Gender-Competence Associations in Computer Science Discourse: A Longitudinal Analysis

Lauren Chen
Syosset High School
laurenkchen@gmail.com

Abstract

Women remain underrepresented in computer science, yet little is known about how gendered associations within the field have changed over time. This study examines longitudinal patterns in gender–competence associations across academic and practitioner discourse from 2010–2025 using 210 Word2Vec models trained on 9.6 billion tokens from arXiv computer science papers as well as Stack Overflow posts and comments. Female-competence associations were measured using relative norm distance (RND) between female, male, and competence wordsets. In academic discourse, relative female-competence associations decreased significantly over time ($\tau = -0.717, p < 0.001$), reflecting a shift towards stronger relative-male competence association. In contrast, practitioner discourse showed no significant longitudinal trend ($\tau = -0.250, p = 0.192$). Academic gender-competence trends were significantly associated with women's representation in the U.S. computer science workforce whereas practitioner trends were not. These findings demonstrate that gendered linguistic associations have not evolved uniformly across computer science and highlight the importance of discourse context when evaluating longitudinal changes in gendered language.

1 Introduction

The underrepresentation of women in computer science is a long-standing issue which has persisted since the 1980s (Abbate, 2012). Women make up only 27% of computer scientists (National Science Board, 2020), are core developers on 5% of open source software projects (Trinkenreich et al., 2022), and author only 16% of computer science research papers (Frachtenberg and Kaner, 2022).

Existing work on gender biases in computer science has examined women's representation through citation disparities (Nakajima et al., 2024), and women's participation in open-source development (Trinkenreich et al., 2022). Qualitative studies provide evidence that gender biases persist (DuBow et al., 2017), while research using the Implicit Association Test (IAT) has provided evidence of implicit gendered associations in software development (Wang and Redmiles, 2019). However, less work has examined how gender and competence are associated within computer science discourse or how these linguistic associations have evolved over time.

Word embeddings have been found to reveal social biases, with men being linked to computer programmers and women to homemakers in general corpora (Bolukbasi et al., 2016). Caliskan et al. (2017) demonstrated that semantic associations encoded in word embeddings reflect human implicit bias, finding through the Word Embedding Association Test (WEAT) that female names are more associated with family than career compared to male names. Caliskan et al. (2022) further identified a male default with 77% of frequently used words being more associated with males than with females in word embeddings. Garg et al. (2018) demonstrated that word embeddings can be used to trace gender and ethnic associations longitudinally by comparing relative norm distance (RND) across embeddings trained on text from different time periods.

Embedding-based approaches have been used to quantify gender associations within various domains including Kenyan politics (Pair et al., 2021) and coming-of-age literature (Koch et al., 2025). Academic publications and practitioner discussions represent two distinct contexts of computing discourse, each serving different purposes in the exchange of technical knowledge. However, previous work has not used embedding-based approaches to longitudinally quantify gender-competence associations and compare their trajectory across contexts, particularly within computer science.

Based on the findings of previous word embed-

Proceedings of the 7th Workshop on Gender Bias in Natural Language Processing (GeBNLP), pages 1–6
November 9, 2026 ©2026 Association for Computational Linguistics

ding research, the present study seeks to answer the following questions: **(RQ1)** How have gender-competence associations in the language of computer science changed from 2010-2025? **(RQ2)** How do trends differ across academic and practitioner computing discourse? **(RQ3)** How are these trends associated with women's representation in the U.S. computer science workforce?

2 Methods

2.1 Data Collection

Practitioner data, including questions, answers, and comments from 2010–2025 was collected from the Stack Overflow data dump. Deleted posts were retained to prevent selection bias introduced by moderation. After deduplication, the final corpus amounted to 89,679,949 posts and 60,371,716 comments. Full-text academic research articles from the same period were collected from arXiv's public Amazon S3 source repository, amounting to a total of 830,538 articles from `cs.*` categories.

Annual estimates of women's representation in the U.S. computer science workforce were gathered through the Integrated Public Use Microdata (IPUMS) American Community Survey (ACS). Computer science occupations were identified using OCC2010 codes 1000–1100, and estimates were calculated using the IPUMS person-level sampling weight (PERWT). U.S. workforce data is used due to its accessibility across the study period, and these estimates provide a U.S.-specific benchmark for comparison with the global textual corpora.

2.2 Preprocessing

Stack Overflow preprocessing removed HTML tags, Code blocks, URLs, DOIs, camelCase tokens, hex strings, non-ASCII characters, and mixed alphanumeric tokens. For arXiv articles, LaTeX tags, references, and bibliographies were removed. Both corpora were lowercased, contractions were expanded, punctuation and tokens containing it were removed, and text was tokenized using NLTK's `word_tokenize`. After preprocessing, Stack Overflow and arXiv corpora amounted to 6.7 billion and 2.9 billion tokens respectively.

2.3 Model Training

For each year of data within each corpus, texts were randomly partitioned into non–overlapping 30-million-token slices. A Word2Vec model was trained for each of these slices using the Gensim library with parameters `vector_size = 300, min_count = 5, epochs = 5, seed = 42, workers = 6, window = 10, sg = 0`. Residual slices with token counts less than 15 million were discarded. A total of 210 models were trained.

2.4 Analysis

For each model, gender–competence association was quantified using relative norm distance (RND) from the Word Embedding Fairness Evaluation (WEFE) library. Following Garg et al. (2018), female, male, and competence word lists were used. Words that were not present in all models were excluded to ensure consistency across years. Final wordlists consisted of: female (she, her, mother, female, sister, women), male (he, him, father, male, brother, men), and competence (adaptable, intuitive, analytical, sage, smart, clever, logical, intelligent, apt, wise). RND measures the difference in distance between competence–word embeddings and the embeddings of female and male target groups. In this case, positive RND scores indicate a stronger association of women with competence.

Temporal trends in annual mean RND scores were assessed separately for each corpus using the Mann–Kendall test (`pymannkendal`). A one–sided Mann–Whitney U test (`scipy.stats.mannwhitneyu`) was used to test whether annual mean RND scores were greater in Stack Overflow than in arXiv. Spearman's rank correlation (`scipy.stats.spearmanr`) was used to assess associations between annual mean RND scores and the percentage of women in the U.S. computer science workforce. Nonparametric tests were used because data were not assumed to be normally distributed.

3 Results

3.1 Female–competence association, 2010–2025

Academic female-competence association decreased from 2010 to 2025 ($\tau = -0.717, p < 0.001$; Figure 1), shifting from a positive association in 2010 to a predominantly negative association after 2015. Conversely, practitioner female-competence association showed no significant change from 2010 to 2025 ($\tau = -0.250, p = 0.192$; Figure 1). Academic discourse exhibited lower relative female–competence associations than practitioner discourse ($p = 0.011$).

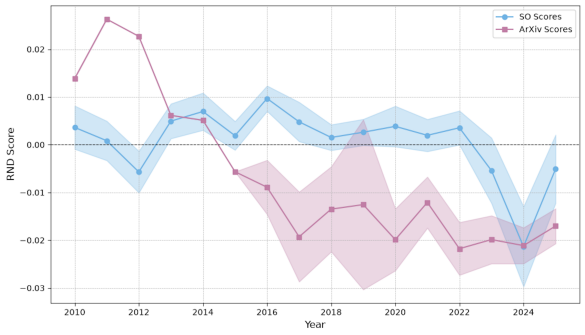


Figure 1: Practitioner vs Academic Associations from 2010–2025. Shading indicates standard error for years where multiple models were available.

	SO		arXiv	
Wordlist / ablation	τ	p	τ	p
Full wordlist (baseline)	$-.250$	.192	$-.717$	$< .001$
Ablations that flip trend classification:				
mother	$-.133$	.500	$\mathbf{-.350}$	**.065**
sister	$\mathbf{-.467}$	**.013**	$-.533$	.005
him	$\mathbf{-.383}$	**.043**	$-.767$	$< .001$
intelligent	$\mathbf{-.417}$	**.027**	$-.717$	$< .001$
Remaining 18 ablations: all consistent with baseline (SO: no trend; arXiv: decreasing, $p < .01$)				

Table 1: Summary of leave-one-word-out ablation results (22 variants) for Stack Overflow (SO) and arXiv. Full results in Appendix A.

3.2 Female–competence association with workforce representation

Academic female–competence associations were positively correlated with the percentage of women in the U.S. computer science workforce over time ($\rho = 0.775, p < 0.001$; Figure 2a). Contrastingly, practitioner associations were not significantly correlated with the percentage of women in computer science over time ($\rho = -0.061, p = 0.8298$; Figure 2b).

3.3 Robustness of Academic and Practitioner Trends

A leave-one-out word ablation was conducted to assess robustness across male, female, and competence terms (22 variants; Table 1). The decreasing academic trend was robust across 21 of 22 ablations, with the sole exception being the removal of "mother" ($\tau = -0.350, p = 0.065$). Of the practitioner ablations, 19 of 22 trials continued to show no significant trend. Three ablations, namely the removal of "sister," "him," and "intelligent", resulted in significantly decreasing female-competence association trends (τ = -0.467, -0.383, -0.417, respectively; $p < 0.05$). Overall, the academic trend was largely robust to individual word removal, while practitioner results showed some sensitivity to specific terms.

4 Discussion

4.1 Shifts in gendered associations

Results indicate that gender–competence associations have not evolved uniformly across computer science discourse. Female-competence association significantly decreased in academic discourse from 2010–2025, whereas practitioner discourse showed no significant longitudinal trend. This decline in academic contexts suggests a shift towards stronger relative male-competence association, consistent with prior work which has identified continued bias against women in computer science (Wang and Redmiles, 2019).

The decreasing association of women with competence in academic contexts is particularly relevant given evidence that gender stereotypes can affect women's sense of belonging and participation in computer science (Master et al., 2016). While the current study doesn't measure stereotype threat, gender-competence associations in language may represent one component of the broader cultural environment in which such stereotypes operate.

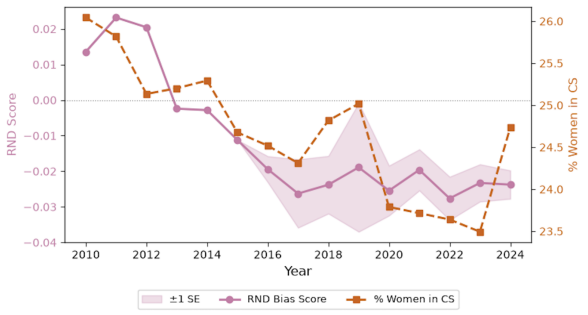
4.2 Corpora-specific Contexts of Gender-Competence Associations

To contextualize quantitative trends, qualitative examples from each corpus were explored to illustrate the linguistic contexts in which gender-competence associations occur (Table 2). These examples are intended to illustrate potential contexts rather than provide a representative sample of either corpus.

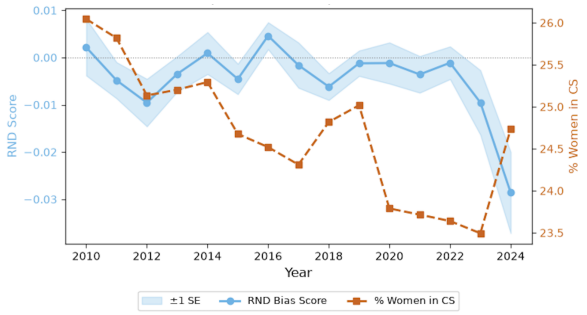
The arXiv examples (Table 2) illustrate that gender-competence associations can emerge when authors discuss or critique gender stereotypes rather than endorse them. In contrast, selected Stack Overflow examples contain direct evaluations of individuals' technical competence. Together, these examples show how gender-competence associations can arise in different linguistic contexts across academic and practitioner discourse, emphasizing the importance of considering corpus-specific contexts when interpreting these measurements.

4.3 Connection to Workforce Trends

Academic gender-competence associations were significantly correlated with women's represen-



(a) arXiv association vs. workforce representation.



(b) Stack Overflow association vs. workforce representation.

Figure 2: Corpus associations vs. workforce trends from 2010–2025. Shading indicates standard error for years where multiple models were available.

Corpus	Year	Example
Stack Overflow	2013	"she is very smart and can read code"
Stack Overflow	2019	"she is a smart engineer"
Stack Overflow	2022	"she used the below which is smart to save some lines"
arXiv	2016	"experiments using name and gender variations on resumes have found that both male and female faculty members tend to rate male applicants as more competent, more hireable, and worthy of more mentoring than female applicants"
arXiv	2021	"whereas hostile work norm involves assuming women incompetent or talking over them in formal meetups, benevolent work practice includes assigning women more mundane or simpler assignments"
arXiv	2024	"women are often seen as intrinsically less capable than men in their companies"

Table 2: Examples of gender-competence language in the Stack Overflow and arXiv corpora.

tation in the U.S. computer science workforce, whereas practitioner associations were not. This matter is of interest because academic researchers constitute only a subset of the broader computing workforce, yet academic rather than practitioner discourse is associated with changes in women's workforce representation. However, this analysis cannot establish whether linguistic change influences workforce representation, demographic change influences language, or both reflect other factors.

5 Conclusion

Relative female-competence linguistic associations decreased significantly in academic computer science discourse from 2010–2025, while practitioner discourse showed no significant trend. Academic but not practitioner associations were correlated with women's representation in the U.S. computer science workforce, highlighting differences in gendered linguistic trends across contexts. Continued examination of gendered language across computing communities is therefore important for understanding how linguistic representations of women evolve alongside broader efforts towards gender equity.

Limitations

This study measures linguistic associations in written text rather than the attitudes or lived experiences of women in computer science. Previous research has identified limitations in using word embeddings to measure social bias. Gonen and Goldberg (2019) demonstrated that reductions in gender bias according to embedding-based metrics may leave underlying gender-related structure intact, while Van Loon et al. (2022) showed that word-frequency differences can influence relationships between embedding-based bias measures and external measures of prejudice. Consequently, changes in RND scores over time may partly reflect shifts in word frequencies rather than changes in underlying associations. These findings highlight the importance of interpreting RND scores as measures of distributional association. Domain-specific polysemy may affect word representations as words like "mother" can have computing-specific meanings that differ from their usage as general terms. In addition, arXiv and Stack Overflow capture only subsets of academic and practitioner discourse and may not represent these communities as a whole. These textual corpora are global, whereas workforce representation is measured using U.S. data, so associations with workforce demograph-

ics shouldn't be interpreted as direct demographic characteristics of the corpus populations. Finally, this study's methodology focuses on a gender binary, while future research should examine linguistic associations in relation to a broader range of gender minorities to ensure that analyses of equity in computer science remain inclusive.

References

Janet Abbate. 2012. *Recoding gender: Women's changing participation in computing*. Mit Press.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.

Aylin Caliskan, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe, and Mahzarin R Banaji. 2022. Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 156–170.

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Wendy DuBow, Alexis Kaminsky, and Joanna Weidler-Lewis. 2017. Multiple factors converge to influence women's persistence in computing: A qualitative analysis. *Computing in Science & Engineering*, 19(3):30–39.

Eitan Frachtenberg and Rhody D Kaner. 2022. Underrepresentation of women in computer systems research. *Plos one*, 17(4):e0266439.

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644.

Hila Gonen and Yoav Goldberg. 2019. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 609–614, Minneapolis, Minnesota. Association for Computational Linguistics.

Mary Kate Koch, Haily Merritt, Misha N Inniss-Thompson, Kathleen C McCormick, and Jane Mendle. 2025. Are you there, god? it's me, gender bias in 100 years of coming-of-age novels. *Journal of Research on Adolescence*, 35(2):e70036.

Allison Master, Sapna Cheryan, and Andrew N Meltzoff. 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology*, 108(3):424.

Kazuki Nakajima, Yuya Sasaki, Sohei Tokuno, and George Fletcher. 2024. Quantifying gendered citation imbalance in computer science conferences. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1011–1022.

National Science Board. 2020. The state of u.s. science and engineering 2020. Technical Report NSB-2020-1, National Center for Science and Engineering Statistics (NCSES), National Science Foundation.

Emma Pair, Nikitha Vicas, Ann M Weber, Valerie Meausoone, James Zou, Amos Njuguna, and Gary L Darmstadt. 2021. Quantification of gender bias and sentiment toward political leaders over 20 years of kenyan news using natural language processing. *Frontiers in Psychology*, 12:712646.

Bianca Trinkenreich, Igor Wiese, Anita Sarma, Marco Gerosa, and Igor Steinmacher. 2022. Women's participation in open source software: A survey of the literature. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31(4):1–37.

Austin Van Loon, Salvatore Giorgi, Robb Willer, and Johannes Eichstaedt. 2022. Negative associations in word embeddings predict anti-black bias across regions–but only via name frequency. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1419–1424.

Yi Wang and David Redmiles. 2019. Implicit gender biases in professional software development: An empirical study. In *2019 IEEE/ACM 41st international conference on software engineering: Software engineering in society (ICSE-SEIS)*, pages 1–10. IEEE.

A Leave-One-Out Ablation Results

	Stack Overflow		arXiv	
Ablation	τ	p	τ	p
Full (no ablation)	-0.250	.192	-0.717	$< .001$
Female: she	-0.233	.224	-0.550	.003
Female: her	-0.217	.260	-0.633	$< .001$
Female: mother	-0.133	.500	$\mathbf{-0.350}$	**.065**
Female: female	-0.233	.224	-0.800	$< .001$
Female: sister	$\mathbf{-0.467}$	**.013**	-0.533	.005
Female: women	-0.117	.558	-0.717	$< .001$
Male: he	-0.250	.192	-0.600	.001
Male: him	$\mathbf{-0.383}$	**.043**	-0.767	$< .001$
Male: father	-0.233	.224	-0.717	$< .001$
Male: male	-0.317	.096	-0.583	.002
Male: brother	0.050	.822	-0.717	$< .001$
Male: men	-0.367	.053	-0.700	$< .001$
Attribute: adaptable	-0.367	.053	-0.700	$< .001$
Attribute: intuitive	-0.217	.260	-0.767	$< .001$
Attribute: analytical	-0.283	.137	-0.667	$< .001$
Attribute: sage	-0.233	.224	-0.667	$< .001$
Attribute: smart	-0.233	.224	-0.717	$< .001$
Attribute: clever	-0.250	.192	-0.733	$< .001$
Attribute: logical	-0.283	.137	-0.733	$< .001$
Attribute: intelligent	$\mathbf{-0.417}$	**.027**	-0.717	$< .001$
Attribute: apt	-0.067	.753	-0.650	$< .001$
Attribute: wise	-0.317	.096	-0.683	$< .001$

Table 3: Leave-one-out ablation results for Stack Overflow and arXiv. Bold values represent ablations where trends differed from the full word-list result for the same corpus.