Gender Equality– Future of work Labour

**Advisory Committee on Equal Opportunities for Women and Men Opinion on Artificial Intelligence – opportunities and challenges for gender equality**

[**https://ec.europa.eu/info/sites/default/files/aid\_development\_cooperation\_fundamental\_rights/opinion\_artificial\_intelligence\_gender\_equality\_2020\_en.pdf**](https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf)

On April 21, 2021, the European Commission proposed a legal framework for Artificial Intelligence that covers all AI systems that impact natural persons within the European Union, regardless of whether they are operated from within or outside the EU.

Main goals

There are a few stated goals that create the foundation for this proposed regulation. First is that AI systems must be safe and must respect the fundamental rights and values of the European Union. This must be true not only for the outcomes of AI systems, but for the entire value chain of AI development.

The commission also recognized that the absence of regulation had a hindering effect on the development of, the investment in, and the implementation of AI. The lack of clear guidance was a risk factor for all stakeholders, including users who would not benefit from the safety requirements, investors who may not be able to recoup their investment if the unregulated space were at some point in the future hit with more restrictive requirements, and of course AI system developers and implementing customers who would face the fines and penalties if found to be non-compliant. Issuing a regulation before massive widespread adoption of AI systems hits the EU market reduces the risk and uncertainty of AI investment.

Then comes the substance of the issue: providing concrete guidance for governing and enforcing safety and protection of fundamental rights as it relates to AI systems. Being the first regulation of its kind, it sets a global example of how to approach the varied designs, applications, and potential risk of AI systems. This proposed regulation takes a risk-based approach, sorting AI systems into different risk categories, and applying more stringent requirements on higher risk systems.

Lastly, the European Market should remain harmonized, and many disparate regulations from each member state could cause major disruption to the development and use of systems across borders. The EU Commission moved relatively quickly to get ahead of and avoid such a muddled regulatory environment.

What to expect

There are a few important aspects to know about this regulation. The proposed regulation will follow a risk-based approach, as we surmised in a previous blog post.

AI applications will be categorized as creating unacceptable risk, high risk, low risk, and minimal risk. An AI system that causes unacceptable risk by violating the fundamental rights of EU citizens will be prohibited. These prohibitions are called out specifically, including:

* Manipulation with the intent to affect human behavior that would result in physical or psychological harm, including subliminal techniques
* Exploitation of vulnerable people groups
* Social scoring systems by public authorities
* Remote ‘real-time’ biometric identification systems in public spaces for law enforcement
* High-risk AI systems are permitted to be on the market, given that they follow stringent requirements to mitigate the extent to which they create risk to the health, safety, or fundamental rights of individuals.
* Systems with low or minimal risk have fewer requirements imposed on them, but notably do need establish transparency. This will be in the form of communication to end users that information has been generated or altered in an automated way. Systems that interact with humans, used to detect emotion or come to other conclusions based on biometric data, or to generate or manipulate content (‘deep fakes’) must account for the possibility that a user will be manipulated by the outcome. A disclosure that automated means were used for in such situations will need to be communicated.
* Lower risk systems are encouraged to voluntarily adhere to the mandatory requirements of higher risk systems, with provisions for environmental sustainability, accessibility, diversity of development teams, and multi-stakeholder approaches.
* Governing bodies will be established at the EU level and at the member state level.

What to consider at this point

The regulation is still only a proposal, and it must cycle through the Council of the European Union and the European Parliament before it is formally adopted. After that point, the regulation will have one and a half years before it comes into full force.

Debates over on the high-level approach to AI governance and minute details will likely arise. The enforcement of this regulation should be not only applied to AI systems but also to member states. Member states should show adequate effort in enforcing this upcoming regulation.

The regulation itself walks a fine line between prohibition and innovation empowerment, with limited success. The legal architecture supports the restrictive aspects, with only voluntary participation in visionary measures, such as committing to generating public value or environmental sustainability. Facilitating both restrictive safety precautions and enabling bold digital advancement for society is a tall order, but could be better articulated and provided for.

It is expected that AI vendors or customers will contest the risk categorization of AI systems, arguing that their systems are in lower risk categories as it will impact the additional regulatory hurdles that they need to overcome. To give an example, evaluating the financial credibility of individuals – as Schufa does in Germany – is by definition a high-risk AI application. Organizations running such high-risk applications must self-assess to find out whether their doing is ethically compliant. This may be the result of strong interest groups, leaving the draft with an unbalanced outcome.

At a detailed level, there may be concern that deep fakes are only subject to minimal transparency obligations, like disclosure that the audio/visual content was altered. Indication to which degree the content was altered is not required. Since deep fakes pose a high risk for social manipulation as well as algorithmic manipulation, the infiltration of deep fakes into training data can undermine the trustability of all algorithms if left with such a minimal and difficult to enforce requirement.

The public debate on the most appropriate way to regulate and govern AI will continue, but the hope that a comprehensive regulation that aims to protect the rights and safety of individuals in a multi-stakeholder manner is a significant step forward.

**What is the impact of AI on gender equality?** We need to connect the digital space, including AI, to the real world as a continuum and a mirror. Therefore, tackling gender inequalities in the real world is crucial to avoid perpetuation of gender stereotypes and inequalities in the digital world. If the data collected and used for the development of AI is gender biased, the outcome/results will be biased as well. In addition, the algorithm itself, which has been coded by humans, may reflect biases of the developers. Rather than being neutral, human assumptions and the use of selected data lead to biased decisions . Due to these biases, AI can affect gender equality in a number of fields. For example, in the field of women’s socio-economic independence and leadership, when AI systems are used to select candidates for possible vacancies, some algorithms may favour or even only select candidates of one sex, for example men, due to a historic imbalance in hiring practices reflected in the data set used for the AI. This could also have an effect in the healthcare sector. At present, there is a gender data gap when it comes to connecting health issues and diagnoses to symptoms. The symptoms for men vary from those of women. The AI-systems, such as symptomatic machines, rely on the current data and may as such be biased and lead to misdiagnosis. Some publicly available AI applications are predominantly gendered. Digital assistants have been found to remain impassive in the face of sexist insults or even consider sexist remarks as compliments (“I’d blush if I could” as a response to “You’re a slut”). This may seem harmless, but in reality, it perpetuates sexist behaviours. Here it is important to identify patterns in the algorithm that lead to such sexist behaviours and to find solutions to stop this, for example by the development of counter algorithms. Harmful behaviour, sexist remarks, and violence against women, such as instances of (sexist) hate speech or the use of lifelike “deepfake” videos and images” also occur online, disproportionately affecting young women and girls, women journalists, politicians, public figures and women human rights defenders . As a result of such online hate speech/abuse, women are more likely to restrict themselves from engaging online. AI can be used to detect harmful online content against women and girls but needs to be monitored and evaluated by humans to understand the context when analysing harmful online content. Here AI can play a role and assist in order to create an online world that is more respectful and embraces a gender equal society.