AI, Human Rights and Engaging in Policymaking

Transcript of Webinar #1:
An Introduction to Artificial Intelligence and Human Rights

01   Welcome & Agenda

This session, delivered by Marlena Wisniak, Senior Advisor for Digital Rights at the European Center for Not-for-Profit Law provides an overview of the following:

- What is Artificial Intelligence (AI).

- The implications of AI for human rights (including privacy, non-discrimination, freedom of expression, and economic, social and cultural rights).

- Specific applications of AI, their societal impact and their human rights implications with a focus on vulnerable and marginalised groups.

The session takes the form of a presentation and opportunity for questions and answers.

02   What is AI?

There are different categories of AI:

- **Symbolic AI** → Also called GOFAI (good old fashion AI): in this case humans are the ones giving direction to the technology, the machine in fact does not learn by itself, instead it is discovered and led by code that guides the machine in what to do. Improvements in its performance are achieved by humans adjusting / adding to the knowledge which is coded directly in the algorithm. E.g. chatbots.

- **Super intelligence also known as AGI (Artificial General Intelligence)** → There are no certainties on whether it will ever exist in the future or not. It would involve a set of algorithms that are considered to be human-like, as it is so intelligent, it can create new tasks itself and break free from human control.

- Machine **learning & data driven AI.** This webinar focuses on Machine Learning/ Data-driven AI. For our purposes, it is not necessary to understand the details of machine learning (in fact, too much emphasis is put on the algorithm instead of on the output of the algorithm when it is applied.) Machine learning is a range of techniques that automate a learning process, the system finds patterns on its own, based on what has been applied previously, and works from it, predicting and making eventual decisions. Input -> algorithm -> output.

Common terminology:

- **Algorithm** → A set of rules defining how to perform a task or solve a problem. In the context of AI, this usually refers to computer code defining how to process data.

- **Big data** → Broader than AI, big data refers to data sets that

are so large and complex that they cannot be effectively stored or processed with traditional methods.

- **Application Programme Interface** → APIs are the access points that apps and third parties can use to engage with larger platforms and systems, like mobile phones, or social media websites.

- **Data Mining** → Automated process for extracting data and identifying patterns and anomalies.

- **Labelled data** → Data that is accompanied with information about the data i.e. content moderation, what to consider as hate speech for instance that needs to be labelled for AI to be able to recognise it.

03   How is AI used?

Some applications of AI systems include:

- **Image recognition** → Facial recognition software can entail challenges regarding terms of human rights, discrimination, privacy and so much more.

- **Voice/audio recognition** → e.g. Siri/Alexa.

- **Predictive text** → Suggesting continuation of the text.

- **Translation** → e.g. Google translate.

- **Targeted advertising** → Business models of big tech- every time information online is consumed, the data is recorded and analysed to make predictions on what else individuals would be interested in purchasing as consumers.

- **Content moderation** → A larger percentage of the moderation is conducted by AI, instead of human beings.

- **Fraud detection** → Includes systems that can identify who has committed fraud, who is getting undue social support etc. This can be controversial as a machine is essentially choosing who is entitled to support which can have a huge impact on people's lives.

Another example is how resumes are reviewed: Automation can be helpful when thousands of emails are being received. Reviewing resumes requires there to be previous applications that will then be processed to create patterns in order to decide the best candidate amongst the pool and then select that one.

McKane Andrus gives a definition for machine learning: "*it uses historical data and a definition of success to find patterns that lead to success*". Thinking of the algorithm, the "successful resume review" is hard to define as the definition of "best" is challenging to quantify without basing it on biases. For example, Amazon created a hiring tool using AI, however it immediately started discriminating against women in 2018- this is a striking example of how AI can impact non-discrimination rights and the right to work. The rationale behind such an impact is linked to the fact that within

the algorithm being a woman was considered a negative feature, therefore the AI would discriminate against resumes that had that feature. Once systems are trained, they make automated decisions that will then be used to retrain the new systems so the loop continues and perpetuates discriminatory practices.

04   Real world challenges

Today the issue is that everything has been datafied - therefore an important point to consider is the massive ecosystem surrounding AI that includes: private and public funding, business development and research, as well as the product development. A high volume of resources are involved in these processes. It is important to remember that the private sector is at the front line in AI design development and deployment. It is also important to consider that algorithms are usually developed by and for the global north actors, which can exacerbate power inequality.

In addition to this, since hardware and infrastructure are the backbone to AI development, it's important to consider its environmental impact.

Furthermore, there is often a lack of transparency from input/output to design and development- as a society we do not have access to the relevant information when all are under confidentiality agreements.

Finally, the exclusion of affected communities is something that needs to be addressed. The narrative of "insufficient technical skills'' is often used to discriminate against specific communities. Multistakeholder participation in policymaking is thus crucial, and it begins with us sharing our concerns and expertise