ALGORITHM WATCH

Submission by AlgorithmWatch

On the European Commission's "White Paper on Artificial Intelligence – a European approach to excellence and trust"

June 2020

AlgorithmWatch welcomes the European Commission's efforts to develop a regulatory framework which is based on European values and in full respect of fundamental rights. We urge the Commission to put public interest, the protection of individual rights, non-discrimination, and equal access to resources and participation at the core of any strategy on Artificial Intelligence; consequently, we call for the establishment of rigorous transparency mechanisms which allow for public scrutiny and contestation, including (1) public registers on ADM systems; (2) remedies for contestation; (3) independent centers of expertise on AI/ADM; and (4) robust, legally-binding data access frameworks to support and enable public interest research.

Our Perspective

AlgorithmWatch (AW) is a non-profit research and advocacy organisation that is committed to watch, unpack and analyze automated decision making (ADM)[1] systems and their impact on society. While the prudent use of ADM systems can benefit individuals and communities, they come with great risks. Therefore, guided by the principles of the protection of fundamental rights, especially non-discrimination, equality and freedom of expression, we consider it crucial to hold ADM systems accountable to

[1] We define algorithmic decision-making systems (ADMS) to encompass
- the design procedures to gather data,
- the collection of data,
- the development of algorithms to analyse the data,
- the interpretation of the results of this analysis based on a human-defined interpretation model, and
- to act automatically based on the interpretation as determined in a human-defined decision making model.

All ADMS are socio-technical systems embedded in societal contexts that need to be taken into account when assessing the implications of their use. E.g. in the - very unlikely - event that a facial recognition system could be trained to accurately identify citizens, it would still be unacceptable if the system is used to indiscriminately mass-surveillance entire populations as the basis for further action.

democratic control. The use of ADM systems that significantly affect individual and collective rights must not only be made public in clear and accessible ways, individuals must also be able to understand how decisions are reached, and, given appropriate options exist, how to contest them if deemed necessary. Our work is dedicated to enabling citizens to better understand ADM processes in order to take informed decisions and action. Hereby, we aim at contributing to a fair and inclusive society and at maximizing the benefit of ADM systems for society at large.

Our Key Positions & Recommendations

In our report Automating Society[2] we have shown how automated decision-making systems are shaping people's daily life in the European Union (EU). Intended to make processes more efficient, **many ADM systems have highly problematic consequences** - they limit people's access to participation, to public goods and services, and infringe fundamental rights[3]. A central challenge in detecting and correcting the outcome of ADM systems is their opaque character. Due to both a lack of adequate and consistent regulation and a lack of knowledge to assess such systems, **most ADM systems remain "black boxes" to the public, inhibiting critical contestation from the outset.**

We therefore welcome the European Commission's efforts to develop a coherent regulatory framework which is based on European values and in full respect of fundamental rights.

I. Prioritizing the public interest

While the European Commission highlights the potential risks of Artificial Intelligence (AI) in the very first paragraph of its White Paper, we share concerns raised by the human rights community that the White Paper's overall narrative suggests a worrisome reversal of EU priorities, putting global competitiveness ahead of the protection of fundamental rights[4]. We **urge the European Commission (EC) to clearly prioritize the public interest**, the protection of individual rights, non-discrimination, and equal access to resources and participation, **as the core concern of any future strategy/ regulatory framework on Artificial Intelligence.**

Consequently, we stress that **any risk-based approach to AI must center on the potential harm for the individual as well as for society** at large, and must **follow clear and transparent rules.** The risk-based approach proposed in the White Paper lacks both clarity and transparency. Moreover, the current approach runs the risk of incorrectly categorizing ADM systems which may harm an individual, impact an individual's access to resources, or concern their participation in society as *low risk*, with the consequence that regulative measures would not apply. To give two examples: *VioGén*, an ADM system to forecast gender-based violence, and *Ghostwriter*, an application to detect exam fraud,

[2] The 2nd volume "Automating Society 2020" is forthcoming. For the German context specifically, see "Atlas of Automation".

[3] See also "Digital technology, social protection and human rights: Report" by the United Nations' Special Rapporteur on extreme poverty and human rights.

[4] Eg. Access Now (2020): Access Now's submission to the Consultation on the "White Paper on Artificial Intelligence - a European approach to excellence and trust".

would most likely fall between the cracks of regulation, even though they come with tremendous risks[5] – simply because they do not meet the sector criterion. We therefore **ask the Commission to revise and realign the risk-based approach and work towards clear and coherent criteria** as to when AI/an ADM system has a significant impact on an individual, a specific group or society at large. As a guidance, we suggest to consider the following aspects: (a) the potential impact an ADM system has on people's life chances and social participation (e.g. e-recruiting systems as opposed to traffic lights); (b) the number of individuals concerned by a decision taken by an ADM system; and (c) whether or not decisions are based on correlation or causality; correlation-based decisions obviously raise more concerns.

II. Establishing rigorous transparency mechanisms and remedies for contestation

Developing risk-based assessments as part of a regulative framework on AI requires that we know what we are dealing with in the first place. We stress that the use of ADM systems that significantly affect individual and collective rights must not only be made public in clear and accessible terms, individuals must also be able to understand how decisions are reached and how to contest and correct them if deemed necessary.

Below we outline **four key recommendations** which we feel are necessary steps in order **to enhance public scrutiny and democratic control.**

1. Establish public registers for ADM systems used within the public sector

Without the ability to know whether AI/ ADM systems are being deployed, all other efforts for the reconciliation of fundamental rights and AI/ADM systems are doomed to fail. AlgorithmWatch and Access Now therefore jointly call for a mandatory disclosure scheme for AI/ADM systems deployed in the public sector. We **ask for legislation to be enacted at the EU level** to mandate that member states **establish public registers of ADM systems used by the public sector.** Such registers should be used to make public the results of Algorithmic Impact Assessments (AIA)/ Human Rights Impact Assessments (HRIA) undertaken by public authorities[6]. They should come with the legal obligation for those responsible for the ADM system to disclose and document the purpose of the system, an explanation of the model (logic involved) and the information on who developed the system. This **information has to be made available in an easily-readable and accessible manner**, including structured digital data based on a standardized protocol. Moreover, we agree with the Council of Europe's Commissioner for Human Rights that an individual who has been subject to a decision by a public authority that is solely or significantly informed by the output of an AI system should be notified without delay[7].

Whereas disclosure schemes on ADM systems should be mandatory for the public sector in all cases, these **transparency requirements should also apply to the use of**

[5] For further details on VioGgén see https://algorithmwatch.org/en/story/viogen-algorithm-gender-violence/. VioGgén and Ghostwriter will both be covered by our upcoming report "Automating Society 2020".

[6] For details, see Council of Europe (2019): Unboxing Artificial Intelligence. 10 steps to protect Human Rights. p. 7

[7] Ibid., p. 10

ADM systems by private entities, when an AI/ADM system has a significant impact on an individual, a specific group or society at large (see above in section I. our guidance on potential criteria).

2. Strengthen remedies for contestation

When **subjected to a decision made with the assistance of an ADM system, individuals must be able to retrieve all relevant information** about what happened and about what has led to the outcome of the decision. This transparency requirement is crucial to be able to contest the automated decision legally, assuming such a basis exists (e.g. in anti-discrimination law). We therefore propose to strengthen people's right to inspect ADM systems, documentation and protocols. Complementary to this, **individuals must have accessible, affordable and effective remedies at hand** to guarantee an impartial review of their claims.

3. Establish independent centers of expertise on AI/ADM

AlgorithmWatch and Access Now jointly call for the establishment of **independent centers of expertise on AI/ADM** at national level to monitor, assess, conduct research, report on, and provide advice to government and industry in coordination with regulators, civil society, and academia about the societal and human rights implications of the use of AI/ADM systems. The overall role of these centers is to create a meaningful accountability system.

As independent statutory bodies the centers of expertise would have a **central role in coordinating policy development and national strategies** relating to AI, and in **helping to build the capacity** of existing regulators, government and industry bodies to respond to the increased use of AI systems.

These centers should **not have regulatory powers, but provide essential expertise** on how to protect individual human rights and prevent collective and societal harm. They should, for instance, **support small and medium-sized enterprises (SMEs) in fulfilling their obligations under human rights due diligence**, including conducting AIA/HRIA and in registering AI/ADM systems in the public register discussed above.

The national centers of expertise should **involve civil society organisations, stakeholder groups and existing enforcement bodies** such as DPAs and National Human Rights Bodies to benefit all aspects of the ecosystem and build trust, transparency and cooperation between all actors.

4. Introduce legally-binding data access frameworks to support and enable public interest research

Holding ADM systems accountable not only requires disclosing information about a system's purpose, logic and creator, as well as the ability to thoroughly analyse and test a system's in- and outputs, but also to **make training data and data results accessible** to independent researchers, journalists and civil society organisations for public interest research. We therefore suggest to **introduce robust, legally-binding data access**

frameworks, focused explicitly on supporting and enabling public interest research and in full respect of data protection and privacy law[8].

Learning from existing best practices[9] at the national and EU levels, such **tiered frameworks should include systems of sanctions, checks and balances as well as regular reviews.** As private data sharing partnerships have illustrated, there are legitimate concerns regarding user privacy and the possible de-anonymization of certain kinds of data. Policymakers should learn from health data sharing frameworks[10] to **facilitate privileged access to certain kinds of more granular data, while ensuring that personal data is adequately protected** (e.g. through secure operating environments). To do so, they should see that governance frameworks integrate the perspectives of multiple competent authorities (e.g. data protection authorities, cyber-security agencies, media regulators).

In the **specific context of platform governance**, we recommend to go beyond databases with information on content moderation and takedown decisions. The aforementioned access frameworks should rather compel platforms to produce high quality, workable, public research APIs and archives with complete, consistent and credible data on key subjects including moderation, advertising, as well as curation and recommender systems.

III. Further recommendations we support

Ban ADM systems that facilitate mass surveillance

ADM systems that are based on biometric technologies, including facial recognition, pose a particular, **serious threat to the public interest and fundamental rights as they clear the path to indiscriminate mass surveillance**. In a draft of the Commission's White Paper on AI, authors considered prohibiting the "use of facial recognition technology by private or public actors in public spaces [...] for a definite period (e.g. 3–5 years) during which a sound methodology for assessing the impacts of this technology and possible risk management measures could be identified and developed." We regret that this idea was dropped from the final version and support European Digital Rights' (EDRi) call on the European Commission and the EU Member States to comprehensively stop all biometric processing in public spaces that could amount to mass surveillance[11] and demand that the Commission's AI strategy clearly and uncompromisingly bans such ADM systems and other applications from the outset.

[8] Our recommendations are mainly informed by our current research project "Governing Platforms", and has been developed on the base of three studies, conducted by the Mainz Media Institute (accessible here) and the Institute for Information Law (IViR) at the University of Amsterdam ("Operationalizing Research Access in Platform Governance: What to learn from other industries?", forthcoming).

[9] See e.g. E-PRTR case study in forthcoming "Operationalizing Research Access in Platform Governance: What to learn from other industries?"

[10] See e.g. Findata case study in forthcoming "Operationalizing Research Access in Platform Governance: What to learn from other industries?"

[11] EDRi (2020): Ban Biometric Mass Surveillance. A set of fundamental rights demands for the European Commission.