“Preventing discrimination caused by the use of AI”

Sarah Burton, Secretary, PACE Committee on Equality and Non-Discrimination

19th European Conference of Electoral Management Bodies
“Artificial intelligence and electoral integrity”

Session 1: AI and fairness in electoral processes

Strasbourg, 14 November 2022

Mr Chair, members of electoral management bodies, distinguished guests, dear colleagues,
As Aldous Huxley famously wrote, “the victim of mind-manipulation does not know that he is a victim. To him the walls of his prison are invisible, and he believes himself to be free.”
Ninety years later, we could equally apply these words to victims of algorithmic unfairness.
Why do I say that, and why is it relevant to electoral processes?

First, some quick background, to put my points for this morning in a broader context:
As we heard in the introductory session, in 2019, the Council of Europe began reflecting on how artificial intelligence was already impacting human rights and democracy, and on the measures that States might need to take to preserve these fundamental guarantees as the use and influence of AI grew.

This led the Parliamentary Assembly also to adopt a series of resolutions and recommendations, in October 2020, on the major issues at stake due to the increased use of AI in fields ranging from justice and discrimination to labour and health care.
Since then, the Assembly has examined the impact of AI on specific aspects of human rights and democracy in some additional resolutions, especially in the field of online communication and media.

So coming back to the topic of this morning’s session on AI and electoral fairness, from the Assembly’s perspective, there are two parts to the question:

- first, how does the use of artificial intelligence affect fairness generally,
- and second, what might this mean in the context of electoral processes?

I’ll be addressing those questions through the prism of non-discrimination, based mostly on Resolution 2343 (2020) “Preventing discrimination caused by the use of artificial intelligence”, prepared for the Assembly by Belgian member Christophe Lacroix.

When we work with machines, we tend to assume that – unlike humans – they are somehow “naturally” objective and fair. After all, they’re machines – no emotions, just mechanical, just executing tasks.

Except that (at least for now), machines only exist because humans create them – and they reflect all the flaws of their human creators.

When it comes to AI, these flaws include bias in the big data that we use to “feed” the machine – because that data comes from the world as it exists today and as it existed previously, with all the discrimination that we know has existed and continues to exist. Discrimination based on gender, age, disability, racialisation, sexual orientation, gender identity and sex characteristics, social origin, religion, family status and so on.

The examples are legion, but two famous ones include online job advertisements, where algorithms only present the highest-paid management jobs to male candidates, because

almost all successful candidates in the past were male, and this data is used to predict who will be successful today; and advertisements for housing, where social origin (for example which town or suburb you come from) is used to determine whether or not you are shown advertisements for housing in wealthier areas, again entrenching existing cycles of social (and often racial) segregation. We might also think here about the example mentioned earlier this morning, of deciding which polling stations should be kept open if their locations are reconfigured.

So biased data is a first, and enormous, part of the fact that using AI tends to perpetuate discrimination.

A second crucial factor is that algorithms reflect the values, beliefs, priorities and convictions of those who design them. They are designed by human beings, and they are optimised to achieve specific objectives identified by human beings.

What they ignore is just as important as what they compute.

A real-world example of the harmful impact of design, with immediate human rights implications, is the way algorithms have been used by public authorities to identify possible welfare fraud. Mr Lacroix's report highlights two quite terrible examples, from the Netherlands and Australia, where thousands of individuals and families were wrongly deprived of welfare benefits because of the way algorithms were designed – leaving some people with enormous and unjustified debts, while reversing the burden of proof, so that it was virtually impossible for them to contest the results of the algorithm.

The most obvious example of the possible impact of algorithm design on the fairness of elections is perhaps the way in which social media algorithms reward hate speech, creating "feedback loops" that amplify polarisation and hate speech, while creating closed circles where people talk only with like-minded others, persuading themselves that they are right and leading to increasingly extreme viewpoints.

At the innocent level of the spectrum, it has been shown that the way in which such algorithms work means that watching videos about vegetarianism will lead you to ones about veganism, and videos about jogging will lead you to ultramarathons, and so on. Applied to the electoral context, we can think about the targeting of political advertising. The US presidential elections and 6 January 2020 come to mind.

I won't go into depth on media issues because there's a full session on that tomorrow morning, but I do want to underline that this polarisation is the result of a human choice to optimise social media algorithms for "engagement" – rather than, say, to prioritise the presentation of a multiplicity of viewpoints or sources.

So we can see that design is crucial, and if those who are designing algorithms fail to question their possibly discriminatory effects, then algorithms that we assume to be fair and objective can in fact imprison us in a world of persisting, and even worsening, unfairness.

Other examples that we might want to think about, where AI could lead to unfairness in electoral systems, might include its use in programmes intended to promote voter enrolment or to strike voters out of electoral lists, for example.

A few further points to flag up quickly, and then I'll move to the Assembly's recommendations for overcoming these issues.

First of all, algorithms are complex and often lack transparency. Second, private companies that design them have an interest in keeping things untransparent, because AI is expensive

to develop and algorithms are often seen as intellectual property to protect. These elements create the so-called “black box” effect, where the lack of transparency makes it extremely difficult for individuals to prove that discrimination has occurred as the result of the use of AI.

A last point to touch on here is that it can also be hard to identify who is responsible for a decision made by algorithm, which again creates obstacles to contesting decisions made by algorithms.

In its series of resolutions on AI, the Assembly outlined five core ethical principles that must underpin all regulatory work in the field of AI. These are: transparency, justice and fairness, accountability, safety and security, and privacy – words which we have already heard quite a lot today.

These principles must also apply when it comes to the use of AI in electoral contexts.

The Assembly also identified a series of measures that can be taken to address the concerns I raised above with respect to fairness and discrimination.

The first is that design teams must take a critical approach to the ways in which algorithms are optimised and the outcomes they produce, from the very earliest design stages and throughout the conception and use of AI.

That means rigorous choice of data, rigorous questioning of the existing bias it might contain and how to eliminate that, and rigorous and constant testing, before the algorithm is rolled out for electoral or other use.

The best way to achieve that is to ensure that AI teams themselves are inclusive – including a diversity of origins, gender, age and so on, as well as a diversity of professional perspectives (not just computer scientists).

It also means ensuring much more diversity in STEM education, but also in companies’ recruitment programmes. And it means that companies also have to improve their capacity to retain a diverse workforce.

It will often mean strengthening antidiscrimination legislation so as to ensure that remedies are accessible despite the obstacles posed by the use of AI.

That means that legislation must cover all grounds of discrimination, with a list of grounds that is as complete as possible, but also open-ended.

Antidiscrimination legislation must also cover direct and indirect discrimination. This is especially important given that some countries do not allow the collection of certain data (for example, data on ethnic origin or sexual orientation), yet algorithms may infer such characteristics from proxies (such as the country of birth of a person or their parents, their postcode, their search history etc).

When it comes to proof, the human rights consequences of requiring individuals to demonstrate their “innocence” in the face of automated decision-making can be immense, as we saw. The shared burden of proof system, set up under the EU Equality Directives, can provide a useful model to follow here.

Finally, I’d like to underline that when it comes to the use of AI by public authorities – and obviously this would apply in electoral contexts too – this is frequently not subject to sufficient parliamentary scrutiny. There are a number of reasons for that, including that parliamentarians may feel out of their depth in this field, or they may feel it is merely scientific or technical, and

not political. But the fact is that AI raises enormous political and societal issues – about due process, the presumption of innocence, the burden of proof and the right to privacy, for example – and one doesn't have to be a specialist in AI to understand and raise those issues.

So there's a really important role for parliaments to play here, to develop effective regulatory frameworks for human-rights-compliant use of AI, and to require that any use of AI by public authorities, in particular, is subject to prior parliamentary scrutiny.

To conclude,

Using AI in electoral contexts doesn't have to lead to unfairness, and it doesn't have to lead us into Aldous Huxley's *Brave New World*.

But to ensure that using AI leads to fair outcomes for everyone, we need to be vigilant, alert, and constantly ready to break down those seemingly invisible walls that can otherwise imprison us in a discriminatory world.