GREEN
EUROPEAN
JOURNAL

Automating Bias? The Risks of the EU's New AI Regulation

Article by Angela Chen, Sarah Chander
May 20, 2021

The European Commission recently unveiled a draft of its highly anticipated artificial intelligence regulation. The legislation comes in the context of the EU's efforts to be the world leader in technology regulation. It contains some strong measures, such as a ban on facial recognition in public spaces (with some exceptions) but there is concern among civil rights groups that the proposal contains too many loopholes and does not go far enough in protecting vulnerable communities.

Angela Chen: The initial reaction to the Commission's draft regulation [on artificial intelligence] from civil rights groups seems to be that it has too many loopholes. Would you agree with that assessment, and could you clarify what those loopholes are, especially regarding the ban on facial recognition?

Sarah Chander: In wording, it is a ban. However, there are exceptions – which for many people invested in the ban on the use of facial recognition in public spaces means it is not good enough, particularly considering how widely the exemptions stand at the moment. Generally, the exemptions are around anti-terrorism and counter-terror measures at the EU level. Many racial justice organisations have pointed to the fact that when we look at counter-terror measures across Europe, you see very potentially discriminatory application of all types of counter-terror policies, particularly over-profiling of Muslim communities.

And there's another exemption which is even wider and it's for "serious crime". That list of crimes is huge: from rape and murder, which are concrete, to things like "swindling" where I don't even know what they mean.

Are there other areas of disparate impact that civil rights experts worry about when it comes to these exceptions?

This concern about disparate impact goes across the different types of "high-risk" AI. It's not just for Muslim communities but for Black communities, Roma communities, people of colour. There's a context that we need to view the deployment of various surveillance technologies in – not just facial recognition but also predictive policing systems and even individual risk assessments in the criminal justice system.

We take a lot of arguments from the U.S. as foresight. We see how predictive policing systems have impacted Black and brown communities, we see how [Immigration and Customs Enforcement] have collaborated with various data-driven organisations to step up their deportation rampage against undocumented communities. Many of these concerns are valid here too if these systems are rolled out more.

There are other technologies that we need to be concerned about too, particularly technologies that purport to automate the recognition of very sensitive identity traits, like race, disability, gender identity. These are not all banned in the draft regulation and are not even considered “high risk”. Not only are they hugely invasive of privacy, but you could also imagine, depending on the political regime of the day, different problematic uses of automated race recognition, for example.

Another criticism I’ve seen is that the draft allows developers of high-risk technologies to self-assess.

Sarah Chander: While this “high-risk” language sounds strong, it requires an internal self-assessment by the developers. So the people that make profit get to determine whether they have sufficiently complied with the transparency or data governance requirements, which are already not so specific. It points to the general ideological underpinning of this regulation. It’s a very commercially minded regulation. It doesn’t have fundamental rights front and centre, even though this is the claim.

Because such systems are steeped in a broader context of racial and class inequality, there is no way you can make a technical tweak […] such that discriminatory results will not ensue from the use of the system.

It very much governs the relationship between providers (AI developers and companies) and users (in this case companies, public authorities, governments) as opposed to governing the relationship between the people or institutions deploying the AI and the people affected by them – which is what I think many people invested in human rights and social justice would have wanted to see. We want to know, what rights do I have when I’m interacting with an AI system? What rights do I have if I’m wrongfully profiled or overly monitored by an AI system?

Is the goal then to have better safeguards and more specific requirements or is the goal a ban?

I can’t profess to have researched every high-risk use in detail, but many of them should be banned. Predictive policing is a really good example. Many people believe that if you de-bias predictive policing systems, they will no longer profile and lead to the over-policing of racialised and poor communities. I disagree. Because such systems are steeped in a broader context of racial and class inequality, there is no way you can make a technical tweak or slightly improve the dataset such that discriminatory results will not ensue from the use of the system. And this leads me to believe that it should be banned. This is one of the areas where the bias debate can be a little bit obscuring.

There may be other uses on that list that could be potentially fixed via safeguards. But I do think the very nature of classifying something as “high-risk” should mean they should be subject to external conformity checks, not internal conformity checks. Otherwise, why

would you bother framing it as “high-risk” if you have complete trust in the entities developing and profiting from them to assess their own compliance?

What interaction does this draft regulation have with the Digital Services Act?

There were multiple points in the regulation where they said they’re specifically not governing what comes under the Digital Services Act. But one potential area where you see overlap is in the prohibitions and the reference to systems that exploit people. In the final version, it’s prohibiting exploiting people specifically on the basis of their age or disability, but you have to show there’s physical or psychological harm that ensues from their exploitation – which is weird wording because it’s basically saying you *can* exploit people on the basis of these things as long as it doesn’t cause physical or psychological harm.

The generalised uptake of AI in any sector should not be a policy goal in and of itself.

However, previously, this was broader; it was [about] exploiting people’s vulnerabilities, which made me think about targeted advertising. Does this mean a ban on targeted advertising? We’re just not sure exactly how this intersects and what the consequences of that are.

This draft regulation will be debated for years to come. What do you think will be the biggest battles?

The prohibitions and list of high-risk technologies was determined by the European Commission, which is an unelected bureaucratic body, highly unrepresentative in terms of our continent’s demographics, highly elite in many ways. They have determined in this proposal not only what constitutes prohibitions and what should constitute “high-risk,” but also that they alone will be the overseers of what is categorised as “high-risk” in the future. There’s not a democratic way in which civil society or people affected could contribute to the categorisation of something which is “high-risk.” I think that will be a key battleground for us. Considering how AI will change, how do we make that process – of arguing for what should be banned and what should be “high-risk” – inclusive and democratic? How do people affected have a say in that process?

It seems like the EU wants to promote the wider adoption of AI generally. Is this too broad a goal?

The generalised uptake of AI in any sector should not be a policy goal in and of itself. The only thing that it does promote is the extension of the private sector into the public sector, which is something I think we should contest. It should be a caveat at least to say, “adopt so far as these things benefit people and comply with human rights” but we don’t see these caveats.

This interview was first published by Heinrich Böll Stiftung.

Angela Chen is a journalist and editor. Her reporting and essays have appeared in The New York Times, The Wall Street Journal, MIT Technology Review, The Atlantic, The Guardian, Smithsonian, Chronicle of Higher Education, National Geographic, Paris Review, Lapham’s Quarterly, and more.

Sarah Chander is a Senior Policy Advisor at European Digital Rights (EDRi), working on AI, anti-discrimination and digital rights. She is co-founder of Equinox Initiative for Racial Justice and an advisor to the Council of Europe on intersectionality and anti-racism. Previously she worked in advocacy at the European Network Against Racism (ENAR), on a wide range of topics including anti-discrimination law and policy, intersectional justice, state racism, racial profiling and police brutality.

Photo credit: Andreea Belu

Published May 20, 2021
Article in English
Published in the *Green European Journal*