European Commission

# The EU Mutual Learning Programme in Gender Equality

# Artificial Intelligence and Gender Biases in Recruitment and Selection Processes

**Online seminar, 12-13 November 2020**

## Summary Report

ösb Consulting
Ein Unternehmen der ÖSB Gruppe.



Justice

*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*

# Introduction

The European Commission Mutual Learning Programme in Gender Equality convened an on-line seminar on 'Artificial Intelligence and Gender Biases in Recruitment and Selection Processes' on 12-13 November 2020. The meeting was hosted by the Ministry of Education, Culture and Science, the Netherlands and was originally intended to be held in The Hague, to share a Dutch innovative practice example of a hackathon. However, both the hackathon and the seminar were transferred to online events because of COVID-19 health measures. Despite these challenges, the meeting was attended with participants from 21 Member States, including Austria, Czech Republic, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Croatia, Ireland, Italy, Lithuania, Latvia, Malta, the Netherlands, Poland, Portugal, Romania, Sweden and Slovenia.[1] Participants were from government agencies, university legal, equality and computer engineering departments, research institutes and the private sector. There were also representatives from the European Commission (EC), the European Institute for Gender Equality and a civil society organisation 'Women at the Table'.

The seminar was an excellent opportunity to discuss the challenges and opportunities of using Artificial Intelligence in recruitment and selection processes. The seminar focused on two main aspects: (i) the potential risk of (gender) discrimination of algorithms in recruitment and (ii) awareness raising of the issue of gender bias in algorithms. The EC had sought to develop a coordinated approach and had published a white paper in early 2020[2], which includes a proposal for AI regulation to ensure that AI is safe, lawful and in line with EU fundamental rights. Algorithms and machine learning, if not transparent and robust enough, risk repeating and amplifying gender and other biases. In March 2020, the EC's Advisory Committee on Equal Opportunities for Women and Men issued an opinion on AI on opportunities and challenges for gender equality[3] and the 2020-2025 gender equality strategy[4], also published in March, recognises the potential of AI as well as its challenges to gender equality and human rights. Hence, the seminar follows on from EC actions and policy objectives in this area.

---

1 For more information on individual countries participating in the seminar, please see the comments papers prepared in the context of this seminar. The papers are available at the programmes website: https://ec.europa.eu/info/publications/artificial-intelligence-and-gender-biases-recruitment-and-selection-processes-online-seminar-12-13-november-2020_en

2 https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

3 https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf

4 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A152%3AFIN

# 1. Setting the scene

Miriam Kullmann, the thematic expert from the Vienna University of Economics and Business, provided an overview of the main types of AI used in labour market recruitment processes and outlined its benefits from a business perspective, particularly when employers experience a high recruitment volume. However, there are many potential sources of gender bias, or other forms of discrimination, and it is therefore crucial to develop tools to address them. An algorithm can be defined as 'a formally specified sequence of logical operations that provides step-by-step instructions for computers to act on data and thus automate decisions'. Amazon is an example of a company that has used AI to select the best candidates for interviews. However, it found that the machine-learning algorithm favoured male candidates because it had been trained on CVs from previous applicants over the last 10 years who were predominantly male. This example clearly illustrates the potential for bias. If the historical data integrated into an algorithm process lacks population diversity, in terms of gender, but also on other grounds such as race and age, there is the danger that such stereotypes will be turned into prescriptions – or as a telling phrase illustrates - if you put garbage in you get garbage out.

The thematic expert provided some pathways to address the potential risks of gender bias starting with the important task of awareness raising at government level, but also among lawyers and information technology (IT) specialists. Creative legal solutions were also needed, in particular to consider how to use EU non-discrimination laws and the General Data Protection Regulation (GDPR). For example, there is a need to envisage how legal remedies can be used when an individual may have been impacted by AI bias or implicit discrimination; and the extent to which a company can be liable for the use of software with embedded biases. Other measures that can be considered include risk assessments, auditing or ethical guidelines to assess AI outcomes.

A thought-provoking discussion followed. Participants recognised the importance of bridge-building initiatives designed to enhance collaboration between the worlds of information, communication and technology (ICT) and law and non-discrimination; and between IT experts and the general public. They noted the persistent under-representation of women in ICT, and in AI in particular, and the importance of introducing issues of gender equality from early years education onwards so as to address gender segregation in the labour market. It was also noted that students and other job-seekers need support to prepare and present their candidacies.

The discussion focussed on how to improve the quality of data sets when they themselves reflect social inequalities. It was noted that in seeking to diversify data sets, there is a danger that marginalised groups could become more vulnerable, for example when race is an important potential area of discrimination. It also raises

privacy and surveillance issues. Participants explored the issue of how to diversify data while at the same time taking into account the inherent risks to vulnerable groups.

Going upstream to question the assumptions behind the model design that uses the data was another potential avenue. In this context, the 2019 Canadian directive to make public the custom sourced code owned by the government was mentioned as a useful example that suggests it might be more strategic to focus on the public rather than the private sector.

Another issue examined was whether there was a need to modify existing legal concepts of discrimination as a result of the emergence of algorithms, possibly with a greater focus on the preventive or restorative functions of non-discrimination law. However, it was noted that most legal frameworks that address the right to remedy are not suitable for use in cases of evolving technologies. The legal concept of non-discrimination is largely tailored to individual cases whereas AI may lead to inter-sectional and collective situations of discrimination. While indirect discrimination can refer to the collectivity, it was noted that it follows rules that are also ill-adapted to AI contexts, such as the burden of proof. In this context, participants discussed whether there is a need to shift the 'risk' of explaining the outcomes of automated decision-making to the business itself and what specific strategies could be employed to do so. It was suggested that it might be necessary either to develop a separate set of non-discrimination rules to address specific AI challenges, or else to amend the existing non-discrimination directive or add specific AI rules.

# 2. The Dutch practice – The hackathon on AI and gender biases in recruitment

## 2.1 Aim and process of the hackathon

The Ministry of Education, Culture and Science of the Netherlands and its Directorate for Gender Equality and LGBTI Equality have set strategic and operational goals for the position of women in the labour market by 2025. The target is to ensure women's financial independence, defined as a woman earning at least 100% of the minimum wage, because the Netherlands has a persistently high rate of women's part-time work relative to other EU Member States. Another goal is to achieve greater diversity on supervisory boards.

The Dutch government is concerned at the potential of new technologies to perpetuate gender inequalities in the labour market. It is now in the process of drafting legislation to ensure that employers using new technologies must ascertain that their outcome does not discriminate. The Ministry has commissioned research on AI and gender in the labour market, due to be completed in 2021, and is providing input into various policy documents, ethical frameworks and design principles. It has a coordinating function and maintains an ongoing dialogue with other government departments and with the private sector on the issue.

The Netherlands Institute for Human Rights issued a seminal report in October 2020 (in Dutch) on AI and discrimination in selection and recruitment processes, highlighting the potentially discriminatory nature of the initial design of algorithms. The report was influential in convincing the Ministry to continue with the plans for the hackathon despite the health emergency.

Designed as a bridge building and knowledge sharing exercise, the hackathon[5] took place as an online event, on October 30 – November 1. It was based around a challenge to develop a solution to prevent biases in the collection and labelling of automated system processes for the purpose of assessing and selecting candidates. The challenge identified five risk factors that can contribute to the creation of bias: a) the definition of 'good' performance or suitability; b) the collection and processing of training data; c) the choice of variables; d) proxies; and e) masking. There were 50 participants divided into ten inter-disciplinary teams, including data scientists, recruiters, business analysts and entrepreneurs; lawyers, social scientists and diversity and inclusion officers. At the end of the hackathon, the teams were invited to present their proposals and a jury selected the winning entries.

### 2.1.1 Discussion

The issue of how to detect discrimination in an individual case and yet respect confidentiality was explored. Because algorithms are so complex, function as a 'black box', and in addition, are constantly changing, it might be more feasible to assess bias through an analysis of the outputs and outcomes. It was also noted that the 'black box' syndrome exists in the real world and that it is a struggle to identify gender discrimination even when algorithms are not used.

The discussion also returned to the importance of the legal framing of discrimination. Protection and remedies against discrimination are reliant on access to information about the selection process. If 'black box' systems lacking transparency are used, victims of discrimination have no effective legal recourse, which is contrary to the basic principles of EU and national law. It was noted that in Finland the equality and data protection monitoring bodies do not coordinate their work and have yet to become involved in the issue although they could potentially carry out impact assessments. In the Netherlands, the Inspectorate of Social Affairs is considering what monitoring role could be envisaged within the framework of the proposed new legislation requiring companies to ensure non-discrimination in the use of AI.

It was noted that detecting gender discrimination in an exhaustive manner is very complicated and that indeed women may themselves self-censure. It was suggested that discrimination in the use of AI generally arises from claims of disparate impact (a US concept) or indirect discrimination. In this regard, the UN Guiding Principles on

[5] https://www.youtube.com/watch?v=3D1oH5gV-80&feature=youtu.be

Business and Human Rights might offer a useful precedent as they shift the perspective of risk so it becomes 'risk to people' rather than 'risk to business'. It was further explained that disparate impact occurs when an apparently neutral policy or practice unduly disadvantages individuals based on protected characteristics. Hence disparate impact assessment could be incorporated into the law or at least into AI impact assessments. One possibility could be that the law require these assessments to be provided by the appropriate regulatory body charged with enforcing anti-discrimination laws.

Participants reflected on the kinds of policies or instruments needed so that companies become aware of the risks or are obliged to take action. The New Zealand government algorithm charter was cited as a useful public policy example.

Auditing mechanisms could also prove a useful avenue so that the type, quantity and quality of data and its training process are made available, which could also benefit companies wishing to prove that their systems are not biased. In Sweden, the Discrimination Ombudsperson has reviewed one case and determined that the company is responsible if the AI systems in use are found to be discriminatory, so some companies are aware of the risks involved and are taking action.

Finally, it was noted that there are a few examples of companies that are carrying out innovative pilot projects. In Austria, one company excluded gender as a criterion in its algorithm and found that the selection process was more equal. The Swedish Vinnova innovation agency is seeking to redesign AI to contribute actively to the promotion of gender equality and the Dutch Institute of Applied Technology is carrying out a pilot using fictive biased and non-biased data to compare outcomes on recruitment and selection processes.

## 2.2 The winning solutions of the hackathon

The programme of the hackathon was based on design thinking and human-centred innovation methodology. The programme included mentor sessions and roundtables with a diverse range of experts. Each team prepared a proposal and the jury awarded the prizes based on four criteria for success: a) added value and impact; b) creativity and innovation; c) feasibility and d) presentation.

The third prize (audience favourite) was awarded to 'Simplify 360', which was designed to assist managers of small and medium enterprises (SMEs) build inclusive and diverse teams based on a personality assessment, with a metrics dashboard for inclusive matching. The second prize was awarded to 'Balancing the Scale', which aimed to ensure a diverse pool of candidates by ensuring that only actionable and accountable data is used during the screening process, based on experience and the measure of knowledge rather than personal criteria such as marital status, citizenship and age. The first prize was awarded to 'FAIr', which proposed developing a transparent modular assessment tool whereby recruiters could choose five skills to assess the job and candidates use a self-assessment against these skills. AI selects

the 20 best candidates, with a gender diverse guarantee, and the final group for interview are selected by human resource managers.

The two winning entries received prize money as well as the opportunity to develop their ideas within an incubation programme sponsored by the Dutch government. Among the lessons from the hackathon it was considered necessary to provide sufficient background information to participants to ensure a common understanding of the complexity of the challenge; and that in order to capitalise on the outcomes, it is important to provide further opportunities for the teams to develop the proposals.

### 2.2.1 Discussion

Participants expressed considerable interest in the hackathon experiment and discussed their strengths and weaknesses. The use of self-assessments or other performance assessments and psychological tests were questioned by some participants as they are not gender neutral but tend to have a bias towards male personalities and masculinities. Another area of concern referred to the proposal to eliminate all personal characteristics as it could be regarded as 'too objective'. It was also noted that some of the proposals left room for subjectivity and could lead to random results and that a focus on SMEs could risk that the pool of candidates was too small. It was pointed out that the criteria of 'added value' used by the jury could be open to interpretation from a 'business and market' perspective or a 'justice' perspective. In the latter case, the goal of AI would be subordinate to principles of equality requiring business interests to respect equality over and above efficiency.

Participants discussed the treatment of maternity leave and how a standard experience algorithmic value could be envisaged in order to balance out the occupational experiences of candidates. They were interested to learn that one of the participating teams had sought to address this issue.

The issue of company culture was discussed, which managers tend to reproduce although it may contain discriminatory practices. In all cases, humans intervene in the recruitment process, which can conflict with the presumed objectivity of the process. On the other hand, it was considered that humans should always take part in the recruitment process, if for ethical reasons only. Participants reflected on whether algorithms could be used to neutralise this company culture or whether it was more feasible to focus on long-term cultural change and enhancing awareness about social responsibility. It was noted that in the context of the Armed Forces the challenge is not only to recruit a more diverse personnel but to also to retain them because of the implicit or explicit culture.

Another area of debate focused on what is meant by a 'fair' algorithm and how to translate ethical discussions into algorithms to reinforce neutral criteria. The complexity and importance of 'de-biasing' algorithms was raised, noting that models are different and there are many mathematical definitions of fairness. Because the concept of 'fairness' differs and is subjective, participants considered whether it is possible to find a common ground from an EU perspective or whether it is better to

use human rights law as a point of departure. In this regard, it was also noted that algorithms written in languages with grammatical genders may have a higher risk of bias. In order to increase transparency and accountability, it was suggested that data sets could be accompanied by an explanatory datasheet, setting out their purpose, composition, collection process and recommended uses, much as an electronic item is always accompanied by an explanatory datasheet.

More information about the use of implicit criteria and proxies was considered important as they can often lead to bias, for example in order to exclude candidates of child-producing age, or on grounds of sexual orientation. It was noted in this context that postal codes in large cities could be used as a proxy to exclude candidates from a migrant background.

Finally, certification systems were mentioned as a useful method to reward good practice. Many public and private companies are working to ensure that their selection processes are fair, reasonable and proportionate and while they understand that AI will increasingly be used, they view it with a high level of caution. A certification process would assist companies to come together and provide evidence that their selection systems are valid.

### 2.3 Outlook: The aftermath of the hackathon and steps for the future

The hackathon was not intended to develop solutions but was viewed as a stepping stone on a path to involve all stakeholders and build greater awareness. The report on the hackathon will be disseminated broadly and it is expected that the major companies will be interested in the results and feel compelled to become involved in the development of new prototypes.

There will be further discussions with the two winning teams on how to develop the incubator programmes with the support of the Ministry of Internal Affairs. The other participating teams will be encouraged to further develop their proposals and data monitor experts will be made available to provide advice.

The hackathon was part of a broader programme designed to generate debate about the role of AI in society and a number of information materials, such as fact sheets, are envisaged as part of this awareness-raising initiative.

## 3. Key issues discussed during the seminar

Many participants noted that the use of AI in recruitment and selection processes did not yet appear to be widespread while recognising that its use was likely to increase in the near future. It was agreed that the **potential risk of gender discrimination in AI recruitment** processes needed to be addressed at various levels and take into account national circumstances and differences in terms of discrimination. Participants recommended an inter-disciplinary approach, as practiced in the online

seminar and at the hackathon. It was also agreed that common definitions on issues such as bias and discrimination, informed by a human rights based perspective, would be useful.

There was a broad discussion about the uses of AI and its potential benefits as well as its impact on the future of work. It was considered important to find the right balance between regulation and auditing on the one hand and allowing for research and innovation on the other hand.

From a technical view, auditing was considered a viable option, although there were constraints to ensuring transparency and obtaining information about coding in part because of commercial confidentiality and preserving intellectual property. Audits could be carried out by external companies, or by existing State monitoring bodies. However, the challenge would be to identify the most hidden cases of discrimination. It was noted that auditors might need training as the issues were complex. Collaboration between State equality and data protection bodies could be encouraged. Participants also considered other approaches, such as a focus on measuring outcomes to ensure they were fair and the use of company certification labels.

There were discussions concerning the legal aspects of auditing and whether the necessary legal tools existed at national or EU level and if not, how they could be created. Some participants considered that the law currently does not address discrimination in the context of AI adequately. Other participants noted that there are draft proposals to develop laws to regulate the use of AI although there is considerable opposition to such proposals from the IT companies. The relative merits of voluntary company regulations and state policy interventions at national or company level were discussed.

Participants noted that trade unions have an important role related to access to information concerning employee assessments and recruitment and that it was important to train trade union representatives or works council members on the role of AI and its potential risks. Collective bargaining, possibly in specific key lead sectors, could prove very valuable.

**Awareness-raising on the issue of gender bias in algorithms** was considered critical as there is a general lack of information and many people consider AI and codes are ‘neutral’. Hence, there is a need for dissemination of information among the general public and all stakeholders. It was also important to ensure that existing company policies on diversity and inclusion are updated and disseminated. In this context, the importance of creating an inclusive corporate culture where disagreement or different voices could be heard and respected was emphasised.

Training in a variety of areas was considered essential, including human resource managers and ICT staff engaged in the recruitment process, as well as among trade unionists, lawyers and equality experts and the general public. The sharing of good practices and examples would be very helpful.

Finally, it was noted that discrimination in AI is an extension of the problems existing in the real world. Participants stressed the importance of strengthening work to promote gender equality throughout the education system, starting with early childhood education and focusing on encouraging more young girls to take up careers in STEM and ICT, as well as providing earmarked support for young women entering these fields.

# 4. Conclusions and recommendations

Given the global nature of the challenge, which involves different aspects, such as social sciences, legal issues and management issues, as a way forward it was recommended that a broad-based methodology, the equivalent of a business plan, be established.

Participants recommended the following set of actions within the framework of an inter-disciplinary and human rights based approach:

- Further research to map more clearly the extent of the use of AI in recruitment and selection processes;
- The dissemination of a checklist or guidelines to assist the public sector or private companies that are considering employing AI for recruitment to help managers understand the critical issues to take into consideration relating to discrimination on grounds of gender or other counts;
- The development of a databank or repository of good practices and pilot projects carried out by companies and other stakeholders;
- Policy development, focusing, among other aspects, on auditing and the need to translate legal frameworks to the new technological realities;
- Data protection and equality monitoring bodies at national and EU level to be encouraged to collaborate together on issues related to biases in AI recruitment and selection systems;
- Training of managers, human resource managers and trade union representatives to create greater awareness about the risks of bias in AI recruitment systems; and training of technical experts in diversity and legal aspects;
- Measures to promote greater inclusivity so that more women take up ICT professions.