Diversity and Impact: A Scientometric Analysis of CHI Papers

HENNING POHL, Department of Computer Science, Aalborg University, Aalborg, Denmark

Diverse teams tend to be more productive and creative and the same is said to hold for science, with evidence that diverse groups of authors publish in more impactful journals. Within HCI, there have been several investigations into the diversity of such author teams, but it is unclear whether more diverse groups of co-authors also produce more impactful HCI papers, and which diversity dimensions matter the most. We investigate this relationship using a dataset of all CHI papers and estimate gender, ethnicity, location, sector, and experience measures of diversity. Through a regression analysis on citations and awards, we determine that co-author diversity indeed affects paper impact, but only some forms of it. Where diversity in experience, ethnicity, sector, and location result in more impactful CHI papers, this does not look to be the case for gender diversity. Through follow-up analysis we further investigate three potential ways how author diversity could be influencing paper impact, by examining research area, novelty, and positionality of these papers. This shows that a paper's research area also contributes to its impact, while varying by diversity dimensions, indicating a link between the two.

CCS Concepts: • **Social and professional topics** → *Computing profession*;

Additional Key Words and Phrases: CHI, diversity, scientometric analysis, human–computer interaction

ACM Reference format:
Henning Pohl. 2026. Diversity and Impact: A Scientometric Analysis of CHI Papers. *ACM Trans. Comput.-Hum. Interact.* 33, 1, Article 2 (February 2026), 51 pages.
https://doi.org/10.1145/3771921

1 Introduction

Diversity has been a prominent aspect in many discussions of computing and its organizations such as the ACM [99, 100]. It contributes to many challenges in the field, such as low participation of women [32] and minorities [53]. Within the area of HCI, three aspects have been the focus of conversations around diversity: (1) the audiences we design for/with and that we study [37, 60, 70, 101], (2) the kinds of research we engage in and structures of the field [21, 103], and (3) the people conducting HCI research itself [49, 63]. In this article, we focus on the latter and examine researcher diversity within authorship teams that published together at the CHI conference. We operationalize diversity to mean that such teams are diverse for a given trait (e.g., gender) when no one trait expression (e.g., male authors) holds an outsized share. Consequently, non-diverse teams will lean toward one or more of such potential biases, and we examine these as well.

For groups of paper authors, diversity varies along many dimensions, such as gender, nationality, ethnicity, and age. There have been many investigations into the composition of research teams along single or multiple of these dimensions. However, such studies do not address the question

Author's Contact Information: Henning Pohl (corresponding author), Department of Computer Science, Aalborg University, Aalborg, Denmark; e-mail: henning@cs.aau.dk.

ACM 1557-7325/2026/2-ART2
https://doi.org/10.1145/3771921

of how author diversity actually affects the work produced. For example, Cheong et al. [14] noted that "*there is a gap in extant research on the status of female academics across computer science communities which look at measures of 'impact' in the context of citations of ranked conferences and papers*" while also stating that "*underrepresentation of women (and other groups) can have a negative effect on the quality and breadth of the work done in the subfields of computer science*." Studies in other areas have generally shown evidence for the hypothesis of a diversity dividend. For example, studies have shown that gender diversity increases task performance [85], cognitive diversity increases creativity, but may come at the cost of team satisfaction [55], and some forms of cultural diversity make for more innovative teams [106]. A meta-review of papers on diversity in different sectors also showed improved outcomes (e.g., financial and quality) with more diversity [33]. Consequently, Nielsen et al. have opined that "gender diversity leads to better science" [69]. We follow a similar interpretation of *impact* in this article, where such "better science" and "quality of work" results in other people assigning higher value to a research paper. Specifically, we use citation and award metrics as proxies for impact, as is common practice in academic assessments [56, 77]. While this is not without issues and other metrics have been proposed (see, e.g., [56, 80, 87, 110]), this approach is amenable to large-scale analyses of HCI publications, given the data available for them.

But how strong is the evidence that diverse teams of researchers actually produce better work? Here we build upon work by Freeman and Huang, who investigated the "link between a team's ethnic mix and highly cited papers" [30]. They analyzed a set of 2.5 million papers, finding that less homogeneous co-author groups published in higher-impact venues and got cited more [31]. With respect to the overall diversity effects and effects within HCI in particular, though, this study has a few problems. First, HCI research was not well covered by their approach as they only included journal articles and HCI is predominantly a conference-focused discipline. Second, they only investigated ethnic diversity, where several other diversity dimensions are needed for a bigger picture, with gender diversity in particular missing, while it has received substantial interest in HCI. We hence conducted an analysis of HCI papers and their authors to investigate whether a wider range of diversity effects are also present within HCI. In other words, we ask: Are HCI papers from diverse groups of authors more impactful?

We focus our investigation on CHI papers and authors, as a representative sample for the wider field of HCI. While there are many more specialized conferences, these do not form a closed set and field boundaries are less clear. Furthermore, focusing on the largest conference has practical benefits, such as easier data collection and reduced data cleaning efforts. In line with previous work, we estimate author gender and ethnicity from their names and combine this with academic age, location (country), and sector (academia, institute, or industry) information for multifaceted author profiles. We use this data to compute a range of diversity indices for each CHI paper. As a proxy for impact, we consider citations and awards and analyze the impact of diversity and several covariates on these through separate **Generalized Linear Models (GLMs)**.

Our main contribution is the evaluation of how author diversity influences the impact of a CHI paper. We do so along a set of diversity measures more extensive than in previous work, which also allows us to determine which forms of diversity matter most. Our results show that overall diversity does influence paper impact, but only some variants and differently for citations and awards. Author ethnicity, experience, and location play substantial roles, but the gender composition of an author team had no influence on citations or awards. These results complement previous studies of who authors papers for CHI and provides evidence that diverse teams author more well-received CHI papers. We supplement this with three follow-up analyses that take a closer look at specific aspects of the papers and how they relate to the metrics used in the main investigation. Specifically, we look at prominent labs, larger region patterns, and the influence of just the first and last authors of a paper. As we find connections between diversity and the impact of the respective papers, this raises the question what the mechanisms behind this are. We posit this is primarily due to network effects, but there are several other potential mechanisms and we hence conduct three further analyses that examine other options. This starts with an investigation of larger research areas where we see that the general topic area of a paper is predictive of its impact. Through a novelty analysis of paper abstract we also find evidence that impact is likely not due to a paper being inherently more novel or different from previously published works. Finally, we analyze CHI papers with positionality statements to see whether a team's self-description, particularly with respect to diversity, relates to impact. Overall, we not only find that diversity does have an effect on the impact of CHI papers, but that this relationship is complex and other factors, such as topic area, factor into it. Our results provide insights to support the discourse around diversity within HCI and shine a light at potential biases and structural issues within the field. In particular, we find large regional differences in how work is received with a substantial bias toward American[1] research.

2 Related Work

Diversity in HCI has mainly been discussed through three lenses: the populations we target, our own community, and our research field in general. We cover these aspects as well as other questions of diversity in academia, previous work on the impact of diversity, and research around impact within HCI in general.

2.1 Diverse Participants

Himmelsbach et al. asked "*do we care about diversity in human computer interaction*" [37] and attempted to answer this question by analyzing 3 years of CHI papers. They investigated how paper authors write about their users and which dimensions of diversity are brought up. Over CHI 2006, 2011, and 2016, the number of diversity dimensions has slightly increased, hinting at an increased interest in these. The common bias in who our users are has been described as WEIRD (Western, Educated, Industrialized, Rich, and Democratic) by Sturm et al. [101] and Linxen et al. [60]. They have noted how the majority of findings at CHI are from studies with these participants and thus might not translate to a more diverse user group. Offenwanger et al. focused specifically on gender biases around study participants and found, for example, an underrepresentation of women [70]. Similarly, Chen et al. [13] focused on ethnicity of participants, something they note only few studies report in the first place. An intersectional approach was taken by Schlesinger et al. [91], who analyzed a selection of CHI papers to investigate how they treat different identities. In contrast to these works, we investigate the diversity of the authors themselves.

2.2 Diverse Scholars

The question of diversity within research itself has come up across disciplines. For the Affective Computing field, for example, Hupont et al. have described female underrepresentation as well as geographical inequities [41, 42]. Buchanan and McKay reported on author diversity within Information Science [9] and together with Zhang also did the same for HCI [63]. In both cases they identified that gender inequities are intertwined with conference location. Bartneck and Hu analyzed the proceedings of CHI 1983–2006 to describe what places authors are from (geographically and by sector) and how this relates to citation and award counts [6]. Among other things, they identified the outsized impact of Canadian research and the lack of an effect of awards on citation counts. Kaye has also looked at gender distributions of CHI authors and found that over time the percentage of male authors has been decreasing [49]. Similarly, Diniz et al. [5] did the same for the Brazilian HCI community, by analyzing authorship of IHC papers. In both cases, author gender/sex

[1]Throughout this article we use "American" to denote work and researchers from the United States of America.

was estimated from names and further information on the authors found online. We build on this work and apply similar methods to estimate characteristics of all CHI paper authors. Where these works have mainly aimed to describe the composition of authorship teams, we relate this to the impact of the resulting papers.

2.3 Diversity in Academia

A broader call for diversity in HCI was formulated by Dankwa and Draude [21], who point to systematic biases and outline "diversity-driven HCI" as a way to tackle inequalities. Diversity in HCI is often discussed along specific dimensions, most prominently perhaps gender. For example, feminist HCI has a long history in the field and has received continuous attention, as described by Chivukula and Gray [16]. Inversely, Seaborn [93] has posited that gender is often looked at through a lens of femininity, obscuring a bigger picture where masculinity should be more critically examined as well. Another common dimension is race, discussed, for example, by Ogbonnaya-Ogburu et al. [72]. Diversity dimensions also intersect and thus intersectional perspectives have been prominent as well, such as with Erete et al. [24] sharing their experiences as Black women in HCI research. Not all diversity dimensions relate to people, such as shown in Wang et al.'s work [108], where they analyzed citation behavior across venues and fields.

Alongside diversity in research, there is also the question of diversity in the operation of academia itself. Most prominent here again are questions of gender, such as in hiring practices. In general, this is an area where bias is likely overestimated, as shown in a large-scale meta-analysis by Schaerer et al. [90]. They only observed discrimination of male applications for female-typed jobs, but not the other way around. Focusing their review on American academia, Ceci et al. [11] investigated the effect of gender along six aspects (e.g., hiring and teaching evaluations). They found, for example, that female tenure-track candidates are more likely to receive offers and that there was no gender bias in grant funding. As they report, this is not congruent with findings in past decades and indicates a shift in sentiment over time. One example of this is from a study of Carlsson et al. [10] who investigated associate professor hiring in the Nordics, a stereotypically egalitarian and gender equal range of countries, and found that female candidates garnered a positive bias in competency and hireability ratings. For hiring on STEM tenure-tracks in the USA, Williams and Ceci [111] found as high as a 2:1 preference for female applicants. Yet, as Ceci et al. [11] also point out, this reduction in gender inequalities is not the case worldwide. For example, they note larger male advantages in grant success outside the USA. Furthermore, as Andersson et al. [3] found, even where there is no overall difference in assessments by gender, there can still be a bias against women at the higher end of the scale.

In general, diversity dimensions are complex and allow for many perspectives. Yet, diversity often gets viewed through a focused lens, with many of the works mentioned here focusing specifically on gender, for example. This diversity factors into the research we do and how we do it, but also into who gets to engage in that research, be it through hiring or funding. Furthermore, efforts to increase diversity in academia are often tied to potential benefits on the work environment, as well as the work itself, as we further discuss below. We adopt a broad view on diversity and focus on the effects of diversity on the papers that get written and how they are received. However, our investigation also provides insights that can in turn feed back into the discussions of diversity in academia. For example, identifying who is under-cited can help focus inclusion efforts and also help those impacted better argue their case.

2.4 Impact of Diversity

Freeman and Huang [30, 31] investigated how the ethnic diversity in author teams influences the impact of their work. They find that homophily (i.e., the tendency of people to associate with others who are similar to themselves) is a predictor for impact factor and citations, with both of them lower as homophily increases. As they put it: "*greater diversity and breadth of knowledge of a research team contributes to the quality of the scientific papers that the team produces*," Furthermore, Liao [59] found that collaboration intensity is a stronger predictor of research quality and impact than diversity, which in fact was not significant in their fitted model. However, this study was for scholars in collaboration networks, not individual papers, so the results are not directly comparable.

Fell and König [27] focused on the gender dimension of collaboration, including whether benefits of collaboration differ by gender. They found that male researchers are more homophilic in their collaboration than female ones. In terms of impact their results show a significant (albeit small) decrease in citations with increased homophily as well as a significant decrease (also small) of journal impact factor for the same.

Hinnant et al. [38] investigated physicists and how the composition of author teams influences productivity and impact, with the latter measured via citation counts They focus on seniority, eminence, and affiliation diversity and fit a model on 123 articles. Their results show a positive effect of first author seniority on citation counts and a negative effect for the number of authors on the same. Furthermore, they found that teams with multiple affiliations also tended to have more senior authors and less diversity of seniority.

Sulik et al. [102] argue that benefits of diversity could be mainly due to cognitive diversity. They note that this then translates to advantages in problem solving, but also problem posing, hypothesis generation, and exploration. However, while they outline potential mechanisms, they also note that, as yet, they "*know of no studies of the specific mechanisms whereby neurodiversity benefits collective problem solving*." Another example comes from Ding et al. [22], who looked at different pathways in which ethnic diversity could result in higher impact. They find that audience diversity was the primary mediator for this effect, such that papers have higher reach if co-authors hail from different populations. On the organizational level, Zhang and Tang [115] have posited that diversity also results in technological diversification with a subsequent increase in innovation. Overall, this suggests impacts of diversity can be due to complimentary effects in groups as well as in improved reach and alignment with recipients.

We build on these analyses, but extend them in several ways: (1) we focus on CHI papers and a combination of citation and award metrics, with the latter specific to the field, (2) we include a larger range of diversity dimensions in order to compare their impact, and (3) we combine the diversity measures with a corresponding range of bias measures.

2.5 Impact in HCI

Beyond the connection of impact and diversity, it is also worth describing how impact in general is described and operationalized in HCI. Horn et al. [40], for example, have investigated the "*impact of CSCW research*" by looking at co-authorship and citation data. However, they also note that a next step would be to use surveys to ask researchers to assess the "*intellectual impact of a paper*," Similarly, Correia [20] also performed a scientometric analysis of CSCW papers, but complemented citation measures of impact with some alternative metrics, such as social media mentions of papers. Visualization can help in this task, as demonstrated by *ImpactVis* [109]. However, the authors also acknowledge that, while popular, this citation-based view on impact has limitations. A similar approach was also followed by Henry et al. [36], when investigating and visualizing the impact of four HCI conferences. Citation analysis also is at the core of Meho and Rogers' [65] comparison of different data sources on HCI research. As they state, this kind of approach is popular because of "*the validity and reliability of the method in assessing, supporting, or questioning peer review-based judgments regarding the impact of a scientist's research output*," Their analysis also covers the h-index of individual researchers, which has been described by Alonso et al. [2], as measuring "*both the quantity and the impact of the researcher's publications*."

However, we can also find other approaches for impact of HCI research in the literature. For HCI toolkit research, for example, Ledo et al. [57] note that some toolkits find large success outside of research with many people adopting them for their projects. Some of this can be measured in number of downloads, but "*success can also lie in enabling new research agendas*." Another perspective can be social impact, as described by Balestrini et al. [4]. However, as they also remark, there is "*a lack of methods to evaluate long-term participation and measure social impact*," which can then "*weaken our chances of producing empowerment through research*." Conversely, Oulasvirta and Hornbæk [74] have commented on the view that "*an important goal of HCI research is not problem-solving but impact on society and industry*." Most recently, Kaltenhauser et al. [48] have asked "*which CHI papers make the most impact?*" While they also draw on citation metrics, they complement this by incorporating expert interviews. Asking seven HCI experts about their own "*impactful work*," their analysis touches upon several meanings of impact. Aside from citations, which still factor prominently into this, the experts also mentioned other aspects, such as inspiring others, advancing visions, and lasting influence on the field. However, they also note that this is a "*complex and sometimes elusive measure*," where "*impact is not something that is easily measured in a short time frame*."

Altogether, citation data constitutes the core of discourses around impact in HCI. While other forms of impact are brought up, these generally are accompanied by caveats such as requiring long time horizons, being very hard to measure, and lacking shared definitions. Interestingly, awards are not commonly framed as a form of impact themselves, but instead mostly investigated as a potential influence on citation counts. For example, Ren et al. [81] have investigated the career impact of awards, measured in terms of publications as well as the citations they receive. From a broader perspective, Penfield et al. [76] reviewed what research impact means and pointed to several further options. This includes narratives, case studies, surveys, testimonies, and metrics such as profit made. However, while this can be useful for individual impact reporting (e.g., for completed grants), development of such approaches for comparative analyses remains pending. As such, citation data still is of central importance for broader analysis of impact in the field. Furthermore, as described by Tijssen [104], citation impact also is "*a key indicator of research excellence*" and thus external impact. We hence also adopt a citation-focused methodology, even though such an approach has limitations, as we will later discuss.

3 Diversity and SIGCHI

Within the SIGCHI[2] community and organization, diversity has been a focus and point of discussion for many years. This is often in connection with equity and inclusion initiatives, but also within work and activism around social justice. Examples of such activities are the *SIGCHI Equity Talks*,[3] initiatives for more diversity at SIGCHI conferences,[4] or new executive committee roles focused on equity.[5]

Correspondingly, there are many statements from such material that exemplify the value and interpretations of diversity within SIGCHI. For example, a SIGCHI blog post on inclusion,[6] makes note of "*diverse points of view*," "*diverse locations*," and "*diversity in conference keynotes*," all toward a goal to "*be more inclusive*." In response to being asked for diversity within the CHI steering

[2]The Special Interest Group on Computer–Human Interaction (https://sigchi.org/) within the ACM.
[3]https://archive.sigchi.org/equity-talks-sigchi/.
[4]https://archive.sigchi.org/internationalisation-diversity-and-inclusion-events-at-sponsored-conferences/.
[5]https://archive.sigchi.org/towards-a-more-equitable-accessible-and-responsive-sigchi/.
[6]https://archive.sigchi.org/the-possibilities-of-inclusion-for-sigchi/.

committee, they pointed to "*multiple members who represent perspectives from various aspects of diversity (e,g,, geographical, racial/ethnic, gender, ability)*",[7] also highlighting specific dimensions of diversity of concern. The same blog posts also describe diversity issues like a lack of people of color in leadership ("`#CHISoWhite`"), representation of people with disabilities, and bridging the academia-industry divide, with all these prompting leadership to further such initiatives. In response to criticisms[8] around racism in SIGCHI's diversity and inclusion initiatives, the term `#SIGCHI4ALL`[9] has been brought up, with the SIGCHI executive committee specifically pointing to current structures "*disadvantag[ing] minority voices*." Race within SIGCHI also has been one of the topics of the equity talks,[10] where a link was made between higher "*racial and ethnic equity*" and "*greater diversity of methodological approaches*." The SIGCHI executive committee has also formulated goals for itself with respect to racial justice, pointing to actions to "*confront racism and other forms of discrimination and structural inequity*."[11] Finally, the SIGCHI execute committee has formulated a set of 10 values for itself,[12] including: "*Inclusive—–we welcome all views and actively seek out input*." The same document describes inclusivity as relating to "*supporting and promoting diversity in all its forms*," in particular pointing to geography and "*different forms of knowledge creation*."

Aside from discussing diversity and equity, SIGCHI as an organization has also made funding and hiring decisions with these aspects in mind. For example, there have been past efforts to support diversity and inclusion events at conferences,[13] noting goals such as to "*build community among underrepresented groups e,g,, women/minorities)*." Alignment with SIGCHI values ("*commitment to diversity, equity, and inclusion*") also was explicitly required in a 2024 call for the SIGCHI awards committee.[14]

With diversity in SIGCHI commonly connected to equity, it perhaps is not surprising that justice, and social justice in particular, heavily play into these discussions. As the SIGCHI executive committee has noted previously:

> *The task that lies before us is one of weaving equity, inclusion, and solidarity into each of our scholarly and professional activities, so that 'diversity and inclusion' are not add-ons, but deeply integrated into our every practice, whether it involves conference processes, volunteering responsibilities, or global community support mechanisms.*[11]

Diversity and inclusion here are goals to pursue in order to achieve equal participation and thus equity. Implicitly, if researchers are not represented, but also not noticed in extension, such an outcome would be inequitable and thus problematic. In turn, justice connotes achieving an equitable state, under the assumption that it has been systemic biases that brought about the current state of affairs.

An example of such a perspective, pertaining to the focus of this article, are endeavors around citational justice. Starting from the *Cite Black Women* movement, citational justice calls out the inequities in who gets cited and thus how credit is conferred in academia.[15] For the field of HCI, Kumar and Karusala [54] have echoed these concerns, followed by the larger *Citational Justice Collective* [18, 19]. As the latter makes clear, "*pursuing citational justice, then, entails moving*

[7] https://archive.sigchi.org/responses-to-the-chi2019-sigchi-townhall/.
[8] https://interactions.acm.org/blog/view/addressing-institutional-racism-within-initiatives-for-sigchis-diversity-an.
[9] https://archive.sigchi.org/sigchi4all/.
[10] https://archive.sigchi.org/sigchi-equity-talks-10-race-sigchi/.
[11] https://medium.com/sigchi/a-time-to-listen-reflect-act-and-represent-7595542e22cb.
[12] https://archive.sigchi.org/sigchi-ec-values-and-strategic-initiatives/.
[13] https://archive.sigchi.org/internationalisation-diversity-and-inclusion-events-at-sponsored-conferences/.
[14] https://web.archive.org/web/20240607214906/https://sigchi.submittable.com/submit/275728/2025-open-call-for-sigchi-awards-committee.
[15] https://www.nature.com/articles/d41586-022-00793-1.

away from individualistic views of authorship and toward a shared, reciprocal understanding of how knowledge is produced" [19]. Hence, this critique of citation practices is not just about ensuring researchers are not ignored, but to broaden what even is considered research (e.g., "*other dialogic practices like storytelling, and transformative feminist and postcolonial ideas*" [19]). Moreover, this shows that such initiatives are not about merely increasing citations counts of some group of scholars, but try to encourage others to more diligently search and engage with a wider set of academic works. Subsequently, the evolving discourse on citation justice has been described by Ogan et al. [71], who further highlight questions of what is considered valid research, but in particular focus on the broader tension between the Global North and South when it comes to research practices, environments, and biases. Towards a goal of "equitable and fair citation of work," the authors ask for broader engagement with research venues and outcomes from the Global South, but mainly to actively search for work that is underrepresented. Yet, as the article also points out, the scholarly archive infrastructure commonly makes it hard to specifically seek out such works, in turn cementing existing biases.

Overall, the position from the SIGCHI organization and activism inside HCI seems to lean toward diversity being a crucial aspect of equity and justice. Everyone should be able to participate (i.e., inclusivity), but expectations are primarily formulated for outcomes (commonly in terms of group ratios). While the former notion appears to have broad consensus in the field, it is unclear whether the same holds for the latter. We will next comment on this aspect and the position we take within this article.

3.1 Positionality

Diversity is not just about how many different identities are represented in a specific group of authors, but also is a core community concern around who is included and seen in the field. While scientometrics can tell us who is publishing and how their work is received in general impact terms, this does not answer the latter question of values around the same. Hence, we here position ourselves in this discussion and also describe the epistemological stance inherent to this work. We do this not with respect to our personal characteristics, but by following the argument of Savolainen et al. [89] that "*the proper locus of intervention is not the individual scholar but the community in which the scholar participates*." It is with respect to these community discourses that we elaborate our position.

An important aspect here is the conflation of *diversity*, *equity*, and *inclusion*, where equity in particular has seemingly emerged as one of the main goals within the SIGCHI organization. Correspondingly, concerns about diversity often seem to focus on injustices holding back different groups. Evidence for such systematic biases is then found in the outcomes, such as with low citations of papers from the Global South being indicative of a bias against work from this region specifically. We share in the support for diversity and inclusion within SIGCHI, but are critical of equity as a target and framing. For example, we believe that work should be evaluated without considering associated author identities, making decisions on whether to accept, cite, or confer awards only based on a paper's content. This is in line with a universalist perspective on science, which requires that "*the acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonist*" [66].

Many of the strategies for supporting equity around the impact of academic work require us to be conscious of who we cite and to more deliberately search out work by scholars who might be structurally disadvantaged. For example, per Kumar and Karusala [54], a potential approach to counter biases are databases that allow researchers to directly search for work from Black, Indigenous, or Queer authors to cite. Similarly, Ogan et al. [71] have called for researchers to "*search proactively for research conducted in the Global Souths*." Yet, even if one were to accept that one should actively cite more scholars who are Queer, Black, or from the Global South, the practicalities of this are daunting. Proposing databases so such researchers can be identified, for example, can easily conflict with data privacy policies, many of which stem from historical issues with such centralized collection of personal data.[16] In many ways this also shows the Anglo-centric nature of SIGCHI, with discourses around equity, for example, primarily featuring researchers from Anglosphere institutions (see, e.g., the special interest group on the SIGCHI Equity Talks [25]).

As mentioned above, another aspect of equity is the idea that some forms of academic work are structurally discriminated against. Consequently, examining a work's impact is marred by the same disadvantages and impact assessment thus comes with inherent issues. Yet, as we described in Section 2.5, citation-based metrics still are the most useful, practical, and common way impact is operationalized in our field. We acknowledge that this has limitations, though, and discuss this further in Section 8.2.

Finally, a contentious point with work like this article is that purely descriptive accounts of citation practices do little to move the field toward goals like equity. Instead, this kind of focused but reduced view on diversity and impact could be said to harm such efforts, given that it looks only at what is, not what could or should be. We disagree with this perspective and instead posit that an in-depth account of the state of the field is a necessary condition for discussing what should change and where problems are the most acute. While there is substantial discourse around diversity in HCI, the data around that very issue unfortunately is lacking. It is here, where we hope to contribute most to that very same discourse. In Section 8.1, we will return to this question and describe the implications following from this work.

4 Data Collection

For our scientometric analysis of CHI we combine data from several sources. This includes publication data from the ACM **Digital Library (DL)**, citation data from Google Scholar, but also automated and semi-manual annotation of the data. Overall, we have collected data on 10,341 CHI papers from 1982 to 2023 and corresponding data on 19,257 authors in 41,662 paper–author combinations. The main data collection was performed in the summer of 2023, and thus, for example, subsequent citations are not counted. In the following sections we describe the data collection process, in particular the challenges around it and choices we made alongside. As the data sometimes is inconsistent, can be incomplete, and sometimes even wrong, these choices matter and are a necessary step for the analysis we set out to run.

Aside from the paper data itself, all other data we collect is from an external perspective on the work and authors. Award and citation numbers encode the reception of the work by others. For gender and ethnicity, we use data-driven methods that estimate these characteristics based on author names. Thus, these are not descriptors of authors' biological sex or genomics, but measures of how they are presenting to a larger audience not personally familiar with them. We argue that such presentation and perception also is what likely matters most in scientometrics, as personal familiarity with authors of work we cite is unlikely in a large and rapidly growing field.

4.1 Paper Metadata from the ACM DL

There are several sources of CHI paper overviews, including DBLP,[17] Scopus,[18] Web of Science, and Semantic Scholar. Among these only DBLP provides a structured list of CHI papers, yet that data also had some issues. For example, we found CHI papers that were not included in DBLP

[16] See, e.g., https://time.com/5290043/nazi-history-eu-data-privacy-gdpr/.
[17] https://dblp.org/db/conf/chi/index.html.
[18] Searching for CHI papers with the following query: CONFSPONSORS(acm) AND CONFNAME(chi) AND (LIMIT-TO(DOCTYPE, "cp")).

(e.g., 10.1145/365024.383749) as well as papers with missing and wrong information. Scopus lists 18,431 conference papers for CHI, many more than the expected number. Furthermore, searching for CHI publications on Web of Science only yielded a small subset of papers with no unified index of CHI proceedings. Semantic Scholar also has no structured list of CHI papers nor provides sufficient and consistent metadata (e.g., author affiliations). We hence decided to use data directly from the ACM DL, as the most authoritative source of CHI metadata. Below we describe this data collection process in detail. We make the collected and cleaned metadata available[19] to aid future analyses of CHI papers.

4.1.1 Conference Listings. We started data collection by scraping the CHI conference pages from the ACM DL. The ACM DL maintains an overview of all CHI proceedings[20] as well as related extended abstract, workshop, and companion proceedings. Note that this overview also lists the proceedings of CHI 1981, which is not considered a CHI conference in other places (e.g., the SIGCHI page on the CHI conference history[21]). We follow this classification and exclude CHI 1981 from data collection. In general, we only retain the paper proceedings from this CHI conference overview for further consideration and disregard other linked proceedings.

We then scraped each CHI conference page on the ACM DL, extracting the included papers as well as listed information on the conference. This results in a collection of 10,579 papers, with most of them from recent years due to the strong growth of the CHI conference. Unfortunately, the data is not perfect, but also has a few issues, such as some proceedings listing things that are not CHI papers. This commonly includes panels (e.g., 10.1145/57167.57215), but also welcome addresses (e.g., 10.1145/1753326.2167156) and awards (e.g., 10.1145/1357054.2181017). While the ACM DL tags each item with a content type (e.g., article, tutorial, panel, or demonstration), we found this field to often be wrong. For example, all the above linked items are listed as "Article." We manually identified 238 non-paper items and removed them from the CHI paper listing, reducing the number of items to collect to 10,341. While this is not relevant to the analysis of this article, we also found several other inconsistencies in the CHI conference listings. For example, we found editor information that does not distinguish between roles (e.g., at CHI 2022) or lists assistants as chairs (e.g., at CHI 2020). Acceptance rate information on the ACM DL also sometimes differs from the ones reported by SIGCHI. For example, CHI 2015 is listed in the former[22] as 486 accepted out of 2,120 (23%), but by the latter as 379 accepted out of 1,520. The SIGCHI history page also reports a 23% acceptance rate, even though the numbers given there add up to a 25% acceptance rate. However, the ACM DL page also only lists 485 papers, so both numbers are off. While we do not make use of acceptance rate information in our analysis, this is representative of the issues we faced with this data.

4.1.2 Paper Pages. For each CHI paper, we then scraped its ACM DL page and extracted the metadata. This constitutes titles, abstracts, references, used keywords, and, most crucially for our analysis, author information. While most of this information is available for all papers, reference information is more spotty. CHI papers before 1998 often lack this information and for 58 papers no references were available at all. In other cases, the listed information is just wrong. For example, for 10.1145/3173574.3174174 only two references are listed on the ACM DL page, yet the actual paper lists 194 items in its bibliography. Furthermore, less than half (204,157 out of 463,118) of the

[19] https://github.com/henningpohl/CHI-Metadata.
[20] See https://dl.acm.org/conference/chi/proceedings.
[21] https://web.archive.org/web/20230606001707/.
[22] https://dl.acm.org/doi/proceedings/10.1145/2702123.

listed references on ACM DL pages of CHI papers contain a DOI, complicating automatic matching and analysis.

4.1.3 Author Identifiers. For our analysis, we make use of author identifiers to distinguish them and link them across papers. Fortunately, the ACM itself tries to maintain its own unique identifier for each author, which is also listed on each paper page in the ACM DL. For example, all papers by Randy Pausch published in the ACM DL are linked to his ACM identifier: 81100493478. This is now complemented with ORCID codes, but these are a fairly recent addition and not available on most papers. Unfortunately, these author identifiers do change sometimes as the ACM merges or updates them. Furthermore, as with paper data, we also found some author information to be missing or faulty. For example, 10.1145/67449.67494 only lists five authors on the ACM DL page, but has eight authors listed in the paper itself. We manually had to fix author identifiers for 12 papers (primarily from CHI 2023) and also added or removed authors from two papers. Furthermore, we backfilled ORCID codes for older papers, also manually fixing duplicate, erroneous, and missing codes.

We also scraped the profiles of CHI authors on the ACM DL. This worked for all but two authors, yielding 19,257 profiles. These profiles list several bibliographic measures, such as the average number of citations per article and the overall citation count.

4.1.4 Author Names. We ran into several issues with the author names listed on ACM DL pages. As mentioned earlier, there were some discrepancies between who is listed there and who is listed on the actual papers. For example, 10.1145/142750.142832 lists one author's academic title as a separate author: 81332505988. Another issue is that name spellings can also differ between ACM DL page and the paper. For example, 10.1145/67449.67472 lists two authors with initials only, where the paper lists their full names. Given that we base much of our later analyses on names, we manually checked all potentially incomplete names against the names listed in the respective papers, resulting in a set of 408 paper author name fixes. We complement this with further fixes to 21 author names where we draw on other information to determine their full name.

Most of the 19,257 unique authors (per their ACM/ORCID identifier) are only linked to one name, yet 1,041 authors have multiple names associated with their identifier. For example, Bill Buxton is listed in five different variations, with varying first names and middle name initials. We hence process the data to normalize names. For each author we pick the most used name, if it is used more than 70% of the time, or the longest name on record. We use a name parsing library[23] to extract first and last names from this name. In cases where the first name is initial-only, but a middle name is available, we default to that as first name instead. Where the name extraction process based on paper information failed, we refer to these authors' ACM DL profiles. These profiles include the authors' names in different places on the page, yet often in different variants. We use the names from the header and title and match them with the last name on file for that author. If a match, we replace initials with the best matching first or middle name from the profile data.

With all these fixes applied, we only had incomplete data for very few authors. Specifically, only 26 authors (0.1% of authors) had an initials-only first name. Only two authors had no first/last name at all: 99658699877 and 99658701783.

Names are a distinguishing characteristic for most CHI authors, with 11,700 last names (18,634 unique full names) across them, for example. However, name diversity is lower in some parts of the world, which is reflected in the data when we look at the most commonly shared names. The ten most common last names, for example, are Wang (237 instances), Lee (211), Kim (198), Chen (174), Li (165), Zhang (162), Liu (121), Yang (98), Wu (85), and Huang (73). However, looking at shared

[23]https://pypi.org/project/nameparser/.

names also uncovers some limitations of the ACM identifier system. Concretely, we found 553 names with more than one ACM identifier associated with them. Spot checking showed that many of these are indeed different authors, but others are not. For example, there are three identifiers in use for Bonnie E. John (81100482886, 99659260312, and 81501657272), which all refer to the same person. Given the low prevalence of potentially conflicting identifiers (about 2% of the unique names in the dataset), we did not manually check and fix all ACM identifiers. In any case, as our goal is to reason about papers instead of authors, duplication here is not a major concern.

4.1.5 Author Affiliations. Another important piece of information are the author affiliations listed on each paper. Unfortunately, this information also had some issues. For starters, this information is available in different places. Author information is found directly at the top of the paper page within a flyout attached to the author names. However, there also is a separate "author info & claims" panel that includes similar contributor information. Yet, this data is often not the same, and thus affiliations disagree on the same page. Furthermore, this information also sometimes differs from the affiliations given in the actual papers (see 10.1145/223904.223939, for an example of wrong data in this panel). In exploring the data, we found the information from the panel to be the least reliable, while extracting affiliation data from pdfs has its own challenges and reliability issues. We hence settled on using the affiliation information on the main paper page in combination with additional sanity checks.

Overall, we collected 41,662 paper–author entries, with 10,207 unique affiliation descriptors used across them. In the process, we applied fixes to 554 affiliations, manually correcting missing, wrong, and incomplete information from the pdfs or external information. After these corrections, we were left with only one (single author) CHI paper with missing data (10.1145/22627.22363).

A problem with the affiliation data is that the named entities used within them are not standardized. Entity linking thus is a hard problem, especially as entity information here is commonly incomplete and not following a common format. While there is a unique identifier system for research organizations,[24] this is not used yet for ACM publications and presently also does not have sufficient coverage of HCI-relevant institutions. While we tried several automated methods for processing the affiliation information, we found that none of these were reliable enough. For example, there is no consistent approach for mentioning multiple affiliations and this could not be detected reliably. We hence manually annotated all unique affiliations, adding country information and classifying affiliations into sectors. The sector of an affiliation can be either "academia," "institute," or "industry," as per Bartneck and Hu's previous work [6].

Adding country information is straightforward for most papers with sufficiently complete affiliations. However, there are also many shorthand (e.g., just "CMU") and incomplete affiliations (most commonly by American authors). Similarly, a classification into academic, institutional, and industry affiliations is complicated with the boundaries between the three not always clear. We classify higher educational institutions like "universities," "schools of," "colleges," "polytechnics" as academic. We tag national academies, national institutes, and independent research institutes as institutional, which covers places like Helmholtz, INRIA, and CWI. Everything else gets grouped under the "industry" label, which hence includes things like high schools and medical centers, but also all the industry research labs. Classifying "institutes" thus is particularly complicated as the name is used across all sectors. Finally, while we did our best to disambiguate affiliations (often only abbreviations are given) and consulted institution websites to aid our classification, some ambiguity in the process remains. However, the vast majority of affiliations posed no problem and the issues are most predominant in the long tail of the distribution.

[24] The **Research Organization Registry (ROR)**: https://ror.org/.

Overall, we have annotations for 41,662 affiliations, with only the one above-mentioned paper missing data. Most affiliations (19,766) are from American authors, followed by authors from the UK (5,050), Canada (2,826), Germany (2,782), and China (1,377). Furthermore, 442 authors are affiliated with institutions from more than 1 country, with 200 of these also partly American. The majority of affiliations are from academic institutions (32,936), followed by authors in industry (6,862) and institutes (691). Many also have multiple affiliations with academia and industry (690), academia and an institute (475), or industry and an institute (6). Finally, in one instance an author was affiliated with all three sectors.

4.1.6 Academic Age. We use the data from the ACM DL author profiles to determine their academic age. These profiles list several bibliographic measures, such as the average number of citations per article and the overall citation count. They also mention the author's publication years, the lower bound of which gives the first year that author published. While this only covers papers indexed by the ACM DL, this does extend beyond papers published with the ACM and we deem this an accurate indicator for authors active at CHI. However, for researchers coming to CHI from other fields, this measure might still underestimate their experience. We use this year of earliest publication to compute the relative academic age for each of the author's papers.

4.1.7 Awards. Every year a selection of outstanding CHI papers receives awards. The policy[25] is that the top 1% of submissions receives a "best paper" award and the remaining top 5% of submissions receive an "honorable mention" award. As there is no actual ranking of all papers, this decision gets made by awards committees that take scores and recommendations of reviewers and the program committee into account. In the current format, each subcommittee designates representatives who—after the program committee meeting—go over a list of candidate papers and determine the award recipients as they see fit.

As-is, the CHI awards do not necessarily go to the most cited papers. In their 2009 analysis of CHI papers, for example, Bartneck and Hu [6] found no link between awards and subsequent citation counts (we check this finding later in this article). With awards distributed by subcommittee split, the quality between subcommittees would also need to be equivalent for the awards to balance out. Yet, acceptance rates and standards vary substantially between subcommittees and thus a paper receiving an award in one might well have been middling at best in another. Furthermore, individual perceptions of paper quality vary wildly and it is unlikely the awards committees' choices are universal. That said, awards do capture a variable of interest in that they denote work that appeals to CHI committees and aligns with appreciated trends and approaches. We should note that awards are not assigned anonymously and author characteristics might well play a role in the selection process.

Initially, we only set out to collect paper award data from the paper pages on the ACM DL itself. However, we then noticed that awards are only listed in the DL for papers from CHI 2016 and later. The same holds for the proceedings pages, which also do not list awards for CHI conferences before 2016. We hence had to collect best paper information from individual conference websites. This worked for CHI 2010–2015, but while the program for CHI 2009 was archived,[26] the award information unfortunately was not. Thus, we only have CHI award data available for 14 years of CHI (2010–2023) and restrict our analysis around awards to those years.

4.2 Author Gender

We use an online service[27] to estimate gender information from the 7,664 unique first names in the dataset, based on the distribution of names between the genders. While this approach is

[25] Mentioned, for example, on https://chi2020.acm.org/for-attendees/chi-2020-best-papers-honourable-mentions/.
[26] See https://web.archive.org/web/20141111021704/http://www.chi2009.org/.
[27] https://genderize.io/.

commonly used for scientometrics (e.g., [27, 34, 88]) it does have some issues (see, e.g., the in-depth discussion on this by Cheong et al. [14]). First and foremost, this kind of gender classification does not sufficiently account for the non-binary nature of gender. Some names are gender-neutral and hence do not allow for classification either. Moreover, after our earlier efforts to fix names and manually disambiguate initials-only names, we were left with 28 authors (0.1% of authors) without a full name. We can still estimate gender for some initials-only names, as first letters of names are not uniformly distributed across male and female names.

For large-scale analysis of author gender, automated methods are a necessity and we thus follow previous work in applying them for our analysis. The adverse effects of automated name processing are also more pronounced on an individual basis, be it through misgendering or deadnaming. Across larger populations, the prevalence of non-binary gender identities is low [82] and thus is not likely to strongly influence the results of our analysis. In general, an author named *Sarah* is as unlikely to be male as an author named *Karl* is to be female. Note that the ACM does retroactively change author names after authors transition and request so from the ACM. We use the latest name on the ACM DL for all authors, even if they earlier might have published under a name that would have resulted in a different gender estimation.

The service we used classifies first names as one of “male,” “female,” and “unknown” and also provides a probability for this estimation. This identifies 12,130 male, 6,693 female, and 432 authors with unknown gender. Most first names for which gender estimation failed had low prevalence and were not shared between more than two authors. Examples of such names are “Shilad,” “Glaudiney,” “Aqueasha,” “Zhengneng,” and “Mjaye.” Asian names are predominant on this list, which is in line with common limitations in gender estimation systems [86]. Overall, however, this process yielded acceptable results, as validated on a small sample (see Section 4.4 for the description of this validation).

4.3 Author Ethnicity

We use the `ethnicolr` library[28] to estimate author ethnicity from their full name. This is similar to the process for gender, and also aligns with previous work. For example, this method has been used in epidemiology [17], scientometrics [84], and economics [50]. The common approach is to use census or other government registration data that includes names and ethnicity information to train a classifier for the latter. This is also the case with the library we used, with which we build on Florida voter registration data [98] for this classification. That dataset contains data for about 15 million individuals [15], split into 5 ethnicities: “non-hispanic White” (9,446,770 individuals), “Hispanic” (2,722,579), “non-hispanic Black” (2,086,582), “Asian” (329,034), and “Other” (424,308). Performance data for the used (LSTM-based) models has shown “excellent accuracy” [15] on par with other solutions [112].

Race/ethnicity data in the Florida voter registration data is self-identified. However, that also means the data underlying our classification is rooted in an American understanding of race and thus has some limitations. For example, distinguishing between “White” and “Hispanic” makes less sense in a global context. Similarly, it is unclear whether people from the Middle East are White, Black, Asian, or Other and self-identification within Florida might not align with how people in other places view this. However, US voter record and census data are representative of a diverse population and has the benefit of being readily available. In contrast, many countries in Europe are “race-muted” [47] meaning no such data is recorded, complicating any analysis of ethnicity.

Just like with gender, ethnicity is fluid and the boundaries can often be unclear, which makes classification inherently imprecise. As Simon and Piché [95] pointed out, ethnic categories are

[28] https://pypi.org/project/ethnicolr/.

socially constructed and different classification systems can put the same person in different boxes. In discussing his own ethnicity estimation method, Harris [35] makes the point that such methods should be complemented with contextual knowledge. However, we are not aware of a dataset that could better guide this kind of estimation for the target group of HCI researchers. As such, while an established method, the results of ethnicity estimation should be taken with a grain of salt. To alleviate the issue of rigid classification systems to some extent, we utilize the probabilities assigned to the ethnicities in our analyses where possible.

In terms of estimated ethnicity, we had 7,537 authors counted as White, 7,256 as Asian, 1,738 as Black, 1,112 as Hispanic, and 1,614 as Other. The model we used, on average, was more sure about some classifications than others. The average assigned probability for Asian authors was 82%, higher than for the other classes. In contrast, the model was, on average, less sure about authors of Hispanic (62%), White (55%), Black (49%), and Other (46%) ethnicities.

4.4 Validating Author Gender and Ethnicity

To validate the quality of gender and ethnicity estimation, we ran three validation studies: (1) an analysis of a dataset with some known author demographics, (2) a comparison with manual annotation of a subset of CHI authors, and (3) a follow-up to the second study.

First, we analyze data from the *CiteHER Bibliography*,[29] a repository, "*designed to collect and make available the published scholarly and creative work of Black women in Computing*," We use the 69 publications tagged as within HCI and clean up the data by completing missing names (i.e., replacing instances of "et al." with individual author names and initials with first names). As these papers are not tagged on the author level, it is only known that at least one of the authors is a Black woman. Correspondingly, we only check whether we can identify any female and Black authors among the authors. Our gender estimation identified female authors for 96% of the papers and Black authors for 70% of them. A more relaxed criterion of identifying non-white authors is true for 91% of the papers. We then combine the two estimators to specifically identify Black women, which we could for 68% of the papers. Again using a more relaxed criterion of identifying non-white women, we could do so for 86% of the papers. Looking at the failure cases, all gender estimation errors were due one gender-ambiguous name (Robin). Failure cases for ethnicity estimation were more varied, but mostly also due to ambiguity. In most cases where a different ethnicity slightly won out, authors were actually assigned a high probability of being Black as well. Consequently, if we treat ethnicity as a spectrum, rather than as distinct labels, the estimator performed well. Overall, we saw sufficiently high accuracy for this dataset. However, given the small scope and distinct focus of this and similar datasets, it is not clear whether this kind of performance generalizes to other author groups.

To investigate a broader range of authors we turn back to the CHI authors we already collected data on. We sample from this data to create a validation dataset which we then manually check as well as gather data on from external validators. Our sample is a random selection of 100 authors, for which we then collected public data on their gender and ethnicity presentation. In order to do so, we made use of resources like LinkedIn profiles, personal homepages, press articles, and research group pages. We determined gender from names and pictures, but also from the pronouns used by the authors themselves (e.g., as statements on preferred gender pronouns). Similarly, we combined text and visual information to estimate authors' ethnicity. Both, especially the latter, are subjective tasks with substantial ambiguity in many cases. As such they also capture authors' public presentation of gender and ethnic identities, not biological measures of them. However, as this public presentation is also what is perceived by others who reference authors' work, this presentation is actually what matters most for our study of diversity effects.

[29] https://blackcomputeher.org/citeher-bibliography/.

From the 100 authors, we identified 59 as male and 37 as female. For four authors, we could not find any information, including a first name beyond initials and hence marked them as unknown gender. With respect to the automatic gender estimates, 97% of female authors and 94% of male authors in our sample were correctly classified. Failure cases include one male author with an initial-only first name that was wrongly classified as female, and another male author with a unisex name.[30] Three mismatches were due to us not being able to find gender-identifying information, where the name-based approach could still estimate a gender.

We identified 59 White authors, where 73% of them were also estimated as such. The 33 authors we counted as Asian were labeled correctly 87% of the time, with one estimated as "Other" and two as White. We only identified two authors as Hispanic, two as Black, and one as "Other," with all being correctly estimated. For three authors we could not determine their ethnicity, while they were estimated as Asian. One aspect this also demonstrates then is that at CHI author ethnicity is more unbalanced than author gender. This is the case in our validation sample, but also in the larger estimated dataset (77% of the ethnicity estimates were either White or Asian).

For external validation we used the Prolific crowdsourcing service to gather another set of gender and ethnicity annotations for our sample. We assembled screenshots of the author pages (i.e., those we found in the previous step) as items to annotate. As we did not find websites for all authors in the sample, this dataset consisted of only 91 screenshots. Raters each saw 13 randomly chosen screenshots from this set, rating them one at a time (being paid GBP 0.90 for their effort). We asked raters to provide their personal estimate of the authors' gender (female, male, non-binary, or unknown) and ethnicity (our previously used categories plus unknown), based on the information shown in the screenshot. They could also answer that they were not able to tell and could leave additional comments with their annotations. We recruited 35 raters (age 18–47, $M = 25.5$, $\mathrm{SD} = 6.5$, 18 male, 17 female) for this task, with raters residing in 11 different countries (7× Poland and South Africa, 5× Mexico, 4× Portugal, 3× Italy and United Kingdom, 2× Greece, and 1× Chile, Estonia, Finland, and Germany).

Overall we gathered 455 annotations (91 authors each assessed by 5 raters). We changed four ethnicity annotations where raters put "Other" as ethnicity, but then further specified their assessment in a comment. Specifically, we changed three instances where raters put South Asian ethnicities to just "Asian" and another one to "White" where that was mentioned in the comment. We then computed gender and ethnicity agreement scores for every assessed author as the percentage of external assessments that agree with the automatic assessment. For gender this agreement is very high (female: 90%, male: 94%) while the results for ethnicity are more mixed (Asian: 65%, Black: 36%, Hispanic: 45%, White: 83%). Given that ethnicity perceptions varied across these raters, we also looked at the alignment of a majority vote (i.e., the prevalent ethnicity assessment across each set of the five raters) with the automatic assessment. This yields high agreement for White (98%) and Asian (73%) authors, but was lower for Hispanic (50%) and Black (40%) authors. Part of the issue here again is that there are only very few instances of Black and Hispanic authors in the sample (as well as the overall dataset). We also find that in the more ambiguous cases the external raters often chose an "Other" annotation (thus resulting in lower agreement scores).

We follow-up this validation with a second round of another 100 randomly sampled CHI authors. The proportions with respect to identified gender and ethnicity roughly aligned with the first sample. Where the first sample was 59% male and 37% female, the second was 69% male and 30% female. In terms of ethnicity, there were similar proportions of White (59 and 60%), and Asian (59% and 31%) authors. Both samples had only a few Hispanic (2% and 5%) and Black (2% and 1%) authors. Similarly, when comparing the manually identified data to the automatically determined gender

[30]For example, per https://www.thenamemeaning.com/jia/.

and ethnicity information, both samples again are in line. Where gender labels were aligned for 94% of the first sample, this was the case for 89% of the second. For ethnicity, the labels were congruent in 76% and 79% of the cases respectively. A chi-square test also shows no significant difference between the two samples: $\chi^2(1) = 0.10, \mathrm{p} = 0.75$. This second round hence overall provides some evidence that the gender/ethnicity proportions as well as the algorithmic performance for their estimation are stable across different samples of CHI authors.

So, are our estimations sufficiently accurate? For gender estimation, this looks to be the case, with the estimate highly in line with the validations. Accuracy also is higher than in previous work on HCI authors, such as by Kaye [49]. The picture is less clear for ethnicity, where there is a higher amount of ambiguity. To a large extent this is because the construct of ethnicity itself is inherently more ambiguous, which we also see reflected in the trouble the external raters had in such cases. For this reason we subsequently make use of the estimated ethnicity probabilities instead of the ethnicity labels, where possible. While this does not completely alleviate the issue, it is an approach more sensitive to the ambiguity in the data.

4.5 Relationships between Author Characteristics

While our subsequent analysis focuses on teams of authors, we here briefly also explore how author characteristics correlate for individual paper authors. Specifically, we look at academic age, gender, ethnicity, and affiliation data and the most relevant combinations of these dimensions. Because this data is per paper author, people with multiple CHI papers are also represented multiple times, albeit with potentially different academic ages and affiliations. We also only look at the top 10 most used country affiliations and group remaining countries together.

Figure 1 shows several relationships between author characteristics. For example, authors from France are much more likely to be affiliated with both academia and an institute. This is due to country-specific structures where members of INRIA and CNRS are commonly also affiliated with a university. In terms of ethnicity, we can see that authors from China, Japan, and Korea overwhelmingly are of Asian ethnicity, while many of the other countries are much more mixed. We can also see that authors from Japan are predominantly male, while American authors, for example, have a more equal gender representation. Finally, while mean academic age does not look to differ much by country or gender, we can see that the most experienced authors are from Canada or the USA and are male.

4.6 Citation Data

We collect citation data for all CHI papers from Google Scholar. While the ACM DL also provides citation numbers for papers, the Google Scholar data is preferable due to the broader scope of what is counted. Especially for cross-disciplinary work it is likely that many citations come from publications outside the ACM, for which the ACM DL citation data only has limited coverage.

We apply a tiered disambiguation approach to link papers to entries on Google scholar. Matches are determined by the similarity to the given paper title, as established through Python's `difflib.SequenceMatcher`. We only parse the first result page and retain the best matching entry if it is above a similarity threshold of 0.9. If no matching entry is found, a different search strategy is chosen. In descending order, we try to find matching Google Scholar entries via (1) *DOI+Year*, (2) just *DOI*, (3) *Title+Year*, and (4) just *Title*. This accounts for the fact that Google Scholar has some papers filed without DOI information or under the wrong year. The vast majority of papers could be matched with *DOI+Year* (10,185), with the other strategies accounting for 122, 18, and 5 papers, respectively. This left 11 papers where the automatic matching failed and for which we manually added citation counts (in one case from the ACM DL as the paper was not listed on Google Scholar).

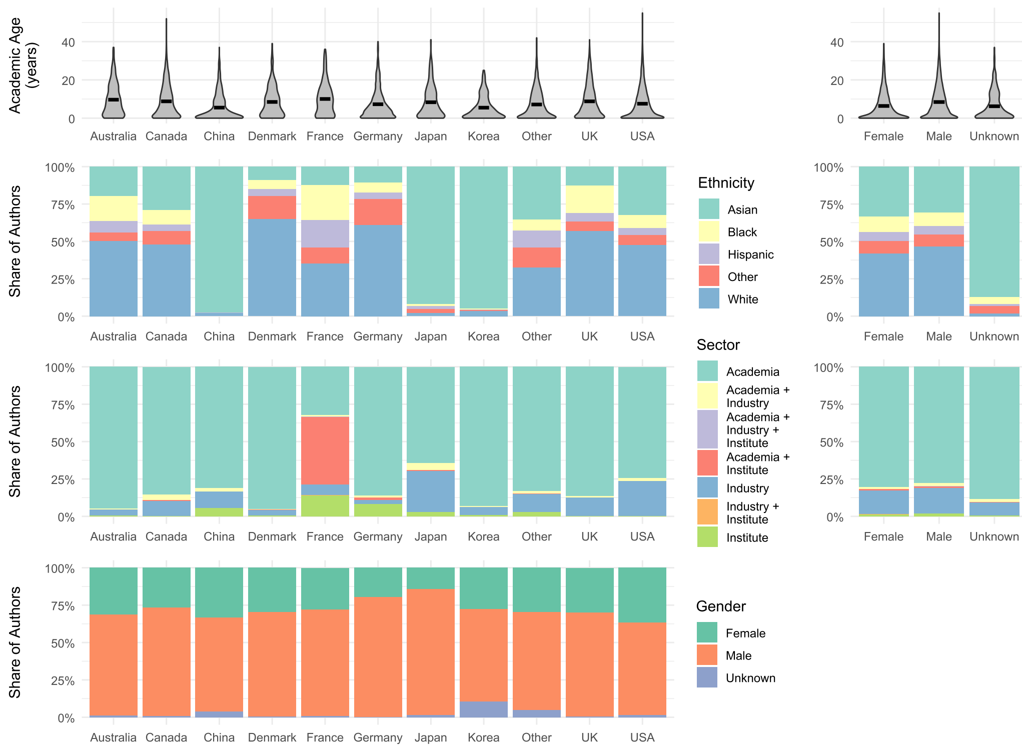

Fig. 1. For each paper author we explore how their characteristics relate to each other, focusing on how country (10 most common ones, with all other ones grouped together) and gender relate to sector, ethnicity, academic age, and each other. Cross-bars show mean academic age.

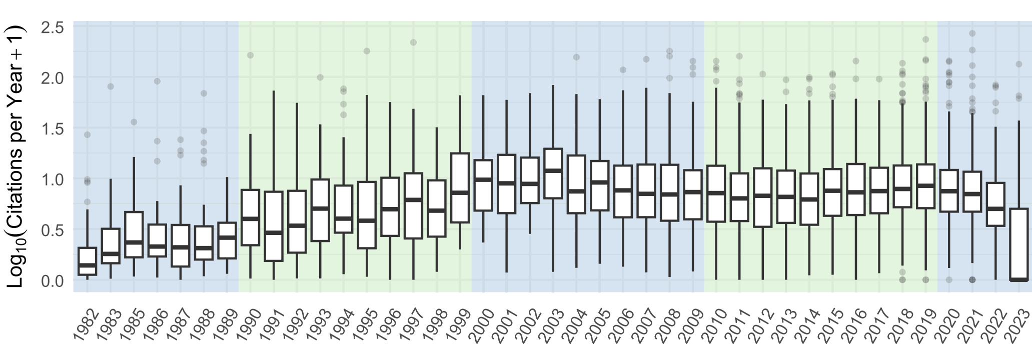

Fig. 2. Citations per year for papers from each CHI conference. Plotted on $\log_{10}$ scale because of heavily-cited outliers. Citation data collected in July 2023.

Number of citations across the dataset ranged from 0 to 5,725, with the average CHI paper receiving 86.5 citations (median: 35). As the amount of time to accrue these citations varies across the papers, we normalize the citation data to a *citations per year* metric. The average CHI paper is cited 8.7 times per year (median 5.2), and the measure overall ranges from 0 to 267 citations per year. Figure 2 shows how citation numbers have changed over time.

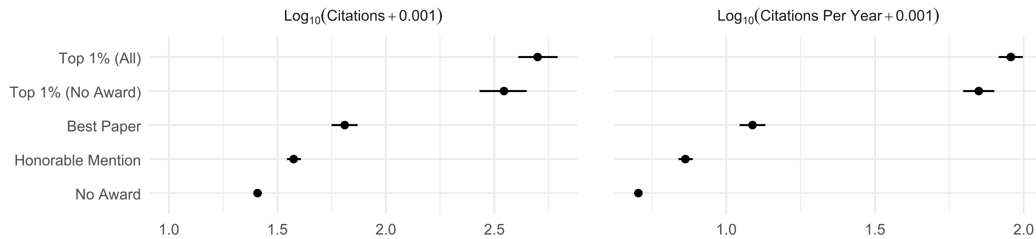

Fig. 3. Relationship between awards and citations for papers from CHI 2010–2022. For comparison, we also show the top 1% most cited papers (once including and once excluding those with awards) of each year. Error bars show 95% CI.

4.7 Do Awards Influence Citations?

With award and citation data collected, we can evaluate the relationship between the two. We do so to confirm they indeed measure different concepts and one is not just a proxy for the other. As we noted earlier, Bartneck and Hu [6] in their 2009 paper found no link between them. Our award data starts the year after that paper got published and it is possible that this relationship has changed over time. We exclude papers from CHI 2023, given that they have not had a chance to be cited much and thus analyze data from CHI 2010–2022.

We compare papers with no award, papers with an honorable mention award, and papers with a best paper award. For these comparisons we use both the number of citations as well as the citations per year. As this data is not normal, we log-transform both responses, adding a small bias (0.001) to account for papers with no citations. One-way ANOVAs show a significant influence of awards on the total number of citations ($F(2, 7{,}950) = 51.63, \text{p} < 0.001, \eta^2 = 0.01$) as well as the citations per year ($F(2, 7{,}950) = 63.74, \text{p} < 0.001, \eta^2 = 0.02$). *Post hoc* permutational t-tests show that for both response variables, all three levels are significantly different from each other (all p = 0.002).

As shown in Figure 3, citation impact increases as papers get an honorable mention award and even more so for best paper awards. Whether this is because these papers are better, or because they were promoted more and received more attention is unclear, though. Also, while there is a significant influence of award on citations, the effect size is small and other factors seem to matter much more. For comparison, we also looked at the actual top 1% of each year's papers (in terms of citations per year). We determine these top papers in two ways: (1) from all papers that year, and (2) from only papers that did not receive an award. As also shown in Figure 3, both groups of the top 1% most cited papers outperform the papers with awards. Hence, while papers with awards are cited more than the average paper, they are by no means the best papers each year in terms of this kind of impact. These results overall are still counter to Bartneck and Hu's, which shows that something in the relationship between awards and citations has changed over the last 20 years. However, the results also show that while there is a relationship between awards and citations, one is not equivalent to the other. Award winning papers receive more citations than average, but award decisions also seem driven by additional factors that are apparently not present with similarly highly-cited papers.

5 Modeling Diversity Impact

Having collected data on the impact of CHI papers as well as the teams who published them, we set out to investigate the connections between these. We chose a regression-based approach with GLMs for this purpose, fitting one model to predict a paper's citations and another to predict paper awards. In both models, we include a range of predictors based on diversity, as well as a few general ones. We first describe these predictors, with Figure 4 showing how they have changed

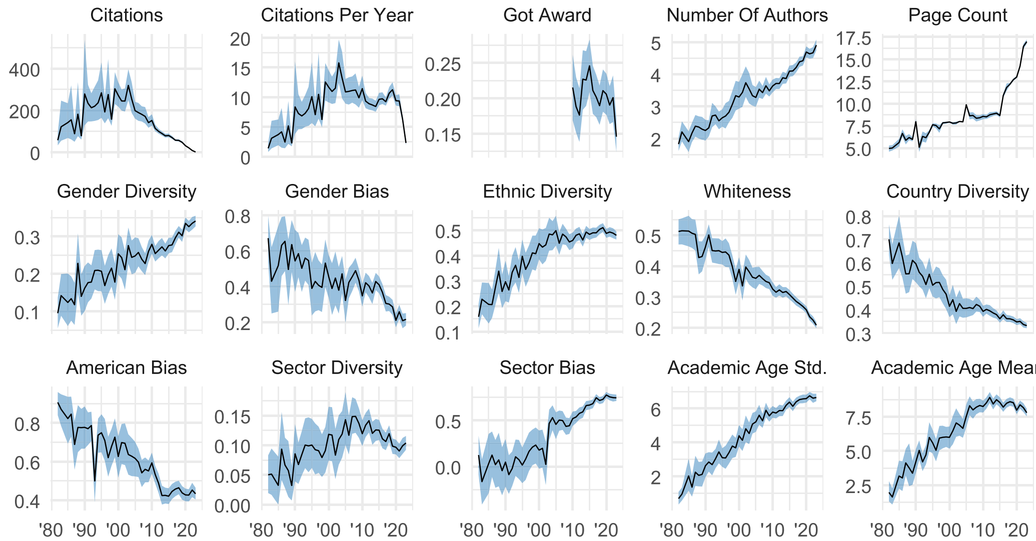

Fig. 4. Overview of the collected and annotated data over the history of CHI. First three panels show the dependent variables. See the description of diversity indices in Section 5.1 for how to interpret their individual ranges. Shaded areas shows 95% CI.

over time, and then describe the models fitted using them. Moreover, descriptive statistics for each predictor are listed in Tables A1 and A2 in the appendix. The data used in our models is available as supplemental material to this paper. We only provide per-paper measures and do not include the previously described personal data we collected for individual authors.

5.1 Diversity Predictors

What exactly diversity means varies and aside from subjective interpretations, there are also different measures of it. Schleuter et al., for example, described different indices of functional diversity [92]—a measure of "diversity of species traits in ecosystems." As they note, such indices can describe richness, evenness, divergence, or combinations thereof. Here richness (or rarefaction) denotes how many different kinds there were observed in a sample, evenness (or heterogeneity) denotes the relative distribution within a sample, and divergence denotes the variance within a sample. While the above definitions are rooted in ecology, they are general definitions also used elsewhere.

Diversity indices are also used for groups of people. For example, Jensen et al. [44] report on the use of a diversity index in assessing ethnic diversity of US census data. They used a definition based on "the probability that two people chosen at random will be from different race and ethnic groups." With respect to teams of authors, we also interpret diversity as that team being heterogeneous. This could mean that authors are of different nationality, political leaning, gender, or age. A previous example along those lines is work by McLaughlin et al. [64], who examined diversity in health professions education. They used Simpson's diversity index [96], which denotes the likelihood that two randomly selected individuals are from the same category (e.g., of the same gender).

We chose a combination of diversity indices, based on the data we collected. Some of the author traits we have are discrete, others are probabilistic, and the range of possible values is not always well defined. For example, we have probability estimates for author gender, overlapping groups for their affiliation type, and an open list of countries for the same. Hence, we made case-by-case decisions on which diversity index to use and describe these choices below. Furthermore, such indices only describe overall group heterogeneity, but not which way a group is leaning if it is biased. For example, an all male and an all female group of authors would have the same diversity index, but clearly vary otherwise. Hence, we complement the diversity indices with bias measures.

Several of our diversity measures are based on the proportions of traits in a group of authors and the probabilities assigned to certain traits. We define $\mathbb{A}$ as the set of a paper's authors, and use function notation in the form of traits (x) to denote the characteristics associated with one author. The proportion of authors with a given trait is then given as:

$$\mathrm{pr}_{\mathrm{trait}} = \frac{|\{x \in \mathbb{A} : \{\mathrm{trait}\} \subseteq \mathrm{traits}\,(x)\}|}{|\mathbb{A}|}. \tag{1}$$

Furthermore, we use $p_{\mathrm{trait}}\,(x)$ to describe the probability of a given trait for an author, where applicable.

We focus on five dimensions of diversity: gender, ethnicity, experience, location, and workplace. This choice is mostly due to data availability and what can be included in this kind of scientometric analysis. We acknowledge that this excludes other forms of diversity that might be relevant as well. For example, a simulation by Hong and Page [39] showed that groups combining diverse problem solving approaches outperform those with just the best-performing solvers. Hence, diversity of opinion, diversity of skills, or diversity of academic field might well be large contributors to the performance of authorship teams. Similarly, Jehn and Bezrukova [43] found group advantages for diverse levels of education and functional background. In an academic context, methodological backgrounds or epistemological perspectives might well also have an influence. Furthermore, more personal characteristics such as age, relationship status, care obligations, and parental status likely also have an influence on how groups of authors collaborate. However, such author attributes can generally not be ascertained from public data, including the authors' publications. This is a limitation of large-scale scientometrics and hence also of the subsequent analysis in this article.

5.1.1 Diversity of Gender. The gender data we collected gives a label for each author as well as a probability for that label. Given that this data is binary, we expect an equal number of male and female authors in heterogeneous teams. We follow McLaughlin et al.'s definition [64] and compute gender diversity as:

$$1 - \left(\mathrm{pr}^2_{\mathrm{female}} + \mathrm{pr}^2_{\mathrm{male}}\right). \tag{2}$$

Hence, this measure peaks at equal representation and then quadratically falls off as either gender is overrepresented. For this measure we exclude co-authors of unknown gender from the calculation. Where all authors are of unknown gender (8 papers), we assign an index value of 0.5, representing equal representation. Note that single author papers in this measure have a diversity value of zero.

Furthermore, we compute a gender bias measure that describes how likely a team of authors is *male* or *female*. This measure is positive for more male groups, negative for more female ones, and zero for authors of unknown gender. Given that we have probabilities for each gender estimate available, we compute this bias as the aggregate likelihood of a team's gender:

$$\sum_{x \in \mathbb{A}} \frac{p_{\mathrm{male}}\,(x) - p_{\mathrm{female}}\,(x)}{|\mathbb{A}|}. \tag{3}$$

Note that as probabilities are only given for the identified gender, it does not follow that an author marked as male with 70% probability also is rated as 30% likely to be female. We normalize by the number of authors so larger teams are not automatically marked as more biased than smaller ones.

5.1.2 Diversity of Ethnicity. Similar to the gender data, we also have ethnicity labels as well as probabilities. Here, we also make use of the probabilities for computing a measure of diversity, to alleviate some of the issues of ethnicity estimation. We define diversity of ethnicity as the average distance between all pairs of authors in the ethnicity space (i.e., authors are described by a vector of each ethnicity's probability). For single author papers, we set this measure to zero, but otherwise compute:

$$\sum_{x,y \in \mathbb{A} \times \mathbb{A}: x<y} \left\| \begin{pmatrix} p_{\text{asian}}(x) \\ p_{\text{black}}(x) \\ p_{\text{hispanic}}(x) \\ p_{\text{other}}(x) \\ p_{\text{white}}(x) \end{pmatrix} - \begin{pmatrix} p_{\text{asian}}(y) \\ p_{\text{black}}(y) \\ p_{\text{hispanic}}(y) \\ p_{\text{other}}(y) \\ p_{\text{white}}(y) \end{pmatrix} \right\| \cdot \binom{|\mathbb{A}|}{2}^{-1}. \tag{4}$$

Hence, as pairs of authors are further apart in estimated ethnicity, the measure increases, while similar authors result in small differences correspondingly. Note that we normalize by the number of author pairs so the size of an author group does not influence the measure.

With five different ethnicity labels, there are multiple possible bias measures. However, as many discussions around ethnicity commonly focus on whiteness [1], we chose to also focus on that aspect of the data. As Erete et al. [23] have also pointed out "*HCI methods tend to center on whiteness*" and thus it is possible a similar focus also manifests in the citation and award structures of the field. Furthermore, White authors form the largest group and could be detected reliably (see Section 4.4). Thus, an analysis of whiteness is also well-supported and stable in the data. We thus compute a whiteness measure of bias, that describes the overall likelihood that authors in a team are White. This utilizes the probabilities from the ethnicity estimation and is defined as:

$$\sum_{x \in \mathbb{A}} \frac{p_{\text{white}}(x)}{|\mathbb{A}|}. \tag{5}$$

Hence, this only considers one dimension of the ethnicity estimation and then again normalizes by the number of authors in a team.

5.1.3 Diversity of Experience. As academic age is a purely numeric measure, we can use descriptive statistics as diversity measures. We take the variance of academic age as a measure of team diversity (setting it to zero for single author papers) and also include the mean academic age as a covariate for overall team experience. On average, the latter is 7.8 years over all CHI papers, indicating that many of the authors at CHI are fairly experienced. In earlier years of CHI, the average academic age was lower, but since about 2006 this measure has stabilized at around eight. The same holds for variance of academic age.

5.1.4 Diversity of Location. From the country information of each author's affiliations we compute a measure of how international a team is. In contrast to the above diversity indices, the set of author countries is less clearly defined. CHI authors have been affiliated with 82 different countries in our data, yet this is but a subset of countries in the world. No team of authors of usual size would encompass all possible countries. As such, we adopt a different index that only takes into account how many countries are represented in a team:

$$\frac{\left| \bigcup_{x \in \mathbb{A}} \text{countries}(x) \right|}{\sum_{x \in \mathbb{A}} |\text{countries}(x)|}. \tag{6}$$

With the numerator the size of the set of affiliated countries and the denominator the total number of affiliated countries, this measure grows the less overlap there is between affiliations. We normalize by the amount of country affiliations within the team (e.g., an author affiliated with KAIST and

POSTECH would only have one country affiliation: South Korea). This ensures, for example, that a single author paper where that author has multiple affiliations is not counted as very diverse.

With respect to bias in author teams, we focus on the dominant affiliation in the data: institutions in the USA. We compute how *American* a paper is by the proportion of authors at least partly affiliated with an institution in the USA:

$$\mathrm{pr}_{\mathrm{american}} = \frac{|\{x \in \mathbb{A} : \{\mathrm{USA}\} \subseteq \mathrm{countries}\,(x)\}|}{|\mathbb{A}|}. \tag{7}$$

We again normalize by the number of authors so author group size does not affect the measure.

5.1.5 Diversity of Workplace. We distinguish between academia, institutes, and industry, where authors can be associated with several sectors. Hence, we slightly adapt the proportion calculation, and instead of dividing by the number of authors divide by the number of affiliations. With this adapted proportion, $\hat{\mathrm{pr}}$, the diversity index is calculated as:

$$1 - \left(\hat{\mathrm{pr}}^2_{\mathrm{academia}} + \hat{\mathrm{pr}}^2_{\mathrm{industry}} + \hat{\mathrm{pr}}^2_{\mathrm{institute}}\right). \tag{8}$$

This measure again peaks at equal proportions (i.e., 33.3% in each sector) and then falls off as different sectors are overrepresented.

We also compute a bias measure that describes whether an author team is more from academia or more from outside of it. For this we assign positive weights to academic affiliations and negative weights for affiliations from industry and institutes. We normalize by number of sectors per author and by number of authors overall:

$$\sum_{x \in \mathbb{A}} \frac{|\{\mathrm{academia}\} \cap \mathrm{traits}\,(x)| - |\{\mathrm{industry, institute}\} \cap \mathrm{traits}\,(x)|}{|\{\mathrm{academia, industry, institute}\} \cap \mathrm{traits}\,(x)|} \cdot |\mathbb{A}|^{-1}. \tag{9}$$

5.2 Other Predictors

We include four covariates into the model that are not linked to aspects of diversity: (1) when the paper was published, (2) how long the paper is, (3) the number of references in the paper, and (4) how many authors there are on the paper. Publication data might matter as the field has grown and changed over time, as have the people in it. As shown in Figure 2, impact in terms of citations has also changed through the years. However, these are broader trends and as such we decided to aggregate publication date to decades, instead of comparing individual years. Hence, we compare papers from the 1980s, 1990, 2000s, 2010s, and 2020s. While the 2020s are still young, they are well represented, with 3,020 papers. The 1980s (357 papers), 1990s (739), 2000s (1,291) all saw fewer papers, with only the 2010s taking the lead (4,933).

The number of pages in a CHI paper has been changing over time as well and could also reasonably affect the impact of a paper. In general, the average number of pages in a CHI paper has been increasing (see Figure 4), with a sharp jump in 2016 when references for the first time did not count towards the page limit (see [73]). The average paper in the 2020s so far is 15.2 pages long, while the average for papers from the 1980s is only 5.7 pages. Similar to Freeman and Huang [31], we also include a variable for the number of references in a paper. However, this data has some reliability issues, as not all papers had their references properly extracted (e.g., 58 papers had no references listed on their ACM DL page).

We include the number of authors as there is a known effect of this on citation counts [12]. The more people are on a paper the more personal networks can push the paper and the more self-citations such a paper can accrue. Furthermore, we include a dummy variable that encodes whether a paper has only a single author (true for n = 494). We do so to account for some oddities of single author papers with respect to our diversity measures. For example, we count these as high

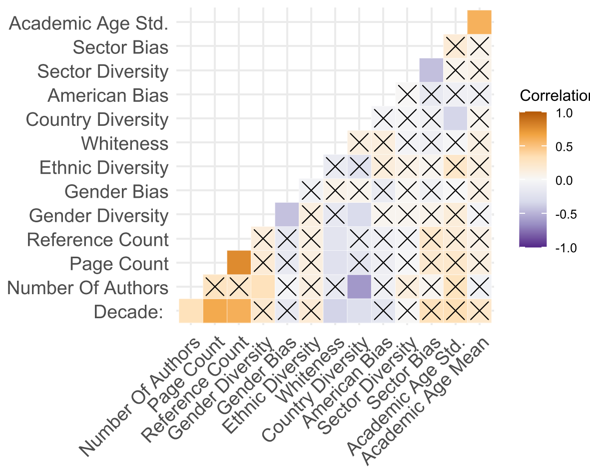

Fig. 5. Correlation between the available predictors. Non-significant correlations are crossed out.

diversity of country, given that such a paper cannot be penalized for several authors being from the same place. We retain these papers as they also serve as baseline data for low diversity.

5.3 Variable Selection

We analyzed the correlations between the different predictors to check whether there is significant overlap among them. As shown in Figure 5, most correlations were not significant. The highest correlation was between the page count and the number of references in a paper (Pearson's $r(10, 338) = 0.77, \mathrm{p} < 0.001$). Because of this overlap and the previously noted reliability issues with the reference data from the ACM DL, we decided to drop the reference count predictor from the model.

Many predictors have changed over time, as indicated by correlations with the decade predictor (and visible in Figure 4). In descending order of correlation this includes page count (papers getting longer over time), whiteness (author teams becoming less White over time), the number of authors (growing over time), and country diversity (decreasing over time). However, we find it crucial to retain the decade predictor as it also covers several other aspects that could have changed over the history of CHI. For example, topics have shifted, new research areas have opened up, technological development has occurred, and new institutions have sprung up.

The only two remaining large correlations are between the number of authors and country diversity as well as academic age diversity and bias. The former is negatively correlated where papers with more authors tend to have lower country diversity. This is mostly due to the metric, which expects the range of countries to increase as the number of authors on a paper increases. Yet, most commonly this is not the case and papers with more authors tend to have most of them from a smaller set of labs. For academic age, the average age and the diversity of age are inherently related. While they describe different aspects, a paper with an overall less experienced team will generally also have less diversity of experience. As the spread in team experience goes up, this inevitably also will increase the mean academic age for that team. Ultimately, we decided to retain both correlated pairs as the correlation was not due to them describing a very similar aspect, but due to side effects of their definitions.

All remaining correlations are at most of medium size. For example, country diversity is also negatively correlated with experience diversity and gender diversity. So as co-author teams get more international, they tend to be of more similar academic age and of similar gender. Furthermore, as teams get bigger they slightly tend to increase in diversity of gender. Just as with academic age, we also find that the diversity and bias measures in the areas of sector and gender correlate. Finally, there was a small, but significant, negative correlation between page count and whiteness. Overall, we follow the same reasoning above, also taking into account the smaller magnitude of these correlations, and retain the corresponding predictors.

5.4 Models

We fit two models to the data, one for citations and one for awards. For both models, we picked predictors based on the previous variable selection process. Some further limitations apply, as citation and award data availability varies across the dataset. We exclude interaction effects from the models, for reasons of feasibility. For example, a model for citations that includes all two-way interactions already did not converge. Hence, our models only include main effects and we report test statistics for individual predictors so the strength of their contribution to the outcome can be ascertained.

For the citation model we exclude all papers from 2023, as too little time has passed since publication for them to accrue citations. The substantially lower citation count of CHI 2023 papers is also visible in Figure 2. Excluding these papers still leaves us with 9,461 papers to fit a model to. Citation data is not normally distributed, with citations per year skewing to the right. Most papers are only cited a little, but some papers in the long tail garner many more citations every year. We thus decided to log-transform the citation data by fitting a log-linked Gaussian GLM. As there are some papers with zero citations overall, we add a small positive bias (+0.001, i.e., one citation every 1,000 years) to the citation data. This bias is an order of magnitude smaller than the smallest non-zero data point of citations per year (0.02).

For the award model we only include papers from CHI 2010 and later, as we have no award data available for earlier years. This still uses the data from 7,953 papers (77% of CHI papers) due to the large growth of CHI over time. Because the number of awards given is tied to the number of submissions (and thus also papers) we exclude year and decade predictors from this model. We do not distinguish between best paper awards and honorable mention awards and fit the model to predict whether a paper got an award at all. We use a binomial GLM (with the default logit link function) as a model.

Both models are chosen so the largest amount of CHI papers could be included in the analysis. However, this does leave out some aspects, such as the keywords and abstracts available for some of the papers. Following this main analysis, we hence zoom in on specific aspects in follow-up analyses (see Sections 6.3 and 7). While these then use only a smaller subset of the data, they offer additional perspectives on what aspects of CHI papers further influence their impact. We bring together these separate analyses in the discussion.

6 Results

Here, we describe the found effects in the citation and award models. We then also present results from follow-up investigations that further shed light on some of the main effects.

6.1 Citation Rate

As shown in Table 1, we found significant influences for several predictors. For starters, citations vary for papers in each decade of CHI, with all but papers from the 2000s being cited less frequently than the reference decade of the 2010s. Citations per year also increased significantly for papers with more co-authors and more pages. As Figure 4 shows, these factors are linked to some extent, with both the number of authors and the page count generally increasing over time, albeit non-linearly.

Table 1. Fitted GLMs for Citations per Year (Left) and Awards (Right).

	Citation Model						Awards Model					
Factor	Coef.	Std. Error	Odds Ratio	t-Statistic	p-Value		Coef.	Std. Error	Odds Ratio	t-Statistic	p-Value	
Decade: 1980s	−1.16	0.23	0.31	−5.00	<0.001	***						
Decade: 1990s	−0.17	0.07	0.84	−2.65	0.008	**						
Decade: 2000s	0.14	0.04	1.15	3.51	<0.001	***						
Decade: 2020s	−0.23	0.04	0.80	−6.22	<0.001	***						
Number Of Authors	0.06	0.01	1.07	9.88	<0.001	***	0.03	0.02	1.03	1.90	0.057	
Single Author: True	−0.17	0.10	0.85	−1.67	0.094		0.40	0.23	1.49	1.69	0.092	
Page Count	0.04	0.00	1.05	23.84	<0.001	***	0.02	0.01	1.02	2.85	0.004	**
Gender Diversity	0.02	0.07	1.02	0.24	0.81		0.03	0.15	1.03	0.21	0.83	
Gender Bias	−0.02	0.03	0.98	−0.79	0.43		0.05	0.06	1.05	0.80	0.42	
Ethnic Diversity	0.14	0.06	1.15	2.17	0.030	*	0.33	0.12	1.40	2.67	0.008	**
Whiteness	0.55	0.09	1.73	6.02	<0.001	***	1.12	0.18	3.07	6.30	<0.001	***
Country Diversity	0.16	0.10	1.17	1.59	0.11		−0.04	0.21	0.96	−0.21	0.83	
American Bias	0.34	0.03	1.40	10.07	<0.001	***	0.16	0.07	1.17	2.39	0.017	*
Sector Diversity	−0.23	0.08	0.80	−2.94	0.003	**	0.19	0.19	1.21	0.99	0.32	
Sector Bias	−0.15	0.02	0.86	−7.42	<0.001	***	−0.06	0.06	0.94	−0.96	0.34	
Academic Age Std.	−0.04	0.01	0.97	−6.44	<0.001	***	−0.03	0.01	0.97	−2.44	0.015	*
Academic Age Mean	0.02	0.00	1.03	6.76	<0.001	***	0.03	0.01	1.03	2.98	0.003	**

The citation GLM is Gaussian with a log link function, while the awards GLM is binomal with a logit link function. The awards model only includes data from CHI 2010 and later, but does not include decade as a predictor. We also report odds ratios for easier interpretation of the models.

*p < 0.05; **p < 0.01; ***p < 0.001.

With respect to the diversity indicators, we found significant influences for diversity of ethnicity, sector, and experience. The more ethnically diverse a team of authors, the more heavily cited their paper. However, papers that include authors from multiple sectors fare worse than papers from just one sector. Finally, diversity of experience was a significant predictor, though with a small effect. Papers with a larger spread of author experience fared slightly worse than those with co-authors of similar academic age.

Finally, we also found significant effects for the influence of several diversity bias measures. With respect to whiteness, papers with "whiter" teams got cited more heavily. Similarly, the more a team is based in the USA also significantly increased citation counts. Moreover, when a paper has a higher share of authors who work in academia, the citation counts significantly decrease. We investigated whether this might be because papers from industry and institutes are from smaller teams, but only found a very small correlation between the two measures (Pearson's $r(9,459) = 0.05, p < 0.001$). Finally, we found a small but significant effect of average academic age where papers of more experienced teams get cited more.

6.2 Awards

As also shown in Table 1, we found significant influences for several predictors. But we can first also note that this model differs from the one for citations, which is also visible when comparing Figures 6 and 7. In general, fewer predictors have an influence on awards. For example, from the non-diversity predictors, only the number of pages had a significant influence on the likelihood for an award, with longer papers having a slightly higher chance. One might assume this is because the award model has less data available for fitting, as it only uses data for CHI 2010 and later. However, this difference is small in terms of actual observations with the award model including 84% of the papers from the citation model.

From the diversity indicators, the award model showed significant influences for diversity of ethnicity and experience. With the former, the more ethnically diverse a team, the higher the likelihood for an award. Yet, as teams are more diverse in experience, the likelihood for an award goes slightly down.

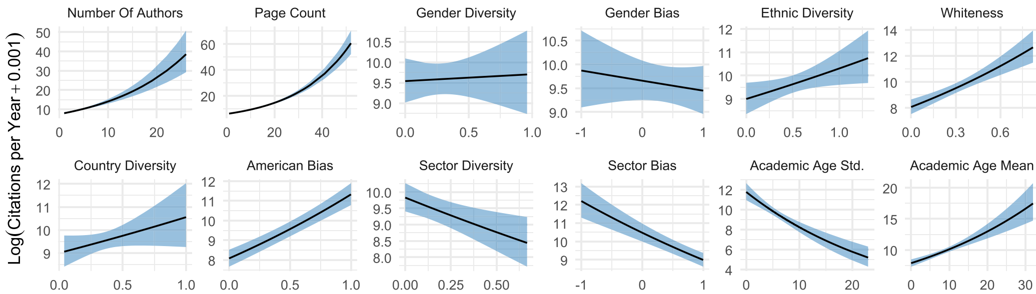

Fig. 6. Marginal effects for the relationship of all numeric predictors to log-transformed amount of citations a paper gets per year. Shaded area shows 95% CI.

Fig. 7. Marginal effects for the relationship of all numeric predictors to the probability of a paper receiving an award. Shaded area shows 95% CI.

The effects for the diversity indices are complemented by effects for several diversity biases. First and foremost, there was a strong effect of whiteness, with "whiter" teams having a much larger likelihood for an award. This likelihood also goes up as teams are more American or more experienced on average. Here, the effect of whiteness is particularly interesting as we also found that more ethnically diverse teams also get more awards. The correlation between these two variables is low (Pearson's $r(7,951) = 0.05, p < 0.001$), so they do capture different aspects of a team's diversity.

6.3 Follow-up Analyses

Some of the effects we have seen so far are counter to what one would expect. For example, if diversity in ethnicity and country increases citation counts, why is the same not true for sectors? Hence, we focus on specific aspects of the collected data to further investigate this and other questions.

6.3.1 Prominent Collaborators. While there was no overall effect for sector diversity, we suspect that some cross-sector collaborations still might stand out. To investigate this, we zoom in closer on the sector data and look at the diversity of collaboration with a set of primary actors. Based on their top rank in the overall share of CHI papers, we separately annotate papers with information on whether they involve authors from Microsoft, Autodesk, Google, Parc, Adobe, Max Planck, INRIA, CNRS, and DFKI. We then fit another citation model to the data with each affiliation as well as their interactions with sector diversity as predictors.

For starters, we note that when not accounting for other influences, increases in sector diversity indeed correspond to higher citation rates. However, just accounting for the influence of the above nine affiliations results in no significant influence of sector diversity (Odds Ratio = 1.08, $p = 0.44$).

For example, the model shows significantly higher citation rates for papers from Microsoft (Odds Ratio = 1.74, $p < 0.0001$), Google (Odds Ratio = 1.96, $p < 0.0001$), and Parc (Odds Ratio = 1.60, $p < 0.0001$). However, a significant interaction effect of sector diversity only showed for Google (Odds Ratio = 0.24, $p = 0.0005$). Papers with co-authors from Google saw a large drop in citation rate for increasing sector diversity. Thus one reading of these results is that generally sector diversity does tend to increase citation rate, however that effect is overshadowed by other, stronger, predictors in the original model and with some of the bigger industry labs we see a counter trend where sector diversity decreases impact when they are co-authors.

6.3.2 Regionality. Given the large effect of American bias, we decided to also take a closer look at other potential effects of region on paper impact. As there are 80 countries in the dataset, we decided to cluster them into broader regions, for which we use UN M49 sub-region names (e.g., "Western Europe" or "Sub-Saharan Africa"). This leaves us with 14 regions, though we subsequently excluded Central Asia from analysis as just one paper had a co-author located there. We then fit another citation model to the data with region and country diversity as well as their interactions as predictors.

This model shows main effects for papers from Northern America (Odds Ratio = 1.91, $p < 0.0001$). Papers from Western Europe (Odds Ratio = 1.25, $p = 0.058$) and Northern Europe (Odds Ratio = 1.13, $p = 0.25$) also had slightly increased odds of being cited more, but not at a significant level. The highest positive influence (Odds Ratio = 5.43, $p = 0.27$) was for papers with co-authors from Northern Africa (all from Egypt), albeit with only eight such papers and not at a significant level either. While not significant, the largest penalties to citation rates are observed for papers from Eastern Europe (Odds Ratio = 0.43), Latin America (Odds Ratio = 0.66), and Eastern Asia (Odds Ratio = 0.81). The latter is particularly noteworthy given the large share of papers from this region. The only significant interaction with country diversity was for Northern America, where larger country diversity is associated with lower citation rates (Odds Ratio = 0.50, $p = 0.002$).

We can also use this opportunity to check a claim by Kumar and Karusala [54], who remark that "*researchers not fluent in English find themselves repeatedly disadvantaged*," For this, we categorize countries by whether English is a *de jure* (e.g., Singapore and Ireland) or predominant (e.g., the USA and Australia) language. For each paper, we can then check how many affiliations are from such English-speaking countries as well as how many are from other countries. Regressing on the ratio of these two using a log-linked Gaussian GLM (also including decade and author count predictors) to predict citation rates, we see a significant influence of English language; $\chi^2(1) = 70.738, p < 0.0001$. On average, as more authors are from an English-speaking country, the rate of citations for that paper increases. The effect here is substantial, with the average paper without any authors from an English-speaking country receiving 6.75 citations per year, while those where all are from English-speaking countries receive 9.37 citations per year. However, we note that the effect of language is still dwarfed by the effect of a paper having American authors. We confirm this with an extended model that also includes the share of American authors; language ratio: $\chi^2(1) = 3.60, p = 0.0578$, American ratio: $\chi^2(1) = 69.57, p < 0.0001$. As earlier, we thus continue to focus on American authorship in the remainder of the analyses.

6.3.3 First and Last Authors. While we looked at the overall author team composition, some authors are usually more relevant than others. Commonly, the first author lead the work and the last author is the most senior and supervised the work. As such, it might well be that the work itself and its reception are most heavily influenced by these two authors. For research in learning analytics this was recently studied by Poquet et al. [79], focusing on the gender of the first and last authors. They found that authors with female last authors, for example, were under-cited by about 9%, but papers with a female first and male last authors over-cited by 8.4%. Moreover, papers

Table 2. ANOVA Tables for GLMs Fitted to First/Last CHI Author Information and Predicting Citation Rates and Award Probabilities

	Citation Model				Awards Model			
Factor	df	*F*	p-Value		df	*F*	p-Value	
Decade	4	46.92	<0.001	***	1	1.58	0.21	
First Author Gender	1	0.42	0.52		1	0.82	0.37	
Last Author Gender	1	0.05	0.82		1	4.16	0.041	*
First Author Ethnicity	4	3.29	0.010	*	4	1.66	0.16	
Last Author Ethnicity	4	4.10	0.003	**	4	3.33	0.010	**
Either Author American	1	129.92	<0.001	***	1	13.24	<0.001	***
First Author in Academia	1	0.92	0.34		1	0.03	0.87	
Last Author in Academia	1	7.41	0.007	**	1	5.10	0.024	*
First Author Academic Age	1	9.26	0.002	**	1	3.47	0.062	
Last Author Academic Age	1	1.90	0.17		1	0.01	0.92	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

with female first and last authors saw big deviations from expected citation rates. Where papers from male first and last as well as male first and female last authors cite them substantially less often, teams of female first and last authors over-cite them heavily. All this suggests that there could be more complex effects at play, depending on who these two authors are. Where Poquet et al. used citation network data for their analysis, we can similarly use our data on overall citations and awards to check whether a similar effect occurs with publications at CHI.

For this analysis, we extract the information on the first and last authors from the collected CHI paper metadata. We exclude papers with a single author and papers where the gender of the first or last author could not be determined. This leaves us with 6,048 papers with a male first author and 3,447 with a female one and respectively with 6,826 male and 2,669 female last authors; 9,495 papers overall.

As in our earlier analysis, we fit the same kinds of GLMs to the data, one again predicting citations per year and one the probability of a paper receiving an award. Both include predictors for all the information we have on the first and last authors, as well as the decade the paper was published in. This includes first and last authors' gender, ethnicity, academic age, whether they are American, and whether they work within or outside of academia. As first and last authors being from the USA was highly associated (Cramer's V of 0.88), we combine the two into a predictor whether any of the authors is American. Hence, in contrast to our previous model, most predictors here are categorical. Correspondingly, we compute ANOVAs for both models, the results of which are shown in Table 2.

As in our previous analysis, when a paper was published has a significant influence on citations. Furthermore, so do the first and last authors' ethnicity, whether the last author is working in academia, whether either of them is American, and the academic age of the first author. As can also be seen in Figure 8, the most notable of those effects is the citation advantage of American authors. The same holds for awards, where papers with American first/last authors also have an advantage (see Figure 9). Awards are also the only dimension where there is a gender effect, specifically with female last author papers receiving slightly fewer awards. Author ethnicity plays into both citations and awards, and again shows a pattern where Asian authors look to be among the most disadvantaged in either regard.

Fig. 8. Marginal effects for the relationship of all predictors available for first and last authors to log-transformed amount of citations a paper gets per year. Shaded area and error bars show 95% CI.

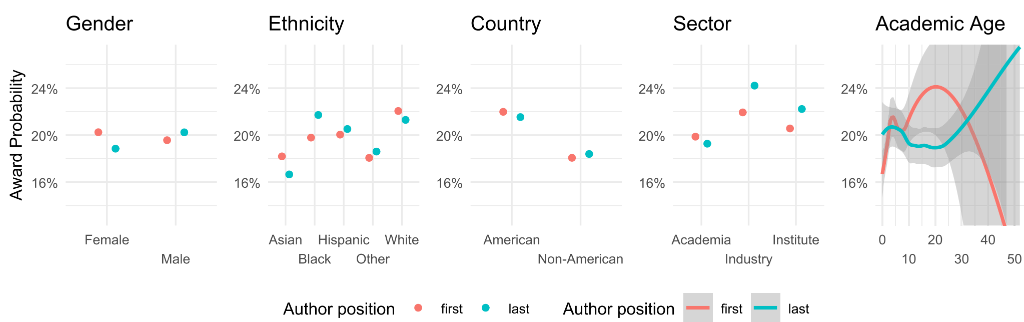

Fig. 9. Marginal effects for the relationship of all predictors available for first and last authors to log-transformed amount of citations a paper gets per year. Shaded area and error bars show 95% CI.

Overall, we see many of the patterns from our main analysis repeated when focusing only on first and last authors. These results also go counter to the findings of Poquet et al. [79] that inspired this angle of investigation. Where they saw gender playing a larger role in citation counts, the same does not hold for CHI papers. Instead, other author properties dominate effects on paper impact, most importantly whether the first or last author are based in the USA.

7 Potential Diversity Mechanisms

Our data does not answer the question *why* some diversity predictors are related to higher impact CHI papers. Potential explanations could be effects of diversity on (1) the quality of the work itself, or (2) the perception of the work. For example, Sulik et al. [102] have argued for cognitive diversity as a likely source of diverse teams producing better work. Moreover, an external effect of diversity would be the larger personal networks attached to the authors and thus also more potential for outreach. Furthermore, the answer might well be a combination of both kinds of effects with the work itself as well as its external reception benefiting from diverse authorship.

We posit that most effects on citation rates are likely due to connectivity, as also underlying the analysis by Chen [12]. Larger personal networks of authors likely lead to improved uptake of a paper, with more diverse teams having a higher chance of such networks. However, such personal connections only result in more impact if the other side has high citing power (e.g., publishes prolifically) and is not shared with other authors (i.e., actually increases reach). The latter point is supported by Ding et al.'s [22] findings around audience diversity. Hence, papers involving authors from non-research areas of industry would not benefit as much, given these authors likely have fewer connections that boost citations. Homophily likely also increases connection strength, which could explain the effects of some of the diversity bias measures we observed. For example, American researchers building strong national networks see benefits while, as our follow-up analysis showed, their citation rates go down with higher country diversity.

Personal network effects likely play a much smaller role for awards, given that the people deciding on them are drawn from the CHI AC population. We posit three different mechanisms that might play a role here: (1) diverse teams actually produce higher quality work and the awards process is able to identify this, (2) award decisions are influenced by committees desire for representation in terms of topics as well as demographics, and (3) award decisions are influenced by homophily effects (i.e., the demographics of the committee relate to the demographics of the recipients). In terms of effect size, whiteness seems to have played a larger role for awards than for citation rates, which could support the latter point. However, we have no data available on the CHI award committees to further investigate this.

We also need to consider the explanation that diverse teams of authors produce better work, which could also result in higher citation and award counts. Going back to Sulik et al.'s [102] work on cognitive diversity, the same mechanisms should apply to CHI papers. Thus, it seems sensible to assume that including authors with more viewpoints, skills, and experiences would allow such teams to better identify critical questions as well as potential solutions to them. Which of our diversity dimensions would result in the most cognitive diversity, however, is not clear and it remains to be investigated more closely.

Overall, while we deem network effects the most likely pathway for diversity at this point, this cannot be directly validated with the data we have available. However, we can more closely inspect the alternative, that diverse teams bring together more viewpoints, skills, and experiences and that this results in papers of higher impact. These papers' quality per se is not something directly measurable, but they should also be more novel and stand out from other papers if indeed there is such a manifested benefit. We further draw on our collected data to investigate whether the work of diverse teams indeed exhibits such increased novelty. Moreover, there might be biases towards or against certain kinds of research, independent of who authored it. We infer research areas from the author keywords used in the papers to investigate how much of a factor they are for the respective paper impact. Finally, we investigate a subset of papers—those that include positionality statements—to see whether reported identities and diversity and considerations around them potentially relate to paper impact. As these are prepared by the paper authors, such statements provide an internal perspective, where the diversity indicators we have used so far are external signifiers.

7.1 Analyzing Diversity and Research Area

Differences in CHI paper impact might not be due to author characteristics, but due to the subject area or method employed in the work. To investigate such a potential pathway, we draw on the keywords associated with papers (inspired by [61]), which commonly encode both those aspects. This data was available for 9,689 of the 10,341 papers, meaning 652 papers (6.3%) are excluded from this analysis. As with other metadata, coverage is not even, with some years having no keywords for any paper (1986, 1988–1991, and 1995). Since the late 90s, however, almost all papers come with this information available in the ACM DL.

Overall, we observed 19,906 different keywords. Keywords follow a power-law distribution with 14,303 of them only used a single time. These numbers already disregard capitalization of keywords, but for further analysis, we perform some additional keywords normalization. Based on automatically identified pairs of similarly spelled keywords, we derive manual corrections we apply to the keywords. This corrects for (1) typos, (2) spellings other than American English, (3) plural forms, and (4) different forms of hyphenation. We employ a manual process as there are many similarly spelled, yet different keywords, such as unsupervised/supervised, motion/emotion, head/hand, or interaction/integration. This leaves us with a smaller number of 18,659 normalized keywords.

The 10 most used tags were “virtual reality” (357 times), “design” (260), “privacy” (235), “social media” (234), “augmented reality” (223), “accessibility” (223), “visualization” (201), “crowdsourcing” (198), “children” (175), and “collaboration” (149). This again shows the rapid drop in number of uses, further indicated with the 99th percentile keyword being used 30 times, but the 90th and 80th percentile keywords already only used four and two times, respectively. It also indicates that some areas likely have more narrow sets of keywords used, while others use a broader set of many different ones.

In a next step, we took the most used keywords and manually grouped them into methodological and topical clusters. We started from the most used ones and then updated and refined the emerging groups as we worked our way down the list. This is hence a subjective process and there is inherent ambiguity as to what exactly the research areas within CHI are, as also visible in the shifting subcommittee options and boundaries. Many of the widely used keywords also are used across much of CHI, such as “user study,” “user interface,” “prototyping,” or “human-computer interaction.” We excluded such keywords from the process. This still left us with a rich set of commonly used keywords, which form a set of 11 research areas, as shown in Table 3. Not all papers with keywords are within these areas, and we were left with a subset of 7,815 out of the 9,689 with keywords. Note also that papers can belong to multiple of these areas, which was the case for 1,351 papers. We treat each paper-area combination as an independent unit in our following analysis, where we investigate the differences between areas.

Plotting the diversity predictors for each of the research areas (see Figure 10), we can see that there is substantial variation across them. For example, authors of papers from A02, which broadly covers design research, are leaning whiter than, for example, those from A11, which includes a range of more technical topics. Similarly, regional diversity of authors was higher in papers from A08, possibly because that area includes the topic of HCI for development. We can also see that the share of American authors varies a lot between research areas. While social computing (A04) or visualization (A06) research lean largely American, the opposite is true for research in the area of extended reality (A01). However, we also see some predictors, such as the average academic age of paper authors, for whom the research area has little to no apparent effect.

With diversity varying by research area, we set out to see how the citation and award models from our main investigation would change when taking the research area into account. We hence included research area as a factor, applying treatment contrasts so one of the areas serves as a baseline. For this we picked social computing (A04), which is well represented in the sample and has been an area of focus for much of the history of CHI. The resulting models (see Table 4) show significant effects for several area coefficients, but also changes to the significance of other factors, now that research area is accounted for. However, we note this is also due to the fact that the papers used for fitting this model are a subset of the papers used earlier, as only those with fitting keywords are included here.

With respect to citations, we see extended reality (A01), privacy and security (A03), and technical topics (A11) significantly outperforming the baseline of social computing research. On the other hand, it was only papers associated with sustainability and HCI for development (A08) that had a significantly higher likelihood to receive an award than the baseline. Including research area in the models removed the significant effect of sector diversity on citations, as well as of ethnic diversity and both academic age predictors on award likelihood. We checked the overall influence of research area with ANOVA tests, changing the corresponding factor to sum-to-zero contrasts, so every area is tested against the mean. This shows significant effects of research area on citations ($F(10) = 8.75, \text{p} < 0.001$), but not award likelihood ($F(10) = 1.69, \text{p} = 0.078$).

Table 3. We Manually Group the Most Used Keywords from CHI Papers into a Set of 11 Research Areas

Area	Contained Keywords	# Papers
A01	virtual reality (357), augmented reality (223), mixed reality (71)	651
A02	design (260), participatory design (142), interaction design (129), ethnography (117), research through design (78), co-design (69), user-centered design (68), creativity (57), design method (56), storytelling (54), design research (48), interface design (48), user interface design (46), game design (46), speculative design (37)	1,255
A03	privacy (235), security (86), trust (85), ethics (61), usable security (40), authentication (35)	542
A04	social media (234), crowdsourcing (198), collaboration (149), cscw (106), computer-mediated communication (99), online communities (76), social computing (62), social network (58), communication (58), facebook (55), twitter (54), social interaction (39), community (38), wikipedia (38)	1,264
A05	accessibility (223), visual impairment (67), assistive technology (61), blind (44), disability (40)	435
A06	visualization (201), information visualization (115), data visualization (50)	366
A07	children (175), education (94), learning (84), older adult (74), gender (63)	490
A08	sustainability (103), hci4d (82), ictd (56)	241
A09	health (91), mental health (63), healthcare (45), social support (38), dementia (36)	273
A10	mobile device (114), ubiquitous computing (113), wearable (112), mobile (96), text entry (85), mobile phone (80), fitts' law (78), smartphone (74), touchscreen (70), gestures (68), navigation (65), touch (56), internet of things (55), mobile computing (55), pointing (53), wearable computing (52), multi-touch (48), smart home (43), smartwatch (43), gesture recognition (39)	1,399
A11	eye tracking (120), machine learning (116), haptics (98), interaction technique (94), 3d printing (76), fabrication (68), input device (66), artificial intelligence (58), computer vision (49), haptic feedback (41), shape-changing interface (40), digital fabrication (37), human-ai interaction (36)	899
	Overall:	7,815

We list the keywords constituting those areas, how often they occur, as well as how many papers fall under each such area. Papers can belong to more than one area.

Overall, this shows that the research area a paper is in indeed has a strong effect on its impact. Models including it still show evidence for diversity effects, though. Furthermore, some diversity measures vary across research areas, meaning that there are interactions that need to be considered. For example, our analysis cannot show whether it is the topic of extended reality that results in those papers being cited more, or the fact that papers from this area are mostly written by male authors. Inversely, while we see an overall strong boost to citation rates when authors are from the USA, these highly cited extended reality papers often are not from such authors.

7.2 Analyzing Diversity and Novelty

Novelty is an important aspect of scientific work and plays a role in its assessment after publication but also already during the review process. For example, reviewers for CHI are asked to rate a submission's *originality*, meaning the "*presence of new ideas or approaches*," CHI papers themselves

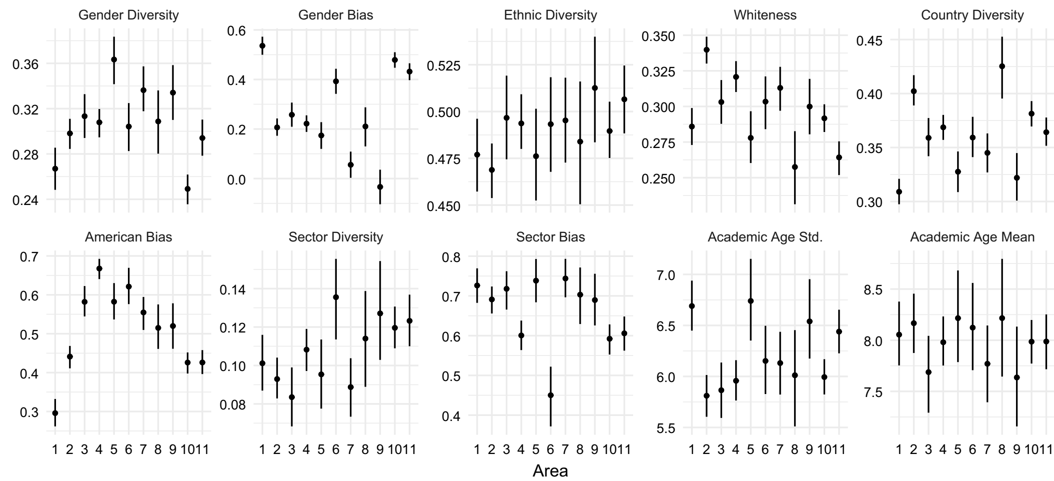

Fig. 10. For 11 areas of CHI research (see Table 3), this shows how the used diversity predictors vary between them.

Table 4. Fitted GLMs for Citations per Year (Left) and Awards (Right)

	Citation Model						Awards Model					
Factor	Coef.	Std. Error	Odds Ratio	t-Statistic	p-Value		Coef.	Std. Error	Odds Ratio	t-Statistic	p-Value	
A01	0.29	0.06	1.33	4.64	<0.001	***	−0.12	0.15	0.88	−0.82	0.41	
A02	−0.07	0.06	0.93	−1.30	0.19		0.12	0.12	1.12	0.94	0.35	
A03	0.18	0.06	1.20	2.92	0.004	**	0.10	0.15	1.11	0.68	0.50	
A05	−0.16	0.09	0.85	−1.75	0.080		0.04	0.17	1.04	0.23	0.82	
A06	−0.04	0.08	0.96	−0.49	0.62		−0.09	0.19	0.92	−0.47	0.64	
A07	−0.14	0.08	0.87	−1.82	0.069		0.09	0.16	1.09	0.56	0.58	
A08	-0.11	0.10	0.90	−1.03	0.30		0.49	0.19	1.64	2.65	0.008	**
A09	0.15	0.08	1.16	1.86	0.062		0.30	0.18	1.35	1.72	0.086	
A10	−0.04	0.05	0.96	−0.80	0.43		−0.01	0.13	0.99	−0.10	0.92	
A11	0.26	0.05	1.29	4.80	<0.001	***	0.24	0.13	1.27	1.85	0.065	
Decade: 1980s	−0.66	0.79	0.52	−0.83	0.41							
Decade: 1990s	0.06	0.07	1.06	0.83	0.41							
Decade: 2000s	0.12	0.04	1.12	2.63	0.009	**						
Decade: 2020s	−0.32	0.05	0.72	−7.09	<0.001	***						
Number Of Authors	0.06	0.01	1.06	7.66	<0.001	***	0.00	0.02	1.00	0.22	0.82	
Single Author: True	0.10	0.11	1.11	0.88	0.38		0.26	0.29	1.29	0.88	0.38	
Page Count	0.03	0.00	1.03	6.72	<0.001	***	0.02	0.01	1.02	2.64	0.008	**
Gender Diversity	0.11	0.08	1.12	1.34	0.18		−0.05	0.18	0.95	−0.26	0.80	
Gender Bias	0.03	0.03	1.03	0.90	0.37		0.04	0.07	1.04	0.56	0.58	
Ethnic Diversity	0.14	0.07	1.15	1.99	0.047	*	0.25	0.15	1.28	1.61	0.11	
Whiteness	0.35	0.10	1.42	3.39	<0.001	***	0.99	0.22	2.69	4.51	<0.001	***
Country Diversity	0.14	0.12	1.15	1.23	0.22		−0.13	0.26	0.88	−0.51	0.61	
American Bias	0.41	0.04	1.50	10.55	<0.001	***	0.20	0.08	1.22	2.50	0.012	*
Sector Diversity	0.08	0.09	1.08	0.85	0.40		0.22	0.23	1.25	0.96	0.34	
Sector Bias	−0.07	0.03	0.93	−2.78	0.006	**	−0.12	0.08	0.88	−1.58	0.11	
Academic Age Std.	−0.03	0.01	0.97	−4.20	<0.001	***	−0.01	0.01	0.99	−1.00	0.32	
Academic Age Mean	0.03	0.00	1.03	6.04	<0.001	***	0.01	0.01	1.01	0.61	0.54	

The citation glm is Gaussian with a log link function, while the awards GLM is binomal with a logit link function. The awards model only includes data from CHI 2010 and later, but does not include decade as a predictor. We also report odds ratios for easier interpretation of the models. In addition to the predictors used in the main analysis, the models here also include a term for which area of CHI research a paper is in. Coefficients for those areas are relative to A04, which was set as baseline.

*p < 0.05; **p < 0.01; ***p < 0.001.

also increasingly highlight what is novel about them [78]. Correspondingly, determining the novelty of a piece of work has been a focus of research, be it around patents [45] or scientific publications [46, 94]. Different methods have been developed for this purpose, such as using co-occurrence of terms in abstracts [29], the reference used [107], combinations of references [8], or the abstracts and titles of those references [94]. Similar methods can also be applied to authors, with Singh et al. [97], for example, describing different patterns between researchers that are *explorers* and *exploiters*. Implementing multiple paper novelty indicators, Pelletier and Wirtz [75] further distinguish between novelty and disruptiveness indicators, with the latter describing "*if a given article behaves as a bottleneck between the knowledge mobilized in a given article and the articles that will cite it*." In general, as shown by Wang et al. [107], research that is more novel has greater impact, and we thus can expect the same to hold for CHI papers. However, they also note that this advantage does not necessarily manifest in the near term and that it often takes longer for such novel work to have this impact.

Deriving a similar novelty measure for our dataset of CHI papers poses some challenges. As described in Section 4.1.2, the reference data given on the ACM DL is unreliable, especially for older papers. This means many of the existing novelty indicators are not suitable for our dataset, as they require citation network information or even the metadata and text of cited works. However, there are some methods that show another path more suitable for the data in question. Specifically, a class of novelty methods operate on paper text and comparison across a corpus. For example, Yin et al. [114] computed word embedding of titles and abstracts and then used these embeddings to derive distances between documents. The distribution of distances from a given document to all others then defines novelty, with papers being more novel the further away they are from others. This is validated through a survey as well as comparisons to existing measures. Similarly, Jeon et al. [46] also used text embeddings but then only considered a subset of papers when determining whether a given one is novel or not. By computing the local outlier factor a paper is novel when it stands out from the papers most similar to it. Another approach was used by Wang [105], who used a language model to compare a given paper text against text predictions. Thus, this captures whether a text is *surprising* or whether it aligns more with what the language model was trained on. We follow Jeon et al.'s [46] approach of identifying papers that differ from what has been previously published (i.e., are novel in some way).

For this analysis, we only consider the abstracts we collected for CHI papers, not their full text. We have abstracts for 9,852 papers, but had to exclude 489 papers from analysis as they had no abstract on their ACM DL page (e.g., 10.1145/22627.22363 and 10.1145/108844.108855). This is not a uniformly distributed issue with all papers from CHI 1991 and 1995–1998 missing abstracts in the ACM DL. CHI in 1986 (2% missing abstracts), 1987 (4%), 1992 (22%), 1993 (20%), and 1994 (99%) also stood out. However, at an overall rate of 4.7% missing abstract and with our analysis not focusing on specific years, there is sufficient data to draw on.

We use OpenAI's embeddings[31] to represent all abstracts in a shared numeric space. Specifically, we use the *text-embedding-3-small* model to produce 1024-dimensional embedding vectors for each abstract. Embeddings are representations that capture the essence of a piece of text or other media in a standardized way, so they can more easily be operated on. For example, in an embedding space distance and similarity measures are mathematical operations on numbers instead of textual operations. The used embedding model is a general-purpose model for any kind of text and designed to facilitate understanding of content relationships as well as clustering and searching of the data.

[31] See https://platform.openai.com/docs/guides/embeddings and https://openai.com/blog/new-embedding-models-and-api-updates.

Table 5. We Generated Embeddings for CHI Abstracts and Then Checked Whether Related Abstracts Are Close in that Embedding Space

Rank	Distance	Paper
Query for: Bartneck and Hu, 2009. *Scientometric analysis of the CHI proceedings*		
(1)	0.899	Caine, 2016. *Local Standards for Sample Size at CHI*
(2)	0.898	McKay et al., 2022. *Who am I, and who are you, and who are we? A Scientometric Analysis of Gender and Geography in HCI*
(3)	0.873	Nikisirat et al., 2023. *Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022*
(4)	0.868	Liu et al., 2014. *CHI 1994-2013: mapping two decades of intellectual progress through co-word analysis*
(5)	0.868	Linxen et al., 2021. *How WEIRD is CHI?*
Query for: Guiard et al., 2011. *Fitt's law as an explicit time/error tradeoff*		
(1)	0.773	Gori and Bellut, 2023. *Positional Variance Profiles (PVPs): A New Take on the Speed-Accuracy Trade-off*
(2)	0.760	Guiard, 2009. *The problem of consistency in the design of Fitts' law experiments: consider either target distance and width or movement form and scale*
(3)	0.753	Bi et al., 2013. *FFitts law: modeling finger touch with fitts' law*
(4)	0.743	Wobbrock et al., 2011. *The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models*
(5)	0.615	Wobbrock et al., 2008. *An error model for pointing based on Fitts' law*
Query for: Pohl et al., 2019. *Charting Subtle Interaction in the HCI Literature*		
(1)	0.894	Hornbæk and Oulasvirta, 2017. *What Is Interaction?*
(2)	0.894	Mekler and Hornbæk, 2019. *A Framework for the Experience of Meaning in Human–Computer Interaction*
(3)	0.891	Bruns et al., 2021. *Expressivity in Interaction: a Framework for Design*
(4)	0.837	Serim and Jacucci, 2019. *Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research*
(5)	0.752	Anderson et al., 2015. *Supporting Subtlety with Deceptive Devices and Illusory Interactions*

The five most similar CHI papers for three example query papers are presented in the table. We find that there indeed is a close relationship in these matches and further validation checks we conducted with other sample papers.

Thus, it is well suited for CHI abstracts that can cover a wide range of topics and our use case of investigating novelty in CHI papers as represented by their abstracts.

We manually explored the resulting embeddings to see whether the inferred relationships are sensible. For this we picked example CHI papers, retrieved the top ranked most similar CHI papers in the dataset, and checked whether the abstracts indeed align. Table 5 shows three example cases with the five most similar papers for each. We find that the embedding overall captured the relationships well, with closely related papers in the embedding space also exhibiting similar closeness in their actual content. Given that this is the case, the same embedding should also be suitable to identify dissimilar (i.e., novel in some way) papers.

For each paper, we then compute its local outlier factor,[32] which is a measure of how much that paper differs from its closest neighbors. Thus, a paper is not compared against all other CHI papers, but just those most similar to it (set to $n = 5$ in our case). Furthermore, we only consider papers from the same or previous years when computing this factor for each paper. Finally, we normalize

[32]Using https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html.

Table 6. GLM Fitted on Metadata and Diversity Data from 9,871 CHI Papers to Predict Novelty (Measured via Local Outlier Factor)

Factor	Coef.	Std. Error	*t*-Statistic	p-Value	
Decade: 1980s	0.014	0.011	1.232	0.22	
Decade: 1990s	0.033	0.012	2.725	0.006	**
Decade: 2000s	0.024	0.006	4.155	<0.001	***
Decade: 2020s	0.008	0.005	1.611	0.11	
Number of Authors	−0.000	0.001	−0.094	0.92	
Single Author: True	0.033	0.013	2.583	0.010	**
Page Count	−0.003	0.001	−5.294	<0.001	***
Gender Diversity	−0.009	0.009	−0.989	0.32	
Gender Bias	0.004	0.004	1.079	0.28	
Ethnic Diversity	−0.024	0.008	−3.203	0.001	**
Whiteness	−0.030	0.011	−2.739	0.006	**
Country Diversity	−0.009	0.013	−0.677	0.50	
American Bias	0.007	0.004	1.740	0.082	
Sector Diversity	−0.020	0.011	−1.710	0.087	
Sector Bias	−0.007	0.003	−2.169	0.030	*
Academic Age Std.	0.000	0.001	0.224	0.82	
Academic Age Mean	−0.001	0.001	−1.052	0.29	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

the resulting outlier factor values per year. Hence, it is not an absolute measure of novelty, but describes how a paper fares compared to its contemporaries.

We then repeat our model-based approach from before, but use a canonical link for the GLM (i.e., equivalent to a normal linear model). Table 6 shows several significant relationships of paper and author properties to novelty. Most of these effects are for metadata such as when a paper was published and how long it is, but single author papers also have higher novelty on average. In terms of diversity, only ethnic diversity had an effect on novelty, albeit a negative one. Furthermore, whiteness and sector bias had a negative influence on novelty as well. However, while there are some significant effects here, the strength of these effects is overall low, as also can be seen in Figure 11. A potential exception is page count, where the marginal effect is more pronounced and longer papers seem to be less novel than shorter ones (per their abstract at least).

Overall, this data does not suggest that more diverse teams of CHI authors also produce more novel work. At least with respect to how much their abstract differs from previous papers, these papers do not stand out. Yet, if such papers indeed explore new topics, methodologies, or target audiences this should manifest in the abstract as well and result in higher outlier scores if such directions have not been explored before. As such, we find this lends further credence to network effects likely being the main driver of author diversity benefits on paper outcomes. However, given we only considered the abstracts of CHI papers, further exploration of the paper content might be a fruitful direction to revisit this aspect more closely in the future. It might also be that diverse authors create papers that are similar in approach to the work conducted in general, but do so at higher levels of quality. Yet, this is ultimately speculation, and we have no data to suggest such differences in paper quality exist.

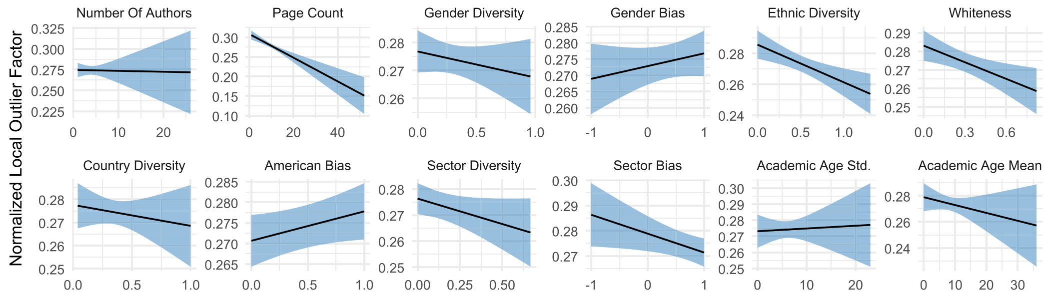

Fig. 11. Marginal effects for the relationship of all numeric predictors to normalized local outlier factor values, which measure how novel a paper's abstract is with respect to previously and concurrently published papers. Shaded area shows 95% CI.

7.3 Analyzing Positionality

Some papers include reflections on the authorship team in the form of a positionality statement. We collect the CHI papers for which this is the case to analyze these statements on their own as well as how they relate to the other data we have already. For this, we searched the ACM DL[33] for CHI papers where the full text includes the term "*positionality statement*" and that have been published at up to and including CHI 2023.

Our search only found 29 papers, from which we later excluded two papers because they mention such statements, but did not include one from the authors. The remaining 27 papers were all published in 2020 or later, with 1 paper from CHI 2020 (759 papers overall, 0.1% with positionality statement), 12 from CHI 2022 (637, 1.9%), and 14 from CHI 2023 (879, 1.6%). Hence, the practice is rare, but has seen some uptake more recently. While our analysis ends with CHI 2023, there were 50 (out of 1,058, 4.7%) and 56 (out of 1,249, 4.5%) results for papers with such statements at CHI 2024 and CHI 2025, respectively.

Given the small number of found papers with positionality statements, we refrain from analyzing overall trends and correlations in the data. Furthermore, given these papers were published very recently, the citation data we have collected is not very informative here either (e.g., this is why we excluded CHI 2023 papers from the citation model earlier). However, we can look at the awards data to see whether these papers with positionality statements fared differently than those that did not include them. Only taking into account the three CHI conferences with such papers, we do see some difference in the data. While 18.0% of the papers without positionality statements received awards, 29.6% of those that include one do so.

We then read through the 28 papers that include a positionality statement to see whether they mention anything in their statements or in other author-related statements that could relate to them receiving or not receiving an award. Diversity in the author team is brought up in several ways, such as "*diverse range of backgrounds and expertise*," "*diverse team composition(racially/ethnically)*," "*diverse socioeconomic backgrounds*," and being a "*racially diverse team*." Instead of explicitly referring to diversity, many such statements report on individual author identities and backgrounds. For example, one statement noted how their "*team includes a mix of cisgender women and men, and a mix of academic researchers and industry practitioners [...] team members hold multicultural backgrounds*." Interestingly, a focus often is not on author diversity, but on alignment of author identities with

[33]Query: https://dl.acm.org/action/doSearch?target=advanced&ContentItemType=research-article&SpecifiedLevelConceptID=119596&BeforeYear=2023&AllField=Fulltext%3A%28%22positionality+statement%22%29+AND+ContentGroupTitle%3A%28NOT+%22Extended+Abstracts%22%29.

those of the studied group. For example, one statement highlighted that "*the paper's authorship team has some significant overlap with the positions of our participants*,"

But are there any diversity-related patterns in these positionality statements that show differences between papers that received an award and those that did not? Based on the given sample of 28 papers, this does not seem to be the case. The issues touched upon, such as privilege, connection to study population, and the authors' experience and background, show up in both paper groups. Furthermore, in both we can find authorship teams that are diverse and others where authors have very similar identities and backgrounds. Compared to the larger population of CHI authors, the reported identities in the positionality statements seems to skew female and non-White, but this also is true irrespective of awards. However, a noteworthy aspect of the papers receiving an award is that all but one of the corresponding 43 authors was located in the US, with the last one located in the UK.

Overall, while authorship team diversity factors into positionality statements, their scope is much wider. For example, the team composition is often directly related to the characteristics of the people who are studied or engaged with. Yet, while this does then sometimes relate author diversity to some forms of impact (e.g., community benefit) no clear connection between such considerations and the CHI award process emerged. An issue here is the small number of papers that include such statements and it might be worth revisiting this question in the future when there is a larger corpus to work with. However, we also note that topic-wise there appears to be a strong bias in which kinds of CHI papers include such statements and thus there likely are confounding effects here that further complicate such an analysis.

8 Discussion

Just as Freeman and Huang [31], we also found an effect of author diversity on paper impact. Their examination of the number of addresses on and references in a paper is similar to our variables of country diversity and page count (which we found highly correlated with reference count, see Section 5.3). Their results and our model showed positive relationships in both cases. Otherwise, their analysis focused on homophily, which is equivalent to our measure of ethnic diversity, but inverted. Where they observed a decrease in citations for papers with more homophily, we similarly found an increase of citations alongside ethnic diversity. While our results are thus overall in line with theirs, we should note that where they investigated the impact on absolute citation counts, we used the rate of citations, which could have had an influence on the results. They also separate their analysis for papers with 2, 3, 4, and 5–10 authors, where we fit models for any number of authors.

Focusing just on awards, we can draw on work by Mubin et al. [67] to complement our results. They investigated how characteristics of CHI papers, readability specifically, influence the likelihood of them receiving an award. Award-winning papers had slightly but significantly worse readability than those not winning an award, for example, containing more difficult words. The only author characteristics they report, experience in terms of years and papers, were not significant. This is in contrast to our awards model, which showed a significant influence of author experience as well as spread on the chance of a paper receiving an award. Yet, the two analyses are not directly comparable as the included predictors vary substantially and the data used also differs. Where we analyzed 7,953 papers, Mubin et al.'s analysis only considered 382 full CHI papers. However, their results do show that content-based measures can indeed influence paper impact, as our own analysis of research areas also indicated.

Fell and König [27] focused on gender differences in collaboration and overall found that these do not predict scientific success even though male and female researchers exhibit different collaboration patterns. They found no correlation of gender diversity in a collaboration with the mean number of citations the resulting paper garnered. Similarly, we also saw no evidence for an effect of gender diversity in a collaboration on citations or on awards received. But they did find a small (Kendall's $\tau \pm 0.04$) bias effect with a higher percentage of male collaborators having a positive impact on citations and a higher female share a negative one. In contrast to that result, we saw no significant effects of bias. In light of Fell and König's, at best, small correlation, the results from their and our analysis suggest that biases in gender distribution have little effect on paper impact.

This lack of an effect for gender diversity is counter to the report of Nielsen et al. [69]. However, they mostly point to increased group performance for gender-diverse teams as well as better group dynamics. Furthermore, they note that there is a gender difference in what topics are worked on (which we also observed in our analysis of research areas) and how. Yet, while they frame this as gender diversity resulting in "better science" this point is not actually supported in terms of improved reception of that science. In a related paper, Nielsen [68] himself summarizes that "*gender differences in citation impact are marginal to non-existing*," Our own data also points to a lack of an effect on impact for gender and gender diversity.

Complimentary to gender diversity effects, we also saw no effects of gender bias in author teams and only a small effect on awards in the analysis of first and last authors. However, when Bikard et al. [7] recently investigated how scientific papers impact patents, they did find an advantage for main authors that are male. This is across several different kinds of controls, and with models otherwise indicating effects similar to our findings, such as teams with more American authors or with industry collaborators being cited more. They also looked at potential explanations and identified inventor (i.e., the ones who do the citing) attention as the most likely one. This was through potential "paper twins" (i.e., similar papers where one is authored by a man and one by a woman), which, for example, also allowed them to determine that research style and university prestige did not differ much by author gender. However, a follow-up analysis where crowdworkers read paper abstracts with the associated author gender manipulated, only found that slightly more time was spent on reading abstracts from male authors. In terms of quality, importance, or usefulness ratings, though, gender was not a significant factor. Overall, this work shows stronger gender effects than our findings, though they also looked at impact in industry, where we looked at impact in academia, and at journal publications from many fields, where we have investigated CHI papers specifically.

Author experience had an effect on both impact measures. This is not a very surprising result, given that one would expect more experienced teams to do better. Furthermore, impact goes down the more diverse in experience a team is. One way to read this is that it is specifically collaborations between more experienced researchers that have high impact. Teams that mix very experienced researchers with very inexperienced ones are less impactful. This could hence also be read as the difference between collaborations of senior researchers and senior researchers working together with students. In the follow-up analysis of first and last authors, we also saw an effect of first author experience on citation rates. Here again, more experience translated to more citations. Also noteworthy in this regard is work by Li et al. [58], who found that collaborations with top senior scientists is a strong predictor for later career success of junior authors. Hence, not only does overall team experience matter on a per-paper level, it also has consequences for the authors.

With respect to non-diversity indicators, we found that the number of authors positively influenced citation counts and the paper length positively influenced citations and the chance for an award. This positive influence on citations of longer papers and papers with more authors is well-established in the literature. For example, while studying the predictive potential of structural variation metrics, Chen also confirmed the effects of author and page counts [12]. Other examples with the same effect are Robson and Mousquès analysis of papers on environmental modeling [83] and Falagas et al.'s analysis of papers in general medicine journals [26].

There are many more non-diversity covariates one could consider and that could help better predict impact. For example, as Pohl and Mottelson [78] showed, the style of writing and titles on papers seem to also have an impact on citation rates. What a paper is actually about, its content, very likely has a big influence on its impact and our analysis has only looked at this aspect from a research area perspective. While this is beyond the scope of this work, future work might want to explore how much the topic, methodology, or presentation in papers matters.

We have already discussed potential diversity mechanisms earlier in Section 7 and how our results fit with previous findings in that respect. Interestingly, our subsequent investigation into paper content and whether diverse teams produce more novel work goes counter to what Sulik et al. [102] suggest. As they note, "*creative or innovative tasks are those where diversity consistently has a positive effect*" and authoring a paper should fall under that definition. However, they also describe those effects being most pronounced in "complex problem solving" which is not necessarily an aspect of work on CHI papers in general. While our results suggest that external factors (e.g., who the authors know) are likely more important than internal (e.g., the novelty of the work they produce) ones, this does not rule out other potential pathways for diversity to work. For example, it might well be that diverse teams produce more creative work, but that such creativity is not captured in comparisons of abstracts and requires closer engagement with the content of individual papers.

As our analysis of research areas indicated (see Section 7.1), impact does vary by area, as do author characteristics. This echoes Henry et al.'s [36] findings from their analysis of HCI conferences, where they note how citation impact varies between them. Specialized conferences, such as InfoVis, here are representative of subsets of CHI research from our analysis. Furthermore, they also identified central authors for these different communities, with one woman in the 20 most central UIST authors, but four women in the same list for InfoVis. While this is not an exhaustive analysis of author characteristics in those communities, it does indicate that some differences exist here, just as we find for the different research areas. Similar to our analysis by research area, Pohl and Mottelson [78] have looked at papers associated with different (inferred) CHI subcommittees. Their analysis indicates smaller differences in readability and citation rates of papers in these areas, where we saw more pronounced effects of different areas.

8.1 Implications for Diversity in HCI

The results from our analysis go counter to how diversity has previously been discussed in other HCI papers, especially with respect to gender. For example, McKay et al. [63] note that "*men are more highly cited than women for reasons of implicit and structural bias*," without providing evidence for this causality. The references that are provided merely describe how author gender in Economics influences who is cited [28], and that men self-cite more than women [52], but do not show an overall under-citing of female authors. Our results do not show a significant bias in citations for papers from teams that are more male. In fact, the trend is reversed with such papers tending to be cited *less* often. A similar effect was shown by Nielsen [68] for research in the management sciences, with women being cited slightly more than men.

Instead, the biggest effects we found for diversity were how White and how American a team is. There are a range of potential explanations for this effect, though we cannot determine the actual cause from the current data. For example, this effect might be rooted in the history of CHI and it having started out in the USA. A head start can be a competitive advantage due to network effects, legacy bias, more senior researchers, but also larger departments that had more time to grow. But, as Figure 4 shows, whiteness and American bias have been decreasing over time. However, even if we limit the analysis to only papers published at CHI 2012–2022, whiteness still is a significant predictor in that model ($p < 0.001$), with an even bigger effect (odds ratio of 1.83).

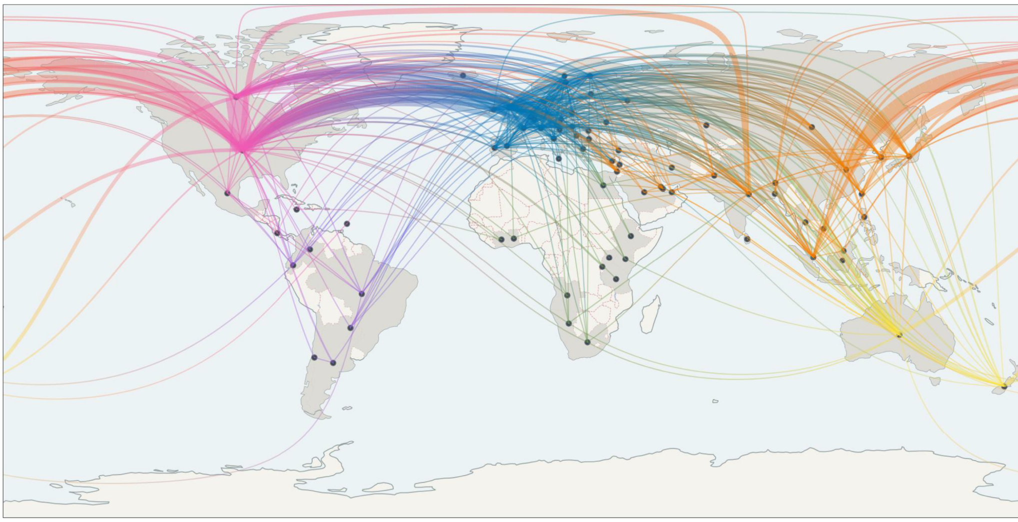

Fig. 12. Co-authorship relations between countries. Line width shows relative prevalence of country connection and line color is assigned per continent and interpolated for intercontinental collaborations.

That non-American research is less visible is not a new phenomenon. Henry et al. [36] already described a similar observation in their analysis of the field of HCI from 2007. They noted that "*Non-American research centers are almost invisible*" and pointed to selection biases and language biases as potential explanations. As our own analysis has shown (see Section 6.3.2), language does factor into impact, albeit less than whether authors are American or not. The roots of CHI also still shine through in the demographics of ACM, SIGCHI, and CHI leadership, which leans American, though with a trend to a more international participation over the years. Another perspective would be to see these indicators as differentiating between papers from the West and papers from Asia. This would relate to the researchers involved in the research, but also to the kinds of research itself, given how methodologies and topics seem to differ between regions. Our follow-up analyses also hint at a structural penalty for papers from East Asia and such a divide between West and East. Analysis of this aspect of diversity is further complicated by the very low share of CHI papers (and thus also authors) from Latin America, Africa, and the wider range of regions in Asia. To illustrate the imbalance in co-authorship, we visualize the prevalence of country pairings in the dataset. As shown in Figure 12, there are strong connections in the Global North, but only weak links to and within the Global South.

Overall, our results suggest that when we discuss diversity in HCI we need to have a broad perspective. It is not enough to look at gender and ethnicity, which have received the bulk of attention, but we need to also consider, for example, national and workplace differences. Opinions will differ on what kind of diversity the HCI community should strive for, and this is likely independent of the effects of diversity investigated here. As Yeager and Nafukho [113] described in their review, for individual teams more diversity does not necessarily mean better performance. But across the whole of our community, more diversity results in higher impact and thus there likely are benefits to increasing it.

How to go about increasing diversity and in which ways, however, is likely a question with many kinds of answers. As we described in Section 3, SIGCHI looks to push for more equity to achieve this goal. However, another potential direction would be to focus more on the biases and inequalities that could have resulted in the present outcomes. For example, there is little reason not to have more blind assessments, where presently author identities might be a factor. This includes awards at CHI, but also in general, as well as reviewing of articles at some journals in the field. However, once work is published there remains the question how biases play into how it is received. Aside from calls for citational justice, there is a larger discourse here around the quality of citation practices in general. Marshall et al. [62] have argued that "*in CHI papers, citation of previous work is typically a shallow, throwaway action that demonstrates little critical engagement with the work cited*." The growing number of references in CHI papers also is not indicative of an increase in this engagement [73]. It thus looks like the current review process is not sufficiently penalizing papers for poor citation practices and engagement with the breadth of prior work.

8.2 Limitations

Most limitations of our work are related to the quality and completeness of the underlying data. For example, while we manually fixed many errors in the ACM DL data, there is a risk that we missed, for example, some publications listed as papers that are not actually papers. The country and sector annotations, which we also did manually, similarly have a risk of coding and other errors. Ideally, high quality metadata for authors and affiliations should be readily available, but unfortunately this is currently not the case. We also only were able to analyze awards for CHI 2010 and later. Given that Bartneck and Hu did a similar analysis for earlier years [6], this data exists, it just was not readily available to use.

Noteworthy here is also the lack of consistent author and institutional identifiers, which was the source for many of the above issues. While the ACM has adopted ORCID for authors, this unfortunately is not retroactively and this information also is at times inconsistent. If the ACM were to also adopt ROR identifiers, this would enable a much better analysis of affiliations. For example, we only annotated country information, but the ROR lists detailed addresses, organization types, and related organizations. Thus, a more fine-grained analysis of organizational and geographical relationships in the data would be possible if this data were linked.

Our analysis of ethnicity and gender relies on automated estimation of both from author names, due to the amount of data to analyze (10,341 papers with 19,257 authors). While this is the established approach within scientometrics and other fields (see, e.g., [27, 31, 84, 88]), it does nonetheless have some issues. Where the gender estimation proved to be reliable in validation, it is the ethnicity data in particular that is problematic. There is a certain amount of irreducible ambiguity around the concept of ethnicity, but there clearly also is potential to do better. For example, improved ethnicity estimation should include global training data, combine this with more contextual information, and output ethnicity along a more universal set of classes. This requires improvements in the methods themselves, but also necessitates higher quality data from the ACM DL or other sources to provide this needed context. Furthermore, gender estimation methods currently do not sufficiently account for non-binary gender expressions, as also discussed by Keyes [51]. We note again that the prevalence of non-binary identities is overall low [82] and thus distorting effects are likely small. However, it certainly would be beneficial to include this aspect as well. In practical terms, however, it is unclear how to account for genderfluid, agender, polygender and other identities in a methodology that requires numeric or at least categorical data for analyses.

While we include a range of diversity measures, these are by no means complete, as we also noted in Section 5.1. For example, disability, author nationality, native language, sexual orientation, religion, socioeconomic status, or political beliefs might well be relevant to co-author dynamics and the work resulting from such collaborations. Yet, much of this data would be very hard to gather and in fact problematic to assemble in the first place.

Our view on impact is also limited as we only consider citations and awards. Instead, one could argue for a wider interpretation of impact that also considers economic, community, policy, social, or professional outcomes of a paper. However, while broadening what kind of impacts are considered would cater to more approaches to research, it also makes the whole notion of impact impossible to compare or measure. Within the value system of academia, citations and awards are supposed to capture the above as well. For example, work that has a noticeable impact on a community should also be garnering interests from others to follow along these lines.

Finally, we would like to make clear that none of the found relationships necessitate a causal link. Writing a longer paper or including more White co-authors will not automatically make that paper more highly cited. Our analysis should be understood as showing broader overall trends in the reception of work published at CHI. Furthermore, there is a strong temporal component in the model and just as what work has what impact has changed in the past, we can expect this to also change in the future.

9 Conclusion

The results from our analysis show that overall diversity in CHI authorship does matter. More diverse teams produced papers that get cited more and receive more awards. But not all aspects of diversity are equally relevant in terms of paper impact. In line with previous work, we identified a positive effect for ethnically diverse teams and no effect for gender-diverse teams. Furthermore, we see several biases in the reception of CHI papers, including advantages for teams located more in the USA and including more White authors.

We complemented the main analysis with follow-up ones where we (1) focus on specific diversity dimensions or subsets and their relationship to our impact measures, and (2) where we investigate alternative explanations for impact. For the former we looked at prominent labs, regionality, and the first and last authors specifically, while for the latter we investigated research areas, novelty, and teams' self-descriptions. We find a range of effects that modulate impact, such as certain research areas outperforming others in terms of citation rates.

Many questions remain or stem from this analysis. For example, global differences in paper topics and reception are an important aspect to investigate. We should also critically examine the ways we assign value to work. As we found, the award process is good at honoring papers that are better than average, but also biased and failing at actually honoring the most impactful work.

With this work we provide data to shed more light on one specific aspect of diversity in the field: how our papers are received. Our analysis does not ultimately answer where these differences come from and different kinds of interpretations are possible, though we provide some indications through our follow-up analyses (e.g., research area likely modulates paper impact). Furthermore, it is important to note that our results should not be read as suggestions for how to optimize potential impact, such as by including more American co-authors. We hope that this work will inspire further inquiry along those lines so we can alleviate unjust biases where present and improve the field of HCI overall. Presently, it provides insights that can support the larger discourse in the field around who participates and to what degree this participation is a fair process.

References

[1] Veronica Abebe, Gagik Amaryan, Marina Beshai, Ilene, Ali Ekin Gurgen, Wendy Ho, Naaji R. Hylton, Daniel Kim, Christy Lee, Carina Lewandowski, et al. 2022. Anti-racist HCI: Notes on an emerging critical technical practice. In *Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. ACM, New York, NY, 1–12. DOI: https://doi.org/10.1145/3491101.3516382

[2] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera. 2009. h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics* 3, 4 (2009), 273–289. DOI: https://doi.org/10.1016/j.joi.2009.04.001

[3] Emma Rachel Andersson, Carolina E. Hagberg, and Sara Hägg. 2021. Gender bias impacts top-merited candidates. *Frontiers in Research Metrics and Analytics* 6 (2021), 594424. DOI: https://doi.org/10.3389/frma.2021.594424

[4] Mara Balestrini, Yvonne Rogers, and Paul Marshall. 2015. Civically engaged HCI: Tensions between novelty and social impact. In *Proceedings of the 2015 British HCI Conference (British HCI '15)*. ACM, New York, NY, 35–36. DOI: https://doi.org/10.1145/2783446.2783590

[5] Simone Diniz, Junqueira Barbosa, Milene Selbach Silveira, and Isabela Gasparini. 2017. What publications metadata tell us about the evolution of a scientific community: The case of the Brazilian human–computer interaction conference series. *Scientometrics* 110, 1 (Jan. 2017), 275–300. DOI: https://doi.org/10.1007/s11192-016-2162-4

[6] Christoph Bartneck and Jun Hu. 2009. Scientometric analysis of the CHI proceedings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, 699–708. DOI: https://doi.org/10.1145/1518701.1518810

[7] Michaël Bikard, Isabel Fernandez-Mateo, and Ronak Mogra. 2025. Standing on the shoulders of (male) giants: Gender inequality and the technological impact of scientific ideas. *Administrative Science Quarterly* 70 (2025), 695–732. DOI: https://doi.org/10.1177/00018392251331957

[8] Lutz Bornmann, Alexander Tekles, Helena H. Zhang, and Fred Y. Ye. 2019. Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data. *Journal of Informetrics* 13, 4 (2019), 100979. DOI: https://doi.org/10.1016/j.joi.2019.100979

[9] George Buchanan and Dana McKay. 2022. You oughta know: Examining author geography and gender in information science. *Proceedings of the Association for Information Science and Technology* 59, 1 (2022), 32–43. DOI: https://doi.org/10.1002/pra2.602

[10] Magnus Carlsson, Henning Finseraas, Arnfinn H. Midtbøen, and Guðbjörg Linda Rafnsdóttir. 2020. Gender bias in academic recruitment? Evidence from a survey experiment in the nordic region. *European Sociological Review* 37, 3 (2020), 399–410. DOI: https://doi.org/10.1093/esr/jcaa050

[11] Stephen J. Ceci, Shulamit Kahn, and Wendy M. Williams. 2023. Exploring gender bias in six key domains of academic science: An adversarial collaboration. *Psychological Science in the Public Interest: A Journal of the American Psychological Society* 24, 1 (2023), 15–73. DOI: https://doi.org/10.1177/15291006231163179

[12] Chaomei Chen. 2012. Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology* 63, 3 (2012), 431–449. DOI: https://doi.org/10.1002/asi.21694

[13] Yiqun T. Chen, Angela D. R. Smith, Katharina Reinecke, and Alexandra To. 2023. Why, when, and from whom: Considerations for collecting and reporting race and ethnicity data in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, 1–15. DOI: https://doi.org/10.1145/3544548.3581122

[14] Marc Cheong, Kobi Leins, and Simon Coghlan. 2021. Computer science communities: Who is speaking, and who is listening to the women? Using an ethics of care to promote diverse voices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. ACM, New York, NY, 106–115. DOI: https://doi.org/10.1145/3442188.3445874

[15] Rajashekar Chintalapati, Suriyan Laohaprapanon, and Gaurav Sood. 2023. Predicting race and ethnicity from the sequence of characters in a name. arXiv:1805.02109. Retrieved from https://arxiv.org/abs/1805.02109

[16] Shruthi Sai Chivukula and Colin M. Gray. 2020. Bardzell's "feminist HCI" legacy: Analyzing citational patterns. In *Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. ACM, New York, NY, 1–8. DOI: https://doi.org/10.1145/3334480.3382936

[17] A. J. Coldman, T. Braun, and R. P. Gallagher. 1988. The classification of ethnic status using name information. *Journal of Epidemiology and Community Health* 42, 4 (1988), 390–395. DOI: https://doi.org/10.1136/jech.42.4.390

[18] Citational Justice Collective, Gabriela Molina León, Lynn Kirabo, Marisol Wong-Villacres, Naveena Karusala, Neha Kumar, Nicola Bidwell, Pedro Reynolds-Cuéllar, Pranjal Protim Borah, Radhika Garg, et al. 2021. Following the trail of citational justice: Critically examining knowledge production in HCI. In *Proceedings of the Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion)*. ACM, New York, NY, 360–363. DOI: https://doi.org/10.1145/3462204.3481732

[19] The Citational Justice Collective, Syed Ishtiaque Ahmed, Sareeta Amrute, Jeffrey Bardzell, Shaowen Bardzell, Nicola Bidwell, Tawanna Dillahunt, Sane Gaytán, Naveena Karusala, Neha Kumar, Rigoberto Lara Guzmán, et al. 2022. Citational justice and the politics of knowledge production. *Interactions* 29, 5 (Aug. 2022), 78–82. DOI: https://doi.org/10.1145/3556549

[20] António Correia, Hugo Paredes, and Benjamim Fonseca. 2018. Scientometric analysis of scientific publications in CSCW. *Scientometrics* 114, 1 (Jan. 2018), 31–89. DOI: https://doi.org/10.1007/s11192-017-2562-0

[21] Nana Kesewaa Dankwa and Claude Draude. 2021. Setting diversity at the core of HCI. In *Universal Access in Human–Computer Interaction, Design Methods and User Experience*. Margherita Antona and Constantine Stephanidis (Eds.), Springer International Publishing, Cham, 39–52.

[22] Jielan Ding, Zhesi Shen, Per Ahlgren, Tobias Jeppsson, David Minguillo, and Johan Lyhagen. 2021. The link between ethnic diversity and scientific impact: The mediating effect of novelty and audience diversity. *Scientometrics* 126, 9 (Sep. 2021), 7759–7810. DOI: https://doi.org/10.1007/s11192-021-04071-1

[23] Sheena Erete, Yolanda Rankin, and Jakita Thomas. 2023. A method to the madness: Applying an intersectional analysis of structural oppression and power in HCI and design. *ACM Transactions on Computer-Human Interaction* 30, 2, Article 24 (Apr. 2023), 45 pages. DOI: https://doi.org/10.1145/3507695

[24] Sheena Erete, Yolanda A. Rankin, and Jakita O. Thomas. 2021. I can't breathe: Reflections from black women in CSCW and HCI. *Proceedings of the ACM on Human Computer Interaction* 4, CSCW3, Article 234 (2021), 23 pages. DOI: https://doi.org/10.1145/3432933

[25] Adriana S. Vivacqua, Andrew L. Kun, Cale Passmore, Helena Mentis, Josh Andres, Kashyap Todi, Matt Jones, Luigi De Russis, Naomi Yamashita, Neha Kumar, et al. 2022. Equity talks @SIGCHI. In *Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. ACM, New York, NY, 1–3. DOI: https://doi.org/10.1145/3491101.3516397

[26] Matthew E. Falagas, Angeliki Zarkali, Drosos E. Karageorgopoulos, Vangelis Bardakas, and Michael N. Mavros. 2013. The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. *PLoS One* 8, 2 (02. 2013), 1–8. DOI: https://doi.org/10.1371/journal.pone.0049476

[27] Clemens B. Fell and Cornelius J. König. 2016. Is there a gender difference in scientific collaboration? A scientometric examination of co-authorships among industrial–organizational psychologists. *Scientometrics* 108, 1 (Jul. 2016), 113–141. DOI: https://doi.org/10.1007/s11192-016-1967-5

[28] Marianne A. Ferber and Michael Brün. 2011. The gender gap in citations: Does it persist? *Feminist Economics* 17, 1 (2011), 151–158. DOI: https://doi.org/10.1080/13545701.2010.541857

[29] Jacob G. Foster, Andrey Rzhetsky, and James A. Evans. 2015. Tradition and innovation in scientists' research strategies. *American Sociological Review* 80, 5 (2015), 875–908. DOI: https://doi.org/10.1177/0003122415601618

[30] Richard B. Freeman and Wei Huang. 2014. Collaboration: Strength in diversity. *Nature* 513, 7518 (Sep. 2014), 305–305. DOI: https://doi.org/10.1038/513305a

[31] Richard B. Freeman and Wei Huang. 2015. Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics* 33, S1 (2015), S289–S318. DOI: https://doi.org/10.1086/678973

[32] Carol Frieze and Jeria L. Quesenberry. 2013. From difference to diversity: Including women in the changing face of computing. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*. ACM, New York, NY, 445–450. DOI: https://doi.org/10.1145/2445196.2445327

[33] L. E. Gomez and Patrick Bernet. 2019. Diversity improves performance and outcomes. *Journal of the National Medical Association* 111, 4 (2019), 383–392. DOI: https://doi.org/10.1016/j.jnma.2019.01.006

[34] Aakar Gupta. 2015. Five years of IndiaHCI: A scientometric analysis. In *Proceedings of the 7th Indian Conference on Human–Computer Interaction (IndiaHCI '15)*. ACM, New York, NY, 56–61. DOI: https://doi.org/10.1145/2835966.2835973

[35] J. Andrew Harris. 2015. What's in a name? A method for extracting information about ethnicity from names. *Political Analysis* 23, 2 (2015), 212–224. DOI: https://doi.org/10.1093/pan/mpu038

[36] Nathalie Henry, Howard Goodell, Niklas Elmqvist, and Jean-Daniel Fekete. 2007. 20 years of four HCI conferences: A visual exploration. *International Journal of Human–Computer Interaction* 23, 3 (2007), 239–285. DOI: https://doi.org/10.1080/10447310701702402

[37] Julia Himmelsbach, Stephanie Schwarz, Cornelia Gerdenitsch, Beatrix Wais-Zechmann, Jan Bobeth, and Manfred Tscheligi. 2019. Do we care about diversity in human computer interaction: A comprehensive content analysis on diversity dimensions in research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing (CHI '19)*. ACM, New York, NY, 1–16. DOI: https://doi.org/10.1145/3290605.3300720

[38] Charles C. Hinnant, Besiki Stvilia, Shuheng Wu, Adam Worrall, Gary Burnett, Kathleen Burnett, Michelle M. Kazmer, and Paul F. Marty. 2012. Author-team diversity and the impact of scientific publications: Evidence from physics research at a national science lab. *Library & Information Science Research* 34, 4 (2012), 249–257. DOI: https://doi.org/10.1016/j.lisr.2012.03.001

[39] Lu Hong and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America* 101, 46 (2004), 16385–16389. DOI: https://doi.org/10.1073/pnas.0403723101

[40] Daniel B. Horn, Thomas A. Finholt, Jeremy P. Birnholtz, Dheeraj Motwani, and Swapnaa Jayaraman. 2004. Six degrees of jonathan grudin: A social network analysis of the evolution and impact of CSCW research. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, 582–591. DOI: https://doi.org/10.1145/1031607.1031707

[41] Isabelle Hupont, Songül Tolan, Pedro Frau, Lorenzo Porcaro, and Emilia Gómez. 2024. Measuring and fostering diversity in affective computing research. *IEEE Transactions on Affective Computing* 15 (2024), 63–78. DOI: https://doi.org/10.1109/TAFFC.2023.3244041

[42] Isabelle Hupont, Songül Tolan, Ana Freire, Lorenzo Porcaro, Sara Estevez, and Emilia Gómez. 2021. How diverse is the ACII community? Analysing gender, geographical and business diversity of affective computing research. In *Proceedings of the 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. DOI: https://doi.org/10.1109/ACII52823.2021.9597426
[43] Karen A. Jehn and Katerina Bezrukova. 2004. A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior* 25, 6 (2004), 703–729. DOI: https://doi.org/10.1002/job.257
[44] Eric Jensen, Nicholas Jones, Kimberly Orozco, Lauren Medina, Marc Perry, Ben Bolender, and Karen Battle. 2021. Measuring Racial and Ethnic Diversity for the 2020 Census. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html.
[45] Daeseong Jeon, Joon Mo Ahn, Juram Kim, and Changyong Lee. 2022. A doc2vec and local outlier factor approach to measuring the novelty of patents. *Technological Forecasting and Social Change* 174 (2022), 121294. DOI: https://doi.org/10.1016/j.techfore.2021.121294
[46] Daeseong Jeon, Junyoup Lee, Joon Mo Ahn, and Changyong Lee. 2023. Measuring the novelty of scientific publications: A fastText and local outlier factor approach. *Journal of Informetrics* 17, 4 (2023), 101450. DOI: https://doi.org/10.1016/j.joi.2023.101450
[47] Philipp Jugert, Marie J. Kaiser, Francesca Ialuna, and Sauro Civitillo. 2022. Researching race-ethnicity in race-mute Europe. *Infant and Child Development* 31, 1 (2022), e2260. DOI: https://doi.org/10.1002/icd.2260
[48] Annika Kaltenhauser, Johannes Schöning, Elizabeth F. Churchill, Hiroshi Ishii, Elisa D. Mekler, and Ben Shneiderman. 2025. The churns and turns of HCI: Which CHI papers make the most impact in an ever-growing sea of HCI publications. In *Proceedings of the Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. ACM, New York, NY, 1–5. DOI: https://doi.org/10.1145/3706599.3716285
[49] Joseph 'Jofish' Kaye. 2009. Some statistical analyses of CHI. In *Proceedings of the Proceedings of the Extended Abstracts of the 2009 CHI Conference* on Human Factors in Computing (CHI EA '09). ACM, New York, NY, 2585–2594. DOI: https://doi.org/10.1145/1520340.1520364
[50] William R. Kerr. 2008. Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics* 90, 3 (2008), 518–537. DOI: https://doi.org/10.1162/rest.90.3.518
[51] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human Computer Interaction* 2, CSCW, Article 88 (Nov. 2018), 22 pages. DOI: https://doi.org/10.1145/3274357
[52] Molly M. King, Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, and Jevin D. West. 2017. Men set their own cites high: Gender and self-citation across fields and over time. *Socius* 3 (2017), 2378023117738903. DOI: https://doi.org/10.1177/2378023117738903
[53] Anagha Kulkarni, Ilmi Yoon, Pleuni S. Pennings, Kazunori Okada, and Carmen Domingo. 2018. Promoting diversity in computing. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '18)*. ACM, New York, NY, 236–241. DOI: https://doi.org/10.1145/3197091.3197145
[54] Neha Kumar and Naveena Karusala. 2021. Braving citational justice in Human–Computer interaction. In *Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. ACM, New York, NY, 1–9. DOI: https://doi.org/10.1145/3411763.3450389
[55] Terri R. Kurtzberg. 2005. Feeling creative, being creative: An empirical study of diversity and creativity in teams. *Creativity Research Journal* 17, 1 (2005), 51–65. DOI:https://doi.org/10.1207/s15326934crj1701_5
[56] Ellie Crookes, Kylie M. Smith, and Patrick A. Crookes. 2013. Measuring research 'impact' for academic promotion: Issues from the literature. *Journal of Higher Education Policy and Management* 35, 4 (2013), 410–420. DOI: https://doi.org/10.1080/1360080X.2013.812173
[57] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for HCI toolkit research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 1–17. DOI: https://doi.org/10.1145/3173574.3173610
[58] Weihua Li, Tomaso Aste, Fabio Caccioli, and Giacomo Livan. 2019. Early coauthorship with top scientists predicts success in academic careers. *Nature Communications* 10, 1 (Nov. 2019), 5170. DOI: https://doi.org/10.1038/s41467-019-13130-4
[59] Chien Hsiang Liao. 2011. How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics Scientometrics* 86, 3 (2011), 747–761. DOI: https://doi.org/10.1007/s11192-010-0309-2
[60] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3411764.3445488

[61] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994–2013: Mapping two decades of intellectual progress through co-word analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, 3553–3562. DOI: https://doi.org/10.1145/2556288.2556969
[62] Joe Marshall, Conor Linehan, Jocelyn Spence, and Stefan Rennick Egglestone. 2017. Throwaway citation of prior work creates risk of bad HCI research. In *Proceedings of the Extended Abstracts of the 2017 CHI Conference on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, 827–836. DOI: https://doi.org/10.1145/3027063.3052751
[63] Dana McKay, Huiwen Zhang, and George Buchanan. 2022. Who Am I, and who are you, and who are We? A scientometric analysis of gender and geography in HCI. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, 1–19. DOI: https://doi.org/10.1145/3491102.3502106
[64] J. E. McLaughlin, G. W. McLaughlin, J. S. McLaughlin, and C. Y. White. 2016. Using simpson's diversity index to examine multidimensional models of diversity in health professions education. *International Journal of Medical Education* 7 (2016), 1–5. DOI: https://doi.org/10.5116/ijme.565e.1112
[65] Lokman I. Meho and Yvonne Rogers. 2008. Citation counting, citation ranking, and h-index of human–computer interaction researchers: A comparison of scopus and web of science. *Journal of the American Society for Information Science and Technology* 59, 11 (2008), 1711–1726. DOI: https://doi.org/10.1002/asi.20874
[66] Robert K. Merton. 1968. *Social Theory and Social Structure*. Free Press.
[67] Omar Mubin, Dhaval Tejlavwala, Mudassar Arsalan, Muneeb Ahmad, and Simeon Simoff. 2018. An assessment into the characteristics of award winning papers at CHI. *Scientometrics* 116, 2 (Aug. 2018), 1181–1201. DOI: https://doi.org/10.1007/s11192-018-2778-7
[68] Mathias Wullum Nielsen. 2017. Gender and citation impact in management research. *Journal of Informetrics* 11, 4 (2017), 1213–1228. DOI: https://doi.org/10.1016/j.joi.2017.09.005
[69] Mathias Wullum Nielsen, Sharla Alegria, Love Börjeson, Henry Etzkowitz, Holly J. Falk-Krzesinski, Aparna Joshi, Erin Leahey, Laurel Smith-Doerr, Anita Williams Woolley, and Londa Schiebinger. 2017. Gender diversity leads to better science. *Proceedings of the National Academy of Sciences* 114, 8 (2017), 1740–1742. DOI: https://doi.org/10.1073/pnas.1700616114
[70] Anna Offenwanger, Alan John Milligan, Minsuk Chang, Julia Bullard, and Dongwook Yoon. 2021. Diagnosing bias in the gender representation of HCI research participants: How it happens and where We are. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, 1–18. DOI: https://doi.org/10.1145/3411764.3445383
[71] Amy Ogan, Frederick van Amstel, Gabriela Molina León, Juan Fernando Maestre, Kristin Williams, Nicola J. Bidwell, Pedro Reynolds-Cuéllar, Saiph Savage, Sushil Oswal, and Vishal Sharma. 2023. Why do We need to learn about citational practices? Recognizing knowledge production from the global souths and Beyond. *XRDS: Crossroads, The ACM Magazine for Students* 29, 3 (Apr. 2023), 12–17. DOI: https://doi.org/10.1145/3589256
[72] Ihudiya Finda Ogbonnaya-Ogburu, Angela D. R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical race theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–16. DOI: https://doi.org/10.1145/3313831.3376392
[73] Jonas Oppenlaender. 2025. Past, present, and future of citation practices in HCI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, USA, 1–16. DOI: https://doi.org/10.1145/3706598.3713556
[74] Antti Oulasvirta and Kasper Hornbæk. 2016. HCI research as problem-solving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 4956–4967. DOI: https://doi.org/10.1145/2858036.2858283
[75] Pierre Pelletier and Kevin Wirtz. 2022. Novelpy: A Python package to measure novelty and disruptiveness of bibliometric and patent data. arXiv:2211.10346. Retrieved from https://arxiv.org/abs/2211.10346
[76] Teresa Penfield, Matthew J. Baker, Rosa Scoble, and Michael C. Wykes. 2013. Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation* 23, 1 (2013), 21–32. DOI: https://doi.org/10.1093/reseval/rvt021
[77] Alexander Michael Petersen, Santo Fortunato, Raj K. Pan, Kimmo Kaski, Orion Penner, Armando Rungi, Massimo Riccaboni, H. Eugene Stanley, and Fabio Pammolli. 2014. Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences* 111, 43 (2014), 15316–15321. DOI: https://doi.org/10.1073/pnas.1323111111
[78] Henning Pohl and Aske Mottelson. 2019. How We guide, write, and cite at CHI. In *Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, 1–11. DOI: https://doi.org/10.1145/3290607.3310429
[79] Oleksandra Poquet, Srecko Joksimovic, and Pernille Brams. 2024. The role of gender in citation practices of learning analytics research. In *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)*. ACM, New York, NY, 807–813. DOI: https://doi.org/10.1145/3636555.3636878
[80] James Ravenscroft, Maria Liakata, Amanda Clare, and Daniel Duma. 2017. Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PLoS One* 12, 3 (2017), 1–21. DOI: https://doi.org/10.1371/journal.pone.0173152

[81] Jing Ren, Yajie Shi, Adrian Shatte, Xiangjie Kong, and Feng Xia. 2022. The significance and impact of winning an academic award: A study of early career academics. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (JCDL '22)*. ACM, New York, NY, 1–11. DOI: https://doi.org/10.1145/3529372.3530913
[82] Christina Richards, Walter Pierre Bouman, Leighton Seal, Meg John Barker, Timo O. Nieder, and Guy T'Sjoen. 2016. Non-binary or genderqueer genders. *International Review of Psychiatry* 28, 1 (2016), 95–102. DOI: https://doi.org/10.3109/09540261.2015.1106446
[83] Barbara J. Robson and Aurélie Mousquès. 2016. Can we predict citation counts of environmental modelling papers? Fourteen bibliographic and categorical variables predict less than 30% of the variability in citation counts. *Environmental Modelling & Software* 75 (2016), 94–104. DOI: https://doi.org/10.1016/j.envsoft.2015.10.007
[84] Philip Roe, Grant Lewison, and Richard Webber. 2014. The sex and ethnicity or national origins of researchers in astronomy and oncology in four countries, 2006–2007 and 2011–2012. *Scientometrics* 100, 1 (Jul. 2014), 287–296. DOI: https://doi.org/10.1007/s11192-013-1223-1
[85] Steven G. Rogelberg and Steven M. Rumery. 1996. Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small Group Research* 27, 1 (1996), 79–90. DOI: https://doi.org/10.1177/1046496496271004
[86] Lucía Santamaria and Helena Mihaljević. 2018. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* 4 (2018), e156. DOI: https://doi.org/10.7717/peerj-cs.156
[87] Cathy C. Sarli, Ellen K. Dubinsky, and Kristi L. Holmes. 2010. Beyond citation analysis: A model for assessment of research impact. *Journal of the Medical Library Association* 98, 1 (Jan. 2010), 17–23.
[88] Ali Sarvghad, Rolando Franqui-Nadal, Rebecca Reznik-Zellen, Ria Chawla, and Narges Mahyar. 2023. Scientometric analysis of interdisciplinary collaboration and gender trends in 30 years of IEEE VIS publications. *IEEE Transactions on Visualization and Computer Graphics* 29, 7 (2023), 3340–3353. DOI: https://doi.org/10.1109/TVCG.2022.3158236
[89] Jukka Savolainen, Patrick J. Casey, Justin P. McBrayer, and Patricia Nayna Schwerdtle. 2023. Positionality and its problems: Questioning the value of reflexivity statements in research. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 18, 6 (2023), 1331–1338. DOI: https://doi.org/10.1177/17456916221144988
[90] Michael Schaerer, Christilene Du Plessis, My Hoang Bao Nguyen, Robbie C. M. van Aert, Leo Tiokhin, Daniël Lakens, Elena Giulia Clemente, Thomas Pfeiffer, Anna Dreber, Magnus Johannesson, et al. 2023. On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions. *Organizational Behavior and Human Decision Processes* 179 (2023), 104280. DOI: https://doi.org/10.1016/j.obhdp.2023.104280
[91] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 5412–5427. DOI: https://doi.org/10.1145/3025453.3025766
[92] D. Schleuter, M. Daufresne, F. Massol, and C. Argillier. 2010. A user's guide to functional diversity indices. *Ecological Monographs* 80, 3 (2010), 469–484. DOI: https://doi.org/10.1890/08-2225.1
[93] Katie Seaborn. 2023. Interacting with masculinities: A scoping review. In *Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. ACM, New York, NY, USA, 1–12. DOI: https://doi.org/10.1145/3544549.3585770
[94] Sotaro Shibayama, Deyun Yin, and Kuniko Matsumoto. 2021. Measuring novelty in science with word embedding. *PLoS One* 16, 7 (2021), 1–16. DOI: https://doi.org/10.1371/journal.pone.0254034
[95] Patrick Simon and Victor Piché. 2012. Accounting for ethnic and racial diversity: The challenge of enumeration. *Ethnic and Racial Studies* 35, 8 (2012), 1357–1365. DOI: https://doi.org/10.1080/01419870.2011.634508
[96] E. H. Simpson. 1949. Measurement of diversity. *Nature* 163, 4148 (Apr. 1949), 688–688. DOI: https://doi.org/10.1038/163688a0
[97] Chakresh Kumar Singh, Liubov Tupikina, Fabrice Lécuyer, Michele Starnini, and Marc Santolini. 2024. Charting mobility patterns in the scientific knowledge landscape. *EPJ Data Science* 13, 1 (Feb. 2024), 12. DOI: https://doi.org/10.1140/epjds/s13688-024-00451-8
[98] Gaurav Sood. 2017. Florida Voter Registration Data (2017 and 2022). DOI: https://doi.org/10.7910/DVN/UBIG3F
[99] Chris Stephenson. 2022. All the Voices: Doing the Work to Improve Diversity, Equity, and Inclusion throughout ACM. Technical Report. Retrieved from https://www.sigcse.org/about/news/2022-09-All-The-voices.html.
[100] Chris Stephenson. 2022. Listening within: ACM employees and leaders articulate the need for a more diverse, equitable, and inclusive organization. *ACM SIGCSE Bulletin* 54, 4 (Dec. 2022), 5–7. DOI: https://doi.org/10.1145/3577871.3577875
[101] Christian Sturm, Alice Oh, Sebastian Linxen, Jose Abdelnour Nocera, Susan Dray, and Katharina Reinecke. 2015. How WEIRD is HCI? Extending HCI principles to other countries and cultures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, 2425–2428. DOI: https://doi.org/10.1145/2702613.2702656

[102] Justin Sulik, Bahador Bahrami, and Ophelia Deroy. 2022. The diversity gap: When diversity matters for knowledge. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 17, 3 (2022), 752–767. DOI: https://doi.org/10.1177/17456916211006070

[103] Harold Thimbleby. 2004. Supporting diverse HCI research. In *Proceedings British Computer Society HCI Conference*, Vol. 2, Research Press International, 125–128.

[104] Robert J. W. Tijssen. 2003. Scoreboards of research excellence. *Research Evaluation* 12, 2 (2003), 91–103. DOI: https://doi.org/10.3152/147154403781776690

[105] Haining Wang. 2024. A content-based novelty measure for scholarly publications: A proof of concept. In *Wisdom, Well-Being, Win-Win*. Isaac Sserwanga, Hideo Joho, Jie Ma, Preben Hansen, Dan Wu, Masanori Koizumi, and Anne J. Gilliland (Eds.), Springer Nature Switzerland, Cham, 409–420.

[106] Jie Wang, Grand H.-L. Cheng, Tingting Chen, and Kwok Leung. 2019. Team creativity/innovation in culturally diverse teams: A meta-analysis. *Journal of Organizational Behavior* 40, 6 (2019), 693–708. DOI: https://doi.org/10.1002/job.2362

[107] Jian Wang, Reinhilde Veugelers, and Paula Stephan. 2017. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* 46, 8 (2017), 1416–1436. DOI: https://doi.org/10.1016/j.respol.2017.06.006

[108] Lucy Lu Wang, Kelly Mack, Emma J. McDonnell, Dhruv Jain, Leah Findlater, and Jon E. Froehlich. 2021. A bibliometric analysis of citation diversity in accessibility and HCI research. In *Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. ACM, New York, NY, 1–7. DOI: https://doi.org/10.1145/3411763.3451618

[109] Yong Wang, Conglei Shi, Liangyue Li, Hanghang Tong, and Huamin Qu. 2018. Visualizing research impact through citation data. *ACM Transactions on Interactive Intelligent Systems* 8, 1, Article 5 (Mar. 2018), 24 pages.DOI: https://doi.org/https://doi.org/10.1145/3132744

[110] Richard Watermeyer. 2014. Issues in the articulation of 'impact': The responses of UK academics to 'impact' as a new measure of research assessment. *Studies in Higher Education* 39, 2 (2014), 359–377. DOI: https://doi.org/10.1080/03075079.2012.709490

[111] Wendy M. Williams and Stephen J. Ceci. 2015. National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences* 112, 17 (2015), 5360–5365. DOI: https://doi.org/10.1073/pnas.1418878112

[112] Fangzhou Xie. 2022. Rethnicity: An R package for predicting ethnicity from names. *SoftwareX* 17 (2022), 100965. DOI: https://doi.org/10.1016/j.softx.2021.100965

[113] Katherine L. Yeager and Fredrick M. Nafukho. 2012. Developing diverse teams to improve performance in the organizational setting. *European Journal of Training and Development* 36, 4 (Jan. 2012), 388–408. DOI: https://doi.org/10.1108/03090591211220320

[114] Deyun Yin, Zhao Wu, Kazuki Yokota, Kuniko Matsumoto, and Sotaro Shibayama. 2023. Identify novel elements of knowledge with word embedding. *PLoS One* 18, 6 (2023), 1–16. DOI: https://doi.org/10.1371/journal.pone.0284567

[115] Guiyang Zhang and Chaoying Tang. 2018. How R&D partner diversity influences innovation performance: An empirical study in the nano-biopharmaceutical field. *Scientometrics* 116, 3 (Sep. 2018), 1487–1512. DOI: https://doi.org/10.1007/s11192-018-2831-6

Table A1. Descriptive Statistics for Numeric Data Available per Paper

Attribute	n	Min	Q1	Mean	Median	Q3	Max	SD	Distribution
Academic Age Mean	10,341	0.00	4.80	7.86	7.43	10.50	36.00	4.30	
Academic Age Std.	10,341	0.00	3.54	5.85	5.61	7.99	23.00	3.40	
American Bias	10,340	0.00	0.00	0.50	0.50	1.00	1.00	0.47	
Country Diversity	10,340	0.04	0.25	0.39	0.33	0.50	1.00	0.22	
Whiteness	10,341	0.00	0.17	0.30	0.31	0.43	0.82	0.18	
Ethnic Diversity	10,341	0.00	0.26	0.47	0.48	0.66	1.31	0.27	
Gender Bias	10,341	−1.00	0.00	0.34	0.38	0.95	1.00	0.57	
Gender Diversity	10,341	0.00	0.00	0.28	0.38	0.48	0.96	0.23	
Sector Bias	10,340	−1.00	0.33	0.59	1.00	1.00	1.00	0.66	
Sector Diversity	10,340	0.00	0.00	0.11	0.00	0.24	0.67	0.19	
Citations Per Year	9,462	0.00	2.79	9.33	5.76	11.07	267.17	13.25	
Page Count	10,341	1.00	10.00	11.12	10.00	13.00	52.00	4.23	
Reference Count	10,341	0.00	23.00	44.78	38.00	60.00	496.00	32.01	
Year	10,341	1982	2010	2013	2016	2020	2023	9	

Q1 and Q3 are the 25% and 75% percentiles.

Table A2. Descriptive Statistics for Categorical Data Available per Paper or Author

Attribute	Per	Values
Award	Paper	None (6,369), NA (2,388), Honorable Mention (1,256), Best Paper (328)
Single Author	Paper	No (9,846), Yes (495)
Country	Author	USA (19,971), UK (5,138), Canada (2,916), Germany (2,865), China (1,430), Australia (1,050), Japan (1,029), Korea (962), France (890), Denmark (650), Finland (623), Sweden (560), Netherlands (527), Taiwan (499), Switzerland (433), Singapore (319), India (304), Austria (295), Ireland (188), Israel (188), New Zealand (158), Belgium (151), Italy (141), Portugal (141), Spain (80), Pakistan (66), Brazil (56), Norway (48), Bangladesh (42), Turkey (38), Greece (30), Qatar (30), South Africa (21), Romania (20), Luxembourg (18), Poland (16), Philippines (15), Ecuador (14), United Arab Emirates (14), Argentina (13), Czech (13), Kenya (13), Malaysia (13), Mexico (11), Egypt (10), Slovenia (9), Namibia (8), Tanzania (8), Russia (7), Cyprus (6), Saudi Arabia (6), Colombia (5), Ghana (5), Chile (4), Ivory Coast (4), Paraguay (4), Rwanda (4), Estonia (3), Lebanon (3), Malta (3), Mongolia (3), Vietnam (3), Bulgaria (2), Costa Rica (2), Ethiopia (2), Jamaica (2), Angola (1), Bahrain (1), Barbados (1), Bhutan (1), Hungary (1), Iceland (1), Indonesia (1), Iran (1), Kazakhstan (1), Serbia and Montenegro (1), Sri Lanka (1), Thailand (1), Uganda (1), Ukraine (1), NA (1)
Gender	Author	Male (27,525), Female (13,315), Unknown (832)
Race	Author	White (18,387), Asian (13,629), Black (3,922), Other (3,385), Hispanic (2,349)
Sector	Author	Academia (32,941), Industry (6,867), Institute (692), Academia & Industry (691), Academia & Institute (475), Industry & Institute (6), NA (1)

Received 28 February 2024; revised 18 September 2025; accepted 21 September 2025