EUROPEAN DIGITAL RIGHTS

Recommendations for a Fundamental Rights-based Artificial Intelligence Regulation

Addressing collective harms, democratic oversight and impermissable use

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

Recommendations for a fundamental rights-based artificial intelligence response

Addressing collective harms, democratic oversight and impermissable use

Published on 04 June 2020 in Brussels

Lead author:
Sarah Chander, EDRi Senior Policy Advisor

Layout by:
Rafael Hernández, EDRi Communications Intern

The EDRi Brussels office would like to express our enormous thanks the whole network for their time, advice, participation and support in producing this paper and to the 28 organisations that particpated in the consultation. In particular, the following organisations and individuals have been instrumental across multiple stages of researching, drafting, discussing and reviewing:

Access Now
Article 19
Bits of Freedom
Digitalcourage e.V.
IT- Political Association of Denmark
Homo Digitalis
Panoptykon Foundation
Privacy International
Statewatch
Vrijschrift

Table of Contents

We are starting to see the impact of artificial intelligence in all areas of public life. As governments, institutions and industry swiftly move to 'innovate' - promoting, investing and incorporating AI into their systems and decision-making processes - **grave concerns remain as to how these changes will impact people, communities and society as a whole**. AI systems have the ability to exacerbate surveillance and intrusion into our personal lives, reflect and reinforce some of the deepest societal inequalities, fundamentally alter the delivery of public and essential services, vastly undermine vital data protection legislation, and disrupt the democratic process itself.

The growth of artificial intelligence is specific and warrants attention because - due to the (designed) opacity of the systems, the complete lack of transparency from state and private actors when such systems are deployed for use in public, essential functions, and the systematic lack of democratic oversight and engagement – AI is furthering the '**power asymmetry between those who develop and employ AI technologies, and those who interact with and are subject to them**.'[1] For some, AI will mean reinforced, deeper harms as such systems feed and embed existing processes of marginalisation. For all, the route to remedies, accountability, and justice will be ever-more unclear, as this power asymmetry further shifts to private actors, and public goods and services will be not only automated, but privately owned.

The European Union's upcoming legislative proposal on artificial intelligence (AI) is an opportunity to protect people, communities and society from the escalating economic, political and social issues, posed by AI. This paper outlines the position of **European Digital Rights (EDRi)** in response to the European Commission's White Paper on Artificial Intelligence.

We argue that the European Union's regulatory response must **reinforce the protectons already embedded in the General Data Protection Regulation** (GDPR), outline **clear legal limits for AI** by focusing on impermissable use, and foreground principles of **collective impact, democratic oversight, accountability, and fundamental rights**.

EDRi argues that regulation is necessary to guarantee fundamental rights in relation to Artificial Intelligence. Regulation is needed for two purposes:

1. Defining the legal boundaries for AI – The EU must set legal boundaries which reflect social and fundamental rights concerns in order to provide certainty as to what AI may be developed and deployed, and for which purposes.
2. Outling clear fundamental rights safeguards – within these boundaries, there must be sufficient safeguards to protect fundamental rights in the procurement, design, development, deployment of all systems.

1 Council of Europe (2019). 'Responsibility and AI DGI(2019)05 Rapporteur: Karen Yeung https://rm.coe.int/responsability-and-ai-en/168097d9c5

Whilst the European Commission has made clear proposals for the latter, the White Paper proposal does not set adequate social, fundamental rights-based boundaries to underpin its regulatory response to AI. The European Commission now has an opportunity to improve its regulatory proposal to ensure a 'human-centred' approach which truly promotes 'trustworthy AI'.

This paper oulines the fundamental rights impacts of aritficial intelligence, making the case for regulation on AI. Further, EDRi outlines general principles to inform the updated regulatory response, and lastly, recommendations for a fundamental rights-based AI regulation.

1. Fundamental Rights Impacts of Artificial Intelligence

The below outlines the main fundamental rights risks of AI underlying EDRi's position on the European Commission White Paper. AI will pose unprecedented challenges for fundamental rights in a number of areas – this must be addressed in AI regulation.

Data-protection: Increased use of artificial intelligence pose inherent risks to existing data protection rights and standards. More structurally, AI relies on the processing of large amounts of data for training and accuracy, raising major questions for consent and personal privacy as general principles. In addition, any regulation of AI must strengthen and 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 or sensitive inferences[2] of personal information about individuals, therefore threatening anonymity and the spirit of the rights enshrined in the GDPR. This poses a challenge for data protection rights and the regulation of AI.[3]

Equality, non-discrimination and inequality: AI and other automated systems are likely to worsen discrimination, due to greater scales of operation, increased unlikelihood that humans will challenge its decisions (automation bias), and lower levels of transparency about how such decisions are made.[4] There remain heightened concerns as to how the deployment of AI in numerous areas pose a risk of discrmination against individuals with charactertistics protected by equality law, with numerous examples in the field of recruitment. In addition to this, however, we see that AI has the potential to pose harms in relation to:

2 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

3 EDRi (2020). 'A human centric internet for europe', https://edri.org/a-human-centric-internet-for-europe/

4 Agata Foryciarz, Daniel Leufer and Katarzyna Symielewicz (2020). 'Black Boxed Politics: Opacity is a choice in AI systems': https://en.panoptykon.org/articles/black-boxed-politics-opacity-choice-ai-systems

a) discrimination on the basis of grounds not covered in existing discrimination law, such as financial status,[5] such as with examples from targeted advertising and financial credit scoring.
b) collective harms, for example systems which disadvantage certain communities, geographic areas, such as with predictive policing tools.[6]
c) exacerbate existing societal inequalities, such as systems which deploy risk scoring in the criminal justice system,[7] biometric recongition systems deployed disproporionately in lower income or minority areas, deployments in the field of social welfare which can have severe financial consequences for people in the case of error, miscategorisation or identification issues[8] , and applications being developed which purpose to estimate or identify sensitive identity traits such as sexual orientation and identity.[9]

Democracy and transparency: The promotion of, and resort to, AI systems for public purposes, whether in the public sector or in *de facto* public domains, such as social media platforms, poses real questions for transparency and democratic oversight of decisions made in the public domain. The 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, consent or have the opportunity to object to or contest decisions made by an automated system. In addition, many AI systems have been deployed in areas of public concern without justification or scientific evidence.

Expression and Disinformation: The use of AI to facilitate profiling and targeted content generation and curation has been increasingly documented as posing a major threat to democratic political processes and exacerbating disinformation.[10] In addition, the use of automated decision making systems for content moderation has demonstrable impacts on rights to privacy and expression, in particular related to decisions made around the handling, removal and prioritisation of content.[11] Regulatory steps to prescribe AI and other automated content moderation and removal systems (so- called upload-filters) are likely to compromise the right to freedom of expression, and encourage censorship of online speech by private actors in order to comply with legislation.[12]

5 Council of Europe (2018). 'Discrimination, artificial intelligence and algorithmic decision-making' https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73
6 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; Access Now (2018) 'Human rights in the age of Artificial Intelligence', available at: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf
7 Liberty (2019). 'Policing by machine' available at: https://www.libertyhumanrights.org.uk/issue/policing-by-machine/
8 UN (2019) Report of the Special Rapporteur on extreme poverty and human rights. https://undocs.org/A/74/493
9 AI Now (2019) 'Disability, Bias and AI' available at: https://ainowinstitute.org/disabilitybiasai-2019.pdf
10 Demos (2018) 'The Future of Political Campaigning'
11 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
12 EDRi (2020). 'Position paper on the digital services act – Platform Regulation Done Righ' https://edri.org/wp-content/uploads/2020/04/DSA_EDRiPositionPaper.pdf

Procedural rights and access to justice: The deployment of artificial intelligence in the criminal justice system and other public areas for the purposes of risk assessment, or the delivery of any process rights pose particular issues for the rights of indivuduals to participate in the justice process and also to challenge and gain information for decisions made about them.

Fundamental rights abuse in migration control: There are increasing examples of AI deployment in the field of migration control, posing a growing threat to the fundamental rights of migrants, EU law, and human dignity. 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.[13] All such uses infringe on data protection rights, the right to privacy, the right to non-discrimiantion, and several principles of international migration law, including the right to seek asylum.

Surveillance: There are grave concerns related to the extent to which AI will both facilitate and necessitate mass surveillance in public and private spaces against the general public.[14] In addition, numerous examples demontrate how AI has been used to facilitate analysis of individuals on the basis of inferences about sexual orientiation, emotion recognition, the veracity of claims made in the processing of visa applications.[15] As such, risks of surveillance, profiling and discrimination are interconected.

Accountability: Features of AI also challenge existing frameworks of legal accountability for rights violations or harms, and are likely to require new systems to regulate and ensure redress for harms emenating from automated decision making systems. The power and information asymmetries specific to artificial intelligence poses a challenge for accountability and redress in the instance of social harms. In addition, characteristics specific to machine learning may lead to unauthorised use or purpose creep. Yet, a tendency of designers and deployers of automated systems to allocate responsibility to the technology poses a severe risk for meaninful accounntability relating to AI. Further, the shift toward 'ethics-based' self-regulation of artificial intelligence can threaten meaningful accountability for real social harms.

13 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

14 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

15 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

2. A Rights-based Approach – General Principles

European Digital Rights (EDRi) argues that the following principles should inform the update of the European Commission's proposal:

- **Upholding fundamental-rights, preventing harm** – the EU must commit to a robust fundamental rights-based approach as the primary priority of the regulation. This should include legislative measures designed to prevent fundamental rights abuses in situations in which AI development or deployment is incompatible with fundamental rights. To prevent fundamental rights such abuses, the EU must outline legal limits for AI and ban impermissable uses.

- **Addressing the collective impact of AI** - Artificial intelligence and other automated decison making sytsems pose serious societal challenges, many of which fall outside the scope of laws designed to protect the rights of indiviudals in society. The EU must acknowledge collective impact posed by AI to people and democracy and adjust its legislative approach accordingly.

- **Ensuring democratic oversight** - Due to the 'power asymmetry between those who develop and employ AI technologies, and those who interact with and are subject to them'[16] the potential for high levels of instrusion and individual and collective impact in many areas of social, economic and public life, it is imperative for the EU to incorporate requirements for real and meaningful democratic oversight and consultation on AI into its legislative approach. This must include specific engagement with civil society organisations, indiviudals and marginalized communities disproportionately impacted by AI systems.

- **Centering accountability** – It is necessary for the EU to establish a system of accountability for rights violations and the social harms resulting from the deployment of AI systems.

PRINCIPLES

16 Council of Europe (2019). 'Responsibility and AI DGI(2019)05 Rapporteur: Karen Yeung https://rm.coe.int/responsability-and-ai-en/168097d9c5

3. Recommendations for a fundamental rights-based AI regulation:

European Digital Rights (EDRi) suggests that the European Commission's regulatory approach incorporates the following recommendations:

1. Conduct and publish a fundamental rights and AI review

The European Commission must demonstrate it has clearly reviewed, assessed and adjusted its coordinated plan on AI in order to address the severe fundamental rights implications of Artificial Intelligence. Such a communication should outline how such risks will be mitigiated in the EU's legislative approach, how artificial intelligence impacts existing legal frameworks (such as the General Data Protection Regulation and anti-discrimination law, and in the implementation) of Member State national strategies.

Legal boundaries for Artificial Intelligence

2. Develop clear criteria for legality of Artificial Intelligence, including:

- Clarity as to the specific use andpurpose of the system in question;
- Standards for scientific and policy evidence demonstrating the use/ purpose;
- Requirements for democratic oversight, control and consulation for the development design, testing and deployment of AI in the public sphere, engaging national parliaments and oversight bodies, but also citizens directly, including exploring the use of 'citizen boards' and other modes of public engagement.[17]

3. Outline uses for which the development and/or deployment of AI are impermissable, to prevent fundamental rights abuses, and outline a system to regulate this, including a **ban on uses of AI which are incompatible with fundamental rights**, fundamental European values, and existing European law including (but not limited to):

- indiscriminate biometric surveillance and biometric capture and processing in public spaces;[18]
- 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;
- predictive policing;
- use of AI systems at the border or in testing on marginalised groups, such as undocumented migrants;[19]

17 Citizen Advisory Boards is but one example of effective ways to engage the public in oversight functions in more comprehensive ways than public consiltation or 'dialogues' which ultimately do not accompany decision-making power. More research is required on metholodies for effective democratic oversight of public deployments of AI.

18 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

19 University of Toronto (2019). 'Bots at the Gate: A Human Rights Analysis of Automated Decision Making in Canada's Immigration and Refugee System'.https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

- autonomous lethal weapons and other uses which identify targets for lethal force (such as law and immigration enforcement);
- general purpose scoring of citizens or residents, otherwise referred to as unitary scoring or mass-scale citizen scoring;[20]

Determining legal limits, impermissable uses or 'red-lines' for AI applications is a necessary step for a people centred, fundamental rights-based AI. Fundamental rights are upheld by focusing on the prevention of systematic harm to individuals, communities and societites.

Further research and democratic engagement is necessary to determine red-lines for AI applications. This debate should explicitly include considerations of the impact of AI applications for meaninfgul democratic engagement, accountability for harms and the role of such systems to automate patterns of inequality and exclusion.

EDRi started the processs by outlining a clear red-line: **the use of biometric processing and capturing in publicly accessible spaces**. Such uses of biometric data signficiantly contribute to unlawful mass surveillance and therefore should be banned, as outlined in EDRis paper '**Ban Biomeric Mass Surveillance: A set of fundamental rights demands for the European Commission and Member States**'. The paper establishes that such uses will transform public spaces into sites of continuous watching and irreversibly compromise fundamental rights to privacy, freedom of assembly, expression, non-discrimination, data protection, fair trials, democracy and the presumption of innocence.

EDRi calls on the European Union institutions and member states to halt all biometric processing amounting to mass surveillance, cease funding for and legislation which condones biometric processing that could lead to mass surveillance, and, specifically, for the European Commission to:

> *"...implement, through legislative and non-legislative means and if necessary, infringement proceedings and Court action, an immediate and indefinite ban on biometric processing that leads to mass surveillance in public spaces. This process must be done under the supervision and/or support of the European Data Protection Supervisor (EDPS), the European Data Protection Board (EDPB), the FRA and DPAs."*
>
> - EDRi (2020) 'Ban Biometric Mass Surveillance: A set of fundamental rights demands for the European Commission and Member States.'

20 European Commission, High-Level Expert Group on Artifical Intelligence, 'Ethics guidelines for 'Trustworthy AI' https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

Outline clear fundamental rights safeguards

Within these boundaries, the European Union must outline sufficient safeguards to protect fundamental rights in the procurement, design, development, deployment of all systems.

EDRi's position is that **the current high/low risk approach in the White Paper is insufficient to prevent, mitigate and remedy fundamental rights violations**. The main problems with this approach are:

- The distinction in itself presents a potential loop-hole for scrutiny of AI applications, poentially presenting a major risk to fundamental rights. Instead, fundamental rights considerations should be the core of AI regulation in general and the application of safeguards in particular;
- The sectoral determination of 'high-risk' applications is overly simplistic and poses the potential to create further exemptions to fundamental rights scruinty and safeguards.
- The sectoral determination will be unable to account for major developments in AI, the secondary use or repurposing of AI applications from one sector to another, and overlooks the central importance on use or purpose of AI applications, which is central to fundamental rights oversight.
- The current approach portrays the issue as a problem of technology application when in essence the main issues with AI relate to governance or decision-making processes. 'Risk' or harm should be determined by clear, fundamental rights-based assessment rather than broad sector-oriented criteria.

In addition to the safeguards proposed in the European Commisison White Paper, EDRi recommends:

4. A clear definition of AI in the scope of the regulation

Any future regulation on AI must be specific about the applications intended to be within its scope. We recommend the regulation defines AI narrowly, excluding mundane and simplistic applications, to ensure horizontal rules to all AI necessary to regulate.

5. Avoid creating exemptions to fundamental rights protections and scrutiny

The European Commission should review the AI coordinated plan with a view to avoiding distinctions which limit fundamental rights protections for AI. There should be no loop-holes or exemptions to scruintybased on sector, size of enterprise, whether or not the system is deployed in the public sector. On this basis EDRi rejects:

- the current high-risk/low-risk distinction, in particular the focus on sector as a determiner of risk;

- that public sector AI uptake should be promoted *prima facie*.

Instead, regulation, must be based on an assessment of levels of harm and outcomes for individuals and society.[21]

6. Mandatory human rights impact assessments for all applications

All systems meeting the legal criteria must complete mandatory human rights impact assessments throughout the design, development, and ongoing development. Following the recommendation of the Council of Europe Committee of Ministers on the **human rights impacts of algorithmic systems**, this assessment should include an evaluation of the collective, societal, institutional and governance implications the system poses, and outlining adequate steps to mitigate this.[22] Such impact assessments must be made publicly available. To implement this recommendation, the Commisison could consider a mandatory disclosure or notification system.

7. Implement a clear oversight and enforcement model

EDRi recommends that the model for oversight and enforcement of AI regulation includes:

- Effective oversight and enforcement of fundamental rights standards and clear EU and national enforcement of AI regulation;
- National Centres of Expertise on AI with strong engagement and involvement with civil society, equality bodies and human rights;
- Individual and collective complaints mechanisms and law enforcement to remain with human rights bodies (DPA, ombudsman, equality body etc);
- Systems of democratic oversight of AI in public sector procurement of AI, incldu-ing citizen boards and other methods of engaging the public;
- Mandatory transparency measures about the use of AI applications, incldu-ing public statistics on deployments, aims, scope and developers.

21 Council of Europe Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

22 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

4. Further resources:

Access Now (2020). 'Access Now's submission to the Consultation on the "White Paper on Artificial Intelligence – a European approach to excellence and trust"'
https://www.accessnow.org/EU-white-paper-consultation

Access Now (2018). 'Human rights in the age of Artificial Intelligence', available at: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

Access Now (2019). 'European Human rights Agenda for the Digital Age: Recommendations.'
https://www.accessnow.org/cms/assets/uploads/2020/02/European-Human-Rights-Agenda-for-the-Digital-Age-Recommendations-on-Artificial-Intelligence.pdf

Algorithm Watch (2019). 'Automating Society: Taking stock of Automated Decision-making in the EU'.
https://algorithmwatch.org/en/automating-society/

Bits of Freedom (2020). 'Facial recognition: A convenient and efficient solution, looking for a problem?'
https://www.bitsoffreedom.nl/2020/01/29/facial-recognition-a-convenient-and-efficient-solution-looking-for-a-problem/

Council of Europe (2019). 'Responsibility and AI DGI(2019)05 Rapporteur: Karen Yeung
https://rm.coe.int/responsability-and-ai-en/168097d9c5

Council of Europe (2018). 'Discrimination, artificial intelligence and algorithmic decision-making'
https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

Council of Europe Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.

EDRi (2020). 'A human centric internet for europe',
https://edri.org/a-human-centric-internet-for-europe/

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

EDRi (2020). 'Position paper on the digital services act – Platform Regulation Done Right https://edri.org/wp-content/uploads/2020/04/DSA_EDRiPositionPaper.pdf

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

European Disability Forum (2018). 'Plug and Pray: a disability perspective on artificial intelligence, automated decision making and emerging technologies' http://www.edf-feph.org/sites/default/files/edf-emerging-tech-report-accessible.pdf

Liberty (2019). 'Policing by machine' https://www.libertyhumanrights.org.uk/issue/policing-by-machine/

Open Society Foundations and the European Consumer Organisation (BEUC) (2019). 'A Human Centric Digital Manifesto for Europe' https://www.opensocietyfoundations.org/publications/a-human-centric-digital-manifesto-for-europe

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

Stanford Global Digital Policy Incubator (2020). 'National Artificial Intelligence Strategies and Human Rights': https://www.gp-digital.org/wp-content/uploads/2020/04/National-Artifical-Intelligence-Strategies-and-Human-Rights%E2%80%94A-Review_.pdf

University of Toronto (2019). 'Bots at the Gate: A Human Rights Analysis of Automated Decision Making in Canada's Immigration and Refugee System'. https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

"Determining legal limits, impermissable uses or 'red-lines' for AI applications is a necessary step for a people centred, fundamental rights-based AI. Fundamental rights are upheld by focusing on the prevention of systematic harm to individuals, communities and societites"

- European Digital Rights