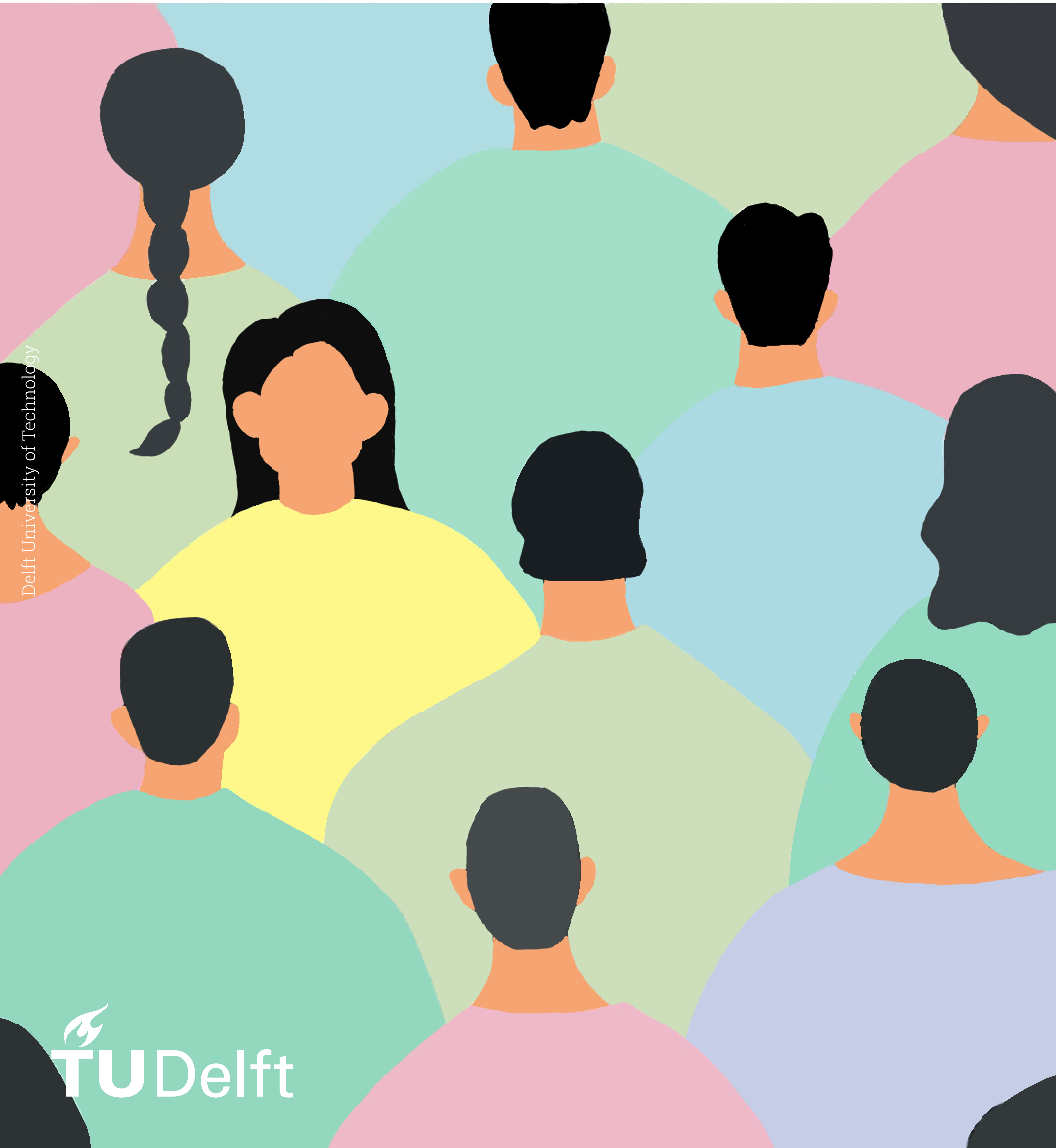
Gender Bias in Computer Science Job Advertisements

Esha Dutta



Gender Bias in Computer Science Job Advertisements

by

Esha Dutta

to obtain the degree of Master of Science
in Computer Science
at the Delft University of Technology,
to be defended publicly on Tuesday May 27, 2025 at 11:00 AM.

Student number:	5915996	
Project duration:	November 5, 2024 – May 27, 2025	
Thesis committee:	Prof. dr. M.M. Specht,	TU Delft, Thesis advisor
	Dr. ir. E. Aivaloglou,	TU Delft, Daily Supervisor
	Dr. J.A. Martinez Castaneda,	TU Delft, Committee Member

TU Delft

Acknowledgments

Writing this thesis has been an incredibly rewarding experience. I have not only improved and utilized the technical skills I have gained in the last few years, but also had the opportunity to read and learn about gender bias and it's manifestations in recruitment and the workplace. This project would not have been possible without the support of many people.

Firstly, thank you to my daily supervisor, Dr. Fenia Aivaloglou. I only knew that I wanted to work at the intersection of gender and technology during our first meeting. Her patience, kindness, and guidance helped me explore, question, and formulate various ideas that eventually culminated in this thesis.

I would also like to thank Dr. Sole Pera for her invaluable feedback which challenged me and enabled me to dig deeper in this project than I had initially planned. Thank you to Prof. Marcus Specht for critically evaluating my research and findings.

I am deeply grateful to my parents and my sister for their constant encouragement and support. Their unwavering belief in me has motivated me throughout this journey. Finally, I want to thank the friends I have made in Delft in the last two years who have made me feel at home, and also the ones who, despite the distances and time-zones, inspire me every day with their intelligence and humour.

Esha Dutta
Delft, May 2025

Abstract

A persistent gender gap exists in computer science occupations which may be partly driven by the subtle use of gendered language in job advertisements and the biased behavior of job recommendation algorithms. This thesis investigates how gender bias in job ads has evolved over time and how it varies across countries, industries, job roles, and work models. It also explores whether recommendation systems expose applicants differently to masculine- or feminine-coded jobs based on gender. A dataset of nearly 470,000 LinkedIn job advertisements related to computer science from 2014 to 2024 was scraped, filtered, and labelled using a gender bias score based on a curated repository of gender-coded words. Synthetic CVs were generated to simulate male and female applicants and used as users for whom job recommendations were generated using five content-based recommendation models (TF-IDF and Word2Vec variants). Results showed that job advertisements are predominantly masculine-coded, though a decline in masculine bias has occurred since 2018. Variation in bias across country, industry, and role is statistically insignificant with low practical effect sizes. The TF-IDF model exhibited the highest disparity in job exposure, while Word2Vec-SkipGram showed more balanced recommendations. A weak correlation was found between applicant gender and the genderedness of recommended jobs.

Contents

1 Introduction

Far from being an emerging problem, gender discrimination has its roots in cultural practices historically related to socio-political power differentials [3]. One of the ways in which gender bias manifests itself is disparate workforce participation by men and women. Although there have been strides in increasing women's participation across many industries [35], a wide gender gap in occupations related to computer science still persists. In the Stack Overflow Survey 2022, 91.88% of professional developers identified as male[1]. A mere 5.17% identified as a woman and 4.06% either identified as neither male nor female or preferred not to say. Gender has not been asked in the survey in subsequent years. This clear gender gap in computer science careers is further exacerbated in certain roles. According to the Stack Overflow Survey 2020, there were about 30 times more men than women employed in job roles like DevOps Specialist and System Administrator[2]. Moreover, according to the United States Department of Labor, only 19.3% of all computer programmers in the USA are women. This statistic is much higher across all occupations; 43.8% of the entire workforce in the USA constitutes women[3]. Thus, there is a marked difference in the participation of women and men in computer science occupations.

This gap in female participation in computer science careers can be attributed to many causes. Individual-level factors like sexist beliefs and attitudes that keep women out of male-dominated areas are well documented [22, 23]. However, research on institutional-level contributors that manifest within the social structure (e.g. public policy, law) to reinforce women's under-representation in male-dominated occupations is more recent. According to the social dominance theory [43], these institutional-level mechanisms exist to reinforce and perpetuate existing group-based inequality. One such institutional-level factor is the gendered language used in job recruitment materials [16]. These recruitment advertisements are the main means for an organization to communicate with potential applicants and persuade them to apply [1].

Like other subtle variations in language can have a causal effect on people's behavior and attitudes [32, 14, 7], subtle variations in the gendered wording used in advertisements may affect people's perception of jobs, such that men and women will find jobs described in language consistent with their gender most appealing because it signals they belong in that occupation [16]. Such unconscious biases may be one of the causes that discourage female applicants from applying to job advertisements [2].

Gendered words capture stereotypes by describing socially desirable traits and behaviors of male or female genders [4, 34]. Compared to other sectors, job advertisements from the technology industry tend to use more typically masculine language [45]. It is also the slowest industry to reduce masculine bias in job advertisements [45]. Thus, it is important to investigate the gendered usage of words in job advertisements in the technology sector in recent years.

Gender bias can also affect the exposure of job advertisements to different applicants. Job search is a task commonly done by applicants on sites like LinkedIn, Indeed, etc. Commonly, a job seeker has two

[1] https://survey.stackoverflow.co/2022/#section-demographics-gender
[2] https://survey.stackoverflow.co/2020#developer-profile-gender
[3] https://www.dol.gov/agencies/wb/data/Employment-and-Earnings-by-Occupation

ways to search for a job using these sites: querying based on some keywords, or creating a professional profile to receive personalized job recommendations. A variety of types of recommender systems, such as content-based filtering, collaborative filtering, knowledge-based, and hybrid approaches are used to recommend jobs to applicants.

However, these systems are not neutral. The biases embedded in real-world data related to gender, race, socioeconomic status, etc. are often inherited by the algorithms trained on them. As a result, recommender systems not only reflect but also amplify existing societal biases and deliver skewed or discriminatory suggestions to users [41, 38].

Historically, the performance of recommender systems has been evaluated primarily on the basis of accuracy [20, 42]. Even when other metrics such as novelty, user satisfaction, or diversity are considered [21, 24], the emphasis remains largely on fulfilling the informational needs of the user. While these metrics are important for ensuring the relevance of recommendations, they fail to address broader ethical concerns like the exposure of items to users belonging to different genders. In recent years, the fairness of artificial intelligence systems, including recommender systems, has been questioned [11]. It is, thus, critical to investigate whether the treatment of different types of users is equitable in recommender systems, especially in sensitive domains like employment.

The goal of this work is to understand the role that job advertisements and job recommendation algorithms play in introducing or exacerbating gender imbalance in computer science jobs. More specifically, this thesis aims to answer the following research questions:

RQ1 How has gender bias in the word usage in computer science job advertisements changed over time?

RQ2 How much does gender bias vary in computer science across the following parameters:

- Country
- Industry
- Job Role
- Work Model (Remote or On-Site)

RQ3 To what extent are feminine-coded and masculine-coded jobs exposed differently through job recommendation algorithms?

RQ4 Is there a correlation between applicant gender and the genderedness of job advertisements recommended to them?

In this work, we have scraped LinkedIn job advertisements from the last 10 years (2014-2024), identified the number of documented stereotypically feminine or masculine words used in them, and labelled each advertisement as feminine or masculine-coded using a genderedness score. We have then analyzed the genderedness of the job advertisements across years, countries, industries, job roles, and work models. To investigate the impact of recommender systems, synthetic CV data was created for male and female applicants. Then, ten jobs were recommended to each applicant based on the highest cosine similarity between applicants' resumes and job descriptions. Five text representations were selected for this task – TF-IDF, Word2Vec Continuous Bag of Words, Word2Vec Continuous Bag of Words with Bigrams, Word2Vec Skipgrams, Word2Vec Skipgrams with Bigrams. Then, the exposure of feminine and masculine-coded jobs and the strengths of correlation between applicant gender and genderedness of job advertisements were calculated for all representations.

The results showed that job advertisements tend to be more masculine-coded across all years, job roles, both work models, and most industries. Between 2014 and 2018, there was a rise in the proportion of masculine-coded advertisements followed by a decline till 2024. The effect sizes of all the dependent variables are small, suggesting that these variables do not strongly predict the genderedness of job advertisements in the data. The TF-IDF model exhibited the highest disparity-based exposure and consistently recommended a disproportionate number of feminine-coded jobs to both male and female applicants. Moreover, applicant gender influences the genderedness of recommendations, but not strongly.

The remainder of the paper is organized as follows. Section 2 presents a literature review on bias in the language of online job advertisements and the use of natural language processing to detect this gender bias. Section 3 presents the methods. The results and analysis are discussed in Section 4. Section 5 presents the discussion and conclusions. The limitations and future work are discussed in Section 6.

2

Related Work

2.1. Gendered Wording Effects in Job Applications

There is an established literature documenting differences in the way men and women use everyday language [31]. Women are perceived as more communal and interpersonally oriented, whereas men are more readily attributed with traits associated with leadership and agency [12, 18, 36]. Moreover, differences in the linguistic style of everyday speech between men and women are well documented [9]. Women use a more communal style of speech than men and make more references to social and emotional words [8, 29]. Language use can also differ based on the gender of whom one is writing about. An analysis of recommendation letters for university faculty jobs within biology found that writers used more “standout words” (e.g., outstanding, unique) when describing male than female candidates [39]. Similarly, in the language used in recommendation letters for university faculty jobs within psychology, women were described as more communal and less agentic than men, suggesting that language use can unintentionally reflect stereotypical gender roles. Furthermore, candidates whose letters contained more communal traits were less likely to be hired, clearly demonstrating that these gender-based differences in language use perpetuate inequality [26].

Gender stereotypes can be captured by gendered words — terms describing socially desirable traits and behaviors of male or female genders [4, 34]. Gendered words are usually extracted from self-reported characteristics through questionnaires given to college students to measure their self-concept and valuation of feminine and masculine characteristics. The Personal Attributes Questionnaire (PAQ) [44] and the Bem Sex Role Inventory (BSRI) [4] are two of the most representative questionnaires in early studies. The words extracted from BSRI and PAQ more typically associate females with more communal attributes (i.e., gentle, warm) and men with more agentic attributes (i.e., aggressive, competitive). Others generalized gendered words into expressive and instrumental traits.

Based on the Social Dominance Theory, which posits that societies contain status hierarchies in which some groups are systematically privileged over other groups [43], Gaucher et al. [16] proposed that words associated with masculine and feminine gender stereotypes may be a mechanism of maintenance of inequality. To examine whether this hypothesis is true, they collected both online and on-campus job advertisements from typically male and female-dominated occupations and gave each advertisement masculine and feminine scores based on a list of masculine and feminine words, representing the percentage of total masculine and feminine words in each. This repository was created based on relevant literature [12, 18, 36, 39, 26]. Their studies indicated that a higher number of masculine wording (i.e., words associated with male stereotypes, such as “leader”, “competitive”, “dominant”) was used in male-dominated fields of employment than in female-dominated fields. This difference in feminine wording (i.e., words associated with female stereotypes, such as “support”, “understand”, and “interpersonal”) was not found in either male- or female-dominated fields. In addition, when women were exposed to job advertisements with a higher number of masculine words, they found these jobs less appealing. They also perceived more men within these occupations. Thus, they showed that gendered language is a factor in the belongingness an applicant feels towards a job. The words used in job advertisements affect whether a candidate applies to a job and subsequently on the gender gap in that profession.

Hentschel et al. [19] further substantiated this claim. In a video-based experiment with 329 university students, they found that when a male recruiter used stereotypically masculine compared to feminine wording, female students reported a lower sense of belongingness, expected lower success of an application, and indicated lower application intentions for career opportunities. Thus, they showed that women's anticipated belongingness mediated the relationship between wording and application intentions. No such effects of wording were found on men.

In Taris and Bok's [46] experiment, 20 female and 20 male undergraduate university students examined 20 personal characteristics and indicated to what degree they felt that these characteristics were typical for the average man and woman of their own age and with their own type of education. This study demonstrated that both male and female applicants recognize the gender specificity of personal characteristics (like self-reliance, communication skills, etc.) mentioned in job advertisements. The male participants felt that they possessed both many male and many female personal characteristics. The female participants, contrary to the men, felt that they possessed relatively few typically male and female characteristics. This finding suggests that men find a personnel advertisement with typically male personal characteristics more attractive than women and that they will think they are more eligible for the job than women will.

2.2. Gender Bias Detection in Job Advertisements

Tang et al. [45] performed a longitudinal analysis of gender bias in job advertisements scraped from LinkedIn belonging to several industries from 2005 to 2016. They scraped a corpus of 17 million job advertisements from LinkedIn spanning these 10 years. To get a list of gendered words, the top 50,000 words with the highest frequency from all English job advertisements were collected. These words were queried through two online services that evaluate words in job advertisements for gender bias — Textio and Unitive — based on the lists encoded by Gaucher et al. [16]. These services assigned a feminine or masculine tone to the words. To measure the gender bias in each advertisement, two metrics were used: Gender Target and Gender Tone. The former was calculated by counting the number of gendered words, with feminine and masculine words canceling each other out, and computing a final score by applying a sigmoid function on the remaining word count. Gender Tone was calculated by first categorizing terms as inclusive or exclusive. The words were weighted based on how gender-specific they were. Gender Target and Gender Tone were used to quantify bias in job advertisements over time.

They found that there is significant gender bias in job advertisements, but bias has been reducing significantly over the decade. The Technology industry was found to be the second-worst-performing industry in terms of gender bias. It is also the slowest industry to reduce masculine bias over the years. They also found that the masculine tone of the job advertisement increases with its seniority ranking. Moreover, jobs with long-term employment tend to be more masculine in tone compared to those with part-time or temporary contracts.

Böhm et al. [5] also performed an analysis of gender bias on a large corpus of job advertisements in German. They developed a tool that automatically highlights words that lead to bias. This tool calculates a gender bias score as a measure of the gender neutrality of the text and offers possible re-wordings to reduce the gender bias.

They collected a corpus of approximately 500,000 job advertisements posted on the jobstairs.de portal from 2010-2020. They created a repository of words that may encourage ("pull" words) or discourage ("push" words) female applicants based on relevant literature on gendered wording in German and English job advertisements [6, 10, 16, 19, 46, 50]. The numbers of "push" and "pull" words in all advertisements were counted and used to calculate a gender score. They calculated this score for jobs from three sectors: Automotive/Automotive Suppliers (AAS), Healthcare/Medical (HCM), and IT Services (ITS). In comparison to the AAS and HCM, ITS jobs have the highest weighted sum of "push" (masculine) keywords. However, they recognized that many of these words are used to describe the specific requirements of the jobs, which are very technically and analytically oriented, like "analyzing". However, the IT job advertisements show the highest mean values for pull keywords. This could be an indication that employers are already making efforts to integrate more pull keywords with the intention of motivating women to apply.

Frissen et al. [15] also classified job advertisements on the basis of gender bias using a keyword repository based on relevant literature. They performed their experiment for five categories of bias: masculine, feminine, exclusive, LGBTQ, and racial. Instead of using bias scores, they used six machine learning algorithms in conjunction with word embeddings to detect bias in job advertisements. The publicly available Employment Scam Aegean Dataset (EMSCAD) was annotated based on a repository of keywords into the aforementioned five categories of bias. Then, six word embeddings in combination with five machine learning classifiers were trained on this annotated dataset. The results indicate that the Random Forest classifier with BERT word embeddings as the textual feature achieved the best performance in categorizing bias.

2.3. Bias in Job Recommendation Algorithms

Previous research has raised concerns about discrepancies in the accuracy of recommendations for different genders [53]. Two prominent studies have focused on gender bias in recommender systems. The work by Shakespeare et al. [41] establishes the existence of bias in the results of music recommender systems, and the work by Ekstrand et al. [13] focuses on bias shown by collaborative filtering algorithms while recommending books written by women authors. Both studies establish that the collaborative algorithms produced biased results after being fed data containing biases from various socio-cultural factors.

Zhang and Kuhn [51] found that jobs recommended only to women and not to men on job boards propose lower wages, require fewer years of work experience, and are more likely to require literacy skills and administrative skills. Such jobs also disproportionately contain words related to stereotypically feminine personality characteristics. Mansoury et al. [27] showed that inconsistent rating behaviours of users, lack of diversity in users' ratings, and inactivity of users may lead to poorer recommendations for women.

3 Methods

3.1. Data Collection

The collection of data took place in three phases. In the first phase, job advertisements were scraped from LinkedIn. In the second phase, the data was filtered on various parameters to retain only relevant job advertisements. Finally, in the third phase, several pre-processing steps were applied. An overview of these phases can be found in Figure 3.1. Scripts used for this process can be found in the GitHub repository[1]. The final data contained 469,967 unique records. This data is available upon request.

3.1.1. Data Scraping

Each job advertisement on LinkedIn corresponds to a unique Job ID. This Job ID was used to scrape job advertisements using the Open LinkedIn API[2]. This API does not allow the retrieval of job advertisements using filters like job title, industry, etc. Thus, initially, job advertisements irrelevant to computer science were also scraped.

Due to the high volume of data from 2014 to 2024, stratified sampling was used with a sampling interval of two years. Job advertisements were only scraped from every alternate year (2014, 2016, 2018, 2020, 2022, and 2024). First, the initial and final Job IDs were identified for each of these years. Then, each of these ranges were divided into ten equal parts to ensure that the scraped advertisements are uniformly distributed across the years. Job advertisements were scraped for every sixth Job ID in each range simultaneously. The scraper was stopped after running for four months, from December 2024 to March 2025.

3.1.2. Data Filtering

Firstly, duplicate records were deleted from the data. Some job advertisements had the same Title, Description, and Company ID, but different Job IDs. In such cases, only the first record was retained. Some job advertisements had the same Job Title and Job Description, but different Company IDs and Job IDs. These different Company IDs were either parent and child companies, or Company IDs belonging to the same company in different locations. In such cases, too, only the first record was retained.

Furthermore, during a manual inspection of the data, many jobs in 2016 and 2020 were found corresponding to a company called “CyberCoders Middleware Test Company” which has 0-1 employees on LinkedIn, contains dummy text in the company description, and has no posts or activity. Since this seems to be a dummy page for a non-existent company, all jobs corresponding to this company were removed from the data.

Then, this data was filtered based on various parameters. Firstly, only jobs relevant to Computer Science were retained. This step was performed by assigning a standard occupation classification (SOC)

[1] https://github.com/eshadutta9/Gender-Bias-in-CS
[2] https://github.com/EseToni/open-linkedin-api/

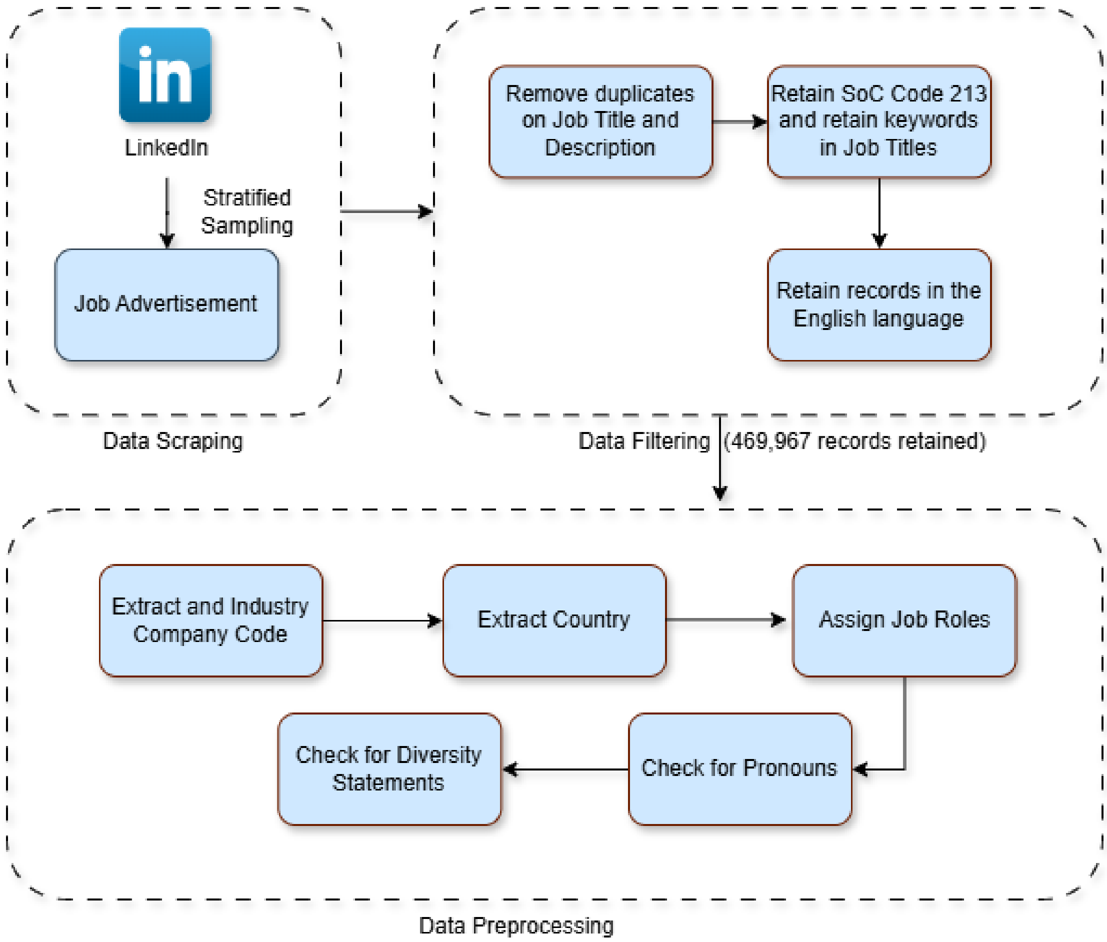


Figure 3.1: Data Filtering and Preprocessing

code[3]—which is a common classification of occupational information used by the UK Office of National Statistics—using a Python library called occupationcoder [48], which returns an SOC code based on a job title and description. All jobs corresponding to SOC code 213 were retained. This code corresponds to Information Technology and Telecommunications Professionals and includes the following jobs: IT specialist managers, IT project and program managers, IT business analysts, architects and systems designers, developers and software development professionals, Web design and development professionals, and Information technology and telecommunications professionals. However, the SOC code did not comprehensively include all job advertisements related to Computer Science in the data. Thus, further job advertisements were retained if the job titles contained any of the following keywords: "devops", "data scientist", "machine learning", "site reliability", "data analyst", "data engineer", "blockchain", "cybersecurity", "database admin", and "system admin". This list of keywords was created using developer roles in the Stack Overflow Survey 2024[4]. However, some roles in the survey were excluded or not used verbatim, as they could lead to incorrect data. For example, "security professional" could also mean security guards and related occupations, so "cybersecurity" was used as a keyword instead. "Project Manager" does not necessarily mean a Project Manager of technology-related projects, so it was excluded. Moreover, the keywords "Scientist", "Educator", "Academic Researcher", and "Research and Development role" were excluded since they could refer to jobs unrelated to Computer Science.

Then, only jobs written in English were retained, since the keyword repository used for data labelling in Section 3.2 is in English.

[3] https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010
[4] https://survey.stackoverflow.co/2024/developer-profile

3.1.3. Data Preprocessing

Firstly, the job advertisements' countries were extracted from their location. The locations in LinkedIn job advertisements contain the names of cities, states, etc. (E.g., "New York" and "Amsterdam"). GPT-4-turbo was prompted to return the countries that each location belongs to.

Then, the company and industry codes for each job advertisement were also scraped. The company code is a unique identifier for a company's account on LinkedIn. Each company on LinkedIn is assigned one of 434 industry codes which indicates the industry the company belongs to[5]. Similar industry codes were grouped together manually. The final list of industries and the mapping used to group them can be found in Appendix A.

The job advertisements were categorized into job roles such as "System Administrator", "Data Scientist", "Data Engineer", etc. This step was performed based on string matching of the titles with developer profiles from the Stack Overflow Survey 2024. However, many job advertisements did not match these strings directly. On further inspection, it was found that many job titles contained names of technologies such as "Python Engineer" or "SQL Developer". Therefore, additional roles were introduced based on the most popular technologies in the Stack Overflow Survey 2024. Uncategorized roles were assigned the category "Other". Finally, similar roles were grouped for analysis. The final list of roles and the mapping used to group them can be found in Appendix B.

Each job advertisement was also classified according to whether it contained an equal opportunity employer statement. If the string "equal opportunity" was present in the job description, or if diversity keywords like "race", "gender", "religion", "age", "disability", "marital status", "veteran status", or "sexual orientation" were present in the job description more than twice, the advertisement was considered to contain an equal opportunity statement.

The number of male ("he", "him", "his", "himself"), female ("she", "her", "hers", "herself"), or gender-neutral ("they", "them", "their", "theirs", "themselves") pronouns present in each job description were also counted.

3.2. Data Labelling

Each job advertisement was annotated as feminine or masculine-coded. Firstly, a repository of feminine-coded and masculine-coded words was created based on relevant literature [16, 25, 46, 19, 40, 33, 47, 6]. The complete keyword repository can be found in Appendix C.

A manual inspection of random samples of the data indicated that some words in the repository were occurring more frequently and in a different context in computer science job advertisements when compared to job advertisements unrelated to computer science. To identify such words, the average frequency of occurrence per job advertisement of each word in the keyword repository was calculated. Then, a z-score for each word was calculated. A z-score higher than a threshold of 3 indicates that it is 3 standard deviations above the mean frequency of occurrence. This means that the word is a CS-specific word that is rare in other job fields.

Moreover, to further substantiate the high frequency of these words, a TF-IDF score was also calculated for each word in the repository, and words with the highest scores were considered to have a significantly higher frequency in computer science-related jobs.

Finally, a list of words occurring in both the z-score and TF-IDF lists was manually inspected by two researchers in two random samples of 50 job advertisements from both general jobs and CS-specific jobs. Upon manual inspection, some of them were found to occur in a different context in CS-specific jobs. Thus, such words were removed. The removed words are: "strong", "best", "analysis", and "lead" from the masculine repository, and "support", "develop", "teams", and "understanding" from the feminine repository. Examples of the different usages of these words in a general context and a computer science context can be found in Appendix C.3.

To label the data as masculine-coded or feminine-coded, the numbers of occurrences of words from the masculine and feminine repositories — denoted as push words and pull words respectively —- were

[5] https://learn.microsoft.com/en-us/linkedin/shared/references/reference-tables/industry-codes-v2

counted for each job description. Then, the following formula was used to calculate a gender bias score, as suggested by Böhm et al. [5]:

$$S = \frac{1}{1+\exp(-x)}, \tag{3.1}$$

where

$$x = \frac{1}{N}\left(\sum_{\text{pull words}} w_i - \sum_{\text{push words}} w_i\right), \tag{3.2}$$

$$N = \left(\sum_{\text{pull words}} w_i + \sum_{\text{push words}} w_i\right). \tag{3.3}$$

where w is the weight which can be 1 or 2. When w = 2, the word has been established by more than one source as being gender-coded, and when w = 1, only one source identifies this word as gender-coded. The score lies in the interval (0, 1). The score is equal to 0.5 when the job description is gender-neutral and contains no gender-coded words. A higher score indicates that more masculine-coded words are in the job description. If the score is below 0.5, there are more feminine-coded words in the job description. Thus, each job advertisement was labelled as “masculine-coded”, “feminine-coded”, or “neutral”.

3.3. Data Analysis

To answer RQ1, a one-way analysis of variance (ANOVA) test was conducted to examine the relationship between the genderedness of the job advertisements and year for the entire dataset. To answer RQ2, a two-way analysis of variance (ANOVA) test was conducted to examine the relationship between the genderedness of the job advertisements and country, industry, role, and work model for the records in 2024.

3.4. Job Recommendations

Due to a dearth of publicly available CV data, synthetic CVs containing various details like gender, education, and skills were generated using GPT-4o-mini. Jobs from the LinkedIn job advertisement dataset were recommended to users based on cosine similarities of various text representations. The top 10 recommendations with highest cosine similarities were selected for each user.

3.4.1. Generation of Synthetic CVs

Since CVs contain personal data like name, gender, and location, it is difficult to find or collect CV datasets. To the best of our knowledge, there is no publicly available dataset of CVs which is not anonymized. Thus, synthetic CV data was created using GPT-4o-mini[6]. The prompts contained values of education, experience, programming languages, databases, frameworks, tools, IDEs, operating systems, and location. These values were selected randomly from the most popular technologies section of the Stack Overflow 2024 survey in the proportion of respondents for each technology. To make these synthetic CVs better mimic how men and women write CVs in the real world, the findings of Ng et al. [30] were used to describe in the prompt the type of language to be used in the CV. They found that women use communal and involved words in their resumes while men use self-promoting and confident words. Similar to the approach taken by Ghosh and Sadaphal [17] to convert unstructured CVs to structured CVs using LLMs, prompts containing gender, skills, and locations were used to create CVs containing the sections Summary, Skills, Experience, and Education. A total of 16,000 CVs — 8,000 belonging to male applicants and 8,000 belonging to female applicants, were generated.

A sample prompt for a male CV is displayed below:

[6]https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

Create a resume from the given unstructured JSON input using self-promoting and confident words:

Education: Master's degree

Experience: 15-17 years

Programming Languages: JavaScript, Java, SQL

Databases: MongoDB, Oracle

Web Frameworks: Angular, jQuery

Other Frameworks: TensorFlow, .NET

Other Tools: Terraform, Gradle

IDEs: Eclipse, Visual Studio Code

Operating Systems: macOS, Windows

Location: United States

Only return the following sections: Summary, Skills, Experience, Education

A sample prompt for a female CV is displayed below:

Create a resume from the given unstructured JSON input using communal and involved words:

Education: Secondary school

Experience: 2-4 years

Programming Languages: SQL, Python, HTML/CSS

Databases: Microsoft SQL Server, Redis

Web Frameworks: Flask, Angular

Other Frameworks: Flutter, React Native

Other Tools: Git, Kubernetes

IDEs: Android Studio, IntelliJ

Operating Systems: Windows, macOS

Location: Germany

Only return the following sections: Summary, Skills, Experience, Education.

A sample female CV is displayed below:

Summary

Dynamic and motivated software developer with a Bachelor's degree and over 1-2 years of hands-on experience in software development. Proficient in a range of programming languages including Java, SQL, and JavaScript, complemented by expertise in cutting-edge web and mobile frameworks like React.js and Flutter. Known for consistently delivering high-quality solutions and for driving innovation through a deep understanding of modern software development practices. A proactive team player eager to contribute to impactful projects and elevate organizational success.

Skills

- **Programming Languages:** Java, SQL, JavaScript

- **Databases:** Firebase, SQLite

- **Web Frameworks:** React.js, Node.js
- **Other Frameworks:** Pandas, Flutter
- **DevOps Tools:** Terraform, Kubernetes
- **IDEs:** Visual Studio, Visual Studio Code
- **Operating Systems:** Windows, macOS

Experience

Software Developer

- Spearheaded the development of scalable web applications using React.js and Node.js, enhancing user experience and engagement.
- Collaborated in agile teams to design and implement cloud-based solutions using Firebase and SQLite, resulting in improved data management efficiency.
- Leveraged Terraform and Kubernetes to streamline DevOps processes, ensuring seamless deployment and operation of applications.
- Developed robust functionality in mobile applications utilizing Flutter, significantly improving application performance and user satisfaction.
- Engaged in code reviews and mentoring sessions, fostering a culture of continuous learning and excellence among team members.

Education

Bachelor's Degree

- Specialized in Computer Science with a focus on software development and database management.
- Completed projects that included full-stack application development and the implementation of data-driven solutions.

3.4.2. Exposure of Recommendation Algorithms

To recommend job advertisements from the LinkedIn corpus to the male and female CVs, five text representations were selected from the study done by Valverde-Rebaza et al. [49]: TF-IDF, Word2Vec Continuous Bag of Words, Word2Vec Continuous Bag of Words with Bigrams, Word2Vec Skipgrams, and Word2Vec Skipgrams with Bigrams.

TF-IDF assigns weights to words as a statistical measure for assessing the relevance of a word in a corpus [37]. This relevance is proportional to the number of times a word appears in the document and inversely proportional to the frequency of the word in the corpus.

Word2vec is a general prediction model for learning vector representations of words [28]. These vector representations, also called word embeddings, capture distributional semantics and co-occurrence statistics. There are two Word2vec models we have used to obtain word embeddings: Continuous Bag of Words (CBOW) and Skip-gram. The former predicts a target word based on the n words before and n words after the target word. The latter predicts the context based on one word, instead of predicting a word based on its context.

The text in job titles, descriptions, and generated CVs were preprocessed. Firstly, stop words and special characters were removed from the text. Stop words are commonly used unimportant words, like "the", "is" and "and", etc. Then, tokenization and lemmatization was performed for these texts. Tokenization is the process of converting a sequence of text into individual words. Lemmatization is a text normalization technique that switches any kind of word to its root word. For example, the word "developer" would be converted to it's root form "develop". Then, the ten nearest job advertisements to each CV were recommended based on cosine similarities.

To answer RQ3, the exposure of the feminine and masculine-coded jobs by each text representation was measured using Disparity-based Exposure. Exposure measures item occurrences in users' recommendations. If an item is recommended to more users, that item has a higher exposure [52]. It is formulated as:

$$Exposure(i) = \sum_{u \in \mathcal{U}} 1(i \in \mathcal{R}_u) \tag{3.4}$$

where i is the item for which exposure is measured, $\mathcal{U}$ is the set of all users in the system, $\mathcal{R}_u$ is the set of recommended items for user u.

Disparity-based Exposure captures the exposure of a group of items and their positions in the recommendations relative to the proportion of that group in the catalog [52]. The underlying fairness assumption is that items from a minority group of providers should be recommended to users proportional to their representation. It is defined as:

$$F(\mathcal{G}_1, \mathcal{G}_2) = \left| \frac{1}{|\mathcal{G}_1|} \sum_{i \in \mathcal{G}_1} \text{Exposure}(i) - \frac{1}{|\mathcal{G}_2|} \sum_{i \in \mathcal{G}_2} \text{Exposure}(i) \right|, \tag{3.5}$$

where $\mathcal{G}_1$ and $\mathcal{G}_2$ are the two groups of items in the catalog, which are masculine-coded and feminine-code jobs.

To answer RQ4, the presence of a correlation between the applicants' gender and the genderedness of the job advertisement was identified. Cramér's V was used to measure the association strength between these two categorical variables. It is based on the chi-squared statistic and produces a value between 0 and 1, where 0 implies no association and 1 implies perfect association. It is is defined as:

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}}$$

where χ^2 is the chi-squared statistic, n is the total number of observations, $k = \min(r, c)$, with r as the number of rows and c as the number of columns in the contingency table. This value has been calculated for the top 1, 5, and 10 job recommendations for each user.

4 Results

4.1. Dataset Statistical Analysis

The final data contained the job title, description, country, industry, role, and work model. Figure 4.1 displays the descriptive statistics of this data set. Figure 4.1a shows that the final data contained 469,967 unique job advertisements. There were 218 distinct countries in the data. More than 50% of the jobs were from the United States of America, as can be seen in Figure 4.1c. The distribution of the most popular industries and roles can be found in Figure 4.1d and Figure 4.1e respectively. There were 20 unique industries and 27 unique roles in the data. The most frequently occurring industries were “IT Services and IT Consulting”, “Staffing and Recruiting”, and “Software Development”. The roles with the highest frequencies in the dataset were “Back-End”, “Manager”, and “Data and Analytics”. Furthermore, Figure 4.1b shows that 89.8% of the job advertisements did not allow remote working.

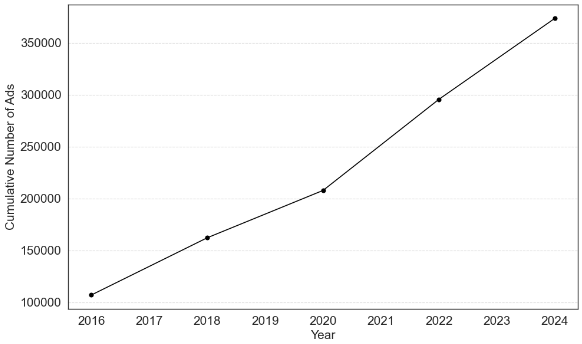
Most job advertisements did not contain male or female pronouns. This can be observed in Figure 4.2f. Moreover, Figure 4.2g shows that most advertisements across all years did not contain equal opportunity statements.

The most frequently occurring masculine words were “quality”, “professional”, “strategy”, “excellent”, “drive”, “individual”, “deliver”, and “control”. The most frequently occurring feminine words were “communication”, “responsible”, “excellent”, “value”, “implement”, “community”, “collaborate”, and “committed”.

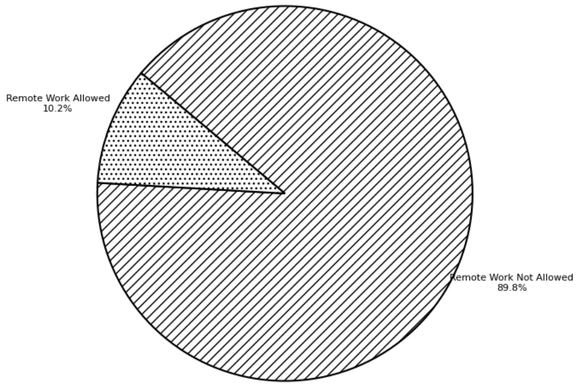
4.2. RQ1: Gender Bias Over Years

Figure 4.2a displays the distribution of masculine-coded, feminine-coded, and neutral job advertisements. In each year, there are more masculine-coded job advertisements than feminine-coded ones. The proportion of masculine-coded job advertisements increased from 2014 to 2018. In 2018, 58.9% of the advertisements were masculine-coded and only 32.9% were feminine-coded. However, there has been a reversal in this trend; since 2018, the proportion of masculine-coded advertisements has been decreasing. Moreover, the lowest proportion of masculine-coded jobs can be observed in 2024. This year also has an almost equal percentage of masculine-coded and feminine-coded jobs (46.3% and 44.9% respectively).

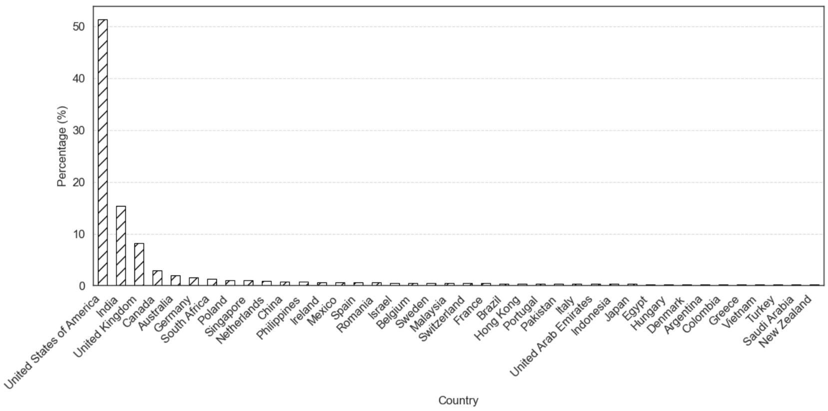
A one-way analysis of variance (ANOVA) revealed a statistically significant effect of year on the score, $F(5, 430{,}826) = 802.97$, $p < .001$. However, the effect size was small ($\eta^2 = .0092$), indicating that year accounts for less than 1% of the variance in scores. This suggests that other factors not included in the model may explain a larger portion of the variation.



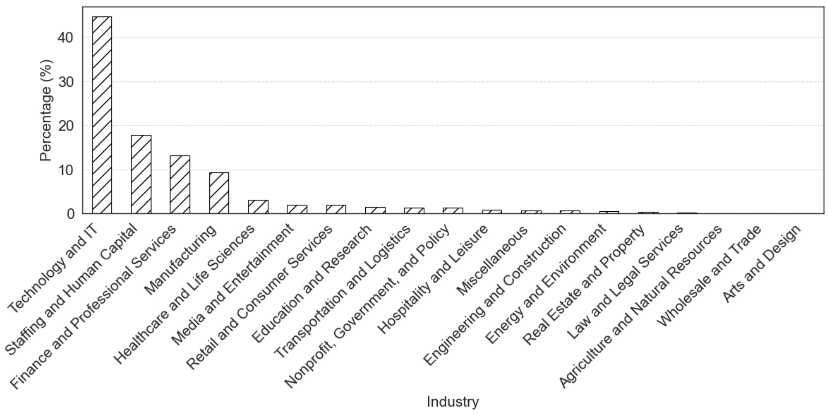
(a) Cumulative Count of Job Advertisements Over Years



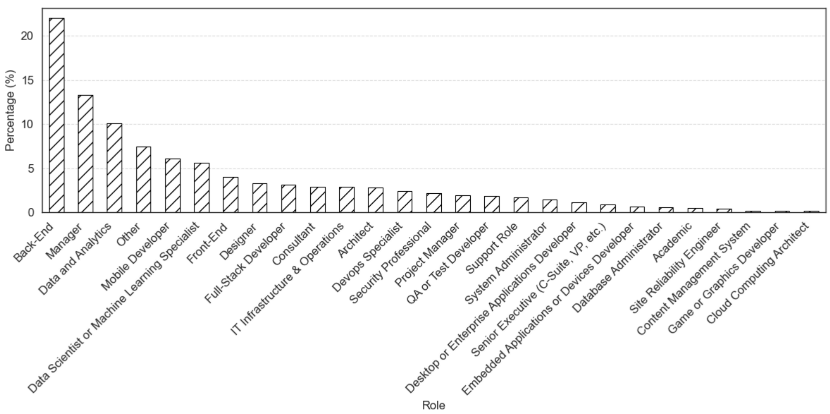
(b) Job Advertisements Per Work Model



(c) Percentage of Job Advertisements Per Top 40 Countries

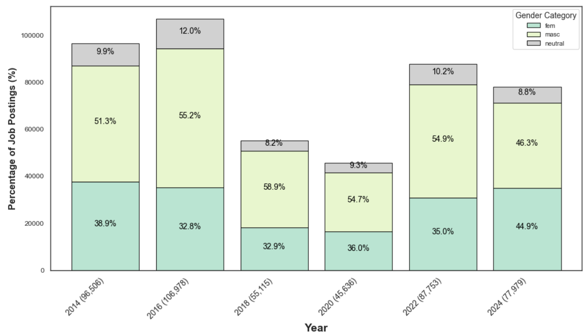


(d) Percentage of Job Advertisements Per Industry

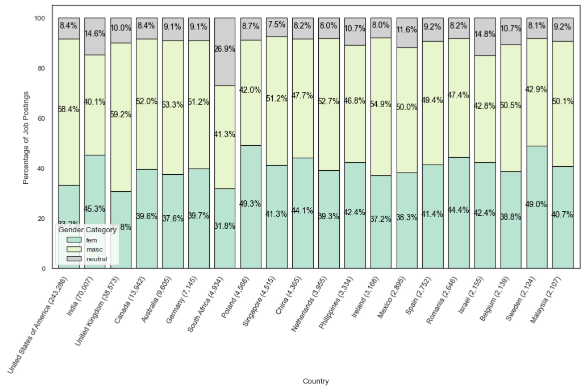


(e) Percentage of Job Advertisements Per Top 20 Roles

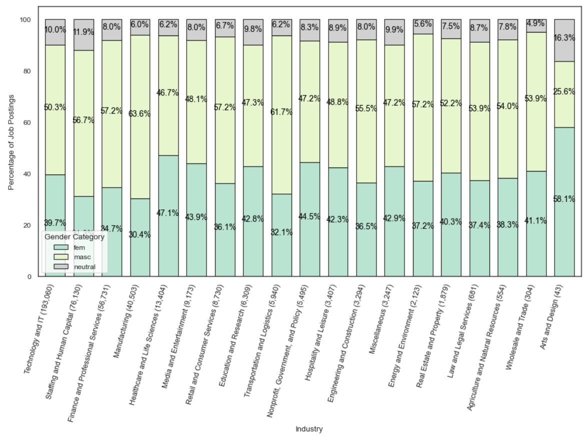
Figure 4.1: Descriptive Statistics of the Dataset



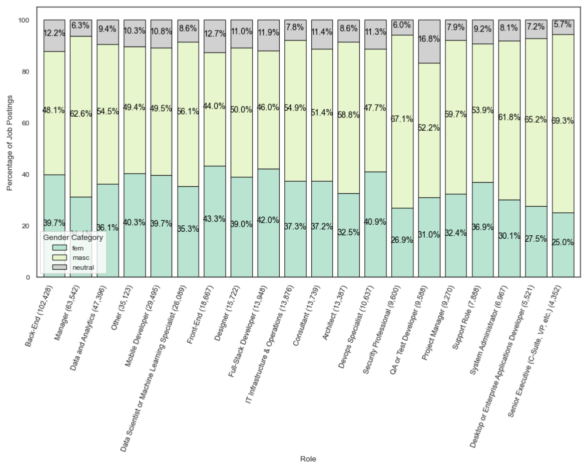
(a) Gender Distribution in Job Advertisements Over Time



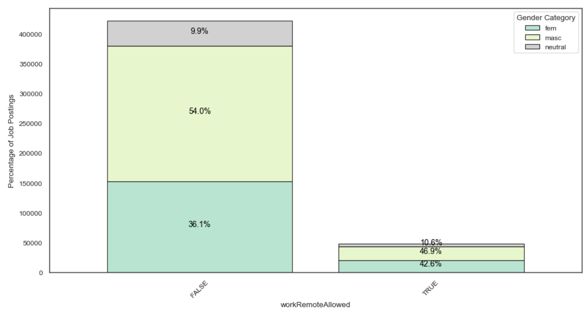
(b) Percentage of Job Advertisements by Gender Category in Top 20 Countries (Sorted by Total Jobs)



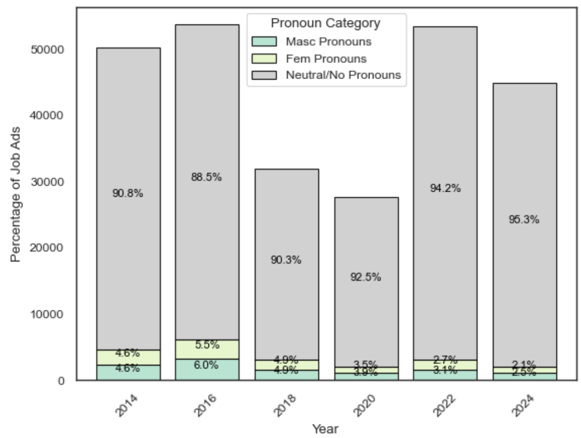
(c) Percentage of Job Advertisements by Gender Category in Industries (Sorted by Total Jobs)



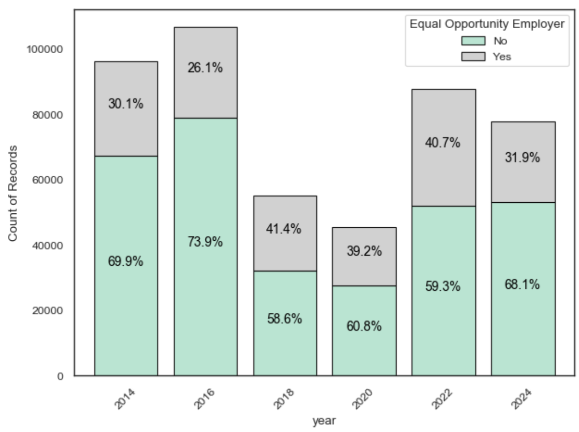
(d) Percentage of Job Advertisements by Gender Category in Top 20 Roles (Sorted by Total Jobs)



(e) Percentage of Job Advertisements by Work Model



(f) Percentage of Job Advertisements by Pronoun Count



(g) Percentage of Job Advertisements Containing Equal Opportunity Employer Statements

Figure 4.2: Analysis of Gender Bias in Job Advertisements

Country (No of Records)	Masc %
United Kingdom (38,573)	59.2
United States of America (243,286)	58.3
Ireland (3,166)	54.9
Australia (9,605)	53.3
Netherlands (3,955)	52.7

Table 4.1: Top Countries with Masculine Words

Country (No of Records)	Fem %
Poland (4,566)	49.3%
India (70,007)	45.3%
Romania (2,646)	44.4%
China (4,365)	44.1%
Philippines (3,334)	42.4%

Table 4.2: Top Countries with Feminine Words

Industry (No of Records)	Masc %
Manufacturing (40,503)	63.6
Transportation and Logistics (5,940)	61.7
Finance and Professional Services (56,731)	57.2
Retail and Consumer Services (8,730)	57.2
Staffing and Human Capital (76,130)	56.7

Table 4.3: Top Industries with Masculine Words

Industry (No of Records)	Fem %
Healthcare and Life Sciences (13,404)	47.1
Nonprofit, Government, and Policy (5,495)	44.5
Media and Entertainment (9,173)	43.9
Miscellaneous (3,247)	42.9
Education and Research (6,309)	42.8

Table 4.4: Top Industries with Feminine Words

Role (No of Records)	Masc %
Senior Executive (C-Suite, VP, etc.) (4,352)	69.3
Security Professional (9,600)	67.1
Desktop or Enterprise Applications Developer (5,521)	65.2
Manager (63,542)	62.6
System Administrator (6,967)	61.3

Table 4.5: Top Roles with Masculine Words

Role (No of Records)	Fem %
Front-End (18,667)	43.3
Full-Stack Developer (13,948)	42.0
DevOps Specialist (10,637)	40.1
Other (35,123)	40.2
Back-End (102,428)	39.7

Table 4.6: Top Roles with Feminine Words

4.3. RQ2: Gender Bias By Country, Industry, Job Role, and Work Model

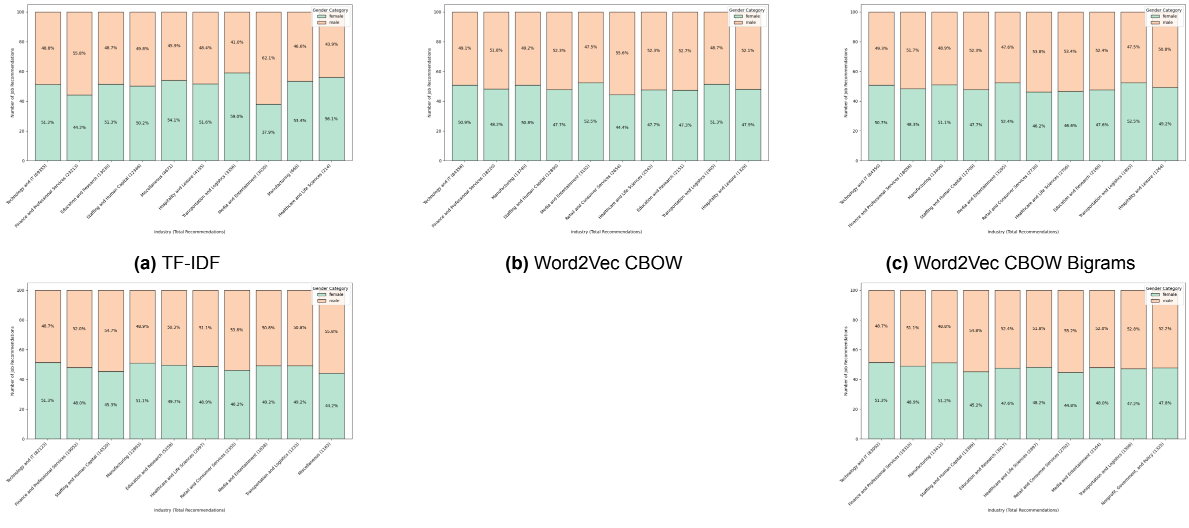
The proportion of masculine-coded and feminine-coded jobs across the top 20 most frequently occurring countries in the data are displayed in Figure 4.2b. Out of these 20 countries, all countries have more masculine-coded job advertisements than feminine-coded ones except for India, Poland, and Sweden. The highest proportion of masculine-coded jobs in these 20 countries can be found in the United Kingdom (59.9%), the United States of America (58.4%), and Ireland (54.9%). Moreover, South Africa has the highest proportion of gender-neutral job advertisements (26.9%).

Tables 4.1 and 4.2 show the top 5 countries in the entire data with the highest proportion of masculine-coded and feminine-coded advertisements respectively, along with the number of records pertaining to each gender category and the percentage of records in that country that constitute the gender category.

Figure 4.2c displays the distribution of masculine-coded, feminine-coded, and neutral job advertisements across the top 20 most frequently occurring industries in the data. All industries have more masculine-coded job advertisements than feminine-coded ones except for two: telecommunications and hospitals and health care.

Tables 4.3 and 4.4 show the 5 industries in the entire data with the highest proportion of masculine-coded and feminine-coded advertisements respectively, along with the number of records pertaining to each gender category and the percentage of records in that industry that constitute the gender category. The top three industries with the highest proportion of masculine-coded job advertisements are: Manufacturing (63.6%), Transportation and Logistics (61.7%), and Finance and Professional Services (57.2%).

The proportion of masculine-coded and feminine-coded job roles across the top 20 most frequently occurring job roles in the data are displayed in Figure 4.2d. In all these roles, the proportion of masculine-coded jobs are more than feminine-coded ones. The roles with the highest proportion of masculine-coded advertisements are Senior Executive (C-Suite, VP, etc.) (69.3%), Security Professional (67.1%), and Desktop of Enterprise Applications Developer (65.2%). Moreover, even the roles with the highest percentage of feminine-coded jobs still have feminine-coded jobs in the minority. Ta-



(a) TF-IDF **(b)** Word2Vec CBOW **(c)** Word2Vec CBOW Bigrams

(d) Word2Vec Skipgrams **(e)** Word2Vec Skipgrams Bigrams

Figure 4.3: Distribution of Industry by Gender of Applicant Recommended

bles 4.5 and 4.6 show the top 5 roles in the entire data with the highest proportion of masculine-coded and feminine-coded advertisements respectively, along with the number of records pertaining to each gender category and the percentage of records in that role that constitute the gender category.

The proportion of masculine-coded job advertisements is higher than feminine-coded ones in both jobs where remote working is allowed and not allowed. This can be observed in Figure 4.2e. However, on-site jobs have a higher proportion of masculine-coded jobs (54%), compared to remote or hybrid jobs (46.9%).

An ANOVA test was conducted to examine whether job advertisement genderedness was influenced by individual predictors and their interactions. None of the main effects — country, industry, role, or work model — were statistically significant predictors of the gender score. Specifically, all had high p-values ($p > .73$), indicating no significant individual effects. Corresponding effect sizes were negligible, with eta squared values ranging from $\eta^2 = .00016$ to $\eta^2 = .00018$ for country, industry, and role, and $\eta^2 = .000002$ for the work model.

However, two interaction effects reached statistical significance. The interaction between country and work model was significant, $F(37, 73599) = 4.54$, $p < .001$, though the effect size was small ($\eta^2 = .0022$). Similarly, the interaction between industry and work model was significant, $F(40, 73599) = 2.40$, $p < .001$, but the effect size was also small ($\eta^2 = .00128$). The remaining interactions were not statistically significant ($p > .35$).

Overall, the predictors and their interactions accounted for approximately 2.02% of the variance in the gender score, leaving 97.98% of the variance unexplained. This suggests that other, unmeasured variables may play a more substantial role in explaining genderedness in job advertisements.

4.4. RQ3: Exposure of Recommendation Algorithms

The disparity-based exposures of the text representations can be found in Table 4.7. TF-IDF has the highest disparity-based exposure compared to all variations of Word2Vec. All variations of Word2Vec perform better, with disparity-based exposure values closer to 1. Word2Vec-SkipGram has the best performance with the closest value to 1.

4.5. RQ4: Correlation Between Applicant Gender and the Genderedness of Job Advertisements

The correlation between applicant gender and the genderedness of the top 1, 5, and 10 recommended job advertisements to them can be found in Table 4.9. Even though the relationships are statistically sig-

RA	Disparity-Based Exposure
TF-IDF	71.246
Word2Vec-CBOW	1.512
Word2Vec-SkipGram	0.964
Word2Vec-bigrams-CBOW	1.670
Word2vec-bigrams-SkipGram	1.740

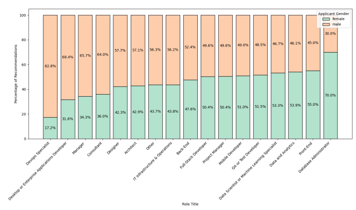
Table 4.7: Disparity-based Exposure for Recommendations

RA	male masc%	male fem%	female masc%	female fem%
TF-IDF	15.9	82.7	12.1	86.3
Word2Vec-CBOW	55.4	40.2	50.9	44.5
Word2Vec-SkipGram	49.2	44.5	45.1	48.4
Word2Vec-bigrams-CBOW	56.8	39.0	52.4	43.1
Word2vec-bigrams-SkipGram	54.0	40.8	50.4	44.2

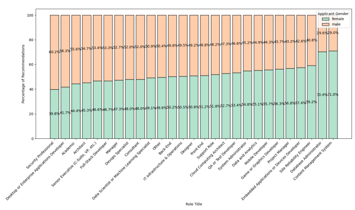
Table 4.8: Recommendation Algorithm User-Based Results

RA	p@1	Cramér's V@1	p@5	Cramér's V@5	p@10	Cramér's V@10
TF-IDF	3.7e-25	0.084	1.5e-62	0.060	1.3e-105	0.055
Word2Vec-CBOW	9.3e-17	0.068	1.2e-51	0.054	2.3e-72	0.045
Word2Vec-SkipGram	7.6e-09	0.048	1.4e-30	0.041	3.7e-60	0.041
Word2Vec-bigrams-CBOW	6.5e-17	0.068	4.3e-41	0.048	8.5e-70	0.045
Word2vec-bigrams-SkipGram	1.1e-03	0.029	2.5e-27	0.039	1.1e-47	0.037

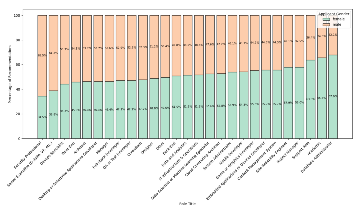
Table 4.9: Cramér's V for Correlation between Applicant Gender and Genderedness of Recommended Job Advertisements for Top 1, 5, and 10 Recommendations



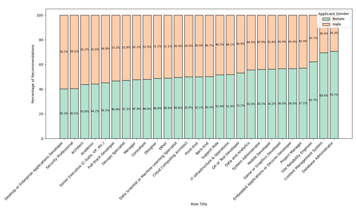
(a) TF-IDF



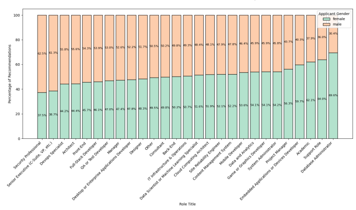
(b) Word2Vec CBOW



(c) Word2Vec CBOW Bigrams



(d) Word2Vec Skipgrams



(e) Word2Vec Skipgrams Bigrams

Figure 4.4: Distribution of Roles by Gender of Applicant Recommended

nificant due to the very low p-values, Cramer's V values are < 0.1 for each recommendation algorithm. Thus, the actual strength of the association is very weak for each recommendation algorithm. Gender appears to play a role, but not a strong one in shaping recommendations. Moreover, Table 4.8 shows that TF-IDF predominantly recommends feminine-coded advertisements to both male and female applicants. All variations of Word2Vec recommend job advertisements to male and female applicants in a more balanced manner. Word2Vec-Skipgram is the only algorithm among these that recommends more feminine-coded jobs to women than masculine-coded ones.

To gain insights into the differences between the industries recommended to male and female applicants, Figure 4.3 shows the distribution of applicant gender per industry in the recommendations for each text representation. The most frequently recommended industry is Technology and IT for all text representations. Despite being a highly masculine-coded industry, Technology and IT is recommended almost equally to male and female applicants by all text representations. TF-IDF recommends jobs belonging to most industries more frequently to female applicants, except for Finance and Professional Services and Media and Entertainment. Manufacturing, another highly masculine-coded industry, is recommended almost equally to male and female applicants by all text representations and even slightly more to female applicants. Similarly, Transportation and Logistics is also recommended slightly more to female applicants by all text representations, except Word2Vec Skipgrams and Word2Vec Skipgrams Bigrams.

Figure 4.4 depicts the distribution of applicant gender per job role in the recommendations for each text representation. Security Professional and Senior Executive, both highly masculine-coded roles, are recommended very frequently to male applicants by all variations of Word2Vec. Moreover, other highly masculine-coded roles like DevOps specialist, Manager, and Desktop or Enterprise Developer are recommended less frequently to female applicants by all text representations. However, some highly masculine-coded roles like Data and Analytics and System Administrator are recommended more frequently to female applicants across the board.

5 Discussion and Conclusion

This work aimed to answer how gender bias in computer science job advertisements changed over time and how country, industry, role, and work model influence the genderedness of job advertisements. The analysis revealed a clear shift in the gendered language of job advertisements over the past decade. Between 2014 and 2018, there was a noticeable rise in masculine-coded advertisements. However, this trend began to reverse after 2018. In 2024, the proportions of feminine-coded and masculine-coded advertisements were almost at par. This suggests growing awareness and corrective efforts by organizations in the computer science jobs, possibly driven by broader Diversity, Equity, and Inclusion initiatives.

According to the USA Bureau of Labour Statistics[1], the percentage of women employed as computer programmers in the last ten years has remained constant between 20-22% till 2023 and reduced to 17.8% in 2024. This can be observed in Table 5.1. While gendered language in job ads has become more balanced in recent years, the actual employment of women in programming roles has not reflected this change. Thus, language reform in job ads may be necessary but not sufficient on its own to improve gender representation in programming roles. One possible explanation is that women may be applying but are not being hired at the same rate as men, possibly due to disparities in recruitment processes.

The results also showed that masculine-coded advertisements were dominant in most countries, industries, and job roles. Countries such as the United Kingdom and the United States of America had the highest proportion of masculine-coded job advertisements, whereas Poland and India had the highest proportion of feminine-coded ones. However, this may be because only English-language job advertisements were retained in the dataset, potentially biasing the representation of countries where English is a primary or widely used professional language.

[1] https://www.bls.gov/cps/tables.htm#otheryears

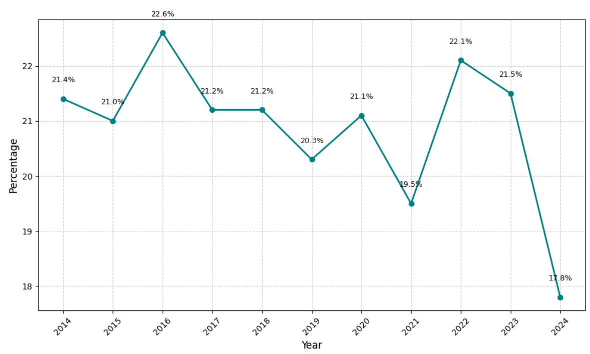


Figure 5.1: Percentage of Women Employed as Computer Programmers in the USA from 2014-2024

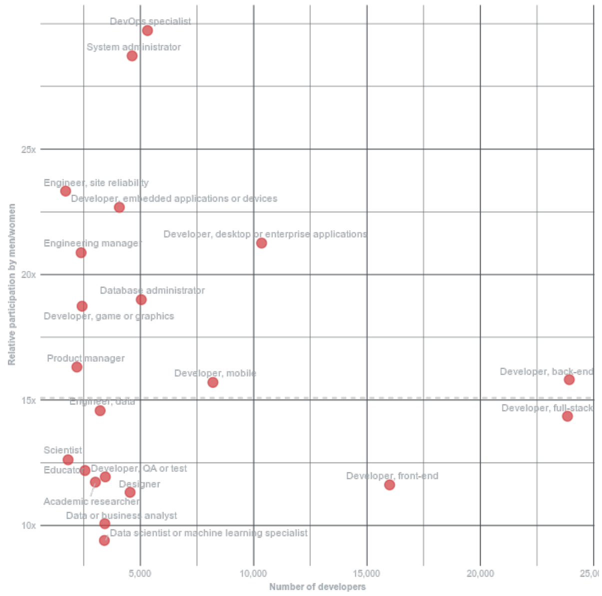


Figure 5.2: Relative Participation by Men and Women (Stack Overflow Survey 2020)

The results showed that gendered language is more prevalent in job postings for specific technical roles. For example, roles such as system administrators and desktop or enterprise applications developers were more frequently masculine-coded. This is in agreement with the statistics of the Stack Overflow Survey 2020[2], which reported that the relative participation by women was lowest for the roles DevOps Specialist, System Administrator, Site Reliability Engineer, Embedded Applications or Devices Developer, and Desktop or Enterprise Applications Developer. There were more than 25 times more men than women in the roles. This can be seen in Figure 5.2. The dashed line in the figure shows the average ratio of men's to women's participation.

The top five roles with the worst ratios also have higher proportions of masculine-coded advertisements in the data. 47.7% of DevOps Specialist roles, 61.8% of System Administrator roles, 61.8% of Site Reliability Engineer roles, 50.3% of Embedded Applications or Devices Developer roles, and 65.2% of Desktop or Enterprise Applications Developer roles are masculine-coded. These values can be observed in Table 4.5.

Equally concerning is the prevalence of gendered wording in job postings for higher-level positions. Roles like "Manager" and "Senior Executive" were often accompanied by language typically associated with masculinity. This likely contributes to the under-representation of women in leadership roles. Gendered wording may be affecting women's self-assessment of fit and reducing female applications. This may be leading to a gender gap at the leadership level.

Gendered wording in job advertisements also plays a significant role in shaping the demographics of various industries. The findings from this study reveal that certain sectors like Manufacturing and Transportation and Logistics tend to use masculine-coded language. These industries also have a gender gap in their participation in the labor force. Only 29.2% of the people employed in Manufacturing and 24.8% of the people employed in Transportation and Utilities in the USA in 2022 were women, according to the Bureau of Labour Statistics[3]. This underscores how subtle cues in language can perpetuate the exclusion of women from entire industries. The persistent use of masculine-coded language reinforces occupational norms that contribute to systemic barriers against women's participation.

However, the influence of these factors is not statistically significant in the results. Although some statistically significant interactions (e.g., between genderedness and the combination of country and work model) were found, their effect sizes were small. Thus, these variables variables do not strongly predict the genderedness of job advertisements in the data.

[2]https://survey.stackoverflow.co/2020#developer-profile-developer-role-and-gender
[3]https://www.bls.gov/opub/reports/womens-databook/2022/

One possible explanation for the negative results is the method used to characterize genderedness. Although the approach used a repository of gender-coded words and weighted scoring, it may lack the sensitivity required to detect subtle, industry-specific patterns in language. For instance, certain words that are considered masculine or feminine in general job advertisements may occur in a more neutral or technical sense in computer science contexts. Although some of these were removed after a frequency and context-based analysis, other nuanced usages may have remained and reduced the discriminatory power of the genderedness metric. This is in agreement with the observations made by Böhm et al. [5] about the the words in such keyword repositories being descriptors of specific requirements of IT jobs. Furthermore, the computer science industry is highly standardized in its language and often uses jargon and technical words that may not strongly correlate with traditional gendered language. This uniformity in vocabulary across job ads might reduce the variation needed for significant effects to emerge across different variables.

Interactions between genderedness and the combination of some variables (e.g., country and work model, industry and remote work) did yield statistically significant results. However, even these had low practical effect sizes. This again reinforces the possibility that either the linguistic features used were insufficiently discriminative, or gender bias in the language may be evenly distributed across countries, industries, roles, and work models, rather than being concentrated in any particular group, reducing the likelihood of detecting statistically significant differences.

This work also sought to answer whether feminine- and masculine-coded jobs are exposed differently through job recommendation algorithms and whether there is a correlation between applicant gender and the genderedness of the job advertisement. Recommendation algorithms vary in how they expose job seekers to masculine and feminine-coded job advertisements, with significant disparities across models. The TF-IDF model exhibited the highest disparity-based exposure and consistently recommended a disproportionate number of feminine-coded jobs to both male and female applicants. In contrast, all Word2Vec-based models showed substantially lower disparity-based exposure, with Word2Vec-SkipGram offering the most balanced exposure of masculine and feminine-coded roles. It is also the only variation of Word2Vec that recommends more feminine-coded jobs to women than masculine-coded ones.

This indicates that TF-IDF amplifies superficial patterns in word use and potentially overfits stylistic features of the synthetic CVs since it is a frequency-based model that does not account for contextual meaning. It counts the frequency of terms adjusted by document frequency, without understanding their meaning or context. In contrast, the Word2Vec model learns word embeddings based on surrounding words and captures deeper semantic similarity between job descriptions and resumes. This may reduce exposure bias.

Interestingly, incorporating bigrams into Word2Vec models also led to increased disparity-based exposure. This suggests that bigrams may amplify existing gender biases embedded in job postings and gendered expressions may be better preserved and emphasized in bigram form. Instead of generalizing across diverse phrases, using bigrams sharpens the model's focus on exact phrase matches which may reflect gender bias in job descriptions. As a result, recommendation models using bigrams may more strongly favor resumes aligned with masculine-coded language.

This suggests that the choice of embedding model in job recommender systems has a direct impact on fairness. Simpler models like TF-IDF may inadvertently expose one group over another. More sophisticated, context-aware embeddings like Word2Vec may lead to more equitable outcomes. This reinforces the importance of embedding selection in the design of fair algorithmic hiring tools.

While there is a statistically significant relationship between applicant gender and the genderedness of recommended jobs, the actual strength of this association is weak across all recommendation algorithms. The low Cramér's V values show that gender influences recommendations, but not strongly.

These findings imply that recommendation algorithms are not entirely neutral: even when using the same input data, different algorithms can produce meaningfully different outcomes in terms of gender bias. Algorithms like TF-IDF, which lack semantic understanding, may unintentionally reinforce or distort bias patterns. In contrast, context-aware models like Word2Vec-SkipGram may minimize the alignment of recommendations with gender stereotypes.

This underscores the importance of algorithmic design choices in mitigating bias. A model that appears fair based on technical metrics (e.g., accuracy) may still introduce or perpetuate subtle biases unless its outputs are also evaluated through fairness-oriented lenses.

These findings carry significant implications for both recruiters and developers of recommender systems. Job descriptions should be audited and refined to reduce inadvertent gender coding, particularly in industries and roles higher gender gaps in participation. Furthermore, recommender systems must be evaluated not only on relevance and accuracy but also on fairness metrics like disparity-based exposure and correlation with user demographics. Developers should consider incorporating fairness-aware embeddings to reduce bias amplification.

In conclusion, to promote true inclusivity, companies and hiring platforms must adopt more gender-neutral language and actively monitor the words used in job advertisements. Addressing gender bias in hiring requires a conscious shift in how roles are described and presented to potential candidates. By deconstructing gendered language patterns, we can pave the way for more equitable representation across all levels of jobs related to computer science.

Several limitations should be considered when interpreting the findings of this study. First, the study assumes a binary definition of gender, which does not account for the full spectrum of gender identities and may exclude important nuances.

Second, the method used to measure genderedness in job advertisements is based on a repository of gender-coded words which may not be fully sensitive to the technical context of the computer science domain. Although some high-frequency words were removed, others may remain.

Duplicates were removed from the data to avoid re-posted job advertisements, but this process may have excluded job advertisements for distinct roles with the same titles and descriptions. Furthermore, while dummy companies with placeholder content were identified and removed, it is possible that other such entries remain in the data. Moreover, some industries were not found while scraping data since some companies' LinkedIn pages have been deleted.

Additionally, the study was limited to English-language job advertisements, which may have excluded a significant number of relevant postings in non-English-speaking countries, thereby narrowing the geographic and cultural scope of the analysis.

Finally, because synthetic CVs were generated using specific types of words to simulate female and male applicants, there is a possibility that these linguistic styles unintentionally aligned with the same gendered words used to label the job advertisements. This may have influenced similarity scores in the recommendation stage.

6 Future Work

This thesis opens several avenues for future research. First, future work can develop a computer science-specific repository. This could involve conducting human experiments to evaluate perceived gender associations of words within technical contexts.

Second, a skill-based analysis of gender bias could be valuable. Instead of focusing solely on word-level semantics, future work could explore whether certain technical skills (e.g., programming languages, frameworks, tools) are more commonly associated with feminine- or masculine-coded job descriptions and whether this association influences applicant behavior or job recommendations.

Additionally, more robust mechanisms for duplicate detection and removal should be developed to distinguish between reposted ads and different roles posted with similar job titles. Improved filtering of dummy or inactive company profiles would also strengthen data reliability.

Another direction involves using real-world resumes or anonymized CV data to validate the patterns observed with synthetic profiles. This would address concerns about whether synthetic CVs inadvertently mirrored the gendered language used to label job advertisements, which may have influenced recommendation outcomes.

Finally, the current study only used content-based recommendation algorithms. Future work could explore collaborative filtering, hybrid approaches, or fairness-aware algorithms, and evaluate them using fairness metrics to better understand their behavior in employment contexts. Incorporating intersectional variables (e.g., race, sexual orientation, etc.) may also reveal how multiple forms of bias interact in algorithmic hiring systems.

References

[1] David Allen, James Van Scotter, and Robert Otondo. "Recruitment communication media: Impact on prehire outcomes". In: *Personnel Psychology* 57 (July 2007), pp. 143–171. DOI: 10.1111/j.1744-6570.2004.tb02487.x.

[2] Inger Askehave and Karen Korning Zethsen. "Gendered Constructions of Leadership in Danish Job Advertisements". In: *Gender, Work & Organization* 21.6 (2014), pp. 531–545.

[3] Simone de Beauvoir. *The Second Sex (Vintage classics)*. 1949.

[4] Sandra L Bem. "The measurement of psychological androgyny." In: *Journal of consulting and clinical psychology* 42.2 (1974), p. 155.

[5] Stephan Böhm, Olena Linnyk, Jens Kohl, Tim Weber, Ingolf Teetz, Katarzyna Bandurka, and Martin Kersting. "Analysing Gender Bias in IT Job Postings: A Pre-Study Based on Samples from the German Job Market". In: *Proceedings of the 2020 Computers and People Research Conference*. SIGMIS-CPR '20. Nuremberg, Germany: Association for Computing Machinery, 2020, pp. 72–80. ISBN: 9781450371308. DOI: 10.1145/3378539.3393862. URL: https://doi.org/10.1145/3378539.3393862.

[6] M. Born and T. Taris. "The impact of the wording of employment advertisements on students' inclination to apply for a job". In: *Journal of Social Psychology* 150.5 (2010), pp. 485–502. DOI: 10.1080/00224540903365422.

[7] Lera Boroditsky. "Does language shape thought?: Mandarin and English speakers' conceptions of time". In: *Cognitive psychology* 43.1 (2001), pp. 1–22.

[8] Sheila Brownlow, Julie A Rosamond, and Jennifer A Parker. "Gender-linked linguistic behavior in television interviews". In: *Sex Roles* 49 (2003), pp. 121–132.

[9] Linda L Carli. "Gender, language, and influence." In: *Journal of personality and social psychology* 59.5 (1990), p. 941.

[10] Dawn R Deeter-Schmelz, Andrea L Dixon, Robert C Erffmeyer, Kyoungmi Kim, Raj Agnihotri, Michael T Krush, and Ellen Bolman Pullins. "Attracting students to sales positions: The case of effective salesperson recruitment ads". In: *Journal of Marketing Education* 42.2 (2020), pp. 170–190.

[11] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. "Fairness in recommender systems: research landscape and future directions". In: *User Modeling and User-Adapted Interaction* 34.1 (2024), pp. 59–108.

[12] Alice H Eagly and Steven J Karau. "Gender and the emergence of leaders: A meta-analysis." In: *Journal of personality and social psychology* 60.5 (1991), p. 685.

[13] Michael D Ekstrand, Mucun Tian, Mohammed R Imran Kazi, Hoda Mehrpouyan, and Daniel Kluver. "Exploring author gender in book rating and recommendation". In: *Proceedings of the 12th ACM conference on recommender systems*. 2018, pp. 242–250.

[14] Gráinne M Fitzsimons and Aaron C Kay. "Language and interpersonal cognition: Causal effects of variations in pronoun usage on perceptions of closeness". In: *Personality and Social Psychology Bulletin* 30.5 (2004), pp. 547–557.

[15] Richard Frissen, Kolawole Adebayo, and Rohan Nanda. "A machine learning approach to recognize bias and discrimination in job advertisements". In: *AI SOCIETY* 38 (Oct. 2022), pp. 1–14. DOI: 10.1007/s00146-022-01574-0.

[16] Danielle Gaucher, Justin P Friesen, and Aaron C. Kay. "Evidence that gendered wording in job advertisements exists and sustains gender inequality." In: *Journal of personality and social psychology* 101 1 (2011), pp. 109–28. URL: https://doi.org/10.1037/a0022530.

[17] Preetam Ghosh and Vaishali Sadaphal. “JobRecoGPT–Explainable job recommendations using LLMs”. In: *arXiv preprint arXiv:2309.11805* (2023).

[18] Madeline E Heilman. “Sex bias in work settings: The lack of fit model.” In: *Research in organizational behavior* (1983).

[19] Tanja Hentschel, Susanne Braun, Claudia Peus, and Dieter Frey. “Sounds like a fit! Wording in recruitment advertisements and recruiter gender affect women’s pursuit of career development programs via anticipated belongingness”. In: *Human Resource Management* 60 (Nov. 2020). DOI: 10.1002/hrm.22043.

[20] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. “Evaluating collaborative filtering recommender systems”. In: *ACM Transactions on Information Systems (TOIS)* 22.1 (2004), pp. 5–53.

[21] Neil Hurley and Mi Zhang. “Novelty and diversity in top-n recommendation–analysis and evaluation”. In: *ACM Transactions on Internet Technology (TOIT)* 10.4 (2011), pp. 1–30.

[22] Aaron C Kay, Danielle Gaucher, Jennifer M Peach, Kristin Laurin, Justin Friesen, Mark P Zanna, and Steven J Spencer. “Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be.” In: *Journal of personality and social psychology* 97.3 (2009), p. 421.

[23] Aaron C Kay and Mark P Zanna. “A contextual analysis of the system justification motive and its societal consequences”. In: *Social and psychological bases of ideology and system justification* (2009), pp. 158–181.

[24] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. “Explaining the user experience of recommender systems”. In: *User modeling and user-adapted interaction* 22 (2012), pp. 441–504.

[25] Linda A Krefting and Philip K Berger. “Masculinity-femininity perceptions of job requirements and their relationship to job-sex stereotypes”. In: *Journal of Vocational Behavior* 15.2 (1979), pp. 164–174. ISSN: 0001-8791. DOI: https://doi.org/10.1016/0001-8791(79)90035-6.

[26] Juan M Madera, Michelle R Hebl, and Randi C Martin. “Gender and letters of recommendation for academia: agentic and communal differences.” In: *Journal of Applied Psychology* 94.6 (2009), p. 1591.

[27] Masoud Mansoury, Himan Abdollahpouri, Jessie Smith, Arman Dehpanah, Mykola Pechenizkiy, and Bamshad Mobasher. “Investigating Potential Factors Associated with Gender Discrimination in Collaborative Recommender Systems”. In: *FLAIRS*. 2020, pp. 193–196.

[28] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. “Distributed representations of words and phrases and their compositionality”. In: *Advances in neural information processing systems* 26 (2013).

[29] Matthew L Newman, Carla J Groom, Lori D Handelman, and James W Pennebaker. “Gender differences in language use: An analysis of 14,000 text samples”. In: *Discourse processes* 45.3 (2008), pp. 211–236.

[30] Linnea Ng, Juan M Madera, Stephanie Zajac, and Michelle Hebl. “When words matter: Communal and agentic language on men and women’s resumes”. In: *Academy of management proceedings*. Vol. 2020. 1. Academy of Management Briarcliff Manor, NY 10510. 2020, p. 20462.

[31] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. “Psychological aspects of natural language use: Our words, our selves”. In: *Annual review of psychology* 54.1 (2003), pp. 547–577.

[32] Margreet Reitsma-van Rooijen, Gün R Semin, and Esther Van Leeuwen. “The effect of linguistic abstraction on interpersonal distance”. In: *European Journal of Social Psychology* 37.5 (2007), pp. 817–823.

[33] Maranda Ridgway, Louise Oldridge, and Sharon Mavin. “‘Leading’ by example? Gendered language in Human Resource job adverts”. In: *Human Resource Management Journal* 35 (Mar. 2024), pp. 1–24. DOI: 10.1111/1748-8583.12549.

[34] Paul Rosenkrantz, Susan Vogel, Helen Bee, Inge Broverman, and Donald M Broverman. "Sex-role stereotypes and self-concepts in college students." In: *Journal of consulting and clinical psychology* 32.3 (1968), p. 287.

[35] Max Roser and Esteban Ortiz-Ospina. *Female Labor Supply*. `https://ourworldindata.org/female-labor-supply`. Our World in Data. Published online. 2019. URL: `https://ourworldindata.org/female-labor-supply` (visited on 05/01/2025).

[36] Laurie A Rudman and Stephen E Kilianski. "Implicit and explicit attitudes toward female authority". In: *Personality and social psychology bulletin* 26.11 (2000), pp. 1315–1328.

[37] Gerard Salton and Christopher Buckley. "Term-weighting approaches in automatic text retrieval". In: *Information processing & management* 24.5 (1988), pp. 513–523.

[38] Shrikant Saxena and Shweta Jain. "Exploring and mitigating gender bias in book recommender systems with explicit feedback". In: *Journal of Intelligent Information Systems* 62.5 (2024), pp. 1325–1346.

[39] Toni Schmader, Jessica Whitehead, and Vicki H Wysocki. "A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants". In: *Sex roles* 57.7 (2007), pp. 509–514.

[40] Mihwa Seong and Simon Parker. "Does gendered wording in job advertisements deter women from joining start-ups? A replication and extension of Gaucher, Friesen, and Kay (2011)". In: *Strategic Entrepreneurship Journal* 18 (Nov. 2023). DOI: `10.1002/sej.1489`.

[41] Dougal Shakespeare, Lorenzo Porcaro, Emilia Gómez, and Carlos Castillo. "Exploring artist gender bias in music recommendation". In: *arXiv preprint arXiv:2009.01715* (2020).

[42] Guy Shani and Asela Gunawardana. "Evaluating recommendation systems". In: *Recommender systems handbook* (2011), pp. 257–297.

[43] Jim Sidanius and Felicia Pratto. "Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression". In: *Cambridge University Press* (1999). DOI: `10.13140/RG.2.2.20242.53440`.

[44] Janet T Spence, Robert Helmreich, and Joy Stapp. "Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity." In: *Journal of personality and social psychology* 32.1 (1975), p. 29.

[45] Shiliang Tang, Xinyi Zhang, Jenna Cryan, Miriam J. Metzger, Haitao Zheng, and Ben Y. Zhao. "Gender Bias in the Job Market: A Longitudinal Analysis". In: 1.CSCW (Dec. 2017). DOI: `10.1145/3134734`. URL: `https://doi.org/10.1145/3134734`.

[46] Toon Taris and Inge Bok. "On Gender Specificity of Person Characteristics in Personnel Advertisements: A Study Among Future Applicants". In: *The Journal of psychology* 132 (Dec. 1998), pp. 593–610. DOI: `10.1080/00223989809599292`.

[47] Rayla Tokarz and Tati Mesfin. "Stereotyping Ourselves: Gendered Language Use in Management and Instruction Library Job Advertisements". In: *Journal of Library Administration* 61 (Feb. 2021), pp. 1–11. DOI: `10.1080/01930826.2021.1883368`.

[48] Arthur Turrell, Bradley Speigner, Jyldyz Djumalieva, David Copple, and James Thurgood. *Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings*. Tech. rep. National Bureau of Economic Research, 2019.

[49] Jorge Carlos Valverde-Rebaza, Ricardo Puma, Paul Bustios, and Nathalia C Silva. "Job Recommendation Based on Job Seeker Skills: An Empirical Study." In: *Text2Story@ ECIR*. 2018, pp. 47–51.

[50] Lien Wille and Eva Derous. "When job ads turn you down: how requirements in job ads may stop instead of attract highly qualified women". In: *Sex roles* 79 (2018), pp. 464–475.

[51] Shuo Zhang and Peter Kuhn. "Understanding algorithmic bias in job recommender systems: An audit study approach". In: *American Economic Association* 12 (2022).

[52] Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C Aggarwal, and Tyler Derr. "Fairness and diversity in recommender systems: a survey". In: *ACM Transactions on Intelligent Systems and Technology* 16.1 (2025), pp. 1–28.

[53] Ziwei Zhu, Xia Hu, and James Caverlee. “Fairness-aware tensor-based recommendation”. In: *Proceedings of the 27th ACM international conference on information and knowledge management*. 2018, pp. 1153–1162.

A

Mapping of Industries in the Dataset

- **Technology and IT:** software development, it services and it consulting, computer and network security, technology, information and internet, technology, information and media, computer games, computer networking products, mobile gaming apps, internet publishing, internet marketplace platforms, desktop computing software products, mobile computing software products, data infrastructure and analytics, data security software products, it system custom software development, it system data services, it system training and support, blockchain services, wireless services, business intelligence platforms, telecommunications, information services
- **Healthcare and Life Sciences:** hospitals, hospitals and health care, pharmaceutical manufacturing, biotechnology research, mental health care, medical practices, medical and diagnostic laboratories, home health care services, veterinary services, alternative medicine, health and human services, physicians, public health
- **Manufacturing:** defense and space manufacturing, oil and gas, semiconductor manufacturing, industrial machinery manufacturing, medical equipment manufacturing, machinery manufacturing, motor vehicle parts manufacturing, computers and electronics manufacturing, textile manufacturing, computer hardware manufacturing, renewable energy semiconductor manufacturing, appliances, electrical, and electronics manufacturing, sporting goods manufacturing, chemical manufacturing, motor vehicle manufacturing, automation machinery manufacturing, food and beverage manufacturing, beverage manufacturing, tobacco manufacturing, dairy product manufacturing, packaging and containers manufacturing, furniture and home furnishings manufacturing, glass, ceramics and concrete manufacturing, manufacturing, paper and forest product manufacturing, plastics manufacturing, climate technology product manufacturing, measuring and control instrument manufacturing
- **Education and Research:** higher education, research services, e-learning providers, education administration programs, education, primary and secondary education, libraries, think tanks, nanotechnology research
- **Finance and Professional Services:** financial services, insurance, accounting, market research, capital markets, investment banking, investment management, venture capital and private equity principals, business consulting and services, human resources services, professional services, strategic management services, executive search services, operations consulting, banking
- **Media and Entertainment:** advertising services, public relations and communications services, entertainment providers, broadcast media production and distribution, media production, online audio and video media, musicians, movies, videos, and sound, animation and post-production, book and periodical publishing, newspaper publishing, internet news, blogs, performing arts, artists and writers, writing and editing, photography, business content
- **Retail and Consumer Services:** retail, retail luxury goods and jewelry, retail apparel and fashion, retail groceries, retail office equipment, retail art supplies, consumer services, restaurants, food and beverage services, food and beverage retail, personal care product manufacturing

- **Engineering and Construction:** civil engineering, engineering services, construction, architecture and planning, building construction, robotics engineering
- **Transportation and Logistics:** aviation and aerospace component manufacturing, airlines and aviation, truck transportation, maritime transportation, rail transportation, railroad equipment manufacturing, urban transit services, freight and package transportation, transportation, logistics, supply chain and storage, ground passenger transportation, vehicle repair and maintenance, transportation equipment manufacturing
- **Energy and Environment:** utilities, renewable energy power generation, solar electric power generation, electric power generation, renewable energy equipment manufacturing, services for renewable energy, environmental services, climate data and analytics
- **Nonprofit, Government, and Policy:** non-profit organizations, civic and social organizations, government administration, international affairs, public safety, public policy offices, government relations services, political organizations, philanthropic fundraising services, religious institutions, fundraising, armed forces, legislative offices, law enforcement, administration of justice
- **Law and Legal Services:** law practice, legal services, alternative dispute resolution
- **Arts and Design:** graphic design, interior design
- **Hospitality and Leisure:** hospitality, travel arrangements, gambling facilities and casinos, recreational facilities, events services, museums, museums, historical sites, and zoos, spectator sports, wellness and fitness services
- **Agriculture and Natural Resources:** farming, horticulture, fisheries, mining
- **Real Estate and Property:** real estate, leasing non-residential real estate, wholesale building materials
- **Wholesale and Trade:** wholesale, wholesale import and export
- **Staffing and Human Capital:** staffing and recruiting, executive offices
- **Miscellaneous:** printing services, design services, animation and post-production, professional training and coaching, outsourcing and offshoring consulting, social networking platforms, individual and family services, desktop computing software products, fire protection, warehousing and storage, home health care services, taxi and limousine services, pet services, shipbuilding, digital accessibility services, security and investigations, international trade and development, holding companies, equipment, translation and localization, facilities services

B

Mapping of Roles in the Dataset

The following roles are present in the data along with all the job titles categorized within them (if any):

- **Front-End:** Front-End Developer, TypeScript Developer, React Developer, JavaScript Developer, Web Developer
- **DevOps Specialist:** Solution Delivery, DevOps Engineer
- **Data and Analytics:** Data Engineer, Data or Business Analyst, BI Developer, Oracle Developer, Data Architect, Database Developer, SQL Developer, Spark Developer, Hadoop Developer, SAS Developer, Statistical Programmer, VBA Developer, Tibco Developer
- **Full-Stack Developer:** Full-Stack Engineer
- **Support Role:** Support Role, Customer Service, Professional Services, Marketing or Sales Professional
- **Data Scientist or Machine Learning Specialist:** Data Scientist, Machine Learning Engineer
- **IT Infrastructure & Operations:** Infrastructure Engineer, IT Engineer, Technician, Systems Engineer, Integration Engineer, IT Administrator, CRM Developer
- **Security Professional:** Cybersecurity Engineer
- **Database Administrator**
- **Site Reliability Engineer**
- **Back-End:** .Net Developer, Java Engineer, Django Developer, GIS Developer, Kotlin Developer, Python Engineer, C# Developer, PHP Developer, C++ Developer, Rust Developer, Ruby Developer, JEE/J2EE Developer, ASP Developer, C Developer, Software Developer, Scala Developer, Back-End Developer
- **System Administrator**
- **Project Manager**
- **Senior Executive (C-Suite, VP, etc.)**
- **Mobile Developer**
- **Game or Graphics Developer**
- **Academic:** Researcher, Academic
- **Other:** All other roles like IT Auditor, Matlab Developer, Scrum Master, etc.
- **Manager**
- **Designer**
- **Architect:** Software/System/IT Architect, Solution Architect, Cloud Computing Architect
- **QA or Test Developer**

- **Desktop or Enterprise Applications Developer**
- **Consultant:** SAP Consultant, Business Consultant, or Other Consultant Role
- **Content Management System:** Drupal Developer, WordPress Developer
- **Embedded Applications or Devices Developer**

C

List of Words in the Repository

C.1. Feminine Words

- advice
- advocate
- affection
- attentive
- care
- child
- cheer
- collaborate
- communicate
- compassion
- compile
- commit
- communal
- communicate
- compassionate
- compile
- commit
- communal
- compassionate
- communicate
- compassionate
- compile
- commit
- communal
- considerate
- conscience
- connect
- contribute

- cooperate
- copy
- counsel
- creative
- dependable
- dialogue
- divert
- emotional
- empathy
- empower
- enable
- encourage
- enthusiasm
- establish
- excellent
- feminine
- flatter
- flexible
- gentle
- heart
- helpful
- honest
- ideal
- implement
- interpersonal
- interdependent
- intuit
- kind
- kinship
- listen
- loyal
- manipulate
- modest
- motivate
- nag
- nurture
- participate
- personal
- persuade
- pleasant
- pleasure
- polite

- quiet
- rapport
- relation
- responsible
- response
- role model
- sensible
- sense
- sensitive
- serve
- service
- shape
- share
- sincere
- social
- society
- sociable
- submissive
- sustainable
- sympathy
- teamwork
- tend
- tender
- together
- trust
- value
- warm
- whine
- yield

C.2. Masculine Words

- achieve
- active
- adventure
- aggress
- ambition
- ambassador
- aspiration
- assert
- athlete
- autonomy
- boast

- business sense
- challenge
- command
- commercial
- compare
- compete
- compute
- confident
- control
- coordinate
- courage
- credible
- decide
- defend
- deliver
- demand
- determine
- detail-oriented
- direct
- dominate
- drive
- entrepreneur
- exceptional
- excellent
- feed
- force
- goal-oriented
- greed
- handle
- headstrong
- hierarchy
- hostile
- impressive
- impulsive
- independent
- individual
- influence
- initiative
- instruct
- intellect
- intelligent
- logic

- masculine
- mentor
- negotiate
- objective
- offbearing
- opinion
- operate
- organizational
- organize
- outspoken
- outstanding
- perform
- persist
- power
- principle
- professional
- proud
- push
- quality
- rational
- reckless
- recruit
- reputation
- resolution
- robust
- self-confident
- self-reliant
- self-sufficient
- seize
- setting-up
- strategy
- stubborn
- success
- superior
- supervise
- synthesize
- task-oriented
- transform
- win
- world class

C.3. Words Removed from Repository

The following words have been removed from the keyword repository due to a higher frequency of occurrence in computer science-related job advertisements. The words are followed by examples of usage in a general context and a computer science context.

C.3.1. Analysis

"...Planning and Financial Analysis: Participate in annual planning, long-term forecasting..." (Strategic planning and forecasting) "...spending initiatives, ad hoc financial analysis and monthly, quarterly and annual planning..." (Financial planning and budgeting)

"Extensive experience as analyst in Systems Analysis and Design..." (Software/system architecture and engineering design) "...incidents, drive investigations, security analysis, monthly metrics/reports..." (cybersecurity operations and threat investigation)

C.3.2. Best

"We know our best asset is our people!" (people-centric language) "One of the best small daily newspapers..." (company reputation)

"Develop routines for end users to facilitate best practices for database use." (technical standards and practices or problem-solving ability)

C.3.3. Develop

"Plan and manage at both the strategic and operational levels..." (leadership and organizational development) "Grow and develop associates..." (Talent development and mentoring)

"Design, develop, maintain, monitor and execute production ETL processes..." (Focuses on software development)

C.3.4. Lead

"Lead by example – Be the change you want to see..." (Inspirational/motivational use of lead as a behavioral value)

"Code and test UI and backend applications. Lead and represent a scrum team..." (Leading in a technical team setting)

C.3.5. Strong

"Strong communication and interpersonal skills..." (interpersonal and communication skills)

"Should have strong Teradata-SQL writing skills..." (technical skills)

C.3.6. Support

"Joining HR Block means you'll have the support of an expert team..." (Emotional and training support) "Product support (online, phone and/or onsite support)"' (Customer-facing support, more service/maintenance focused)

"The successful Support Engineer does much more than plug computers together or track changes." (technical support role requiring systems knowledge) "Support IT and Operations in database monitoring, capacity planning, and performance tuning." (technical support for backend infrastructure)

C.3.7. Teams

"Work closely with product/solutions, digital, content teams, field marketing and the sales teams to inform and scale your efforts" (cross-functional coordination, involving marketing, sales, and support functions)

"Proven ability to lead a team of Frontend Web Engineers." "Provide technical guidance for junior members of the team..." (technical collaboration in software development, testing, and deployment processes)

C.3.8. Understanding

“Understanding of food production and fundamental cooking techniques...” (knowledge in culinary skills and food preparation processes)

“Understanding of Collaboration, Security, SDN, server and desktop virtualization...” (technical knowledge in various IT domains)