Gender Equality Index 2020: Digitalisation and the future of work

Digitalisation and equal rights – the role of AI algorithms

AI is being developed at an unprecedented rate, with decision-making algorithms becoming an intrinsic part of our everyday lives. AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-linked, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or internet of things applications) (European Commission, 2018b). AI systems have the power to create an array of opportunities for European society and the economy, but they also pose new challenges.

The increasing use of AI in every aspect of people's lives requires reflection on its ethical implications and assessment of potential risks, such as algorithmic gender bias and discrimination.

AI has been high on the EU agenda since the European Commission launched the European strategy on artificial intelligence, which set the basis for discussions on a coordinated EU approach to addressing the challenges and opportunities of these new technologies (European Commission, 2018b). In her political guidelines, the Commission President highlighted the need for a coordinated European approach to the human and ethical implications of AI while prioritising investments (von der Leyen, 2019).

In 2020, the Commission's White Paper on Artificial Intelligence proposed a policy framework for the creation of a dynamic and trustworthy AI industry. It recognised the need to increase the number of women trained and employed in this area, as well as the risk of bias and discrimination against women by AI systems (European Commission, 2020d).

In the EU gender equality strategy 2020–2025, the Commission reiterated the importance of AI as a leading driver of economic progress and the importance of women as creators and users in order to avoid gender bias (European Commission, 2020c).

Gender bias in AI puts gender equality at risk

There is growing concern that AI tools may be harmfully biased against certain groups, determined by characteristics such as gender, ethnicity, age or disability. Existing biases within society, organisations and individuals – particularly those engaged in the development of AI – can be built into the systems and algorithms, with or without intent.

The lack of gender diversity in the science and technology workforce (see subsection 9.1.3), especially in sectors developing digital technologies, has been credited with enabling and aggravating explicit and implicit gender biases embedded in digital services and products (Wang and Redmiles, 2019). Recent research into gender biases in software development points to the fact that the needs of users whose characteristics (gender/age/disability) match those of the design team tend to be best served by the software (Burnett et al., 2018).

Algorithms, an automated data processing technique, are the basis of AI and require the right governance mechanisms. Automated decision-making is certainly helpful, but when it produces a gender-biased (or otherwise wrong) decision, detection can come too late or its decision could be impossible to change. The term 'black box' is used to describe how algorithms work, neatly encapsulating the fact that, while inputs and outputs can be seen and understood, everything in between – what happens inside the 'black box' – is unfathomable.

The complexity of an algorithm is such that even full access would not bring any clarity as to how the output was created, not even for the developers of the algorithm themselves (Bathaee, 2017). This lack of transparency poses considerable challenges for the evaluation and regulation of algorithms, which are important, particularly for the community that will be ultimately affected by an algorithm's decisions (Al-Amoudi and Latsis, 2019; Goodman and Flaxman, 2017).

The quality of data is an important risk factor for bias in AI. Unprecedented data availability, especially through online collection, has seen much attention paid to the quantity of data available rather than their quality. Problems may arise, such as accurate representation – when data does not represent the population intended – or in measurement – when data does not measure what it aims to (FRA, 2019).

When it comes to algorithms, the correct input is a prerequisite for a correct output (known in data science as the 'garbage in, garbage out' principle). The use of data that reflects existing biases can lead to unfair treatment of certain individuals, resulting in discrimination based on gender, age, dis/ability, ethnic origin, religion, education and sexual orientation (LIBE Committee, 2018).

Use of AI may have gendered consequences in a wide range of settings

Word embedding (a type of algorithm) is used to power translations and autocomplete features in everyday technology. This technology is trained on a body of data of ordinary human language, usually from online sources such as news articles (Bolukbasi et al., 2016; Caliskan et al., 2017). The real novelty of word embedding is that it tries to understand and calculate the relationship between words, instead of taking a word-by-word approach (Nissim et al., 2020).

Leaving aside its innovative nature, word embedding is an example of how the blind application of machine learning risks amplifying gender bias. For instance, one study testing a system's ability to complete analogies resulted in 'man is to computer science as woman is to homemaker' (Bolukbasi et al., 2016). Another study found that use of this tool can result in gender bias in relation to occupations that should be considered gender neutral, with different results given when the system was fed 'he' (doctor) and 'she' (nurse) (Lu et al., 2018).

It is not gender bias alone that surfaces, but other problematic cultural associations, too. Fortunately, there is a push to develop tools to detect and eliminate such bias (Bolukbasi et al., 2016; Chakraborty et al., 2016; Lu et al., 2018; Prates et al., 2019).

AI is increasingly used in hiring or pre-employment assessments, which constitute a clear determinant of economic opportunity for any individual (Bogen and Rieke, 2018; Metz, 2020). AI hiring tools not only offer employers reduced costs but may also help to address or mitigate bias, giving (more) equal opportunities to future and current employees.

One of the selling points of such technology is the ability to assess candidates objectively, without human bias. However, if the algorithm is built without taking into account sensitive characteristics or learns from data on previous biased hiring decisions, it will reproduce institutional and systematic bias while providing the appearance of objectivity (Bogen and Rieke, 2018; Raghavan et al., 2020).

Such cases have already occurred in the labour market: recently, several US companies were found to use algorithms that disadvantaged women candidates, having learned from the hiring history of the company and failed to identify relevant and sensitive characteristics from the data, thus reinforcing gender bias and segregation (Dastin, 2018). The potential for AI to correct discrimination and deliver workplace diversity is undeniable, but it can be fully realised only with awareness, transparency and oversight.

AI has substantial potential to change healthcare through the increasing availability of data and analytical techniques. AI can learn from a large volume of healthcare data, self-correcting to improve its accuracy and the accuracy of medical diagnoses and therapy, all while providing the latest medical information to health professionals (Jiang et al., 2017).

However, medical research is a field historically lacking gender sensitivity, where the lack of representation of women in clinical research has translated into gender-blind or biased healthcare services (EIGE, 2020a). When applying AI to the healthcare sector, bias may arise from the data used to create, train and run the algorithms, while the limitations of an AI tool can easily translate into inaccurate, incomplete or skewed results.

The complexity of the systems makes it difficult to identify and regulate discriminatory practices, a serious concern given their widespread use and the potential to worsen lives. The absence of gender analysis in designing, implementing and evaluating the application of AI in health policy can result in existing health and gender inequalities being overlooked, or new inequalities being created (Sinha and Schryer-Roy, 2018).