



The EU Mutual Learning Programme in Gender Equality

Artificial Intelligence and Gender Biases in Recruitment and Selection Processes

12-13 November 2020

Comments paper - Spain



The information and views set out in this paper are those of the author(s) and do not necessarily reflect the official opinion of the Commission. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.



This publication is supported by the European Union Rights, Equality and Citizenship Programme (2014-2020).

This programme is implemented by the European Commission and shall contribute to the further development of an area where equality and the rights of persons, as enshrined in the Treaty, the Charter and international human rights conventions, are promoted and protected.

For more information see: http://ec.europa.eu/justice/grants1/programmes-2014-2020/rec/index_en.htm

Gender Bias in Human Resources from a Spanish Perspective

Carmen Fernández-Martínez

CETINIA-Universidad Rey Juan Carlos

1. Introduction

Countries are implementing national strategies for Artificial Intelligence (AI) and introducing AI concepts progressively in institutions; however, doubts arise when it comes to tackling ethical issues related to bias. The aim of the paper is assessing Spanish trends concerning AI in Recruiting and offering an insight into algorithmic bias and more specifically, AI gender bias. Understanding why bias is produced, both technically and sociologically, is a big step for mitigating it and offering solutions.

Lately, we have witnessed a tendency towards Artificial Intelligence (AI) in Human Resources (HR) from CV screening, video-interviewing technology and turnover prediction. The choices to introduce AI in Spain in Recruiting and Human Resources are for the time being slightly behind Anglo-Saxon countries. At European level, we are witnessing further uses of AI in corporate environments, e.g. to predict extraversion and employee assistance through AI-powered conversational agents [1] [2]. Video interview software has increased considerably in Spain due to the coronavirus pandemic. Alexandre Coffin, Director of Walters People notes 62% increase in demand in recent months (Comillas, blogs). More and more enterprises are embracing the trend. Companies like Telepizza, Porcelanosa, Five Guys and Securitas Direct, Econocon or Sprinter are active users of technology in Spain (El País, Tecnología). Other multinationals have been using AI for years globally, IBM, Procter&Gamble, Goldman Sachs, Unilever, Vodafone for instance. They are companies that already receive hundreds or thousands of applications, so the inclusion of AI improves their HR production processes and saves recruiting time.

Bias problems are caused, in the first instance by the very nature of the technology and incomplete datasets. Some are inherent to human factors or other particularities, e.g. the fact that the Spanish language has gender. For this reason, there is a gender bias inherent in our language. Some translators still maintain bias translating from English, for example, to Spanish. Another reason for bias is gender disparity at all levels of society. Women are less traditionally related to technology positions and we even have less presence in higher ranks in corporate environments. That is why we must encourage girls to be active users and designers of technology and promote the presence of women in technology and at all levels of business, from technicians to managers. The promotion of STEM and technological vocations and computer literacy is the first step.

This is our breeding ground to begin formally assessing gender bias in AI from a Spanish perspective. Concerning education, Madrid was one of the first regions in Europe to introduce a compulsory course on Technology, Programming and Robotics in middle school¹. The topics are not obligatory in the national curriculum, though, but other regions like Catalonia implement similar measures. Gender bias is a topic very frequently discussed in Spain. For instance, *Elcano Talks* or the ongoing initiatives presented by the Women's Institute (*Instituto de la Mujer*). Researchers from *Pompeu Fabra* have researched the question 'What role do general gender stereotypes play when employers recruit new staff?' addressing the term 'bias' recruitment processes [3]. With relation to corporate initiatives to track bias specifically in AI, it stands out BBVA Next Technologies shared initiative, called *Feminist Bot Design Canvas*, to evaluate bias among design teams. *BBVA and the University of Barcelona* have achieved a model to classify feminine and masculine superheroes without any track of information about race or gender.

The rest of the paper is organised as follows. Section 2 explains the sources and causes of algorithmic bias from a gender perspective and deals with the legal consequences of bias. Section 3 and section 4 describe the Spanish legal framework and initiatives at corporate and public level. The final section presents a discussion and a set of recommendations or better practices to eliminate or mitigate gender bias at a national and European level.

2. Algorithmic bias in Business. Legal aspects of bias in Recruiting software

Algorithmic bias and more specifically, gender bias can manifest in several ways. Bias is created by reproducing past prejudices, by using inappropriate variables for training data and imperfect, limited and flawed datasets. Traditionally, AI played no more than an assistant role in HR, e.g. resume and CV scanning. But lately, apps and systems like HireVue², Montage, SparkHire and WePow and emotion-recognition software Affectiva [4] have been changing how recruitment is carried out internationally. These companies do not have a presence in Spain so far, but a growing number of companies are using HireVue, for example. First, when we speak about video-interviewing systems and its inherent bias, we do mean all those involving facial recognition technology. Models do not work well for unrepresented groups in training data. Technologists are a majority of men. For women, there is an under-representation in the training data. Learning algorithms have been traditionally developed and tested by men. Thus, gender bias is reinforced.

We have seen the issue of gender bias in text recognition and CV screening (e.g. Amazon bias against word women's). As an illustration of the advances of sexual

¹ Spanish schools play catch-up on high-tech technology (2018). Financial Times.

² <https://www.hirevue.com>

orientation and gender recognition in other areas, like images and sound, it is important to mention some recent experiments carried out in Italy and the USA. Both studies needed ethical supervision. Kosinski and Wangs' work [5] shows theories based on a morphologic study of homosexuality. The study argues that there is a relationship between homosexuality and exposure to a particular concentration of hormones in the womb and that morphological features can determine the sexual orientation (for example, more masculine or feminine jawline and forehead). The datasets were controversial. The research was contrasted with what users disclosed in dating apps and Facebook groups. [6]. The *voice-based categorisation study* [7] reminds us, however, that every language has their particularities. A voice that sounds more feminine in a man or vice versa may be due to anatomical reasons or exposure to feminine agents during childhood. The sexual orientation guessing proved to be more explicit for listeners in Italian or Romanic languages; still, the categorisation works similarly both in German and Italian. In these cases, some gay speakers were consistently misidentified as heterosexual and vice versa [6].

Miriam Kullmann suggests some other potential risks of gender discrimination and especially, legal issues and has commented on the possibility of discrimination by proxy. Legal frameworks in the US protect basically against racial discrimination at the workplace, but it is more lenient for gender discrimination. Incorrect datasets led to direct or indirect discrimination by proxy. Employers can consciously and deliberately choose -or discriminate- according to protected attributes, "gender", "race", "childbearing age". Inaccurate AI fed by systemically biased datasets leads to indirect discrimination and the extension of it, the so-called side effect *Disparate impact*. This phenomenon comes up when biased data lead to chain and unfair decisions against minorities or protected groups at a societal level. Therefore, it is highly essential addressing the issue of intentional and unintentional proxy discrimination related to the effects of AI and Big Data analysis [8, 9]. Although law scholars have used the term discrimination by proxy extensively for years, we see its effects recently in all types of sectors, from banking to employment. What does Proxy discrimination mean? It leads to discriminate in favour or against a protected class, or a protected group (e.g. women, race, people with higher incomes and so on) but based first on legitimate or innocuous grounds (e.g. zip code). These reasons for discrimination are highly correlated with belonging to a protected class (e.g. racial minorities living in the same areas or ghettos). So, this segregation ends up affecting the protected groups and being the reason for more discrimination or unfair business decisions. When there is a human actor behind this favourable treatment or less favourable, we speak of "intentional" discrimination by proxy [8]. However, the nature and mix of the datasets used for testing and training could also lead to "unintentional" discrimination by proxy. Relying on external sources often causes more algorithm bias. It is often the case, that the company holds that they can justify this discrimination as a business necessity if it is a predictor of employee's future performance. Therefore, they are not infringing Equal Employment legislation even though the practices resemble historical patterns of discrimination or use limited datasets [10].

3. AI in Recruiting: Spanish legal framework and latest trends

There is no specific legislation addressing gender bias in the recruitment process in Spain, although several working groups are working in standards for it. The most critical piece of legislation regarding gender is the Law of Equality Act or 'Gender Equality Act' (Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres), passed on March 24th, 2007 by the Spanish Parliament. The latest decisions at the national level regarding gender inequality in the labour market are limited to the wage gap between men and women, a visible problem. The Spanish Data Protection Agency (Agencia Española de Protección de Datos) has prepared several documents where it talks about the challenges of the result of automated decisions by AI that affect people (Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción). At European level, Article 22° of the GDPR states that European citizens have the right not to be subject to a decision based solely on automated means, including profiling, whether the decision produces legal effects, which affect them significantly in a similar manner.

What can we do to protect against discrimination resulting from bias? What are the latest trends? It is noted that algorithms can detect, and sometimes protect against discrimination Pedreschi et al. [11] describe well in their work the possibility of finding evidence of discrimination in Automatic Decision Support Systems (DDS), used for screening purposes insensitive social tasks, including access to credit or mortgage, tracking if the input could lead to a wrong yes/no decision. Disparate impact liability is difficult to prove too. Many software companies and academics like Haijan and Domingo-Ferrer [12] propose techniques and software, which are effective at removing direct and/or indirect discrimination biases. So, they reinforce Pedreschi's proposal in their paper, of inducting patterns that avoid discriminatory decisions despite drawing from training biased datasets [6]. Other methods to prevent the discrimination of algorithm in Recruiting from a technical point of view, are comparative studies by a software tool. Even though reports have shown that the results of HireVue Systems are highly diverse and show even more presence of non-white candidates and gender equality, there is disinformation about what precisely this software targets for every position.

4. Recommendations

- **Fostering model auditing and compliance.** A good idea would be a fostering model auditing by third-party non-profit entities. Some cities like Toronto already audit and disclose smart-cities data relevant data to third-party independent auditing bodies. Proper auditing in domain-specific AI could avoid manipulative uses. Fostering blind auditions that have proved to be useful for gender bias.
- **Awareness about gender equality.** The way to integrate the entire gender perspective into the system, increase transparency and reduce bias is to work to

promote STEM vocations and fight to reduce the gender gap, which is also an economic gap as a consequence.

- Educational Challenges STEM
- Awareness of proportion gender/men in corporate environments
- **Allowing companies to be active stakeholders in mitigating bias.** As a decisive initiative in the business world, Accenture Spain research team has launched a tool called AI Fairness in collaboration with the Alan Turing Institute. It is a novel tool that allows companies and public entities to verify that their AI-powered systems do not incorporate biases inherited from past decisions. The tool makes it possible to measure the impartiality of an algorithm quantitatively.
- **Finding a balance between privacy and corporate intellectual property and model explainability.** As stated in Miriam Kullmann's paper 'Gender Biases in Selection and Recruitment Software', explainable models are important for critical sectors, like health care. We should avoid using "black box" AI software without compromising trade secrets.
- **Improvement of neutral audit tools and datasets for test and training. Review model patterns to avoid bias inherited from older data and strange variables. Considering different outcomes for model results. [13]**

5. Conclusion

The problem of bias is an urgent problem that concerns us all, companies, users, regulators and researchers. AI cannot afford to harbour bias errors. Only the right AI will survive, and this is extremely important for business competitiveness. The understanding of the issue of bias seems to be in correlation with the time spent on projects with AI technology. Recent attempts carried out in Spanish middle Schools to introduce Artificial Intelligence in Spanish middle schools show no remarkable gender differences in results. Gender equality starts with education. To date, the low representation of women in the world of technology is an endemic problem. World Economic forum presented a report recently. In sectors such as Cloud Computing, Engineering, there is around 15% female presence worldwide, whereas in Big Data and Artificial Intelligence, this representation increases, 26%. Equality and awareness could be useful to mitigate bias. Most companies already have a set of best practices in the same way that there are public policy recommendations for ethical AI. Future works and initiatives involve AI ethic as a priority in AI national strategy. Coincidentally AI ethics is last in the list, number six.

References

- [1] Chandar, P., Khazaeni, Y., Davis, M., Muller, M., Crasso, M., Liao, Q. V., & Geyer, W. (2017, September). Leveraging conversational systems to assist new hires during onboarding. In IFIP Conference on Human-Computer Interaction, 381-391, Springer, Cham.
- [2] Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., ... & Silva, E. (2020). Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review*, 36, 100239.
- [3] González, M.J., Cortina, C., and Rodríguez, J. (2019). The Role of Gender Stereotypes in Hiring: A Field Experiment. *European Sociological Review*
- [4] Fernández, C., & Fernández, A. (2019). Ethical and legal implications of ai recruiting software. *Ercim News*, 116, 22-23.
- [5] Wang Y. and M. Kosinski, M. "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images," *Journal of personality and social psychology*, vol. 114, no. 2, pp. 246-257, 2018
- [6] Fernández-Martínez, C., & Fernández, A. (2020). AI and recruiting software: Ethical and legal implications, *Paladyn, Journal of Behavioral Robotics*, 11(1), 199-216. doi: <https://doi.org/10.1515/pjbr-2020-0030>
- [7] Sulpizio, S., Fasoli, F., Maass, A., Paladino, M.P., Vespignani F., Eyssel, F. and Bentler, D. (2015), "The Sound of Voice: VoiceBased Categorization of Speakers' Sexual Orientation within and across Languages," *PLOS ONE*, vol. 10, no. 7, 1-38
- [8] Prince A. and Schwarcz, D. "Proxy Discrimination In The Age Of Artificial Intelligence And Big Data," *Iowa Law Review*, Forthcoming, 2020.
- [9] Datta, A., Fredrikson, M., Ko, G. Mardziel, P. and Sen, S. "Proxy discrimination in data-driven systems," *arXiv preprint arXiv:1707.08120*, 2017.
- [10] Barocas S. and Selbst, A. "Big Data's Disparate Impact," *California Law Review*, vol. 104, no. 3, pp. 671-732. 2016. [43] D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating induction and deduction for finding evidence of discrimination," *Proceedings of the 12th International Conference on Artificial Intelligence and Law (ICAIL '09)*, 2009.
- [11] Pedreschi, D., Ruggieri, S. and Turini, F. "Integrating induction and deduction for finding evidence of discrimination," *Proceedings of the 12th International Conference on Artificial Intelligence and Law (ICAIL '09)*, 2009
- [12] S. Hajian S. and Domingo-Ferrer, J. "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1445-1459, 2013
- [13] Davenport, T. (ed.) (2018), *The AI Advantage. How to put the AI Revolution to Work*, MA, USA: MIT