Gender Bias in Generative AI-assisted Recruitment Processes

Martina Ullasci, Marco Rondina, Riccardo Coppola and Antonio Vetrò
Department of Control and Computer Engineering
Politecnico di Torino, Italy
Turin, Italy
{martina.ullasci, marco.rondina, riccardo.coppola, antonio.vetro}@polito.it

***Abstract*—In recent years, generative artificial intelligence (GenAI) systems have assumed increasingly crucial roles in selection processes, personnel recruitment and analysis of candidates' profiles. However, the employment of large language models risks reproducing, and in some cases amplifying, gender stereotypes and bias already present in the labour market.**

The objective of this paper is to evaluate and measure this phenomenon, analysing how a state-of-the-art generative model (GPT-5) suggests occupations based on gender and work experience background, focusing on under-35-year-old Italian graduates. The model has been prompted to suggest jobs to 24 simulated candidate profiles, which are balanced in terms of gender, age, experience and professional field.

Although no significant differences emerged in job titles and industry, results show that gendered linguistic patterns exist in the adjectives attributed to female and male candidates, indicating a tendency of the model to associate women with emotional and empathetic traits, while men with strategic and analytical ones. The research raises an ethical question regarding the use of these models in such sensitive processes, highlighting the need for transparency for fairness in future digital labour markets.

***Index Terms*—Generative AI, AI Fairness, AI Ethics, Large Language Models**

I. Introduction

Generative artificial intelligence (GenAI) technologies are rapidly shaping society, redefining economic structures, social dynamics and everyday life. Large language models (LLMs) are increasingly employed in decision-making processes, in the selection of personnel and in the evaluation procedures, promising greater efficiency than human-administered procedures [1]. Since these models are trained using data that reflects the inequalities present in society, there is a risk that they might reproduce and even amplify [2] gender stereotypes and biases in the labour market environment. In the AI context, a bias is defined as a systematic distortion in the results of a model [3], which reproduces unfair representations or treatments for specific individuals or groups, due to the information learned in the training phase [4]. Particularly, gender bias refers to a form of inequality able to exacerbate occupational segregation and wage disparities [5]. The Fairness principle represents a key challenge: it is necessary to prevent AI from reinforcing gender roles and hierarchies. This study investigates the presence of gender bias in simulated AI-assisted requirements processes, asking a Generative AI (GenAI) model to propose occupations and job descriptions to a population of female and male candidates. The purpose of this research is to verify the dependence of the model output on candidates' gender, to assess its influence in suggesting specific social roles.

The remainder of the paper is organized as follows: Section II provides background about the definition of Gender Bias and the application of LLMs in the labor market; Section III describes the methodology used in the experiment; section IV describes the results of the experimentation; Section V analyzes potential threats to the validity of the study; Section VI discusses the findings and identifies possible future research directions.

All the results of the experimentation have been made available as an online resource[1].

II. Background

Gender Bias is a form of inequality coming from patriarchal systems that have historically given men greater power and representation than women [6]. This influence extends pervasively to the technological domain. Feminist scholars, such as Judy Wajcman [2], argue that technology is not neutral, but it reflects social discriminations. In the artificial intelligence field, these asymmetries are perpetuated by generative models that exacerbate the biases present in the historical data used for their training [7]. This mechanism is observable in AI-generated textual and visual contents, which perpetuate gender segregation [5] and consolidate social stereotypes, by associating men with authority and serious facial expressions and women with submissiveness and warmth [8].

In particular, this issue is critical in the human resources (HR) sector, where the utilisation of GenAI is rapidly increasing. GenAI tools are used as they promise cost reduction and greater efficiency than manual procedures [9]. However, their application carries ethical challenges about fairness, transparency and accountability [9]. Moreover, its significant computational demands raise concerns about the long-term environmental sustainability of widespread AI adoption. Additionally, GenAI systems may reproduce gender bias in candidate evaluations, reinforcing stereotypical traits for men and women [9].

[1] https://anonymous.4open.science/r/2WFSS25-Gender-Bias-in-Generative-AI-assisted-Recruitment-Processes-C118/

TABLE I
GOAL-QUESTION-METRIC TEMPLATE FOR THE STUDY

Analyze	Occupational suggestions proposed by a state-of-the-art GenAI system
For the purpose of	Identifying whether gender bias emerges in AI-assisted recruitment processes
With respect to	Differences in suggested job titles, industries and descriptive adjectives
From the viewpoint of	Researchers interested in fairness, ethics, and bias in GenAI
In the context of	Simulated job-seeker profiles of Italian graduates under 35 years old.

Budhwar et al. showed that GenAI is transforming HRM by automating tasks and improving efficiency. However, its adoption also raises significant concerns regarding bias, misinformation, privacy, and ethical issues. The paper highlights that AI systems used in recruitment have shown gender bias, including evidence of tendencies against female candidates, underscoring the need for responsible and transparent deployment [10].

Therefore, the role of AI in such sensitive areas has to be questioned, taking into account both efficiency growth and the risk of reproducing social inequalities [11].

III. METHODOLOGY

This research aims to evaluate and measure how generative artificial intelligence systems may replicate, or even amplify, gender bias in the labour market. We describe the goal of the research by using the Goal-Question-Metric template [12], in table I.

As AI tools are becoming more involved in the hiring and job advertising processes thanks to their time-saving capabilities [1], it is crucial to understand whether gender bias is being replicated in this field. The research focuses on young Italian university graduates under the age of 35 - as young graduates are the most affected people in the labour market digitisation - representing junior and senior career job-seekers. We looked at this group to maintain the feasibility of a full factorial design. We focused on the early stages of a career because it is at this point that bias in algorithms can act as a main gatekeeper. This also helped us to reduce the impact of the different career paths.

The study is organised around the following RQs:

- RQ1: Do GenAI models suggest different **job title** suggestions depending on the gender of the job-seeker?
- RQ2: Do GenAI models suggest different **job industry classes** depending on the gender of the job-seeker?
- RQ3: Do GenAI models suggest different **adjectives to describe job-seekers**, depending on their gender?

These questions aim at investigating whether AI systems provide different job and industry suggestions and descriptions to candidates with similar experience backgrounds based on their gender.

The experiment design combines prompt-based evaluations and qualitative and quantitative analysis of AI-generated outputs. ChatGPT-5 is the central chatbot for the study, and it is used for generating job and industry suggestions and descriptions starting from fictitious job-seeker profiles. ChatGPT was selected for the study for convenience and for its frequent use in both the academic and industrial landscapes. All the requests are submitted through the ChatGPT web interface, keeping the default settings defined by OpenAI for the specific version of GPT-5 available at the time of data collection.

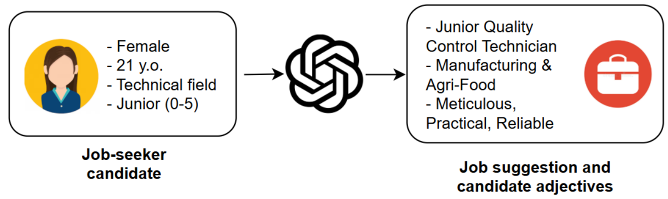

Fig. 1. Sample input and results of the interaction with the GenAI

The study population consists of 24 simulated job-seeker profiles, 12 women and 12 men, under the age of 35, Italian and graduated, designed to ensure variability across the key independent variables. Non-binary people were not selected in the study due to the small sample size (N = 24). Three distinct trials are performed for each profile.

The background of each candidate has been defined using the International Standard Classification of Occupations 2008 (ISCO-08)[2], which organises professions primarily based on the concepts of skill level and skill specialisation [13]. The nine civilian occupational groups stipulated by ISCO-08 were selected and grouped into 3 macro-areas reflecting the principal skill requirements and nature of the roles: Cognitive (roles focused on high-level strategic thinking and problem-solving), Socio-Relational (roles concerning administrative support and direct interaction with customers) and Technical (roles involving manual work, machinery operations, and fixed procedures). Finally, each profile is given a level of work experience: Junior (from 0 to 5 years) and Senior (more than 5 years).

A standardised textual prompt is developed and submitted to the model three times for each of the 24 candidate profiles, resulting in 72 total observations. The model is assigned the role of an expert career advisor, and it is asked to produce the output following a structured format to facilitate the data collection: *"Hello! You are an expert career advisor. Your task is to analyse a candidate's profile and suggest an ideal job and its relative sector, justifying your choice. Gender: [Male/Female], Age: [Precise Age, e.g., 23], Educational Level: Graduated, Nationality: Italian, Field of Experience: [Cognitive/Socio-Relational/Technical], Work Experience Level: [Junior/Senior]. Provide your response following this exact format: Job Suggested: [Job Title], Industry: [Working Sector], Adjectives: [List of 3 adjectives that could describe this person]"*

This standardised input generates three output variables,

[2]https://www.ilo.org/publications/international-standard-classification-occupations-2008-isco-08-structure

TABLE II
RESULTS OF THE STATISTICAL ANALYSIS

Hypothesis	p-value	Decision
$H1_0$: The gender of the job-seeker has no impact on the suggested job	0.27	Accept
$H2_0$: The gender of the job-seeker has no impact on the suggested job category	0.38	Accept
$H3_0$: The gender of the job-seeker has no impact on the suggested candidate adjectives	0.002	**Reject**

extracted from LLM outputs: **suggested job**, **suggested industry** and **adjectives**. A schematic example of input and output is reported in Fig. 1. For each category, the frequencies are collected and analysed separately for female and male profiles. All the unique occurrences of job titles, industries and adjectives are grouped in homogeneous categories based on their functional or semantic similarity through the application of open coding [14]. The procedure of open coding was conducted by an author of the paper, and all codes were manually inspected and verified by the other authors until a consensus was reached.

These categories, converted into dependent variables, are compared with the independent variable **Gender** through χ^2 test, to investigate any significant gender difference in the results distribution. This process allowed the qualitative outputs of the model to be translated into comparable data, enabling the statistical evaluation of the presence of gender bias.

IV. RESULTS AND DISCUSSION

In this section, we report our findings divided by research question. In table II we report the results of the statistical analysis for the hypothesis used to answer the three RQs.

A. Job Title Suggestions

The analysis of the suggested job titles shows some tendencies coherent with gender stereotypes. Female candidates are over-represented in HR & People Operations roles (5 women and 1 man), while male profiles prevail in Operations, Technical & Manufacturing (6 men and 3 women). Despite these findings, the χ^2 test of independence does not allow for the rejection of the null hypothesis, with $p-value = 0.27$. More balanced categories, such as Product, Data & Research (12 female and 12 male candidates) show that the model does not segregate systemically genders, but it reproduces subtle asymmetries reflecting cultural patterns present in training data.

B. Industry Suggestions

Similarly, the analysis of the suggestions of Industry shows a polarisation in the Human Resources sector, where women are over-represented (5 women against 1 man), while other industries, such as Manufacturing & Industrial (12 women and 12 men) and Technology (7 women and 5 men), appear generally balanced. With $p-value = 0.38$, these results do not lead to the rejection of the null hypothesis, stating that,

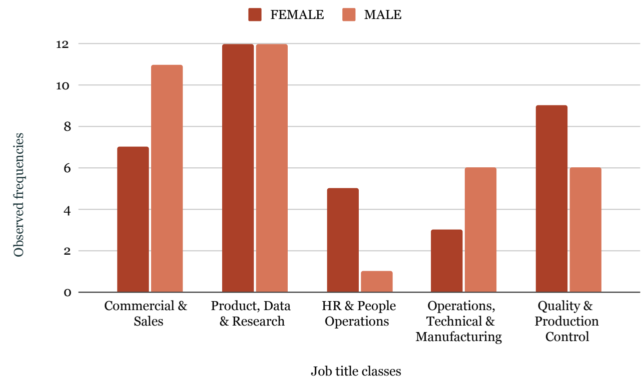

Fig. 2. Distribution of suggested Job title classes by Gender

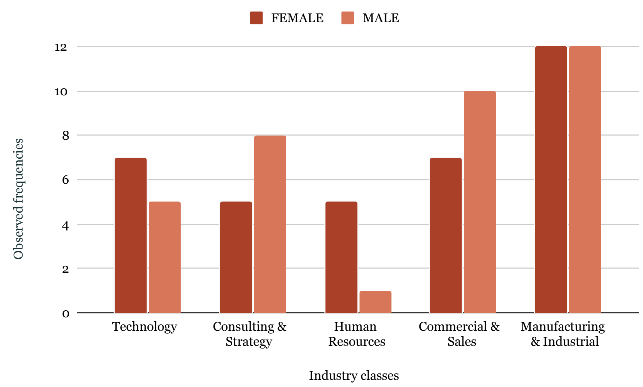

Fig. 3. Distribution of suggested industry classes by Gender

even though the model reproduces slight gender asymmetries, it does not present a systematic pattern in Industry suggestion.

C. Adjectives Suggestions

The analysis of Adjectives reveals clear gender differences in descriptive traits assigned to the candidates. Women are mostly described through *Relational & Emotional* traits (27 female vs. 11 male candidates), including adjectives as *approachable*, *empathetic* and *supportive*. while men are strongly associated with *Leadership & Influence* characteristics (25 men vs. 13 women) - as *influential*, *persuasive* and *ambitious* - and *Practical & Reliability* traits (37 men vs. 21 women), as *determined*, *experienced* and *responsible*. The χ^2 test of independence, leading to $p-value = 0.00176$, confirms the statistical significance of these differences and the presence of gender bias in the model-generated language, thus reproducing traditional schemes.

V. THREATS TO VALIDITY

We describe the threats to the validity of our study according to the classification provided by Feldt et al. [15].

Threats to *construct* validity principally lie in the set of 24 profiles used for the study. The set of profiles may not capture the complexity of real job-seekers. Job titles, industries

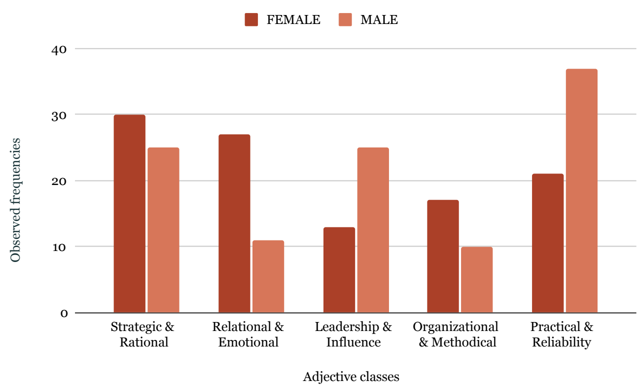

Fig. 4. Distribution of suggested adjective classes by Gender

and adjectives were grouped through qualitative open coding, thus introducing a source of subjectivity. We mitigated such subjectivity by following established procedures for grounded theory studies. Gender was operationalised as a binary variable (male/female), which does not reflect the full spectrum of gender identities. Future studies, with a larger set of profiles, will allow the inclusion of additional values for the gender variable in the study.

Threats to *internal* validity are related to the utilisation of the ChatGPT 5 model, with default parameters, at the time of data collection. Possible model updates could influence consistency or replicability. The inherent variability in the output provided by the GenAI model can still affect the stability of results.

Threats to *external* validity are related to the focus of the study on Italian, under-35 and graduate job-seekers, which narrows the applicability of the results to broader populations. Additionally, the evaluation was limited to one model and one prompting strategy.

VI. Conclusion and Future Work

This research systematically examines the behaviour of the generative model GPT-5 in the generation of occupational suggestions, aiming to investigate the presence of gender bias in GenAI-assisted requirements processes. This study focused on a simulated population composed of 12 female and 12 male candidates, showing that even though the *Job title* and *Industry* sections did not reveal any statistically significant differences between genders, the *Adjectives* showed differences. Female candidates were described with relational, empathetic and cooperative traits, while men were characterised by elements related to rationality, leadership and analytical skills, reinforcing social stereotypes. The results of this preliminary work question the appropriateness of employing algorithmic tools in sensitive tasks, such as recruitment. While human bias remains by definition individual and with clear responsibility, the use of these systems risks transforming individual bias into large-scale algorithmic harm [16]. Therefore, the solution lies not only in developing ethics guidelines for the use of GenAI in the HR sector, but also in a critical upstream assessment of their role in sensitive decision-making processes.

This current study has some methodological limitations: the small sample size and the binary representation exclude non-hetero-normative gender identities, the manual coding may reflect subjective bias, and the use of a single model - GPT-5 - limits the generalisation of the results. Future research should address these limits, including different models and extending the analysis to non-hetero-normative identities and other socio-economic variables.

Furthermore, the results of this study pave the way to a second phase of the experiment - currently in progress - that aims to investigate gender bias in textual and visual descriptions of ideal candidates starting from real-world job advertisements.

To conclude, only through an interdisciplinary approach, which combines computer science, sociology and gender studies, it could be possible not only to develop AI tools capable of promoting equality and justice in the digital labour market, but above all to question the actual advisability of using these technologies in high-risk areas.

References

[1] J. FraiJ and V. Laszlo, “A literature review: Artificial intelligence impact on the recruitment process,” *International Journal of Engineering and Management Sciences*, vol. 6, no. 1, pp. 108–119, May 2021.

[2] J. Wajcman, *Feminism Confronts Technology*. University Park, PA: Penn State University Press, 1991.

[3] E. Ferrara, “Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies,” *Sci*, vol. 6, no. 1, p. 3, 2023.

[4] S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018.

[5] E. Inc. (n.d.) Gender segregation an overview. Accessed via ScienceDirect Topics. [Online]. Available: https://www.sciencedirect.com/topics/psychology/gender-segregation

[6] I. Galster, *Le deuxième sexe de Simone de Beauvoir*. Presses Paris Sorbonne, 2004.

[7] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

[8] L. Sun, M. Wei, Y. Sun, Y. J. Suh, L. Shen, and S. Yang, “Smiling women pitching down: auditing representational and presentational gender biases in image-generative ai,” *Journal of Computer-Mediated Communication*, vol. 29, no. 1, Nov 2023.

[9] R. Kotczezki, D. Csikor, and B. E. Balassa, “The role of generative ai in improving the sustainability and efficiency of hr recruitment process,” *Discover Sustainability*, 2025.

[10] P. Budhwar, S. Chowdhury, G. Wood *et al.*, “Human resource management in the age of generative artificial intelligence: Perspectives and research directions on chatgpt,” *Human Resource Management*, 2023.

[11] S. Chowdhury, P. Budhwar *et al.*, “Generative artificial intelligence in business: Towards a strategic human resource management framework,” *British Journal of Management*, 2024.

[12] R. Van Solingen, V. Basili, G. Caldiera, and H. D. Rombach, “Goal question metric (gqm) approach,” *Encyclopedia of software engineering*, 2002.

[13] International Labour Office. (2012) International standard classification of occupations: Isco-08. volume i: Structure, group definitions and correspondence tables.

[14] S. H. Khandkar, “Open coding,” *University of Calgary*, vol. 23, no. 2009, p. 2009, 2009.

[15] R. Feldt and A. Magazinius, “Validity threats in empirical software engineering research-an initial survey.” in *Seke*, 2010, pp. 374–379.

[16] C. O’Neil, *Weapons of Math Destruction*. Crown Publishing Group (NY), 2016.