed that 986 (49.7%) were from the USA, 142 (7.2 %) were from Canada, 117 (5.9 %) were from the UK, 115 (5.8 %) were from China, and 83 (4.2 %) were from India (Figure 3). From income levels, 1631 authors (82.2%) were from high-income countries, 209 (10.5 %) were from upper-middle-income countries, 135 (6.8%) were from lower-middle-income countries, and 9 (0.5 %) were from low-income countries.

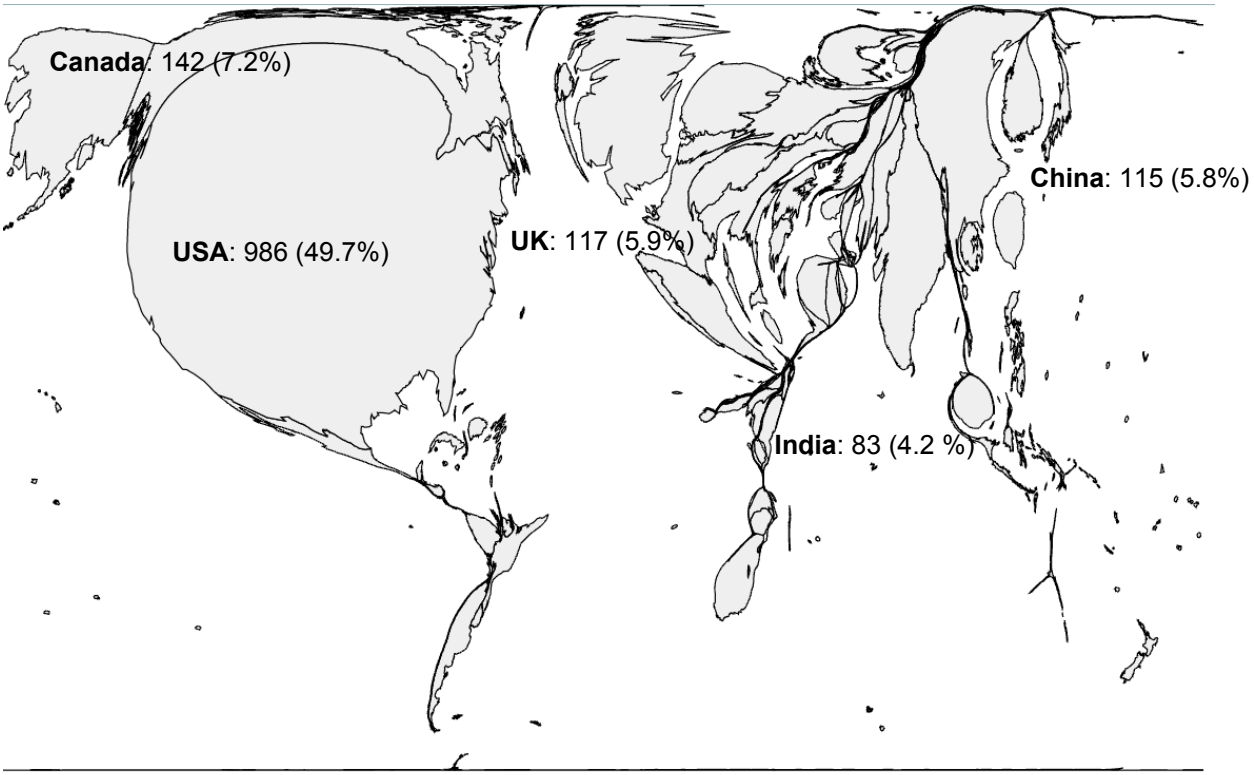


**Figure 3.** Distribution of each nationality– AI fairness community in Healthcare

When looking at the 375 first authors, 175 (46.7 %) were from the USA, 27 (7.2 %) were from Canada, 22 (5.9 %) were from the UK, and 21 (5.60 %) were from China (Figure 4, A). Among the last authors, 179 (50.7 %) were from the USA, 29 (8.2 %) were from Canada, 20 (5.7 %) were from the UK, and 14 (4.0%) were from China (Figure 4, A). For the first authors, 318 (84.8%) were from high-income countries, 32 (8.5 %) were from upper-middle-income countries, 24 (6.4 %) were from lower-middle-income countries, and 1 (0.3 %) were from low-income countries. For the last authors, 312 (88.4 %) were from high-income countries, 27 (6.8 %) were from upper-middle-income countries, 15 (4.2 %) were from lower-middle-income countries, and 2 (0.6 %) were from low-income countries (Figure 4, B).

Figure 4A, Global Dispersion of First Author Countries


Figure 4B, Global Dispersion of Last Author Countries


**Figure 4.** Global Dispersion of first (A) and last (B) author's nationality– AI fairness community in Healthcare

**Citations and Funding:**

By investigating citations across gender and ethnicity, it was observed that there was an overrepresentation of white-male authors. The data indicates, as illustrated in Figure 5, that papers with more male and white authors tended to receive more citations than those with fewer male and white authors.

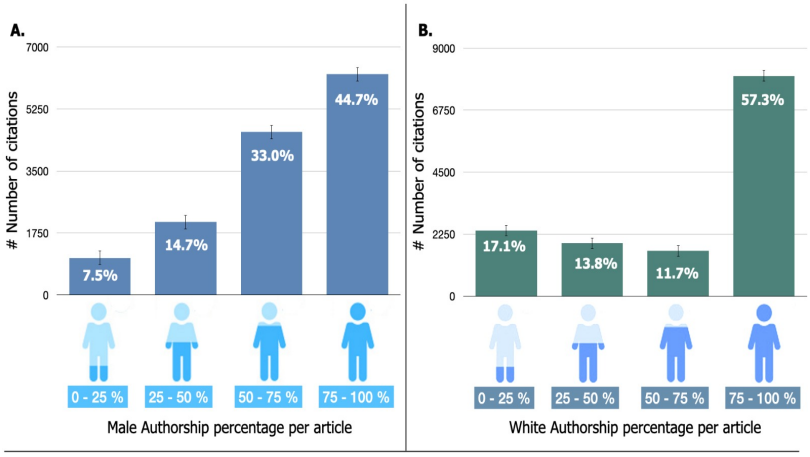


**Figure 5. A.** Distribution of Male authorship percentage per article with Number of citations and percentage - AI Fairness community in healthcare.
**B.** Distribution of Male authorship percentage per article with Number of citations and percentage - AI Fairness community in healthcare

On further investigation, it became evident that on average for both first and last authors, non-white and female authors receive fewer citations than white male authors, which is illustrated for last authors in Figure 6.

From here, the data revealed that articles with male last authors accounted for a substantial 76.4% of all citations, with white male last authors alone responsible for 58.3% of the total citations for all articles (Figure 6).

The findings from Figure 5 and Figure 6, prompts the argument that male authorship, particularly that of white males, may be associated with higher-impact articles published in high-impact journals.

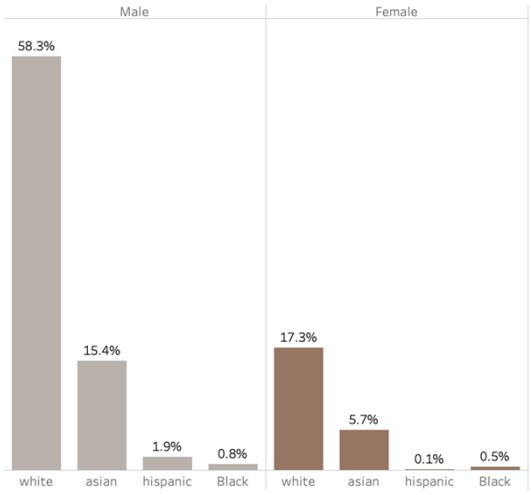


**Figure 6.** Distribution of citations among last authors based on gender and ethnicity - AI fairness community in Healthcare

The analyses in the study also suggest that higher-income countries may have a higher likelihood of being funded and producing higher-impact articles in terms of citations (Figure 7). This could be due to higher-income countries having greater access to resources and funding, which could contribute to the production of higher-impact articles. Additionally, the research

253 culture, infrastructure, and collaboration networks in higher-income countries may also play a role in
254 producing impactful research. However, it is also important to consider the potential biases that may be
255 present, such as language bias, publication bias, citation practices and funding. Recent scholarship on
256 citational practices and politics draws attention to the ways that structural inequities among authors are
257 reflected in citation practices, noting that scholars can take an active role in upending these hierarchies
258 through an intentional transformation in their own citational practices [10, 11]. These biases could impact
259 the analysis and interpretation of the results, and their access to funding, leading them to make less
260 impactful articles.
261
262 Articles were also evaluated according to how often they
263 get cited, in which year of publication was another
264 variable in classification, which showed that more recent
265 (closer to 2022) articles were most likely to be cited.
266
267 OLS Regression Results (Citations)
268
269 Regression analyses revealed that the percentage of
270 female authors, percentage of white authors, race of
271 authors, gender of authors, and the number of authors is
272 not correlated with the likelihood of being cited.
273 Publication year was the only factor affecting to be
274 cited.

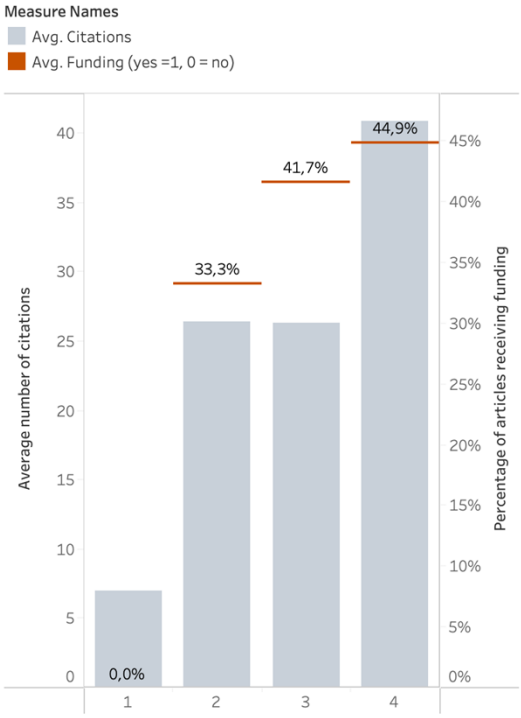


**Figure 7.** Distribution of citations among for last authors' income level - AI fairness community in Healthcare

275
276 OLS Regression Results (Funding)
277
278 Regression analyses revealed that the following parameters: percentage of female authors, percentage of
279 white authors, gender of first and last authors, and first and last authors, whether white or non-white, do
280 not affect funding. The number of authors and years of publications does not affect being funded.

Predictor factors of Citation and Funding

Regression analyses revealed similar trends for citation and funding. The percentage of female authors, the percentage of white authors, and the gender and race of the first and last authors did not have a statistically significant effect on whether a paper was cited or whether it was funded. Instead, publication year was the only factor with a statistically significant effect on the number of citations a paper received.

## DISCUSSION

The performed bibliometric study highlights the relative homogeneity of the authors of the AI fairness community, most notably seen in the distribution of gender, ethnicity, and countries of authorship. Male authorship, particularly that of white males, may be associated with higher-impact articles published in high-impact journals.

There is an overrepresentation of white, male authors, particularly in the first and last author role. 60% of the total number of authors were male, and the same proportion continued within the distribution of first authors and increased to 69% when looking at last authors as shown by Figure 1. Moreover, papers that had more male authorship, particularly that of white males, had a higher impact factor via the number of citations, as demonstrated by Figures 5 and 6. Articles with male last authors accounted for a substantial 76.4% of all citations. White male last authors alone were responsible for 58.3% of the total citations for all articles.

The increasing numbers of health disparities in underrepresented ethnic groups, and the underrepresentation of minority ethnic groups and women in academia has been well documented. Multiple studies and reports have addressed this trend and this study shows that it persists within the AI fairness community as well [12-17]. Women are less likely to be first or last authors and are less likely to be cited compared to male authors. This trend aligns with the findings of [12-14], which shows that there

is a significant disparity in gender and ethnicity in critical care medicine, particularly with regards to women being underrepresented in leadership roles, which can limit their opportunities for publication and recognition [16]. However, from the data, there was a slight indication that papers with female last authors were more likely to get a funding source, compared to the male counterpart, counterbalancing the notion of male predominance [18]. This finding highlights the positive impact that could be achieved, when equality measures are implemented by regulatory institutions in AI research within healthcare.

The analysis also shows that higher-income countries are more likely to produce higher-impact articles, most likely a reflection of the amount of funding received as demonstrated by Figure 7. Massuda et al show that underfunding a survey can lead to a significant reduction in quality, perpetuating the status quo [19]. This difference in funding perpetuates the current power dynamic where countries from underfunded institutions in low- and middle-income countries are less likely to produce high-quality research that is widely cited and well-regarded. Promoting research from underrepresented groups and communities is essential to promoting fairness and equity in research.

Possible actions to ensure proper representation include supporting research capacity development in lower-income and lower-middle-income countries and promoting research conducted by researchers of underrepresented gender identities and ethnic minorities. By supporting this development, the global research landscape becomes more inclusive. This in turn helps to advance and strengthen medical knowledge and promote social justice within the scientific community. In addition, promoting collaboration and cooperation between researchers from diverse backgrounds and locations can also lead to more innovative and impactful research [20].

Another way to promote diversity and inclusion in research is to establish guidelines for diversifying the composition of authors based on their ethnicity and sex. Providing formal training on equity issues and the importance of diversity in the research process can help educate researchers and promote greater awareness and understanding of these issues. This can be incorporated into the syllabi of academic institutions to ensure that the next generation of researchers is equipped with the knowledge and skills

necessary to promote diversity and inclusion in their work. Diversity should be highlighted in published work and working groups. Disclosing authors' nationalities, races, ethnicities, and sexes can promote diversity and inclusivity. Inclusivity also begins at the door. Institutions should develop initiatives that can help to attract more diverse scholars through transforming institutional cultures and priorities, as well as recruitment, hiring, and promotion policies.

In addition to promoting diversity in the composition of working groups and authorship of published work, it is also important to consider diversity in the content of the work, for example, including diverse perspectives and experiences in the research or addressing issues that affect diverse communities. AlShebli et al. found that ethnic diversity had the strongest correlation with scientific impact [20]. Recruiters should always strive to encourage and promote ethnic diversity, be it by recruiting candidates who complement the ethnic composition of existing members, or by recruiting candidates with proven track records in collaborating with people of diverse ethnic backgrounds.

Researchers should seek to understand their own group composition and how it should coincide with the communities which the research may impact. Representativeness and collaboration with communities can result in better science [21] and as such, greater understanding and awareness of these groups' challenges and issues can ultimately lead to more effective solutions. It is also worth noting that groups with higher cognitive diversity are often more effective at complex problem-solving and can help to reduce biased judgment in strategic decision-making [21-22].

Journals, editors, reviewers, and grantors can mandate that the author teams disclose their goals for achieving such diversity. Doing so would promote transparency and accountability and encourage authors to prioritize diversity and inclusion in their research. The National Institutes of Health (NIH) actively promotes diversity within the scientific community by encouraging conference grant applicants to include plans to enhance diversity in the selection of organizing committees, speakers, other invited participants and attendees [23]. These plans will be assessed during the scientific and technical merit review of the application and will be considered in the overall impact score. The underrepresented groups

include individuals from nationally underrepresented racial and ethnic groups, individuals with disabilities, individuals from disadvantaged backgrounds, and women. Encouraging authors to highlight their efforts to promote diversity in their groups can raise awareness of the importance of diversity and inclusion in the scientific field and promoting diversity and inclusion in all aspects of research can ensure that the work is more representative and relevant to a broader range of people, ultimately leading to more equitable and effective outcomes.

Limitations

Although several significant publications resulted from PubMed and Google Scholar searches, some were excluded. We used the third-party package PyPaperBot when selecting papers resulting from Google Scholar searchers which enabled us to extract 497 papers out of potentially hundreds of thousands. A large portion of the articles was removed from further analysis through manual vetting. However, third-party APIs are the only ways to parse through Google Scholar results. PyPaperBot was used, but a limitation of all the APIs seen is that they can only fetch the first 1,000 results, even if there are more. The extent of the literature is more vast than what we were able to extract, and it is crucial to scale up this analysis to capture more of the literature base in the future.

While manually vetting and extracting the authors' demographic information, information may differ from the authors' preferred identities, specifically for gender, race and countries of origin, as race, ethnicity, and gender are social constructs that can vary significantly depending on their socio-political context, and it may not accurately reflect the author's personal identity. Our analysis may have mischaracterized this vital information if their identities were not clearly stated on the internet. The use of predetermined categories for race and ethnicity made it particularly difficult to capture authors who may identify as multi-racial, or as belonging to several of these categories. It is also important to recognize that some people may not have the freedom or opportunity to publicly express how they identify. Similarly, assessments of whether an individual identified as non-binary were particularly challenging if not explicitly stated, and as such were not included in this study.

Moving forward, it will be important to develop better methodologies for representing a diverse range of possible identifications in order to better study questions of diversity, and a preferred methodology would involve interviewing each author in order to accurately record their nationality, self-identified race, and sex, as well as expanding the categories, however due to the scale of the study, it was not possible to obtain self-identified information in all cases. Systemic changes that allow for proper expression of identification are also necessary. Despite the presence of some inaccuracy within the data, as a necessity to perform statistical analysis, the overall trends revealed within this data are clear.

## Conclusion

As progress is made in both AI and healthcare, equity and inclusivity must be prioritized as it can lead to more innovative and impactful research, and a science that works for all. Thus the composition of the AI fairness research community is of the utmost importance  as whether AI will be a tool which only those who meet certain criteria can benefit from or a platform that serves all communities no matter their demographics, depends heavily on those who have a say in its design.

**REFERENCES**

[1] Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health. 2019 Dec;9(2):010318. doi: 10.7189/jogh.09.020318

[2] European Commission, Directorate-General for Communications Networks, Content and Technology, (2019). Ethics guidelines for trustworthy AI, Publications Office. https://data.europa.eu/doi/10.2759/346720

[3] Propublica. (2008) ProPublica - Journalism in the Public Interest. United States. [Web Archive] Retrieved from the Library of Congress, https://www.loc.gov/item/lcwaN0007149/.

[4] Harding, Sandra G. 2015. *Objectivity and Diversity: Another Logic of Scientific Research*. Chicago: The University of Chicago Press.

[5] The World Bank Group. GDP per capita (current US$) 2022.

[6] Rutgers Graduate School of Education. LIST OF MINORITY SERVING INSTITUTIONS. 2021.

[7] The Python Package Index. PyPaperBot 1.2.2 2022. https://pypi.org/project/PyPaperBot/ (accessed December 27, 2022).

[8] Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng 2007;9:90–5. https://doi.org/10.1109/MCSE.2007.55.

[9] Waskom M. seaborn: statistical data visualization. J Open Source Softw 2021;6:3021. https://doi.org/10.21105/joss.03021.

[10] Ahmed, Sara. 2012. *On Being Included: Racism and Diversity in Institutional Life*. Durham: Duke University Press.

[11] Liboiron, Max, and Rui Li. “Citational Politics in Tight Places.” Civic Laboratory for Environmental Action Research. https://civiclaboratory.nl/2022/03/02/citational-politics-in-tight-places/

[12] Vincent, JL., Juffermans, N.P., Burns, K.E.A. et al. Addressing gender imbalance in intensive care. Crit Care 25, 147 (2021). https://doi.org/10.1186/s13054-021-03569-7

[13] Goode, C. A., & Landefeld, T. (2018). The Lack of Diversity in Healthcare: Causes, Consequences, and Solutions. Journal of Best Practices in Health Professions Diversity, 11(2), 73–95. https://www.jstor.org/stable/26894210

[14] Salsberg E, Richwine C, Westergaard S, et al. Estimation and Comparison of Current and Future Racial/Ethnic Representation in the US Health Care Workforce. JAMA Netw Open. 2021;4(3):e213789. doi:10.1001/jamanetworkopen.2021.3789

[15] Chatterjee P, Werner RM. Gender Disparity in Citations in High-Impact Journal Articles. JAMA Netw Open 2021;4:e2114509. https://doi.org/10.1001/jamanetworkopen.2021.14509.

[16] Filardo G, da Graca B, Sass DM, Pollock BD, Smith EB, Martinez MA-M. Trends and comparison of female first authorship in high impact medical journals: observational study (1994-2014). BMJ 2016:i847. https://doi.org/10.1136/bmj.i847.

[17] Valantine HA, Collins FS. National Institutes of Health addresses the science of diversity. Proc Natl Acad Sci U S A 2015;112:12240–2. https://doi.org/10.1073/pnas.1515612112

[18] Sela N, Anderson BL, Moulton AM, Hoffman AL. Gender Differences in Authorship Among Transplant Physicians: Are We Bridging the Gap? Journal of Surgical Research 2021;259:271–5. https://doi.org/10.1016/j.jss.2020.09.037.

[19] Massuda A, Hone T, Leles FAG, de Castro MC, Atun R. The Brazilian health system at crossroads: progress, crisis and resilience. BMJ Glob Health 2018;3:e000829. https://doi.org/10.1136/bmjgh-2018-000829.

[20] AlShebli BK, Rahwan T, Woon WL. The preeminence of ethnic diversity in scientific collaboration. Nat Commun 2018;9:5163. https://doi.org/10.1038/s41467-018-07634-8.

[21] Aminpour P, Schwermer H, Gray S. Do social identity and cognitive diversity correlate in environmental stakeholders? A novel approach to measuring cognitive distance within and between groups. PLoS One 2021;16:e0244907. https://doi.org/10.1371/journal.pone.0244907.

[22] Meissner P, Wulf T. The effect of cognitive diversity on the illusion of control bias in strategic decisions: An experimental investigation. European Management Journal 2017;35:430–9. https://doi.org/10.1016/j.emj.2016.12.004.

[23] Office of The Director, National Institutes of Health (OD). (2021, January 22). *Not-OD-21-053: Updated guidelines for enhancing diversity and creating safe environments in conferences supported by NIH grants and cooperative agreements*. National Institutes of Health. Retrieved February 26, 2023, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-053.html

## APPENDIX

### Search String:

("Artificial Intelligence"[mesh] OR "Pattern Recognition, Automated"[mesh] OR "Data Mining"[Mesh] OR artificial intelligence[tiab] OR computational intelligence[tiab] OR machine intelligence[tiab] OR intelligent automation[tiab] OR intelligent system*[tiab] OR machine learning[tiab] OR deep learning[tiab] OR deep network*[tiab] OR supervised learning[tiab] OR natural language process*[tiab] OR neural net*[tiab] OR perceptron*[tiab] OR algorithmic decision making[tiab] OR predictive care tool*[tiab] OR predictive medicin[tiab] OR predictive model*[tiab] OR data mining[tiab]) AND ("Health Equity"[mesh] OR "Social Discrimination"[Mesh] OR "Healthcare Disparities"[mesh] OR fairness[tiab] OR egalitarian*[tiab] OR distributive justice[tiab] OR ((inequalit*[tiab] OR disparit*[tiab] OR inequit*[tiab] OR equity[tiab] OR equality[tiab] OR underrepresent*[tiab]) AND (health*[tiab] OR healthcare[tiab] OR racial[tiab] OR ethnic[tiab] OR sex[tiab] OR sexual[tiab] OR socioeconomic[tiab] OR economic[tiab]))) **mesh terms removed for the Google Scholar searches.

| | First author | Last Author | Other Authors | Total |
|---|---|---|---|---|
| **Gender** | | | | |
| Male | 220 (58.8%) | 243 (68.8%) | 727 | 1190 |
| Female | 155 (41.3%) | 110 (31.2%) | 529 | 794 |
| **Race** | | | | |
| White | 234 (62.4%) | 234 (66.3%) | 802 | 1270 |
| Non - White | 141 (37.6%) | 119 (33.7%) | 436 | 696 |
| *Hispanic* | *14 (3.7%)* | *11 (3.1%)* | 49 | 74 |
| *Asian* | *103 (27.5%)* | *96 (27.2%)* | 334 | 533 |
| *Black* | *24 (6.4%)* | *12 (3.4%)* | 53 | 89 |

**Table 1.** Distrubtion of authors across Gender and Race– AI fairness community in Healthcare

Fig x1. Citation OLS Results.

| Variable | Coefficient | Standard Error | t | P>\|t\| | [0.025 0.975] |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| % Female | 7.1467 | 16.706 | 0.428 | 0.669 | [-25.762 40.056] |
| % White | -16.9713 | 19.172 | -0.885 | 0.377 | [-54.738 20.795] |
| First author (white) | 4.1721 | 11.288 | 0.370 | 0.712 | [-18.065 26.409] |
| First author (female) | -7.9310 | 8.206 | -0.966 | 0.335 | [ -24.096 8.234] |
| Last author (white) | -0.8426 | 10.010 | -0.084 | 0.933 | [-20.560 18.875] |
| Last author (female) | -6.1935 | 8.299 | -0.746 | 0.456 | [-22.542 10.154] |
| # of authors | 1.7810 | 0.821 | 2.170 | 0.031 | [0.164 3.398] |
| Year | 0.0133 | 0.004 | 3.044 | 0.003 | [0.005 0.022] |

Fig x2. Citation Logit Regression Results

| Variable | Coefficient | Standard Error | z | P>\|z\| | [0.025 0.975] |
|---|---|---|---|---|---|
| % Female | 0.0804 | 0.656 | 0.122 | 0.903 | [-1.206 1.367] |
| % White | 0.2175 | 0.751 | 0.290 | 0.772 | [-1.254 1.689] |
| First author (white) | 0.1087 | 0.442 | 0.246 | 0.806 | [-0.758 0.975] |
| First author (female) | -0.2146 | 0.322 | -0.666 | 0.506 | [ -0.846 0.417] |
| Last author (white) | 0.2955 | 0.391 | 0.756 | 0.450 | [-0.471 1.062] |
| Last author (female) | 0.5383 | 0.327 | 1.647 | 0.099 | [-0.102 1.179] |
| # of authors | 0.0031 | 0.032 | 0.098 | 0.922 | [-0.060 0.066] |
| Year | -0.0003 | 0.000 | -1.538 | 0.124 | [-0.001 7.32e-05] |

| | First author | | | |
|---|---|---|---|---|
| | White | Non-White | Male | Female |
| # of papers | 146 | 102 | 141 | 107 |

| | | | | |
|---|---|---|---|---|
| % Funded | .521 | .441 | .503 | .467 |
| P-value | .220 | | .573 | |

| | Last author | | | |
|---|---|---|---|---|
| | White | Non-White | Male | Female |
| # of papers | 148 | 100 | 164 | 75 |
| % Funded | .534 | .420 | .470 | .572 |
| P-value | .0792 | | .137 | |

Updates Visuals:
https://public.tableau.com/app/profile/tiffany.chua/viz/AIintheHealthcareFairnessCommunity/SpatialDash