
EUROPEAN DIGITAL RIGHTS

# Artificial Intelligence & Fundamental Rights

*How AI impacts marginalized groups, justice and equality*

European Digital Rights 12 Rue Belliard, 1040 Brussels www https://edri.org twitter @edri_org tel +32 (0) 2 274 25 70

*This short explainer details how the growing use of Artificial Intelligence (AI) in a number of areas of public life may pose risks for equality, justice and fundamental rights.*

The European Commission has launched a **public consultation** on the regulation of AI (**deadline 14th June 2020**) following an initial proposal to introduce a "risk-based" framework in their White Paper on Artificial Intelligence. It is important that **organisations working for human rights, equality and non-discrimination are heard** in this process.

## WHAT IS ARTIFICIAL INTELLIGENCE?

AI is a broad term encompassing a range of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decisions and solving problems.[1] AI systems are built up of complex **algorithms**, which are processes or sets of rules to be followed by the machine.

When we speak of AI today, we typically mean **machine learning**, a subtype of AI that involves the computerised statistical analysis of large data sets to train a system to make predictions (e.g. when your word processor makes suggestions to autocomplete sentences) or spot patterns (e.g. recognising faces in CCTV footage).[2] In essence, machine learning is about automating decision-making processes by learning from data on past trends or behaviours to make predictions about the future.

## USES AND RISKS OF AI

**Law enforcement**: 'Predictive policing' tools are increasing bought and deployed by police across Europe.[3] Such technologies claim to forecast where, and by whom, crime is likely to be committed. Evidence is emerging showing that these systems repeatedly score working class, migrant and racialised communities with a higher likelihood of future criminality. The use of apparently "neutral" factors such as postal code in practice serve as a proxy for race, reflecting histories of over-policing of certain communities, exacerbating racial biases and affording false objectivity to patterns of racial profiling. This also undermines the presumption of innocence by treating people as individually suspicious based on inferences about a wider group.

There is an added concern about how AI, used in combination with facial recognition and other biometric processing technologies can facilitate unlawful **mass surveillance** in public spaces. Risks of surveillance, profiling and discrimination are interconnected, particularly when likely to be deployed disproportionately in lower income or minority areas.[4] EDRi is calling for an outright ban on biometric mass surveillance, or the untargeted processing of data (such as facial and speech recognition) in public places.

1 UN General Assembly, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc. A/73/348, 29 August 2018.
2 Acess Now 'Mapping Regulatory Proposals for Artificial Intelligence in the EU' https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf
3 European Network Against Racism (2019). 'Data-driven policing: hardwiring discriminatory profiling' Available at: https://www.enar-eu.org/Data-driven-policing-is-leading-to-racial-profiling
4 Liberty (2019). 'Policing by machine' https://www.libertyhumanrights.org.uk/issue/policing-by-machine/

**Justice sector:** The use of automated decision making in the criminal justice system and other areas for the purposes of risk assessment, such as to predict risk of re-offending, is a huge cause for concern. These systems of risk calculation popular in the US pose severe risks of racism and discrimination, with black and brown defendants systematically scored as 'more likely to reoffend', even in identical scenarios. Risk-scoring systems like COMPAS also overpredicts the risk for women to re-offend, leading to unfair sentencing of women. Automation bias – the propensity for humans to favour suggestions from automated systems - in these cases threatens the rights of individuals to a fair trial and to participate in the justice process. It also creates a barrier to their legal right to challenge and gain information for decisions made about them.

**Migration control:** There is a growing danger that the deployment of AI in the field of migration control will exacerbate the vulnerability of migrants in a context of wide-scale data sharing about migrations to facilitate deportation[5], impeding access to vital services such as healthcare, social security and other state support. AI is being tested to detect lies for the purposes of immigration applications at European borders, allocate resources at refugee camps through iris scanning, and to (inaccurately) monitor deception in English language tests through voice analysis. In addition, plans to revise the Schengen Information System, will use AI tools such as facial recognition to help facilitate the return of migrants.[6] These uses of AI must be seen in context of increased data processing and sharing, cross-border merging of data-sets, and the denial of data protection rights to migrants, in particular people without documents.

**Recruitment and in the workplace:** The growing use of AI software and other algorithmic systems for recruitment poses concerns for historically discriminated groups in employment and is likely to exacerbate existing inequalities experienced by women, racialised groups, those living with disabilities, LGBTQ communities. Such systems purport to find a good fit for a particular role by screening candidates' applications, based on pre-designed specifications of the ideal candidate. A key concern here is that the "ideal candidate" is often modelled on previous successful employees, likely to reflect and deepen existing privileges, hierarchies and hiring biases.[7] One highly concerning example is the development of technology for hiring which purports to identify whether applicants have a disability, as recently patented by the AI company HireVue.[8]

There are also concerns for workers' rights with the growing trend of AI tools for **worker surveillance**. Such systems have been used in a variety of ways to make automated calculations about worker performance, 'mood assessment', monitoring of task productivity and more.[9]

5 PICUM (2019) "Data Protection, Immigration Enforcement and Fundamental Rights: What the EU's Regulations on Interoperability Mean for People with Irregular Status"

6 Ana Beduschi (2020) 'International Migration Management in the age of Artificial Intelligence' *Migration Studies*, available at: https://academic.oup.com/migration/advance-article/doi/10.1093/migration/mnaa003/5732839

7 Institute for the Future of Work 'AI in hiring: Assessing Impacts for Equality' Available at: https://static1.squarespace.com/static/5aa269bbd274cb0df1e696c8/t/5ea831fa76be55719d693076/1588081156980/IFOW+-+Assessing+impacts+on+equality.pdf

8 Loren Larsen, Keith Warnick, Lindsey Zuloaga, and Caleb Rottman, "Detecting Disability and Ensuring Fairness in Automated Scoring of Video Interviews," United States Patent Application Publication, August 20, 2018

9 https://www.tuc.org.uk/news/6-10-workers-say-being-snooped-their-boss-fuels-distrust-and-discrimination

**Social welfare:** AI systems have been deployed in contexts of social welfare resource allocation, eligibility assessment and fraud detection. In a famous case the Dutch government deployed SyRI, a system to detect fraudulent behaviour in benefits creating risk profiles of individuals. In 2019 a Dutch court found that this system violated human rights and privacy law. The court noted that the SyRI program, primarily deployed in poor and migrant neighbourhoods also could lead to discrimination.[10] There are more and more examples of how automated decision-making, profiling and digitalisation more generally is affecting poor and working class people. For example, for many years the Polish government has used data-driven systems to profile unemployed people.

**Online advertising:** There is growing evidence demonstrating that AI and other forms of algorithmic decision-making pose additional risks of discrimination. Here the business model of tailoring advertising to users based on the collection of intimidate personal data and sensitive inferences about their identities has had serious consequences for the content marginalised groups are likely to see (or not see). Evidence of discriminatory exclusion of women from seeing STEM jobs online, censoring of Muslim and LGBTQ content, to drastically different advertising on recruitment, housing and other results delivered in Google searches by people of colour has led to people experiencing 'filter bubbles', and may contribute to election manipulation and even forms of excluding or segregating communities.[11]

**Social media:** Greater oversight is needed of the use of AI in content moderation on social media platforms. Marginalised groups and political activists face heightened risk of censorship, content take-downs, and account suspension, and are at the same time more vulnerable to hate speech, online harassment and threats.

**Domestic violence:** "Smart home" technologies that automate various facets of household management have enabled gender-based violence (GBV) and domestic abuse, facilitating remote control of household objects consequently used as tools for violence and control.

## CROSS-CUTTING FUNDAMENTAL RIGHTS RISKS OF AI

**Compounding discrimination and inequality:** AI presents huge potential for exacerbating discrimination in society, at a scale and to a degree of opacity that goes beyond non-automated or 'human' processes. In addition, automated decision making has often been wrongly portrayed as neutral and 'objective', when in fact it embeds and amplifies the underlying structural biases of our societies. This creates a high risk of automation bias and can lead to difficulties for humans to challenge discrimination which is perpetrated by machines or complex systems. In addition to this, however, we see that AI has the potential to pose harms in relation to:

a) discrimination on the basis of grounds not covered in existing discrimination law, such as financial status, such as with examples from targeted advertising and financial credit scoring.

10 https://pilpnjcm.nl/en/dossiers/profiling-and-syri/
11 Latanya Sweeney (2013) "Discrimination in online ad delivery"

b) collective harms, for example systems which disadvantage certain communities, geographic areas, such as with predictive policing tools.
c) the deepening existing societal inequalities, such as systems which deploy risk scoring in the criminal justice system, biometric recognition systems deployed disproportionately in lower income or minority areas, or deployments in the field of social welfare.

**Classification marginalisation and human dignity:** Many automatedsystems rely on making classifications, some of which bear heavily on identity categories resulting from historical processes of marginalisation (race, gender identity, disability status, etc). AI has been used to attempt analysis of individuals on the basis of inferences about disability, sexual orientation, emotion and behaviours, whether they might be 'victim' to child abuse, or the veracity of claims made in the processing of visa applications.[12] For use in these cases, AI systems necessarily reduce or overlook the complexity of identity categories in favour of rigid classifications. There is a potential for harmful consequences, e.g. the denial of medical care; being evaluated as having a high risk of criminal recidivism. Further, these uses inherently negate our basic dignity – to self-define our identities and for this information to be private if we wish. Making distant, un-transparent and often unchallengable decisions based on these inference is incompatible with our dignity and basic freedoms.

**Accountability:** The use of AI applications may lead to a lack of accountability when harms are produced by apparently "neutral" applications. In addition, characteristics specific to machine learning may lead to unauthorised use or purpose creep. It is likely that new systems of accountability for the impact of the AI-based technologies are required. Yet, a tendency of designers and deployers of automated systems to allocate responsibility to the technology poses a severe risk for meaningful accountability relating to AI. Further, the shift toward 'ethics-based' self-regulation of artificial intelligence can threaten meaningful accountability for real social harms by presenting these issues as vague ethical problems rather than legal rights-based problems – with rules and requirements for mitigating and redressing harm.

**Data-protection:** AI relies on the processing of large amounts of data for training and accuracy, raising major questions about consent and personal privacy as general principles. In addition, any regulation of AI must complement the enforcement of the GDPR, addressing severe issues posed by AI for the enforcement of meaningful consent, objection, data minimisation, purpose limitation, explanation. Further, many uses of AI function through the use of "non-personal" data (not covered by the GDPR) or sensitive inferences[13] of personal information about individuals, therefore still in essence infringing their rights and privacy, but with unclarity about how they pursue a remedy.

**Democracy and transparency:** The promotion of AI systems for public purposes, whether in the public sector or in *de facto* public spaces, such as social media platforms (even though in reality privately owned), poses real questions for transparency and democratic oversight. The

12 Parliamentary question :iBorderCtrl: False incrimination by and discriminatory effects of video lie detector technology https://www.europarl.europa.eu/doceo/document/E-9-2020-000152_EN.html
13 Sandra Wachter (2019) 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' Columbia Business Law Review, 2019(2), 494–620. Retrieved from https://journals.library.columbia.edu/index.php/CBLR/article/view/3424

procurement, design, testing, and deployment of AI systems in areas such as healthcare, social services, housing, policing, migration and other areas demonstrates real issues relating to the influence of private actors in public governance, opacity, and a real potential impact on many fundamental rights of people who may not know about, consent to or have the opportunity to object to/ contest decisions made by an automated system. In addition, many AI systems have been deployed in areas of public concern without adequate justification or scientific evidence. The main reason given is often cost-saving and increased efficiency, yet on very few occasions are these decisions transparent or made with meaningful public consultation, nor can such reasons justify violations of human rights.

**Expression and Disinformation:** The use of AI to facilitate profiling and targeted content generation has been increasingly documented as posing a major threat to democratic political processes and exacerbating disinformation.[14] In addition, the use of automated decision making systems for content moderation has demonstrable impacts on rights to privacy and freedom of expression, in particular related to decisions made around the handling, removal and prioritisation of content.[15]

## WHAT CAN WE DO? – *UPHOLDING FUNDAMENTAL RIGHTS IN AI*

It is vital that these fundamental rights risks are further explored and addressed in our advocacy – at the European, national and local levels. How can we defend human rights in the use of AI? Here's what human rights organisations can do:

1. **Engage our communities** – the human rights risks of automated systems and data-driven tools are relatively unknown. The more we raise awareness and make the connections between movements working for digital rights, equality,and justice in different areas, the more informed and effective we will be.

2. **Demand preventative measures and 'red-lines'** - To protect human rights, and diginity, there must be a strong focus on **preventing abuse** and prohibiting certain government and private uses of AI – where they infringe on rights, equality and justice. EDRi's main recommendation toward the EU institutions is to **set the legal limits for AI, based on fundamental rights and the impact on individuals, communities and society**. We propose possible red lines at:

    - indiscriminate biometric surveillance and biometric capture and procesing in public spaces[16]
    - use of AI to solely determine access to or delivery of essential public services (such as social security, policing, migration control)
    - uses of AI which purport to identify, analyse and assess emotion, mood, behaviour, and sensitive identity traits (such as race, disability) in the delivery of essential services

14 Demos (2018) 'The Future of Political Campaigning'
15 Privacy International and Article 19 (2018) 'Privacy and Freedom of Expression in the Age of Artificial Intelligence' Available at: https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf
16 EDRi (2020). 'Ban Biometric Mass Surveillance: A set of fundamental rights demands for the European Commission and Member States' https://edri.org/wp-content/uploads/2020/05/Paper-Ban-Biometric-Mass-Surveillance.pdf

- predictive policing
- autonoumous lethal weapons and other uses which identify targets for lethal force (such as law and immigration enforcement).

Law-makers must engage the public – in particular members of marginalised communities likely to be affected by such technological changes – meaningfully in decisions about AI.

3. **Feed into political processes on digital rights** – Communities and civil society need to have a say in decision making related to technology, particularly in the public sphere. The tech and security industries have heavily invested lobbying resources into political processes around tech and AI, and human rights and justice centred voices need to counteract this. Here's some ways to do this currently:

EDRi shares with you its response to the European Commission's consultation on Artificial Intelligence and encourages you to also give your input: your expertise is greatly needed.

The full EDRi consultation response, recommendations, and answering guide for the public on AI regulation to the European Commission can be found here.