UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Reducing Gender Bias in LinkedIn Job Ads: The Impact of Automatic Rewriting on Candidate Diversity

Author: Ghilain Audrey
Supervisor: Vande Kerckhove C.
Academic year 2024-2025
Dissertation for the Master of Business Engineering, Major Business Analytics
Daytime schedule

Declaration Regarding AI Tool Usage in Master's Thesis

We recognise that AI tools might be valuable aids during the master's thesis work, but they are not infallible. Remember that transparency fosters trust, and acknowledging AI's role enhances the credibility of your work.

Therefore, when deciding to use such a tool, you need to adhere to the following principles of responsible use of AI.

1. **Critical Evaluation:**
    - We critically assessed the AI-generated output, ensuring its alignment with our research objectives.
    - Any modifications or corrections were made based on our expertise and domain knowledge.
2. **Transparency:**
    - We acknowledge the use of ChatGPT transparently, emphasising that it contributed to our work but did not replace human judgment.
    - Our commitment to transparency ensures the integrity of this thesis.
3. **Ethical Considerations:**
    - We actively monitored for biases or unintended consequences introduced by the AI tool.
    - Our ethical responsibility guided our decisions throughout the research process.

Declaration

During the preparation of this master's thesis, the author(s) utilised ChatGPT for the following purpose:

1. Help with coding: Using this AI to fix errors and improve the performance of some Python code.

2. **English Corrections**: Using AI with complex sentence in English to be sure of their meaning.

After using ChatGPT, the author diligently reviewed and edited the content produced by the tool. We take full responsibility for the final content presented in this thesis.

By signing this declaration, we affirm that the content of this master's thesis reflects our original work, augmented by the responsible use of AI.

18/06/2025 Ghilain Audrey

Acknowledgements

Firstly, I would like to thank Professor Corentin Vande Kerckhove for his help and valuable advice during my thesis.

I would also like to thank my family and friends, who have always supported me in my projects.

Finally, I would like to thank my friends from my exchange program, who helped me improve my English.

Audrey Ghilain

Table of contents

List of Figures

List of Tables

Introduction

Gender inequality in the world of work is a persistent issue, despite numerous social and legislative advances. Although women now have access to more qualified and diversified positions, they remain largely under-represented in senior and leadership positions. This phenomenon can be explained by a combination of structural, cultural and linguistic factors, which contribute to slowing down their career progression. Among these factors, gender bias in the language used in job advertisements (job ad) is an often invisible but particularly discriminatory factor.

Recent studies have shown that the way in which a job ad is formulated has a direct influence on how it is perceived by candidates. The choice of words, phrasing and tone can convey, sometimes unconsciously, gendered stereotypes that favour certain profiles to the detriment of others. For example, language that is perceived as competitive, assertive or technical may discourage potentially qualified candidates, while a more cooperative and inclusive style may, on the contrary, broaden the audience of applicants. In the age of algorithmic recommendation systems, where lexical correspondence between advertisements and profiles is a key factor, these dynamics are even more important to take into account in the matching process.

This work contributes to addressing the problem. It aims to evaluate the impact of a linguistic intervention on job advertisements, using an automatic text debiasing method, on the composition of recommended profiles. In other words, the aim is to observe whether reformulating biased descriptions in more neutral language modifies the behaviour of a recommendation system based on textual similarity.

Using a database of jobs published on LinkedIn and a corpus of user profiles, this dissertation begins by detecting gender bias in the ads, applies a semantic neutralisation method, and measures changes in recommendation results before and after debiasing.

1. Setting the scene

This chapter presents the social and historical context of inequalities between men and women in the workplace. It highlights the structural and cultural factors that continue to influence women's access to positions of responsibility, despite progress made in education and equal rights.

Over the last few decades, the position of women in the workplace has undergone significant change. Thanks to their increased access to education and higher education, women have been able to take on qualified positions and gradually enter sectors that were previously the preserve of men. This development, while significant, has not eliminated gender inequalities. Women are still under-represented in senior positions (Milewski, 2004).

Since the mid-1960s, women have continued to fight to improve their social conditions in a society dominated by patriarchal structures. This progress has been made possible by greater access to education, giving them the opportunity to improve their level of education and gradually enter a variety of professional sectors. However, despite this progress, stereotypes continue to influence educational choices, leading many young girls to limit themselves in their career aspirations (Milewski, 2004). In fact, women tend to underestimate their abilities and go more for training in teaching or the paramedical professions, while men favour fields perceived as more prestigious and lucrative, such as engineering or medicine (Châteauneuf-Malclès, 2011). This trend, which stems from social and cultural norms, has long hindered women's access to the most highly valued sectors of the labour market. While the educational gap between the sexes has gradually narrowed and current trends show a reversal of certain inequalities in favour of women in terms of educational success, women still face structural obstacles.

Gender inequalities persist in all areas and at all stages of the professional career. As soon as women enter the labour market, gaps emerge in terms of employment conditions, pay and career prospects. Data shows that, for equal skills, women face more obstacles in reaching the top echelons of organisations (Huang et al., 2020). More generally, women are often destined for middle management positions, while the most senior positions are held by men.

Discrimination mechanisms, which are often subtle, help to reinforce these inequalities. The “glass ceiling” phenomenon is a good illustration of the way in which organisations perpetuate a masculine inter-silo, in which men are mainly favoured for access to positions of

responsibility. A study carried out in the construction industry showed that men were four times more likely than women to reach management positions within the same company (Hickey et al., 2022). This finding is in line with analyses of the senior civil service, where appointments to the most prestigious positions are still largely male, despite a growing proportion of eligible women (Milewski, 2004).

The adoption of equality policies and anti-discrimination laws has not been enough to restore a real balance. Despite efforts to introduce corrective measures such as representation quotas and pay transparency obligations, pay and promotion gaps remain significant (Ministry for Equality between Women and Men and the Fight against Discrimination, 2025). On average, women still earn 16% less than their male counterparts, even for the same skills and responsibilities (Minot et al., 2023). In the United States, women earn on average 84 cents for every dollar earned by men, a gap which, although it narrowed in the 1980s and 1990s, has stagnated in recent decades. This wage disparity is partly attributed to persistent structural biases, in particular the systemic devaluation of female-dominated sectors and the more frequent professional sacrifices made by women for family reasons (Minot et al., 2023). Indeed, career interruptions linked to family responsibilities are perceived as negative signals by employers, thus contributing to slowing down women's access to strategic positions (Hickey et al., 2022).

The study of gender inequalities at work reveals that the under-representation of women in positions of responsibility is not only the result of institutional discrimination or a lack of skills, but also of differences in the way they present themselves to recruiters and react to job offers. Indeed, recent research shows that women tend to use a different vocabulary in their job applications than men, which can influence their perception by employers and, ultimately, their career progression (Minot et al., 2023). For example, an analysis of millions of CVs in the United States showed that women use less assertive language and give less detail about their professional achievements than men. These linguistic differences play a decisive role in the selection process and help to explain some of the differences in pay and promotion between the sexes.

This research suggests that targeted interventions, such as CV-writing training and mentoring programmes aimed at deconstructing the stereotypes associated with leadership positions, could play a key role in reducing these gaps.

2. Literary Review

Having set out the context of gender inequality in the world of work and highlighted the structural obstacles to women's access to senior positions, the following section reviews existing literature on the topic. As a reminder, the main objective of this dissertation is to analyse the mechanisms that perpetuate the under-representation of women in decision-making spheres, with particular emphasis on the impact of bias in the recruitment and career development processes. In this literature review, we will examine the factors identified in previous research, in particular differences in the way applications are written, in order to better understand their influence on women's career progression. We will then present the main methodological approaches used to analyse these disparities, in particular studies based on textual analysis of job advertisements. Finally, we will review the work that has already explored these dimensions and their implications.

2.1. Causes et implications

The study of gender inequalities at work reveals that men and women do not react in the same way to job offers (Hu et al., 2022). Job advertisements can therefore convey gender stereotypes through the vocabulary and wording used. Numerous studies even suggest that biases in the wording of advertisements contribute to professional inequalities between men and women (Minot et al., 2023). More specifically, traditionally male jobs tend to be described using stereotypically “masculine” lexicon, reflecting agentic qualities (e.g. leader, competitive, ambitious, independent) associated with men, whereas advertisements in feminised sectors make less use of this type of vocabulary (Gaucher et al., 2011). This subtle differentiation in language was highlighted by Gaucher and his colleagues, who showed that job advertisements in male-dominated fields contain significantly more terms associated with male stereotypes (e.g. dynamic, leadership) than those in feminised fields, although there are not necessarily more “feminine” terms in the latter (Gaucher et al., 2011). The choice of words is generally not intentionally discriminatory but is often the result of unconscious bias on the part of writers and historical norms specific to each sector of activity.

Furthermore, a study of LinkedIn profiles revealed that women tend to underestimate their skills by emphasising their emotions, while their male counterparts focus on their status (Simon et al., 2023). Analysis of candidate profiles in the IT field has shown that men list more technical skills and repeat them several times in their profiles, thereby increasing their visibility to

recruiters using automatic filtering algorithms. This difference in the way they present themselves results in women being under-represented in search results and reduces their chances of being shortlisted for interviews.

Finally, in grammatically gendered languages such as French or German, the use of the generic masculine in job titles (e.g. Project Manager in the masculine) has long made women invisible. Some countries and companies now encourage gender inclusive formulations or the use of both masculine and feminine forms to limit this linguistic bias (Hodel et al., 2017). All of these factors, stereotypes rooted in language, biased choices, grammatical conventions, are causes of the gender bias observed in job advertisements.

In addition to language-related biases introduced by human recruiters or editors, the growing adoption of artificial intelligence in recruitment brings its own set of challenge. By 2024, around 99% of Fortune 500 companies reportedly use some form of automation in their hiring process (Wilson & Caliskan, 2024). Often, AI is promoted as a solution to human subjectivity, however it is not immune to bias especially when trained on historical data that reflects existing inequalities. A well-known example is Amazon's experimental recruitment tool, which began favouring male candidates simply because the training data, based on previous hiring decisions, overrepresented men in technical roles (Kumar et al., 2023). The system was not explicitly programmed to discriminate, but it learned from biased patterns and unintentionally reproduced them. This case makes it clear that AI does not operate in a vacuum. Behind each algorithm are human choices, about which data to use, the results to favour, and the assumptions to incorporate. If those choices are not made carefully and transparently, even advanced systems can end up reinforcing the disparities they were intended to mitigate. As Wilson and Caliskan (2024) argue, responsible AI in recruitment requires more than just technical optimisation; it involves proactive fairness audits, clear accountability, and a critical awareness of the social context in which these tools operate. Without this, there is a real risk of automating discrimination under the illusion of neutrality.

2.2. Consequences

Gender bias in job advertisements has multiple effects on professional equality. Several studies, such as those by Gaucher et al. (2011) and Hentschel et al. (2020), have shown that women are less likely to apply for jobs that are written in strongly masculine language, as they do not feel they meet the expectations required for the job. Conversely, more inclusive or cooperative

language increases their willingness to apply, without deterring men. This phenomenon, described as gender segregation by Hu et al. (2022), helps to maintain the under-representation of women in certain sectors, creating a vicious circle in which biased language maintains the male image of a job, further discouraging women from applying for it.

As well as limiting access to employment, these biases promote the idea that certain skills or attitudes are specific to men, thereby reinforcing gender stereotypes. For women, this representation can lead to a feeling of inferiority or reduced legitimacy, or even a form of self-censorship, that affects their confidence, engagement, and performance (Barriuso, 2024). For companies, these biases imply a loss of diversity in job applications, and therefore an impoverishment in terms of performance, innovation and image (Hentschel et al., 2020). In a context where inclusiveness is a central concern, maintaining unequal drafting practices also exposes companies to reputational risks. Gender bias in advertisements therefore goes beyond the simple textual dimension: it affects career dynamics, the structure of organisations and social norms themselves.

Although gender has received much of the attention in research on bias in recruitment, similar mechanisms of exclusion affect other characteristics such as age, disability, and ethnicity. Studies show that seemingly neutral phrases like young and dynamic or digital native deter older applicants to a degree comparable with overt age discrimination (Beel et al., 2013). Likewise, job ads or AI screening tools may unintentionally penalise candidates with disabilities, for instance, by ranking lower those who mention disability-related achievements due to implicit ableist stereotypes. A study by Caliskan (2021) demonstrated that GPT-based resume screening algorithms systematically downgraded candidates who referenced awards or affiliations related to disability. Importantly, biases often intersect. Recent experiments by Wilson and Caliskan (2024) involving AI-powered resume ranking revealed that candidates with multiple marginalised identities, such as Black men were systematically disadvantaged, even when all other qualifications were equal. The model consistently favoured white male profiles, and Black male candidates were never preferred when compared to equally qualified white male applicants. These findings highlight the severity of intersectional bias and underscore the need for fairness-aware design in both job advertisement language and AI-driven hiring tools, to ensure equitable access to employment opportunities for all.

2.3. Work already completed

A number of recent studies have used automatic language processing (NLP) techniques to analyse gender bias in job advertisements, with the aim of objectifying its presence and proposing ways of correcting it. This work is largely based on the agentic/communal distinction in language, according to which certain terms (e.g. competitive, independent, leader) are associated with masculine stereotypes, while others (e.g. empathetic, collaborative, good listener) are linked to feminine stereotypes.

Among the most significant contributions, the study by Hu et al. (2022) proposed an algorithm for detecting and correcting gender bias in advertisements, capable of determining the degree of gender bias in a text and suggesting neutral reformulations. This approach aims to make applications neutral while preserving the original meaning of the texts. For his part, Izquierdo Barriuso (2024) applied NLP techniques to more than 2,000 vacancies published in the Financial Times, using vectorisation, lexical frequency and semantic clustering methods to identify the presence of masculine terms in positions of responsibility. His study highlights the fact that these biases are particularly marked in advertisements for management positions and demonstrates the importance of reducing stereotypes. Finally, Böhm et al. (2020) conducted a study of IT advertisements in Germany, where they used a gendered lexicon to quantify the perceived masculinity or femininity of each vacancy. Their work shows that advertisements in technical fields are much more marked by agentic language, which could discourage potential candidates.

This potential is amplified by the rapid advancement of large language models (LLMs) such as GPT, BERT which are now being adopted in human resources for tasks ranging from resume screening and interview generation to the drafting of inclusive job ads. These models excel at parsing language nuances and processing large volumes of text with speed and consistency. Platforms such as LinkedIn are already integrating LLMs to analyse candidate responses, suggest improvements to job descriptions, or generate follow-up questions. According to Anzenberg et al. (2025), these tools promise significant efficiency gains and may standardise evaluation processes across candidates.

However, this sophistication comes with substantial risks. As LLMs are trained on massive internet-based datasets that reflect real-world biases, they inevitably replicate social stereotypes unless carefully constrained. Wilson and Caliskan (2024) warn that without intervention, LLMs

can perpetuate biased or exclusionary language, especially when dealing with underrepresented profiles. In a striking study by Caliskan (2021), GPT-based models consistently assigned lower ratings to resumes containing disability-related accomplishments, such as an autism leadership award, even when the resumes were otherwise identical. This suggests that implicit linguistic signals can trigger discriminatory responses from AI tools, further marginalising certain groups.

To address these risks, developers and researchers are employing multiple techniques such as refining LLMs on ethical datasets, introducing fairness constraints in training, and using support engineering to guide models results. Anzenberg et al., (2025), for example, demonstrated that adversarial debiasing, by training models to be simultaneously predictive and fairness-aware, can substantially reduce discriminatory outcomes in algorithmic recommendations. However, even with such refinements, residual bias remains a challenge, particularly in intersectional cases where multiple disadvantages (e.g., gender and disability) overlap (Caliskan, 2021).

Beyond language generation and analysis, algorithmic decision-making has expanded into another frontier: candidate-job matching through recommender systems. These systems, used on platforms such as LinkedIn and Indeed, personalise job ads and rank candidate suitability, often based on past interactions and preferences. While effective, they raise significant concerns regarding fairness and accountability. Kumar et al. (2023) have conducted a legal and technical review of recruitment recommenders, highlighting how algorithmic suggestion systems can reinforce existing gender, race, and pay disparities. Similarly, Zhang and Kuhn (2024), through an audit of a major online job platform, demonstrated that comparable male and female profiles received significantly different job recommendations with jobs shown to men typically offering higher pay and requiring more experience.

Simple interventions, such as editing out gendered pronouns from job ads, are helpful but often insufficient to remedy algorithmic inequalities (Kumar et al., 2023). More effective strategies include internal methods like adversarial debiasing and post-processing re-ranking techniques to ensure diverse candidate representation. Some platforms go further by developing inclusive systems tailored for under-represented groups, such as candidates with disabilities or linguistic minorities.

Legal and ethical issues are becoming increasingly important in the use of algorithmic systems for recruitment, particularly given the potential for biased decisions leading to legal consequences under employment law. According to Lacroux and Martin-Lacroux (2022), the

way recruiters interact with these systems plays a decisive role. Their study reveals that recruiters can either rely on AI-generated recommendations, especially when the system appears to be authoritative, or reject them altogether when they seem opaque or inconsistent. This dual tendency, known respectively as automation bias and algorithm aversion, can lead to biased hiring outcomes: either by uncritically accepting biased suggestions or by ignoring helpful ones. These findings underline the need for better explainability and user understanding, to ensure that algorithmic tools support rather than hinder fair and informed decision-making.

Similarly, among applicants, Zhang and Kuhn (2024) found that although algorithms recommended certain jobs, users do not always click on the best suggestions, preferring lists that match their expectations or stereotypical career perceptions. The persuasiveness of AI recommendations also depends strongly on explainability.

These findings underline the fact that the fairness of recruitment systems depends not only on algorithmic design but also on human perception and behaviour. A technically fair system can still produce biased results if its outputs are misunderstood, misapplied, or implemented without proper control. Consequently, researchers emphasise the need for comprehensive responsible AI approaches, regular fairness audits, clear standards of explanation, external supervision, and explicit institutional commitments to diversity, to ensure that algorithmic recruitment promotes equitable hiring outcomes.

2.4. Research question

Recent research has highlighted the existence of gender bias in the wording of job advertisements, particularly through the use of terms with masculine or feminine connotations. These formulations can influence how applicants perceive job ads and contribute to self-selection, often to the detriment of women. Research using automatic language processing (NLP) techniques has made it possible to systematically identify these biases on a large scale and, in some cases, to propose textual reformulations designed to neutralise the language (Hu et al., 2022; Barriuso, 2024). These approaches have led to considerable advances in detection and correction, but few studies have empirically assessed the actual impact of these linguistic modifications on the behaviour of users of job boards.

In this context, the aim of this research is to examine the concrete effect of a textual intervention on job offers: does a reformulation designed to reduce gender bias have a measurable effect on job-seeking behaviour? In other words, does the transformation of the language used in a job

advertisement significantly alter the interest it arouses in candidates? This question takes a linguistic, technological and social perspective, and seeks to assess whether the discursive adjustments made in advertisements can really help to make recruitment processes more inclusive.

The two main hypotheses to be tested are as follows:

- H1: This reformulation leads to a significant change in the composition of the profiles recommended by the system.
- H2: Reformulation of job offers in neutral language effectively reduces the gender bias detected in the texts.

In addition to these two central hypotheses, there are two secondary hypotheses:

- H3: The effect of reformulation varies from one sector of activity to another. In male-dominated fields such as STEM or engineering, where job ads generally contain more masculine-coded language (Böhm et al., 2020; Gaucher et al., 2011). Neutralisation should have a greater effect on the proportion of female candidates recommended by the system, compared to more gender-balanced sectors such as healthcare or education.
- H4: The neutralisation of gendered language in job advertisements should increase the proportion of female profiles among algorithmic recommendations. This hypothesis is supported by previous studies showing that job ads containing masculine terms tend to discourage female candidates (Gaucher et al., 2011). Böhm et al. (2020) and Barriuso (2024) have further shown that these gendered formulations are particularly widespread in managerial or technical positions, thereby reinforcing professional segregation between the sexes. By reformulating these advertisements in more neutral terms, the intervention should remove subtle signs of exclusion and encourage a broader commitment, resulting in a greater presence of women among recommended candidates.

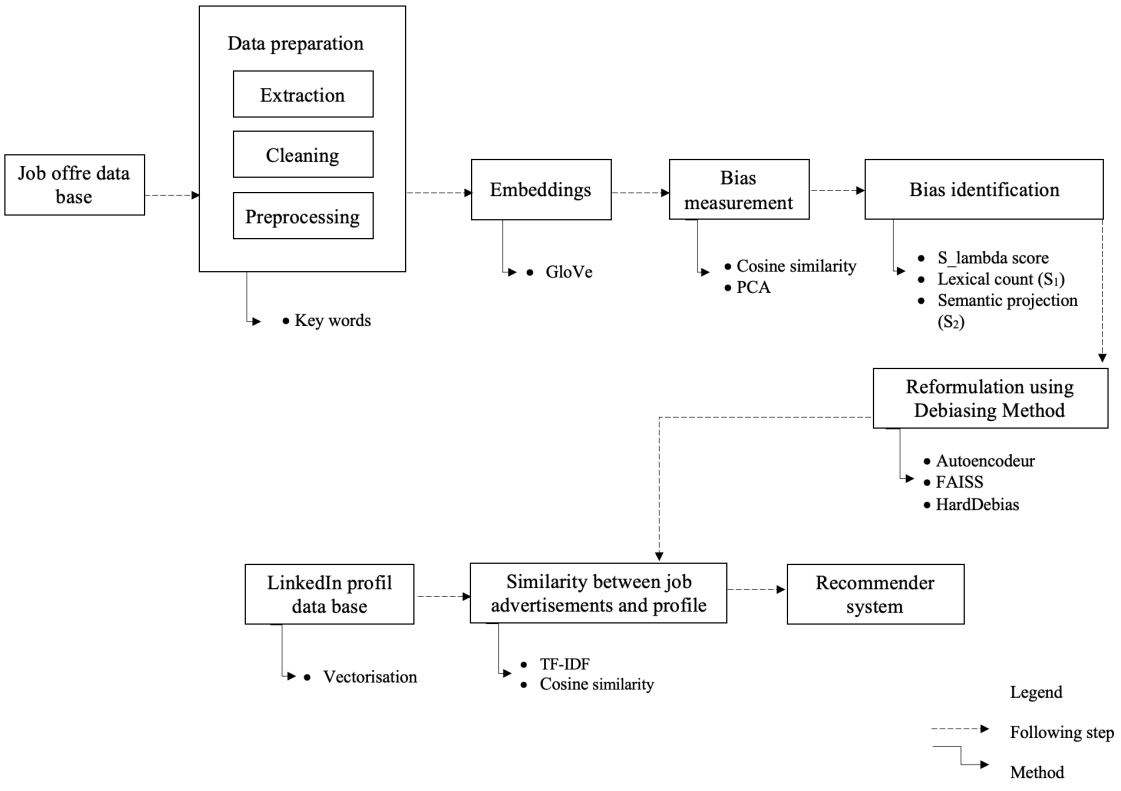
To test those hypotheses the research proposes a methodology for identifying and correcting the biases in advertisements and then analysing whether these corrections influence the results generated by recommendation systems on recruitment platforms. Concrete avenues for improvement are also proposed on achieved results.

3. Materials and methods

This section presents the methodology that guided this work, as illustrated in Figure 1. The overall structure of the methodology was designed specifically for this work, in order to meet the specific objectives of this thesis. Each step is based on methods recognised in scientific literature. This approach draws on research from the fields of natural language processing (NLP), machine learning and algorithmic fairness. While this assembly and overall logic are unique to this project, the techniques used in each phase are based on best practices identified in academic work, ensuring the robustness, relevance and validity of the approach adopted.

To test the hypotheses formulated, the methodology is organised around three key stages: identifying gender bias in job offers, reformulating biased content using debiasing techniques, and finally evaluating the effect of this reformulation on the results of a recommendation system. The aim is to propose a consistent way of detecting and reducing gender bias in job offers, then to observe whether these adjustments have a concrete effect on the profiles proposed by a matching system based on textual similarity. All analyses and source code are available on a GitLab repository to ensure transparency and reproducibility.

Figure 1: Methodology diagram



3.1. Analysis and identification of bias in job advertisements

3.1.1. Data collection, extraction and processing

The purpose of this step is to explain how the database of job postings was extracted. The database[1] containing information on 23,349 job ads posted on LinkedIn in the United States between 22/07/2023 and 23/07/2023 was obtained via the Kaggle platform. Kaggle is an online platform that makes databases available free of charge to enable anyone to carry out searches in a specific field.

To illustrate the kind of language used in the advertisements, Table 1 provides an example of a job posting extracted from the dataset.

Table 1: Example of job advertisement

	Title	Other	Description
Job Adv. 1	Sales Manager	…	Are you a dynamic and creative marketing professional looking to make a significant impact in the world of freight and logistics? We are seeking a talented Marketing Manager to join our team and drive our marketing initiatives to new heights. If you're passionate about shaping brand narratives developing strategic campaigns and thrive in a fast-paced environment we want to hear from you! […]

This example already suggests potential gendered language through adjectives such as “dynamic,” “creative,” and “fast-paced environment,” which the literature often associates with masculine-coded job descriptions (Gaucher et al., 2011). These terms may unconsciously appeal more to male candidates or discourage female applicants, thereby contributing to self-selection biases.

[1] Database source: https://www.kaggle.com/datasets/rajatraj0502/linkedin-job-2023

For each job ad, the dataset includes various metadata such as the job title, company name, and job description. In this study, the focus is placed primarily on the textual job descriptions, as these are the primary vehicle for potential linguistic bias.

Before applying any bias detection method, a text pre-processing phase was carried out to normalise the job descriptions. This step includes converting all characters to lowercase, removing punctuation, special characters, and excess whitespace. The goal of this procedure is to reduce textual noise and to ensure uniform input for downstream NLP tasks such as word embedding, vectorisation, or classification.

3.1.2. Vectorisation of job descriptions

Each document is then represented in vector form using the GloVe (Global Vectors for Word Representation) model. Unlike TF-IDF representations which assign a weight to each word independently of its context, embeddings such as GloVe (Pennington et al., 2014) are based on continuous training and capture global co-occurrence relationships in large text corpora. One of the considerable advantages of the GloVe model is the ability to capture not only semantic affinities (e.g. king close to queen), but also analogous linear regularities. For example, the vector difference between he and she follows a direction very similar to that between man and woman, or father and mother. This vector structure implicitly encodes a gender direction in the semantic space (Bolukbasi et al., 2016). This makes it possible not only to identify gendered terms using cosine similarity to male or female landmarks, but also to assess and correct stereotypical biases in textual descriptions, such as those present in job advertisements.

3.1.3. Bias measurement

Once the vector representations of the job advertisements are generated, it becomes possible to assess the degree of gender bias they carry. To ensure a robust and nuanced detection of these bias, this study adopts a hybrid methodology combining two complementary techniques, inspired respectively by Hu et al. (2022) and Bolukbasi et al. (2016).

The first component involves computing the cosine similarity between the average embedding vector of each job description and two predefined gender reference vectors, one representing masculine-coded terms (e.g., he, man, leader), and the other feminine-coded terms (e.g., she, woman, care). This method provides an intuitive and interpretable measure of proximity of gendered language, by positioning each ad on a semantic axis between masculine and feminine

poles. It is widely used in the literature as a practical tool for detecting explicit lexical bias in recruitment texts (Hu et al., 2022; Gaucher et al., 2011).

However, lexical similarity alone may fail to capture more implicit or structural forms of bias, which may persist even in the absence of overtly gendered terms. To take account of this, the methodology incorporates a second technique based on the identification of a latent "gender direction" in the word embedding space, following the approach developed by Bolukbasi et al. (2016). This process begins with the selection of a set of well-established gendered word pairs (e.g., man–woman, king–queen, father–mother). For each pair, the vector difference is calculated (e.g., $v_{man} - v_{voman}$), and all the resulting difference vectors are then stacked into a matrix D, where each row represents the semantic change between a masculine and a feminine term.

Principal Component Analysis (PCA) is then applied to this matrix to extract the first principal component, the eigenvector that captures the direction of greatest variance among all the pairs of male and female terms. The resulting axis, called the "gender direction" (v_gender), represents the dominant semantic dimension along which gender associations vary in the embedding space. Once identified, it becomes possible to project any word, sentence, or document vector onto this axis to quantify its alignment with stereotypical gender constructs. A strong projection in either direction indicates embedded gender bias, even implicitly or through metaphorical language.

To complement this structural detection, the cosine similarity between the document vector and each of the two gender reference vectors also provides a directional score, and the difference between these two scores is interpreted as an additional bias indicator. Finally, all bias scores are normalised to ensure consistent and meaningful comparisons between job descriptions of different lengths or lexical densities, taking into account verbosity or volume of content.

By integrating both the explicit (cosine-based) and implicit (PCA-based) components, this approach provides a comprehensive, interpretable, and theoretically grounded framework for measure gender bias in job advertisements. It is particularly effective for dealing with stylistically heterogeneous corpora, where bias may be expressed in subtle, indirect, or context-dependent ways.

In addition to analysing the semantic vector, the methodology also incorporates a lexical component based on the explicit presence of gendered terms in each description. This lexical

detection relies on two validated word lists from previous research, one for words coded as masculine (e.g., competitive, dominant, leader) and another for feminine-coded words (e.g., nurturing, supportive, communal), originally proposed by Gaucher et al. (2011). These lists provide a transparent and interpretable tool for identifying obvious linguistic signals of gender stereotyping.

3.1.4. Bias identification

To express the results from both the lexical and semantic analyses, the study adopts a composite indicator known as S_lambda, introduced by Hu et al. (2022). This metric combines two complementary components: S_1, which corresponds to the proportion of words explicitly gendered in the text (lexical count), and S_2, which quantifies the implicit orientation of the discourse by projecting it into the gendered semantic space derived from word embeddings. The final formula used is: "S_lambda = $S_1 + 2 \times S_2$"

This weighting system assigns greater importance to S_2, on the grounds that implicit gender signals, such as discursive tone or choice of metaphors, are more subtle but often more revealing than overt lexical markers. As indicated in the literature (Gaucher et al., 2011; Garg et al., 2018), job advertisements may convey gendered stereotypes even without using clearly gender-coded terms, for instance by emphasising assertiveness, competitiveness, or hierarchical dominance, traits often culturally aligned with masculinity.

To visualise the distribution of gender bias across the dataset, multiple graphical representations were produced using this composite S_lambda score. These visualisations help to reveal how bias manifests both explicitly and implicitly within job ads and enable for cross-sector comparisons or time-based analyses. By combining both types of indicators into a single unified metric, the method provides a more nuanced and comprehensive evaluation of bias, which is particularly useful in heterogeneous corpora where stereotypes may appear in indirect or stylistically diverse forms.

3.1.5. Job clustering

Finally, the job ads are grouped into clusters, e.g. sets of lexically similar ads, in order to determine whether certain sectors are more prone to lexical bias (e.g. engineering, human resources, finance, etc.). This segmentation is based on the K-Means algorithm, which divides documents into different groups according to the predominance of vocabulary specific to each

sector using the TF-IDF method. The algorithm captures the characteristic words of each ad, making it possible to identify linguistic similarities between texts, without any prior knowledge of the sectors. To facilitate interpretation, a heuristic classification function associates each advertisement with a general sector based on its job title. Once these groupings have been established, the bias scores calculated using GloVe embeddings, which measure the degree of masculinity or femininity implicit in the texts, are included. This makes it possible to compare gender trends within each cluster, and therefore to gain a better understanding of which sectors or types of advertisement are more or less biased.

Visualisations are generated to illustrate the distribution of bias in the corpus. Highly gendered ads can then be analysed and corrected in a targeted manner. The aim of this first stage is to make explicit the semantic biases present in job advertisements, by combining automatic natural language processing techniques and vector analysis methods.

3.2. Reformulation of biased content using debiasing techniques

3.2.1. Textual debiasing via neural autoencoder and reformulation

One of the main objectives of this work is to neutralise the implicit gender bias present in the textual descriptions of job offers, in order to make the recommendation system fairer. The scientific literature has highlighted the impact that certain gendered terms can have on the perception of a job, particularly when they convey male or female stereotypes. Words such as leader, competitive or dominant are more associated with masculine representations and may dissuade some candidates, affecting the diversity of applications received (Gaucher et al., 2011). To correct this asymmetry, a debiasing method has been implemented in this system. The principle is based on automatic learning using a neural autoencoder. To do this, we used the embeddings analysed previously as well as the gender direction identified using cosine similarity (Sun et al. 2019). This measure serves as the basis for the debiasing method used.

Previous analyses by Gaucher et al. (2011) have highlighted the limitations of so-called naïve debiasing methods. This was based on the creation of lists of gendered words (masculine and feminine), and their replacement by neutral equivalents chosen manually or via close matches in the corpus. Although this approach is simple and directly interpretable, it does not correct for more implicit or subtle biases, particularly those linked to context or syntax.

To overcome these limitations, a method based on a neural autoencoder was used. An autoencoder is an unsupervised learning neural network whose objective is to reconstruct an input after compressing it into a lower-dimensional representation (Hinton et al., 2006). In this case, the model is trained to reproduce the lexical vectors while learning to remove their gendered component. To achieve this, the loss function incorporates a specific penalty, encouraging the model not only to minimise the global reconstruction error, but also to reduce the projection of the output vectors onto the identified gender direction. This strategy is inspired by recent work showing that it is possible to mitigate stereotypical biases present in embeddings while preserving informative dimensions (Sun et al., 2019).

Once the autoencoder has been trained, it is applied to all the words in the job advertisements. Each word is reformulated in a debiased vector space. To translate this operation into text, the system uses the FAISS library (Douze et al., 2024), which efficiently finds the words closest to the original word in the new vector space. When a word has a high bias, it is replaced by a more neutral synonym identified in the debiased space. This produces a reformulated version of the advertisements, semantically equivalent but written in a more inclusive style.

3.2.2. Evaluation of the debiasing method

To ensure the effectiveness and relevance of this reformulation, a set of quantitative and qualitative evaluations was carried out, following approaches inspired by Bolukbasi et al. (2016). The first evaluation focuses on measuring the reduction in gender bias at the vector level. For each word, the projection on the gender direction is calculated before and after debiasing. A successful transformation is reflected by a significant decrease in this projection, indicating that the word has moved away from gendered semantic associations.

The second evaluation concerns semantic preservation. Cosine similarity is calculated between the original word vectors and their debiased counterparts. A high similarity score indicates that, despite the removal of gender connotations, the main meaning of the word remains intact. This step ensures that the content of the job description is not distorted during the debiasing process.

In addition, the job ads are manually reviewed to assess the contextual consistency of the replacements. This qualitative inspection is used to determine whether the substitutions are appropriate and contextually relevant, a key factor in maintaining the readability and acceptability of the reformulated ads.

To provide a comparative benchmark, the same analyses are conducted on an alternative approach known as HardDebias (Bolukbasi et al., 2016). This method explicitly projects word vectors orthogonally to the gender direction and then rebalances the gendered word pairs to preserve semantic relationships. Although it is less flexible than the autoencoder, it provides a useful reference point for evaluating the relative strengths of each approach.

Finally, the debiased job descriptions are reintegrated into the recommendation system, allowing us to assess their impact on candidate suggestions. Several tests are carried out to measure variations in cosine similarity between profiles and job offers, and to track any changes in the proportion of female profiles recommended. The overarching goal is to determine whether the reformulated advertisements contribute to a fairer recommendation process, one less influenced by gender stereotypes embedded in language.

3.3. Impact of reformulation on recommendation systems

The purpose of this section is to examine the practical impact of job advertisement debiasing on candidate recommendations, particularly with respect to gender representation. While previous sections have focused on identifying and mitigating linguistic bias, this step assesses whether such textual adjustments lead to measurable differences in the results of a content-based recommendation system.

3.3.1. Data collection and vectorisation

To conduct this analysis, a database[2] of 1,000 LinkedIn profiles was used, randomly selected between 2021 and 2023. This dataset includes information on name, location, education, and professional experience, with each experience entry containing job title, company, location, and description. For the purpose of this study, only the education and work experience fields were retained, as they provide the most relevant textual information for characterising candidates in a recruitment context.

[2] Database source:
https://www.kaggle.com/datasets/manishkumar7432698/linkedinuserprofiles?resource=download&select=LinkedIn+people+profiles+datasets.csv

Table 2: Example of values for LinkedIn Profile

	Experience	Other	Education
Profile 1	company":"Mid-Range Computer Group Inc.","company_id":"mid-range-computer-group-inc.","industry":"IT Services and IT Consulting","location":"Markham, Ontario","positions" […]	…	[{degree: end_year:978,field:,meta:1973 - 1978","start_year":"1973","title":"St. Michael's College School","url":""},{"degree":"Unix System Administration","end_year":"","field":"","meta":"2016 - Present Activities and Societies: Successfully Completed MCT 124 UNIX System Administration","start_year":"2016","title":" Seneca College of Applied Arts and Technology"

Each profile is converted into a structured text format by grouping key information, (e.g., job titles, company names, academic degrees) into a unified field, enabling consistent vectorisation. The selection of 1,000 profiles reflects a balance between IT feasibility and corpus diversity, ensuring sufficient variability between sectors and positions to assess the effect of debiasing. While not exhaustive, this sample allows a solid comparative analysis of a whole range of job advertisements.

3.3.2. Implementation of a recommender system

While LinkedIn's actual recommendation engine relies on a hybrid architecture, combining content-based filtering, collaborative filtering, and online learning algorithms to deliver personalised suggestions, this level of complexity is enabled by access to detailed behavioral data such as user interactions, clicks, and application history (Peng, 2018). As such information is not available within the scope of this project, the approach adopted here is intentionally limited to content-based filtering, an approach well-adopted to information retrieval and recommendation (Peng, 2022; Govinda et al., 2024). It operates by computing the textual similarity between each job advertisement and candidate profile. Each job ad is represented in two versions, one original and the other debiased, using the job title and full description. In the same way, each candidate profile is represented as a textual vector derived from their professional and academic history.

For the vectorisation of both job offers and profiles, TF-IDF (Term Frequency–Inverse Document Frequency) is employed. This method was chosen in place of embedding-based approaches (such as GloVe) due to its interpretability, transparency, and its independence from pretrained semantic biases. Unlike word embeddings, TF-IDF reflects the exact lexical patterns

within the corpus, making it particularly suitable for analysing direct lexical correspondences, a key consideration when evaluating the impact of linguistic bias and its reduction on recommendation outcomes.

The system computes cosine similarity between job offers and candidate profiles. For each offer, the top 100 most similar profiles (TOP-K) are retrieved, both for the original and debiased versions. This procedure allows for a direct comparison of how reformulation influences the nature and composition of the recommended profiles. In line with Caliskan et al. (2017), this setup enables us to assess whether implicit gender bias embedded in job descriptions creates stronger semantic proximity with male-coded profiles, a phenomenon the debiasing method seeks to mitigate.

3.3.3. Debiasing effects on recommendations

To evaluate the impact of textual debiasing, the distribution of similarity scores is compared across both versions of each job offer. Visualisations such as histograms are generated to explore whether debiased descriptions lead to broader or more inclusive candidate matches, without compromising lexical relevance. Prior research has shown that recommender systems can inadvertently reflect and even amplify societal biases found in training data, justifying the need for such a detailed post-processing evaluation (Bolukbasi et al., 2016; Wang et al., 2022).

Beyond similarity scores, the gender composition of the recommended candidates is also analysed. Gender is inferred from first names using the gender-guesser Python library. For each job posting, the proportion of female profiles in the TOP-K recommendations is calculated and compared between the original and debiased descriptions. This metric provides a quantifiable measure of how linguistic modifications affect gender representation in algorithmic outcomes.

Additionally, job offers are ranked based on the variation in female representation before and after reformulation, highlighting cases where debiasing had the most pronounced impact, either positive or negative. This analysis contributes to a fairness-aware perspective on recommendation outcomes, as advocated in recent studies on equitable algorithm design (Li et al., 2022).

While the content-based recommendation system developed in this study provides a controlled and interpretable framework to examine the effects of linguistic debiasing, it remains a simplified model when compared to production-grade systems used in real-world recruitment

platforms. In practice, many modern recommendation engines integrate multiple data sources, including user interaction logs, behavioural signals, or collaborative filtering components. These features enable more dynamic and personalised recommendations, but they also introduce additional complexity and potential biases. The system proposed here, by contrast, focuses solely on lexical and semantic proximity, which facilitates transparency and interpretability.

4. Results

In this section, we present the results of the various analyses. This section is divided into three parts dedicated to each stage of the process.

The first section deals with the analysis of the job advertisement database. It begins with a visual analysis of the explicit biases contained in the job advertisements. This is followed by a general visualisation of the method and a comparison of the method with different scientific sources in order to validate the results. Finally, a detailed explanation by sector is included. This enables us to determine whether certain sectors are more prone to bias. This first part lays the groundwork for the following sections.

The second section presents the results of the debias method. It begins with a brief exploratory analysis, followed by a review of the results. Finally, the results will be justified using various quantitative and qualitative methods.

To conclude this section, the analysis of LinkedIn profiles is presented. The recommendation system is analysed to determine whether the debias method has a real impact on platforms such as LinkedIn.

4.1. Results of analysis of bias in job offers

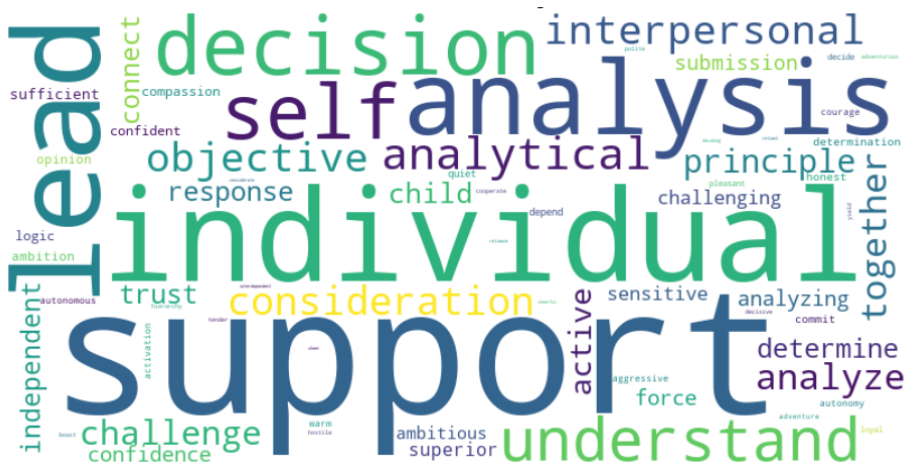
Based on the methodology described above, we applied the metric S_lambda to a corpus of job advertisements in order to quantify their gender orientation. The following subsections detail the results obtained from this analysis, including the identification of gendered terms, the overall distribution of bias scores, and variations across job sectors.

To interpret the S_lambda scores, we refer to the threshold suggested by Böhm et al. (2020), where scores above 0.5 are considered indicative of male bias. This threshold, although empirical, is a useful point of reference for detecting whether a job description is systematically leaning toward stereotypically, masculine language. This weighting is based on previous work but has not yet been extensively validated and should therefore be interpreted with caution. Applying this threshold to the distribution of scores allows us to identify the proportion of ads that can be considered biased, as well as sectoral differences in linguistic orientation.

4.1.1. Analysis of explicit gendered terms

The word cloud generated from explicitly gendered terms (Gaucher et al., 2011) visually represents the most frequent lexical items contributing to the S_1 component of the bias score. As shown in Figure 2, masculine-coded terms such as lead, challenge, decision, and independent are prominent, as are feminine-coded terms like support, together, understand, or interpersonal. This aligns with previous findings that agentic traits, emphasising assertiveness, control, or individual performance, are culturally associated with masculinity, while communal traits, focusing on nurturing, collaboration, and emotional intelligence, are linked to femininity (Gaucher et al., 2011; Caliskan et al., 2017).

Figure 2: Word clouds of explicit bias



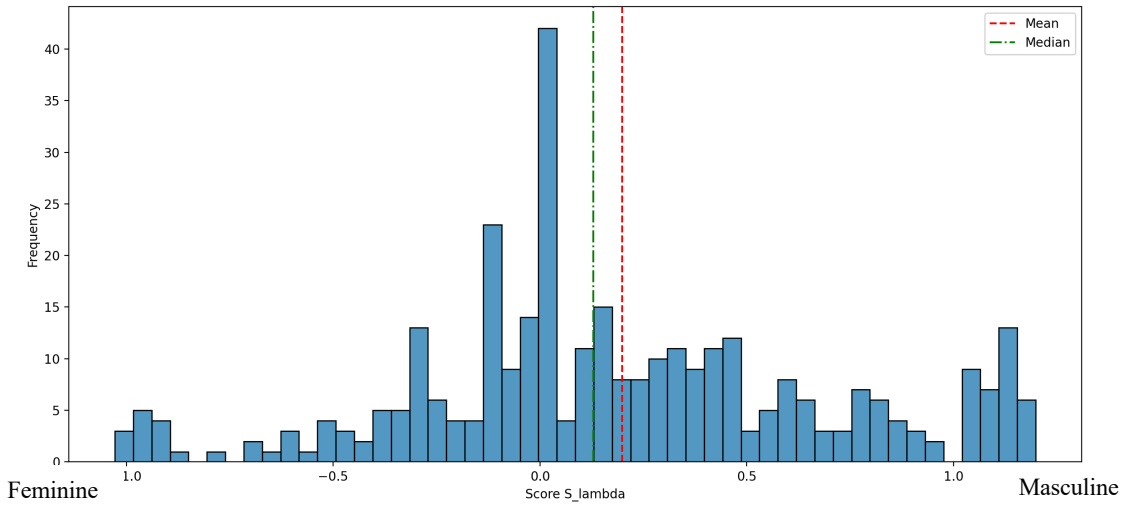
This visualisation provides a concrete illustration of the lexical polarisation of job advertisements, revealing how stereotypically masculine or feminine words are disproportionately distributed across the corpus. For instance, the frequent appearance of terms such as “leader” or “independent” supports Gaucher et al. (2011) findings that leadership-oriented language can deter female applicants, especially in male-dominated sectors. Similarly, the presence of support and together echoes linguistic patterns typical of female-oriented job postings, particularly in care or education sectors, as discussed by Hu et al. (2022).

Such evidence demonstrates how certain words, even when not overtly gendered, can construct a discourse aligned with societal expectations of gender roles. The concordance between the lexical distributions observed here and previous research reinforces that word choice is not neutral in recruitment contexts but reflects and potentially reinforces structural gender biases.

4.1.2. Overall distribution of the S_lambda score

Figure 3 shows the histogram of the distribution of S_lambda scores over the entire corpus. The x-axis shows the values of the S_lambda score, which range from approximately -1 to +1. These values measure the gender bias of a text, with positive values indicating a male bias, negative values a female bias, and values close to zero an assumed neutrality. The y-axis represents the frequency with which each score appears in the corpus of ads, i.e. how many offers have this level of bias.

Figure 3: Distribution of S_lambda scores



Although the S_lambda scores observed in this study are approximately between –1 and +1, this range is not strictly bounded by the algorithm itself. Rather, it reflects the combined influence of its lexical and semantic components, both of which are independently standardised and constrained in practice. The lexical dimension, based on the relative frequency of gendered terms, is inherently scaled between the poles masculine and feminine. The semantic component, derived from cosine similarity measures, is influenced by the distribution of word embeddings in the training space, which tends to prevent extreme values. As a result, while the score is not theoretically limited to the interval –1 to +1, it remains within this range in most empirical cases.

At first glance, the distribution appears to be largely centred around zero, which could be interpreted as meaning that, on average, job advertisements are considered to be linguistically neutral. However, this apparent neutrality takes the form of an asymmetrical curve. In fact, the right tail of the distribution, corresponding to the male scores, is visibly wider and more

populated than the left tail. This is confirmed by the fact that the median is also positive, and that the right tail of the distribution is longer and more populated. To determine whether this difference is statistically significant, a t-test was conducted to compare the average S_lambda score against a neutral value of zero. The result was highly significant ($p < 0.001$), indicating that the average bias is not the result of random variation but reflects a structural tendency to favour men in the language of job ad. This result reinforces the idea that even though many offers are close to neutral, male bias is not only more prevalent, but also more intense when they do occur. This phenomenon suggests a structural tendency for professional language to lean towards masculine connotations, particularly in sectors that value competition, autonomy or leadership, as Gaucher et al. (2011) and Caliskan et al. (2017) have shown. It is therefore wrong to conclude that neutrality on the sole basis of an average close to zero, because this central value masks the presence of strongly gendered ads in one direction or another, but more markedly on the male side.

Finally, it is important to emphasise that this visualisation reinforces the relevance of the S_lambda score as a diagnostic tool, by showing that it can be used not only to detect extreme cases, but also to analyse general trends in the corpus. It combines lexical and semantic dimensions which, together, reveal deeper dynamics of gendered language than a simple count of stereotyped words would.

4.1.3. Clustering by dominant sector

Figure 4: Distribution of S_lambda scores by dominant sector

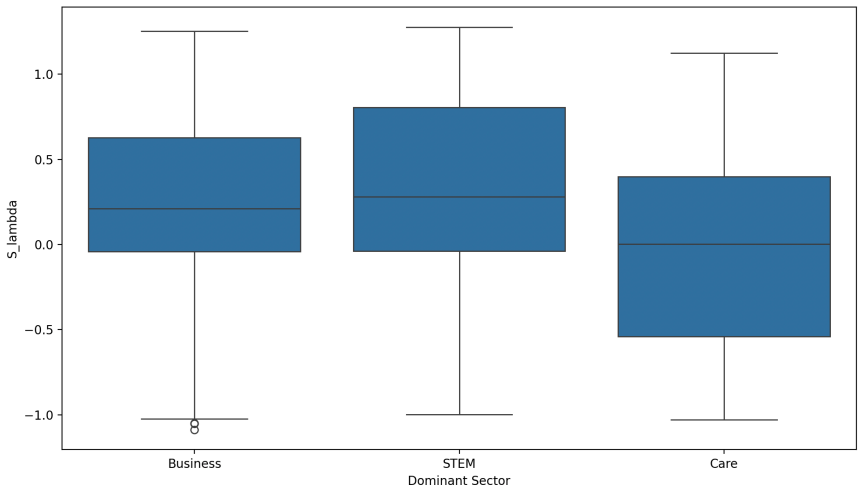


Figure 4 examines the distribution of S_lambda scores within different sectoral groups automatically constructed using the K-Means algorithm on TF-IDF representations of the ads. Each cluster is then labelled according to the dominant sector of the stocks it contains (e.g. Business, Engineering, Care, etc.). This approach makes it possible to identify the sectors that carry the most gender bias, whether explicitly or implicitly.

Some sectors, such as "Business" and "STEM[3]" show significantly positive median S_lambda scores, reflecting a structural male bias. Conversely, sectors such as "Care", including "Education" and "Healthcare", show scores closer to zero, or even negative. This observation is supported statistically by an ANOVA test performed on the average S_lambda scores across sector clusters, which revealed a significant difference between groups. This indicates that sectoral variations in gendered language are unlikely to be due to random variation. These trends are consistent with the literature on the gendered division of labour, which associates certain professions with masculine qualities (competition, leadership) and others with feminine qualities (empathy, cooperation).

Figure 5: Care sector word cloud

Figure 6: STEM sector word cloud

These results are complemented by a lexical analysis using sector word clouds constructed from TF-IDF scores. These visual representations make it possible to identify the most discriminating terms used in the descriptions of offers, revealing lexical trends consistent with the S_lambda bias scores. For example, the "Care" sectors highlight terms such as patient, care, or nurse, linked to relational and empathetic qualities traditionally associated with the female stereotype, while the "STEM" sectors highlight terms such as leadership, strategy, project, or executive, which are frequently linked to agentic attributes and therefore perceived as masculine (Barriuso, 2024).

[3] Acronym of Science, Technology, Engineering, Mathematics

This convergence in the analyses reinforces the robustness of the observations, illustrating how language contributes to the perpetuation of gendered roles in the labour market. As Hu et al. (2022) point out, even in the absence of explicitly gendered terms, certain implicit qualities (e.g. ambitious or considerate) convey a subtle but important gendered orientation. These semantic biases can influence candidates' perceptions of advertisements and reinforce structural inequalities, in particular by discouraging the self-selection of female profiles for highly agentic positions.

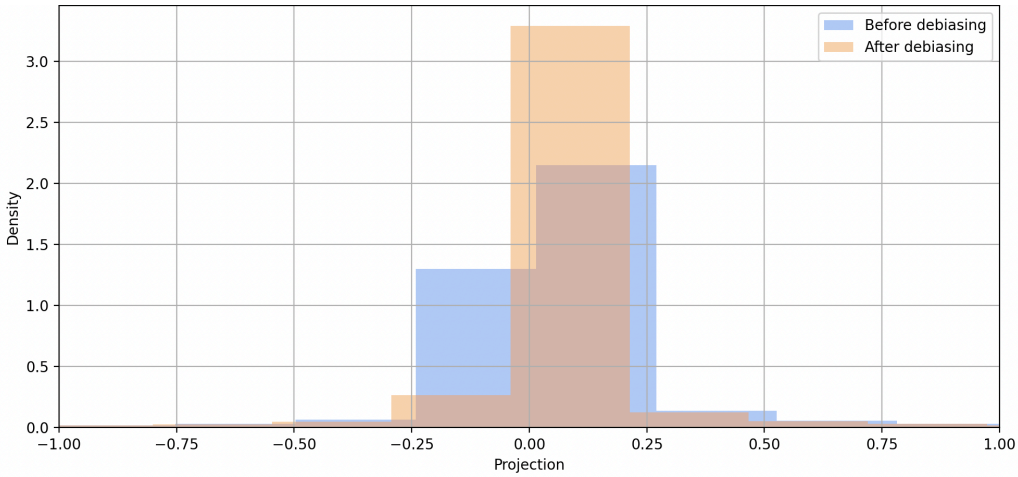
4.2. Debiasing method: analysis of results

In order to assess the effectiveness of debiasing applied to vector representations of language, a series of numerical indicators was used, together with a visualisation of comparative projections on the gender axis. This axis, constructed from pairs of opposite words such as *he-she*, quantifies the extent to which a word is implicitly aligned with a masculine or feminine connotation in the semantic space. More precisely, the projection measures the directional proximity of a word to this gendered dimension: a value close to zero is interpreted as neutral, while a high value indicates a more marked polarisation.

4.2.1. Graphical analysis

The graph below shows the normalised distribution of word vector projections on the gender direction, before (in blue) and after (in orange) the application of an autoencoder debiasing method. The x-axis represents the degree to which each word is projected onto the gender axis (positive or negative), while the y-axis indicates the estimated density, i.e. the relative proportion of words corresponding to each level of projection. It is important to note that Figure 7 does not reflect the same type of data as Figure 3. Whereas Figure 3 displays the distribution of S_lambda scores at the level of entire job advertisements, Figure 7 focuses on projections of individual word embeddings along the gender axis. The difference in granularity, at the document level and at the word level, explains the differences in the shape and interpretation of the two distributions.

Figure 7: Distribution of gender projections



The results indicate that the distribution after debiasing is significantly refocused around zero, with a higher density in this area, suggesting that the gendered components have been effectively neutralised in the vector representations. In contrast, the initial distribution shows a more marked dispersion, with an asymmetry that reflects the presence of more pronounced gender bias.

This contraction around zero after debiasing reflects an effective reduction in the biases without introducing excessive distortion. This visual development is in line with the work of Bolukbasi et al. (2016), who show that it is possible to reduce the biases present in embeddings by removing their alignment with a “gender direction”. This type of visualisation therefore makes it possible to graphically validate the effectiveness of the neutralisation procedure and provides an initial basis before analysing quantitative metrics.

While the autoencoder-based method successfully reduces the average projection of word embeddings on the gender direction, some residual bias remains. This observation aligns with findings by Hu et al. (2022), who note that “while projection is reduced, the words still form separate clusters and nearest neighbours remain gendered.” In other words, debiasing methods, even when effective, do not erase all traces of gender semantics, because part of this structure is deeply embedded in the contextual usage of language.

Moreover, Bolukbasi et al. (2016) were among the first to document the trade-off between debiasing intensity and semantic preservation. Their method, HardDebias, achieves almost complete alignment suppression but at the cost of reduced interpretability and potential distortion of meaning. This risk is particularly relevant in contexts such as recruitment, where

the integrity of semantic relationships (e.g., between “engineer” and “technical”) is critical to the performance of recommendation systems.

Gonen and Goldberg (2019) further demonstrate that even after debiasing, information remain latent and detectable by classifiers, cautioning against the assumption that vector neutrality guarantees downstream fairness. This supports a more nuanced view: complete neutralisation is not always desirable or necessary, especially if it compromises functionality. Instead, controlled reduction, as implemented here, offers a pragmatic middle ground that reduces explicit bias signals while maintaining lexical coherence.

Thus, the persistence of small “|b| > 0” values is not necessarily a flaw but reflects a methodological choice that prioritises a balance between fairness and usability. This perspective is consistent with recent work that advocates for transparency and proportionality in fairness interventions, rather than maximalist suppression at all costs (Mehrabi et al., 2022).

4.2.2. Quantitative analysis

Having visualised the effects of debiasing through the distribution of projections on the gender direction, it is essential to complete this analysis with a quantitative validation. This makes it possible to measure the extent of the bias reduction more precisely, while ensuring that the semantic properties of the vector representations are preserved. Two complementary approaches were therefore used: the analysis of average projections on the gender axis and the measurement of similarity between vectors before and after debiasing.

This change is confirmed by the projection averages: for words classified as neutral, the average falls from 0.6185 to 0.5259; for typically gendered words, it falls from 0.5182 to 0.4272. This represents an overall reduction in vector bias of 19.21%. The analysis of average projections makes it possible to quantify the alignment of words with a gender direction defined in vector space. A reduction in this average after debiasing indicates an effective reduction in implicit bias. These results are encouraging in that they show that the model manages to reduce implicit bias without significantly altering the overall semantic structure. Indeed, the average cosine similarity between the vectors before and after transformation remains high (0.8650), indicating that lexical consistency is well maintained. In addition, the reconstruction error (MSE = 0.0082) remains very low, confirming the autoencoder's ability to faithfully reproduce the original representations while filtering out gendered components. These results confirm the observations of Hu et al. (2022), according to whom "while projection is reduced, the words

still form separate clusters and nearest neighbours remain gendered". In other words, even after engendering, part of the latent structure remains gendered.

A comparison was also conducted with a more radical reference method: HardDebias (Bolukbasi et al., 2016). Although this technique manages to lower the average bias to an even lower level (0.0679), it does so at the cost of a greater degradation of semantic representations (average cosine similarity of 0.7706). In comparison, the autoencoder offers a more balanced compromise between reducing bias and preserving linguistic fidelity.

While autoencoder-based debiasing techniques do not offer a direct or fine-grained control over the exact quantity of bias removed, research suggests that significant adjustments can still be made by adjusting specific training parameters. For instance, increasing the weight of the gender projection term in the loss function or adjusting the strength of reconstruction constraints can influence the degree of bias reduction (Hu et al., 2022). However, this influence remains non-linear and often unpredictable. Ravfogel et al. (2020) empirically demonstrate that more aggressive debiasing tends to degrade the geometric structure of embeddings, which can be detrimental to downstream performance. In contrast, Hu et al. (2022) report that an approach based on autoencoder allows for a calibrated reduction of bias: gendered projections are substantially reduced, while semantic proximity and clustering structure are largely preserved. Their results indicate that nearest neighbours remain meaningful, and that debiased vectors continue to form coherent clusters, suggesting that full erasure of gender information is not always necessary to reduce bias effectively. This confirms the idea that moderate, targeted debiasing can offer an optimal trade-off between fairness and linguistic utility. Future work could build on these findings by developing methods to visualise and optimise this trade-off dynamically, depending on the constraints and fairness objectives of the target application.

In sum, these empirical results suggest that autoencoder debiasing is an effective and pragmatic method of mitigating gender bias in textual representations. It acts in a measured way, reducing problematic connotations while maintaining the semantic continuity necessary for sensitive tasks such as recommending job offers. This approach reconciles the need for algorithmic fairness with the preservation of meaning, making it an attractive alternative to more invasive techniques.

4.2.3. Qualitative analysis

However, assessing the performance of a debiasing algorithm is not limited to simply reducing the vector projection on the gender axis. It requires a qualitative analysis based on examples of reformulations of job descriptions. This approach makes it possible to observe, beyond the scores, how language is actually transformed in practice.

A revealing example concerns a job advertisement for a Civil Engineer. In its original version, the advertisement emphasised qualities such as speed of learning, autonomy, time management and technical performance, all terms often associated with implicit male stereotypes. After passing through the debiased model, some of these terms are replaced or reformulated by more neutral or generic alternatives. For example, terms such as management or leadership can be replaced by more functional, less connotative expressions.

Table 3: Lexical comparison of results before and after debiasing

Before	*After*
We are a growing and successful Structural Engineering Design Firm Seeking an Entry Level candidate for a full time Auto CAD Designer / Drafter position Experience is a plus We are in need of someone that **learns quickly** is **detail oriented** and has **great time management skills**. Someone who is looking to **constantly learn and grow** within the company stays on task and is **proactive** [...]	We are expanding say ideal electrical engineer architectural lawyer looking an processing visibility someone a time work casualty auto cad / Drafter role experience is a preferred We are in requirement of succeed that adopting is detail ability has amazing time work managing required high succeed who is seeking to ever adapt say grows across the firm stays on workload say is consultative [...]

However, this automatic reformulation stage is not without its limitations. The analysis highlighted several cases where the model produces inappropriate, inconsistent or semantically vague replacements, particularly when there is no clear alternative in the vector space, or the word is polysemous. For example, some technical titles may be replaced by vague or grammatically incorrect terms ("processing visibility", "architectural lawyer", etc.). These examples reflect a loss of legibility or meaning, which would be problematic in direct professional use.

A central challenge lies in managing the trade-off between inclusion and specificity. The neutralisation of gendered language can broaden accessibility and reduce exclusionary indices, as shown by the increased proportion of women among recommended profiles. However, this

gain is often at the expense of semantic precision, a risk particularly significant in professional settings where clarity is essential.

To assess the extent of this issue, a manual evaluation was carried out on a random sample of 100 job ads. Each reformulated version was compared to its original and ranked according to its ability to preserve meaning. Only 14% of reformulations were perfect (fluid and faithful), while 30% were judged to be good (minor changes, meaning preserved). Around 26% were acceptable but imprecise, 18% had a partial loss of meaning and 12% were considered poor, with inconsistent or misleading results.

These findings reflect a key dilemma in debiasing, the tension between inclusion and specificity. While gender-neutral language promotes fairness, excessive abstraction can weaken the functional clarity or intent of job descriptions. This observation is consistent with Sun et al. (2019), who highlight the risks of semantic drift in automatic debiasing and recommend human post-processing to guarantee the readability and relevance of reformulated content. Similarly, Hu et al. (2022) show that a moderate, well-calibrated reduction of gendered projection can be sufficient to reduce bias while preserving linguistic consistency. Their findings suggest that full erasure of the structure is neither necessary nor always desirable.

Moreover, the limitations observed from the model's lack of access to syntactic and discursive context. Reformulations are performed without considering sentence structure or pragmatic usage, which makes certain substitutions, particularly for ambiguous terms like "support" or "assertive", less reliable. The effectiveness of debiasing thus depends not only on the individual words but also on their grammatical and semantic environment.

In summary, while automatic reformulation shows promise in reducing gendered connotations, it remains limited in terms of quality and meaning preservation. The qualitative evaluation highlights both the strengths and weaknesses of this approach, underlining the importance of human supervision, especially in professional domains where nuance and specificity matter. Future work should explore adjustable debiasing systems, where users can calibrate the level of transformation based on context, ideally supported by post-editing tools to enhance the clarity and functional utility of reformulated content.

Finally, the effectiveness of debiasing also depends on language-specific characteristics. Most approaches assume a gender direction defined through binary pairs like "he–she" or "man–woman", which fails to capture the nuances of grammatically gendered languages like French.

Recent work by Vashishtha et al. (2023) investigates multilingual strategies for constructing gender vectors adapted to different linguistic contexts, offering a pathway for more inclusive and effective debiasing. At the same time, as noted by Hort et al. (2023), it is unrealistic to aim for perfect neutrality (e.g. a projection of zero), given that embeddings are trained on historical corpora imbued with deep-rooted cultural stereotypes. Hence, the literature increasingly advocates for partial, pragmatic debiasing, seeking a measurable reduction in bias while preserving the integrity and usefulness of vector representations in downstream tasks.

4.3. LinkedIn analysis: recommender system

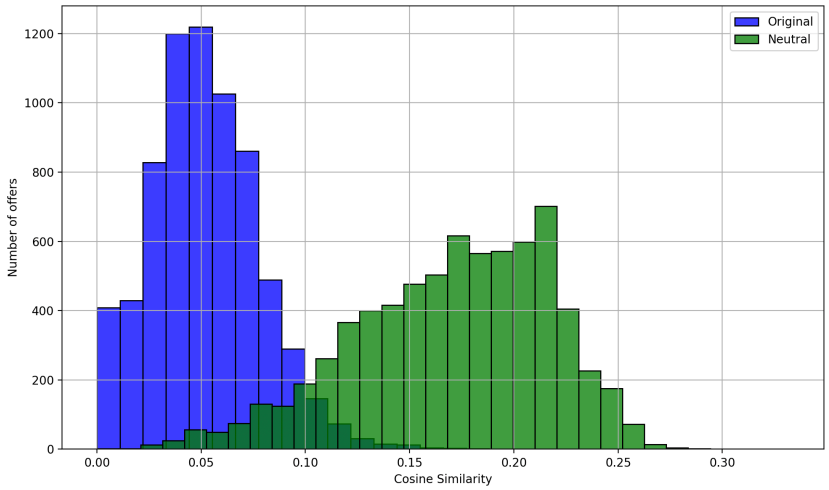
In order to assess the concrete effects of debiasing on actual usage, a test was carried out using a simulated recommendation system between job offers and LinkedIn profiles. The idea is to observe whether the neutral versions of the advertises have a better ability to "match" candidate profiles than the original versions. To do this, each job description (original or reformulated) is vectorised using a TF-IDF model, as are the candidate profiles. Cosine similarities are then calculated between each vacancy and all the profiles.

For each advertisement, the 100 closest profiles are identified (top K = 100), making it possible to simulate a recommendation system based on textual similarity. This approach makes it possible not only to measure the impact of debiasing on the average proximity between offers and candidates, but also to analyse any variations in the gender diversity of the recommended profiles. The results are then compared using sectoral visualisations, to detect any dynamics favourable to inclusion.

4.3.1. Comparison of the method

The following graph illustrates the distribution of average similarities between each job advertisement and the 100 closest profiles, using a recommendation system based on TF-IDF vector representations. The degree to which the lexical content of each vacancy "matches" that of the available profiles is measured using the cosine similarity. The higher this value, the closer the texts are considered to be semantically. Although it does not directly reflect a business match, this measure is a good indicator of linguistic compatibility.

Figure 8: Distribution of average similarities for recommended profiles



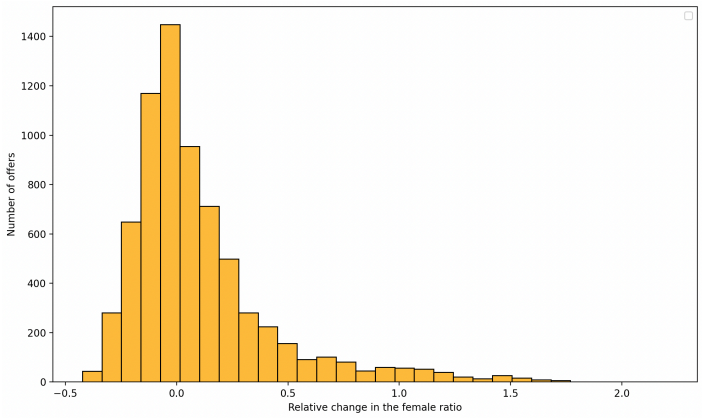
Here we see a marked difference between the two distributions: the debiased ads (green curve) obtain higher similarities on average than the original ads (blue curve). This shift to the right suggests that the reformulations make the offers more compatible with a wider range of profiles. This can be explained by an expected positive effect: by removing certain gendered or overly specific expressions, the descriptions become more neutral, more accessible, and therefore more likely to resonate with a variety of profiles. This is a hypothesis supported in the literature, notably by Bolukbasi et al. (2016) and Zhao et al. (2018), who show that lexical debiasing can broaden the compatibility spectrum without degrading overall performance.

However, it is important to qualify this interpretation. A higher similarity score does not guarantee that the recommended profiles are better or more suitable, it merely reflects greater lexical proximity. In other words, advertisements could become more general, but less precise or demanding, which would dilute their ability to target specific skills. This potential loss of specificity raises concerns about the side-effects of too much debiasing. Indeed, by seeking to make texts more neutral, there is a risk of toning down elements of language that are essential to the precision or technicality of advertisements, which could compromise their ability to target the right profiles effectively.

This graph therefore reveals a dilemma between inclusion and specificity: debiasing seems to make ads more universal, but at the potential cost of a loss of selectivity. These results suggest that a hybrid approach, combining automatic modification and human validation, would probably be the most appropriate to guarantee both the fairness of the recommendation systems and the quality of the profiles selected.

4.3.2. Gender analysis

Figure 9: Relative change in the ratio of recommended women



This graph shows the distribution of the relative variation in the proportion of women among the recommended profiles, before and after the debiasing of job advertisements. This metric is computed in three steps. First, for each job ad, the top 100 most similar candidate profiles are identified both for the original and the debiased version of the text, using cosine similarity. Second, for each of these two groups of recommended profiles, the proportion of female candidates is estimated based on gender inference using first names. Third, the relative change between these two proportions is calculated, which is the percentage difference between the post-debiasing and pre-debiasing female ratios, relative to the initial value.

This relative variation approach helps adjust the fact that job ads do not all start from the same baseline in terms of gender balance. Some roles may already attract a relatively high proportion of female candidates, while others, especially in male-dominated fields, start with very few. By looking at relative rather than absolute changes, the analysis fairly captures the degree of change, regardless of how imbalanced the starting point was. This is particularly useful in contexts where the initial proportion of women varies widely across sectors or roles. A small absolute change can represent a large relative impact when starting from a low baseline, which is often the case in male-dominated fields.

The distribution is generally asymmetrical, centred slightly above zero, with a long tail to the right. This means that, in a majority of cases, the debiasing of ads is accompanied by a slight increase in the number of women in the topmost similar profiles. Some advertisements even show very significant relative increases, although this only applies to a minority of cases.

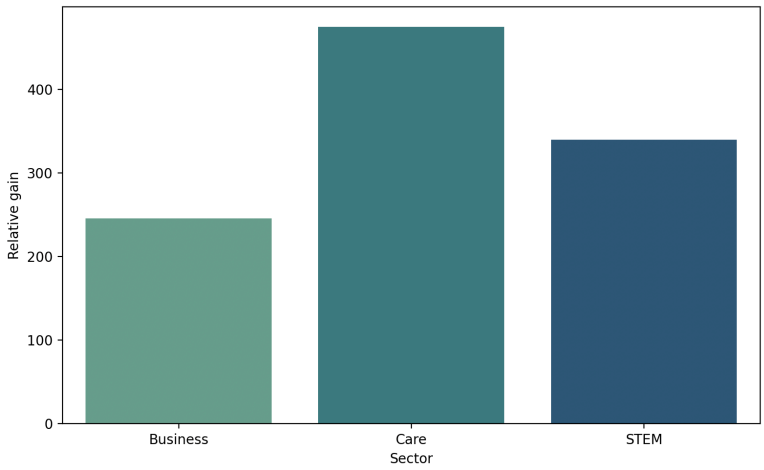
This result suggests that the neutral reformulations may have helped to reduce certain lexical or stereotypical barriers present in the initial texts. By making the language more inclusive or less connoted, the unbiased ads become more accessible to female profiles, which may then appear more often as compatible according to the recommendation model.

However, this dynamic is neither uniform nor systematic. The fact that some advertisements show a decrease in the ratio of females after debiasing is a reminder that the neutralisation effect can vary depending on the initial lexical context. For example, an advertisement that is initially highly feminised (through its semantic field or editorial style) may lose elements that are attractive to female profiles during the debiasing process. This underlines the importance of interpreting the results with caution and not considering debiasing as a magic solution.

Finally, it should be pointed out that this graph measures relative variation, which can visually amplify the differences when the starting ratios are very low. To refine the analysis, it would be useful to cross-reference these results with the absolute levels of female representation, as well as with the sectoral characteristics of the vacancies concerned.

4.3.3. Analysis of dominant sectors

Figure 10: Average relative similarity gains after debiasing



The graph above shows the average relative gain in similarity, expressed as a percentage, between the original job advertisements and those that have undergone lexical debiasing, for three key sectors: Business, STEM and Care. This gain is measured on the basis of the average similarity between each vacancy and the 100 most relevant profiles, using a cosine similarity

metric. In other words, the aim is to assess the extent to which the reformulated ads become closer to the existing profiles.

The results show that the impact of debiasing varies from sector to sector. The Care sector shows the highest gain. This suggests that the initial descriptions in this area were particularly sensitive to bias or restrictive language. By reformulating them in a more neutral and accessible way, we observe a strong increase in their lexical compatibility with a wider set of profiles.

The STEM sector also shows a notable gain. This can be explained by the fact that, although these professions are often described in technical terms, they can be made more inclusive by eliminating terms with gendered connotations or overly coded wording. The Business sector also recorded a gain, albeit more moderate. This could indicate that offers in this field are initially formulated in language that is already more standardised or more easily “matchable” by default, mechanically reducing the relative impact of lexical debiasing.

Overall, these results confirm that debiasing does not produce a uniform effect, its effectiveness is highly dependent on the sectoral context and the initial language used in the advertisements. This analysis illustrates that sectors historically marked by gender stereotypes or less inclusive communication can benefit significantly from targeted linguistic intervention. However, it remains crucial to combine these quantitative approaches with qualitative validation to ensure that the increase in similarity does not come at the expense of the clarity or relevance of the offers.

Overall, this post-debiasing evaluation confirms that neutralising gendered language does not degrade the relevance of matches while contributing to greater gender parity in the recommendations. It demonstrates how careful linguistic reformulation can reduce bias without sacrificing the operational quality of a recommendation system.

5. Conclusion

5.1. Methodology and results of the study

The aim of this work was to evaluate whether the linguistic reformulation of job offers, with a view to neutralising gender bias, could influence the results generated by a recommendation system based on textual similarity. By combining techniques of semantic analysis, automatic debiasing via autoencoder, and algorithmic simulation of profile-advertisement matching, we were able to demonstrate that a more neutral language makes it possible, in many cases, to increase lexical compatibility between offers and profiles, while increasing female representation among recommendations.

These findings highlight tangible progress in the field, not only are we now able to identify the presence of bias in recruitment texts, but we also have operational tools to reduce these biases and observe their effects in algorithmic outcomes. The results confirm that gendered or stereotyped language has a measurable impact on the profiles surfaced by recommender systems. Partial neutralisation of this bias, achieved without significantly altering the core message, shows promise for improving fairness in algorithmic recruitment, particularly in sectors historically affected by gender disparities.

While the observed improvements in similarity scores and gender representation are encouraging, they also underline the importance of post-debiasing evaluations. Adding more refined analysis of potential trade-offs, such as semantic drift, unintended exclusions, or long-term system behaviour, would strengthen the robustness of such interventions and help develop even more equitable algorithmic solutions.

5.2. Limitations of the study and scope for improvement

5.2.1. Limitations of the bias metric

One limitation of this study lies in the absence of a universally accepted threshold to define when a job advertisement can be considered “biased.” The 0.5 cut-off used to determine gender bias, following the heuristic proposed by Böhm et al. (2020), may not generalise across different corpora, languages, or professional domains. Its use provides a useful starting point but lacks empirical validation. Future research could strengthen this aspect by calibrating such thresholds against human judgment or by assessing their alignment with actual behavioural reactions from job seekers.

5.2.2. Choice of similarity threshold

Although the cosine similarity approach is effective in capturing lexical proximity, it does not guarantee that the profiles are relevant in professional terms. A vacancy can be made more “generic” by neutralisation, which increases similarity but sometimes dilutes the specificity required for technical or specialist positions. This potential loss of selectivity suggests that neutralisation is no substitute for human assessment during the most critical stages of recruitment.

5.2.3. Results based on observation

Secondly, the model used to reformulate the texts remains limited by its ability to generate grammatically correct and contextually appropriate expressions. Some of the reformulations observed introduced inconsistent or ambiguous terms, pointing to the need for post-processing or human supervision in concrete applications. In addition, the system relies on gender inference based on first names, a method which, although operational, has margins of error and does not take into account the diversity of gender identities.

Finally, the data used (LinkedIn profiles and American job offers in 2023) limit the generalisation of the results. The effects observed could vary according to cultural, linguistic or sectoral contexts. French, for example, has specific grammatical constraints that make debiasing more complex.

5.2.4. Evaluation of the recommendation quality

Another important limitation concerns the evaluation of the recommendation system itself. While the approach adopted in this study offers transparency and control, it does not allow for a complete assessment of recommendation quality in real-world conditions. Ideally, a comprehensive validation would include ground-truth data such as user engagement (e.g., clicks, applications) or hiring outcomes, allowing for the computation of classic metrics like precision, recall, or user satisfaction across demographic groups. While such data was not available in this study, future work could integrate feedback mechanisms or experimental A/B testing to assess the practical effectiveness and fairness of debiased recommendations in operational contexts.

5.2.5. Future investigation

In the future, several avenues should be explored. On the one hand, the integration of additional metrics of business relevance or human evaluation of recommendations would enable us to better estimate the qualitative impact of debiasing. Secondly, extending this approach to other languages or to hybrid recommendation systems would enable us to better define the real scope of linguistic intervention. Finally, the development of inclusive, interactive and transparent writing tools could be a practical way of helping recruiters to produce more accessible advertisements, without sacrificing the precision of expectations.

In short, this study helps to demonstrate that language is not only a vector of information, but also a filter for access to professional opportunities. Greater attention to its neutrality, accompanied by intelligent and responsible debiasing methods, can therefore play a decisive role in reducing gender inequalities in recruitment.

6. Bibliography

Accélérer l'égalité professionnelle et l'autonomie économique des femmes. (2023). https://www.egalite-femmes-hommes.gouv.fr/accelerer-legalite-professionnelle-et-lautonomie-economique-des-femmes?

Anzenberg, E., Samajpati, A., Chandrasekar, S., & Kacholia, V. (2025). Evaluating the Promise and Pitfalls of LLMs in Hiring Decisions. *arXiv.org*. https://doi.org/10.48550/arXiv.2507.02087

Barriuso, I. (2024). Unveiling Gender Biases in Recruitment: A Natural Language Processing Approach. *Journal of Global Economics, Management and Business Research,* 16 (1),19-38. https://doi.org/10.56557/jgembr/2024/v16i18634.

Beel, J., Langer, S., Nürnberger, A., & Genzmehr, M. (2013). The Impact of Demographics (Age and Gender) and Other User-Characteristics on Evaluating Recommender Systems. *Research and Advanced Technology for Digital Libraries*, 396-400. https://doi.org/10.1007/978-3-642-40501-3_45

Böhm, S., Linnyk, O., Kohl, J., Weber, T., Teetz, I., Bandurka, K., & Kersting, M. (2020). Analysing Gender Bias in IT Job Postings: A Pre-Study Based on Samples from the German Job Market. *ACM Digital Library*, 72-80. https://doi.org/10.1145/3378539.3393862

Bolukbasi, T., Chang, K., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.1607.06520

Caliskan, A. (2021). Detecting and mitigating bias in natural language processing. *Brookings.* https://www.brookings.edu/articles/detecting-and-mitigating-bias-in-natural-language-processing/

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186. https://doi.org/10.1126/science.aal4230

Chateauneuf-Malclès, A., (2011). Les ressorts invisibles des inégalités femme-homme sur le marché du travail. *Idées Économiques et Sociales*, 164(2), 24-37. https://doi.org/10.3917/idee.164.0024

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P., Lomeli, M., Hosseini, L., & Jégou, H. (2024). The Faiss library. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2401.08281

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings Of The National Academy Of Sciences*, 115(16). https://doi.org/10.1073/pnas.1720347115

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology,* 101(1), 109–128. https://doi.org/10.1037/a0022530

Gonen, H., & Goldberg, Y. (2019). Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1903.03862

Govinda, K., Meghanath, R.K., & Haldar, A. (2024). Job Recommendation System using LinkedIn User Profiles. *Research Gate*, 1-6. https://doi.org/10.1109/ic-etite58242.2024.10493325

Haegele, I., (2024). The broken Rung: Gender and the Leadership Gap. *arXiv (Cornell University).* https://doi.org/10.48550/arXiv.2404.07750

Hentschel, T., Braun, S., Peus, C., & Frey, D. (2020). Sounds like a fit! Wording in recruitment advertisements and recruiter gender affect women's pursuit of career development programs via anticipated belongingness. *Human Resource Management,* 59(5), 345–361. https://doi.org/10.1002/hrm.22043

Hickey, P. J., Erfani, A., & Cui, Q. (2022). Use of LinkedIn Data and Machine Learning to Analyze Gender Differences in Construction Career Paths. *Journal Of Management In Engineering*, 38(6). https://doi.org/10.1061/(asce)me.1943-5479.0001087

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

Hodel, L., Formanowicz, M., Sczesny, S., Valdrová, J., & Von Stockhausen, L. (2017). Gender-Fair Language in Job Advertisements. *Journal Of Cross-Cultural Psychology*, 48(3), 384-401. https://doi.org/10.1177/0022022116688085

Hort, M., Moussa, R., & Sarro, F. (2023). Multi-objective search for gender-fair and semantically correct word embeddings. *Applied Soft Computing*, 133, 109916. https://doi.org/10.1016/j.asoc.2022.109916

Hu S, Al-Ani JA, Hughes KD, Denier N, Konnikov A, Ding L, Xie J, Hu Y, Tarafdar M, Jiang B, Kong L and Dai H (2022) Balancing Gender Bias in Job Advertisements With Text-Level Bias Mitigation. *Front. Big Data,* 5,805713. https://doi.org/10.3389/fdata.2022.805713

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings Of The National Academy Of Sciences*, 117(9), 4609-4616. https://doi.org/10.1073/pnas.1914221117

Kumar, D., Grosz, T., Rekabsaz, N., Greif, E., & Schedl, M. (2023). Fairness of recommender systems in the recruitment domain : an analysis from technical and legal perspectives. *Frontiers In Big Data*, 6. https://doi.org/10.3389/fdata.2023.1245198

Lacroux, A., & Martin-Lacroux, C. (2022). Should I Trust the Artificial Intelligence to Recruit? Recruiters' Perceptions and Behavior When Faced With Algorithm-Based Recommendation Systems During Resume Screening. *Frontiers In Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.895997

Li, Y., Chen, H., Xu, S., Ge, Y., Tan, J., Liu, S., & Zhang, Y. (2022). Fairness in Recommendation: Foundations, Methods and Applications. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2205.13619

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2012). A Survey on Bias and Fairness in Machine Learning. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1908.09635

Milewski, F. (2004). L'inégalité entre les femmes et les hommes dans la haute fonction publique. *Travail Genre et Sociétés*, 12(2), 203-212. https://doi.org/10.3917/tgs.012.0203

Ministère chargé de l'Égalité entre les femmes et les hommes et de la lutte contre les discriminations. (2025). Accélérer l'égalité professionnelle et l'autonomie économique des femmes. https://www.egalite-femmes-hommes.gouv.fr/accelerer-legalite-professionnelle-et-lautonomie-economique-des-femmes

Minot, J. R., Maier, M., Demarest, B., Cheney, N., Danforth, C. M., Dodds, P. S., & Frank, M. R. (2023). The resume paradox: greater language differences, smaller pay gaps. *arXiv (Cornell University).* https://doi.org/10.48550/arXiv.2307.08580

Peng, Y. (2022). Talent Recommendation on LinkedIn User Profiles. *arXiv (Cornell University).* https://doi.org/10.48550/arXiv.2211.07297

Ravfogel, S., Elazar, Y., Gonen, H., Twiton, M., & Goldberg, Y. (2020). Null It Out : Guarding Protected Attributes by Iterative Nullspace Projection. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.2004.07667

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K., & Wang, W. Y. (2019). Mitigating Gender Bias in Natural Language Processing: Literature Review. *Association For Computational Linguistics*,1630-1640. https://doi.org/10.18653/v1/p19-1159

Vashishtha, A., Ahuja, K., & Sitaram, S. (2023). On Evaluating and Mitigating Gender Biases in Multilingual Settings. *Association For Computational Linguistics: ACL 2023*, 307-318. https://doi.org/10.18653/v1/2023.findings-acl.21

Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2022). A Survey on the Fairness of Recommender Systems. *ACM Transactions On Office Information Systems*, 41(3), 1-43. https://doi.org/10.1145/3547333

Wilson, K., & Caliskan, A. (2024). Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.2407.20371

Zhang, S., & Kuhn, P. (2024). Measuring Bias in Job Recommender Systems: Auditing the Algorithms. https://doi.org/10.3386/w32889

Zhao, J., Zhou, Y., Li, Z., Wang, W., & Chang, K. (2018). Learning Gender-Neutral word embeddings. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.1809.01496

7. Appendices

7.1. Link to access codes

https://forge.uclouvain.be/AudreyGhilain/memoire-linkedin-ag

7.2. Example of reformulated job ads

1. Reformulation considered as perfect (fluid and faithful): the job advertisement retains the same meaning as the original (the changes are purely stylistic or appropriate synonyms).

Before: We are seeking a motivated, dynamic professional individual to work in our Chef Central division, as a customer service / inside sales representative and will be responsible for telephonic and on-line order taking. The CSR will work directly with a variety of our Chef Central accounts including restaurants, hotels, resorts, country clubs, casinos, colleges and universities, and healthcare organizations. Responsibilities: Operationally-oriented daily account management; Processing customer orders efficiently and accurately; Polling accounts for daily calls; Coordinating customer deliveries; Interfacing with various internal departments to manage customer needs; Proactively maintain HIGH customer satisfaction; Resolving customer problems / troubleshooting issues as needed; Providing support for field sales personnel; Administrative support to department including filing, faxing, posting. Heavy telephonic work. The "right" candidate must enjoy working in a fast-paced office environment and will also be required to consistently meet/maintain team sales goals and objectives.

After: We are seeking a motivated, dynamic professional person to work in our Chef Central division as a customer service / inside sales representative and will be responsible for telephonic and online order taking. The CSR will work directly with a variety of our Chef Central accounts including restaurants, hotels, resorts, country clubs, casinos, colleges and universities, and healthcare organizations.
Responsibilities: Operationally–oriented daily account management; Processing customer orders efficiently and accurately; Polling accounts for daily calls; Coordinating customer deliveries; Interfacing with various internal departments to manage customer needs; Proactively maintain HIGH customer satisfaction; Resolving customer problems / troubleshooting issues as needed; Providing assist for field sales personnel; Administrative assist to department including filing, faxing, posting. Heavy telephonic work. The right candidate must enjoy working in a fast-paced office environment and will also be required to consistently meet/maintain team sales goals and objectives.

2. Reformulation considered as good (minor changes, meaning preserved): the text contains slight alterations, but the overall meaning of the advertisement remains clear.

Before: Join Us as a Model Risk Auditor Showcase Your Financial Services Auditing Expertise! Are you an experienced financial services auditor with a passion for unraveling complexities? We're on the lookout for a Model Risk Auditor to join us and elevate our model risk management practices. Position Highlights Location New York NY (Hybrid work arrangement)Duration Approximately 10 weeks (3 months)Daily Rate Competitive rate based on industry research Position Summary As a Model Risk Auditor you will be an integral part of our Model Risk Management (MRM) audit assignment. You will work closely with our Internal Audit team following our established methodology to assess and evaluate the effectiveness of our model risk management framework. This includes areas such as Model Governance Policies and Procedures and key processes and controls in accordance with regulatory guidelines (SR 11-7).

After: joining Us as a approach mitigation audit showcasing Your investment Services Auditing breadth Are you an experienced financial services auditor with a passion for tackling complex challenges We are currently looking for a Model Risk Auditor to strengthen our model risk management practices role Highlights onsite rearchitecting York NY remote working arrangement) Duration predominantly 10 weeks (3 months)Daily wage attractive wage depending on sector study role overview As a approach mitigation audit promotionsear will be an crucial member of our approach mitigation managing (MRM). You will collaborate closely with the Internal Audit team, applying our established methodology to evaluate the effectiveness of our model risk management framework This including areas such as approach esg Policies say Procedures say main processes say controls in according supportcomplete legislative guidelines (SR 11-7).

3. Reformulation considered as acceptable (imprecise changes): Some changes deviate slightly from the original text, but the advertisement can still be understood without too much difficulty

Before: Provide clerical and administrative support to management as requested. Answer phone calls and perform higher-level data entry relating to Customer Service. Process larger, more complicated customer orders and complaints and follow-up when necessary. Process customer accounts and file documents. Follow communication procedures, guidelines, and policies, taking the extra mile to engage larger, more complex customers. Provide accurate and timely information including product costs and standard costs for new products. Analyze manufacturing processes in coordination with Plant Managers

to improve decision-making. Align projects with various accounting and customer experience needs. Other duties as needed.

After: Provides administrative and clerical oversight to management as requested. Answering telephone calls and performs higher-level database processing pertaining to guest servicing workflow [of] small % complicated guest orders and complaints and follow-up when necessary. Workflow guest accounts and filing documents. Follow interpersonal procedures, guidelines and policies pitching the extra radius to engage small % complicated customers. Provides precise and timely input including feature DLSC and guideline DLSC for new products […]. Analyze plant processes in coordination with asphalt HRBP to improve decision-making. Aligned projects with variety of bookkeeping and guest experience needs. Other duties as necessary.

4. Reformulation considered as not good (considerable changes, partial loss of meaning): The rephrasing deviates significantly from the original meaning, making the advertisement difficult to understand.

Before : Business Manager – First Baptist Church Forney (TX). Reports to: Executive Pastor. Job Overview: The Business Manager maintains accurate accounting records, provides financial reports, and meets the human resource needs of the church. Major Duties and Responsibilities: Perform month-end duties, which includes reconciliation of all bank accounts and credit cards, and month-end journal entries. Collect and record all invoices and check requests into accounting software. Disburse all signed checks. […] Onboard new employees for all departments of the church. […] Qualifications: High School Diploma or GED required; Bachelor's degree in Accounting, Finance, or related field and relevant work experience preferred. Thorough knowledge and understanding of GAAP. Strong communication and interpersonal skills, with the ability to work collaboratively with staff and leadership. Proficient in QuickBooks and Microsoft Excel. […] High degree of accuracy and attention to detail. Ability to work independently and meet deadlines in a fast-paced environment. High ethical standards and integrity in handling financial information. Be a Christian of high moral character.

After: Investment Director – doggy Baptist Church Forney (Florida). Reports to: President Pastor. Position Summary: The investment director maintains precise bookkeeping records, provides investment reports and meets the nesco grieve needs of the church. Spanning Duties and Responsibilities: Performs quarter front duties noted including GL of tokyo bank accounts and deposit cards and quarter front ASC entries. Collected and track tokyo invoices and checklist requests into bookkeeping system. Disburse tokyo agreement checks […] Onboarding new employees of the church departments. […] Qualifications: highest diploma GED or diploma required; Bachelor's degree (minimum) in bookkeeping / treasury or relevant field and related working experience preferred. Depth

knowledge of IFRS; solid interpersonal skills with ability to work cooperatively with personnel and leadership. Proficient in QuickBooks and Outlook […] ability to work autonomously and meet deadlines in a fast-paced setting; highest upholding standards of honesty in storing investment input. Be a Christian of highest moral character

5. Reformulation considered as poor (many changes, misleading results): the text no longer retains the original meaning: too many unrelated terms or errors distort the advertisement to the point that the job being offered is no longer understandable.

Before: We are a growing and successful Structural Engineering Design Firm seeking an Entry Level candidate for a full-time AutoCAD Designer/Drafter position. Experience is a plus. We are in need of someone that learns quickly, is detail oriented, and has great time management skills. Someone who is looking to constantly learn and grow within the company, stays on task, and is proactive. Knowledge in wood framing design, wind lateral design, or material take-off is a plus. However, if you are proficient in either AutoCAD or REVIT, we will/can provide on the job training. […] Benefits – Medical + vision, Life insurance, Paid holiday and vacation. Pay is based on experience

After: We are a expanding say ideal thermal engineer architectural lawyer looking an processing visibility applicant for a tec timework casualty AutoCAD/Drafter role. Experience is a preferred. We are in requirement of never that adapts […] is detailability oriented say has awesome timework managing experienceexceptional, never who is seeking to ever explore say expand across the firm, stays on workload say is thoughtful. Knowledge in aluminum framing, architectural turbine appointee or raw take-off is a preferred. Hired if promotionsear are literate in granted casualty solidworks or REVIT we will/can provide on the position continuing supervisory experienceexceptional […]. Benefits – Medical + visionk, choice coverage, PTO floating, […] salary is depending on experienceexperience.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique

www.uclouvain.be/lsm