



The EU Mutual Learning Programme in Gender Equality


Artificial Intelligence and Gender Biases in Recruitment and Selection Processes

12-13 November 2020

Comments paper - Malta



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

Artificial Intelligence and Gender Biases in Recruitment and Selection Processes

Claudia Borg

University of Malta

1. AI and Gender within the Maltese context

Malta is the smallest EU Country and currently ranks 5th in the Digital Economy and Society Index in 2020¹. This was an increase from the previous year in 2019 when Malta ranked 10th place². However, the analysis also places Malta at the 17th rank when analysing the Women in Digital Scoreboard for 2019³. The latter report analyses the use of internet, users' digital skills and specialist skills and employment. This last category is where the difference stems when it comes to Malta's ranking. Fewer STEM female graduates, fewer ICT specialist and unadjusted gender pay gap.

There are very clear indications that Malta is investing heavily in technology and AI. Witness to this is the setting up of the Malta Digital and Innovation Authority⁴ which is tasked with seeing that Malta becomes a centre for excellence in technological innovation, whilst simultaneously enforcing standards and compliance obligations. To date, the primary regulations are aimed at System Auditors and Innovative Technology Arrangements (which encompass distributed ledgers and smart contracts). However, to date, standards and compliance obligations aimed specifically at AI systems have not yet been proposed.

The Maltese government also published a national AI Strategy⁵ in 2019 with the objective to gain a strategic competitive advantage in the global economy in the field of AI. The strategy sets out a number of case studies that are AI-centric which can be developed and deployed in the public or private sector. The long-term vision is for Malta to become the "Ultimate AI Launchpad" by 2030, setting short-term projects to be deployed by 2022.

The AI Strategy also set the tone for the development of a legal and ethical framework, publishing a Strategy towards trustworthy AI⁶. The strategy uses as its basis the Ethics Guidelines for Trustworthy AI (2019) published by the High-level Expert Group on AI set up by the EC, and the recommendations of the Council on AI adopted in 2019 by

¹ Digital Economy and Society Index (DESI) 2020 Malta - Report

² Digital Economy and Society Index (DESI) 2019 Malta - Report

³ Women in Digital Scoreboard 2019 Malta Report

⁴ <https://mdia.gov.mt/about/>

⁵ https://malta.ai/wp-content/uploads/2019/11/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf

⁶ https://malta.ai/wp-content/uploads/2019/10/Malta_Towards_Ethical_and_Trustworthy_AI_vFINAL.pdf

OECD countries. The Strategy towards trustworthy AI does provide direction in terms of possible future regulation to deal with bias, with one of the main conditions outlined being: “AI should be designed and deployed in a manner that is equitable and mitigates bias the greatest extent possible.”

1.1 Gender Bias in Malta

There are several indicators that demonstrate the lack of equality in Malta, even if this might not always be the result of a direct bias. Malta currently ranks in the 15th place in the Gender Equality Index for 2020⁷. The gender gap in employment remains the largest in the EU, less women choose STEM subjects and the pay gap persists, with women earning 15% less than men in Malta. When it comes to specific ICT specialist jobs, 89% of the positions are held by men, with only 11% represented by women.

The data shows that there is a clear underrepresentation of women in several sectors but with a special focus on the ICT sector. It is difficult to attribute this underrepresentation directly to inherent bias. However, one encounters several anecdotes where women are asked about their family plans during job interviews or promotion interviews⁸. Women representation in Boardrooms continues to be low⁹, leading to a lesser representation in decision making scenarios. The situation is further challenged by the fact that such types of appointments are generally allocated not through open advertising and recruitment through merit, but rather personal networks, leaving out the talent pool of women¹⁰.

Although the above sets the scenario in particular for Malta, there are similar situations in other European countries. Therefore, it makes sense to seek policies at European level, through which EU countries can share their successes in mitigating gender bias in general.

2. Policy Debate

The starting point in terms of policy direction needs to be based upon two factors; (i) there is a disparity in gender representation, especially in the field of ICT and in decision-making situations, and (ii) although there are proposals in terms of how AI can be regulated, there are concrete difficulties due to the potential stifling of research progression and the actual difficulty of implementing and enforcing such regulations.

2.1 Understanding bias in recruitment

The key understanding of bias in AI is that it is mainly derived from the data that is fed into machine learning algorithms. The data is generally historic, and therefore contains any inherent human bias. It is important to understand why this issue is

⁷ <https://eige.europa.eu/publications/gender-equality-index-2020-malta>

⁸ <https://timesofmalta.com/articles/view/Two-complaints-about-job-interview-bias.532473>

⁹ <https://timesofmalta.com/articles/view/Gender-equality-in-the-boardroom.608387>

¹⁰ <http://womendirectors.org.mt/article/gender-equality-boardroom/>

becoming prevalent now, when machine learning has been used for a number of decades. Machine learning algorithms were initially applied on smaller datasets and the type of algorithms used required the developer to specify the features that the algorithm would extrapolate information from. Neural networks, a subclass of machine learning algorithms, do not require such specification. They simply need to be supplied with the data and the goal, and the algorithm will decide what features should be given more importance to arrive to the desired goal. As our computational resources and datasets became larger, this meant that these algorithms could be applied on more sophisticated and complex problems.

Focusing on the use of AI in recruitment, and in particular, the selection process, this is one such complex problem. The way human recruiters interview potential candidates can be unstructured, thus giving rise to the possibility of bias to occur. The selection criteria might also contain inherent biases that recruiters might not be fully aware of. Looking at only these two small aspects, we can already conclude that avoiding bias in this complex task would be challenging for an AI system, no matter how well it is designed and how well the dataset might be balanced. The goals that are set out for the AI system might contain that inherent bias to begin with.

A study conducted by Totaljobs¹¹ finds that a majority of job adverts contain gender biased words, with male-coded words such as lead, competitive, chief, confident; whilst female-coded words such as responsible, support, dependable, understanding and committed. If the starting point, the specification of the job requirements, already contains bias, and the resulting selection further strengthens that bias, an AI system will simply replicate the path from beginning to end.

It would be very helpful to understand and document best practices carried out by human recruiters in the different processes. If these are shared with software developers, they might be in a better position to recognise possible bias in AI systems and try to make the necessary adjustments to mitigate the bias.

2.2 Possible legal implications

An important aspect in terms of policy making when it comes to AI is to recognise that it is simply a tool, albeit a powerful one. It is important to define who is responsible for the making of this tool since it contains several components, and to which extent each stakeholder is liable for any bias that ends up as part of the final system. It is equally important to understand that certain systems do not have the mechanisms to explain the decision-making process easily.

When it comes to raising awareness about these types of systems, it is important to have lawyers who are specifically trained and exposed to knowledge with respect to all these different components. There must be a clear understanding that software developers make use of (i) external libraries, (ii) code that has been built by other people/companies, and (iii) datasets that are curated by other people. A software

¹¹ <https://www.totaljobs.com/insidejob/gender-bias-decoder/insights/>

developer can take certain measure to ensure that datasets used are balanced and have a better representation of the diverse data points. However, certain choices made in terms of selection of which library packages to use are made on the basis of the type of documentation available, ease of use, and other such variables. From a legal perspective, we must also consider the difference in treatment between software developed in the EU and outside of the EU, and ensure that there isn't a positive discrimination towards any category.

An important discussion from a legal perspective is to conduct a thorough analysis of existing laws and deciding specifically which cases are not catered for by existing laws. The analysis should be an honest and in-depth exercise that identifies whether existing gaps and loopholes have been negatively used already in situations where humans implemented certain biases. In this case, it is important that the laws are modified to ensure that both human and machine cannot carry out such biases.

An aspect to keep in mind is that AI can be used to highlight our inherent biases. The developer of the AI should not be made to hide those biases, but rather encouraged to produce a much-needed evaluation to support changes in policy and mentality by human recruiters.

2.3 Future risks and future opportunities

In determining policies with respect to bias that emanates from AI, we must be conscious of the fact that the development of AI is soaring at an incredible speed. Therefore, we must not limit to current situations and knowledge, but rather explore further possibilities, risks and opportunities.

Although a mechanism for classifying risks of potential applications and software can be useful, we must also envisage how AI will develop through the information that is currently available. It is often the case that different fields in technology borrow ideas from each other and apply them to new areas. This results in innovative exploration and sometimes improvement of results in different sectors. One important aspect to keep in mind is that there are areas where AI has actually surprised humans by the choices that it made – trying out completely new paths that were not at all present in the observed data provided to the machine. An example of this is in the scenario of Game AI, with AlphaGo being at the forefront of this type of innovation.¹² Go is a strategic board game popular throughout the world and very predominant in Asia. As a game, Go is considered to be a profoundly complex game that requires multiple layers of strategic thinking. Deepmind employed deep learning techniques to build a software, AlphaGo, that would teach itself how to play Go. Eventually AlphaGo beat the human Go world champion. The moves that AlphaGo learnt by playing against itself surprised researchers and Go players alike, developing moves that had never been seen in a game of Go before. This was something that no researcher could have thought would be possible until it actually happened. Till then, machine learning

¹² <https://deepmind.com/blog/article/innovations-alphago>

always provided solutions based on the data that it observed. However, with deep learning and allowing the algorithm to create its own learning scenario, we see that the algorithm learns to optimise solutions to reach the desired goals and that these solutions were never thought of by humans before.

It is very much a possible scenario that an AI system (or a number of individual, but connected, AI components) learns how to measure meritocracy in the workplace and thus learns to define the requirements of a new position and selection in a much better (unbiased) way than a human would. Of course, we cannot wait until such systems become a reality and therefore it is important to implement the necessary precautions and guidelines from now to obtain, protect and preserve equality.

2.4 Sharing data and decision outcomes

In moving towards more transparency in decision-making and sharing of outcomes and justifications, we must realise the need for protecting individuals and their privacy. Existing research looks at the automatic anonymisation of names and other personal details so that data can be processed without such information. This can also be a way of minimising bias that might occur in the training, but also facilitate the sharing of information when it comes to auditing and controlling what the outcomes are.

3. Recommendations

It is clear that irrespective of how AI continues to develop, there must be a clear focus on eliminating the source of the problem – human bias. Encouraging diversity, ensuring equal pay, eliminating reasons why people are discriminated against, and ensuring better representation at all levels of the workforce are essential key components so that the situation can begin to change. It is also essential to document best practices used by human recruiters to eliminate any inherent bias that might be present.

An extensive analysis of existing legal frameworks should be carried out with the aim to identify any existing gaps and loopholes that allow for human and/or AI bias in recruitment to go unchecked. It is also important for lawyers to have a clear idea of current IT systems, how they are built and how they work so that the legal implications can be properly catered for.

Society in general needs to understand why and when AI makes incorrect decisions. Investing research funds in explainable AI will encourage and facilitate the development of this area, whereby the focus is to provide understanding for AI's decision-making processes. In the case of recruitment, these explanations should be transparently available and mimic best practices used by human recruiters.

AI can be considered to be in its infancy. It has now reached a stage of maturity whereby AI applications are found everywhere. Laws must consider the potential advances in AI and must ensure not to stifle research and advancements, but to place the necessary safeguards towards society.

We need to ask the hard questions about AI. We must not see this as a means of instilling fear of AI and creating a doomsday scenario, but rather to act responsibly and precisely define the type of AI systems that we want. Simply making broad statements like '*AI should serve humans*' is not sufficient since serving humankind can take many different forms when seen from different angles.