Consultation on the White Paper on AI - a European approach

Submission

Fieke Jansen

June 2020

Table of Contents

1. Introduction

Data Justice Lab's Fieke Jansen in her capacity as PhD candidate and Mozilla public policy fellow is submitting this response to stress the need to put human rights and environmental well-being at the centre of Europe's Trustworthy AI approach. The Data Justice Lab is a space for research and collaboration at Cardiff University's School of Journalism, Media and Culture (JOMEC). It seeks to advance a research agenda that examines the intricate relationship between datafication and social justice, highlighting the politics and impacts of data-driven processes and big data.

We are in alignment with the EU approach of human-centric AI, where the development and use of AI technologies should benefit people, democracy, the planet and our economy. However, we are worried about the notion of promoting indiscriminate uptake of AI across the public and private sector as proposed throughout the White Paper. The EU should be mindful of the fact that the benefits of AI technologies are at this point merely a promise, that builds on generic and unsubstantiated claims about its potential. When we look at the history of technology, similar promises have carved out unregulated spaces for unfettered innovation and 'progress', which in turn has economically benefited a small number of people, while posing significant risks to how we organize our democratic society and has disproportionately harmed under-served and underprivileged communities.

As such a European approach should avoid technological solutionism, which is characterized by the promotion of indiscriminate uptake of AI technologies and the fact that not societal well-being but the fear of missing out and the desire to 'remain in the race' with the US and China is shaping the investment in and regulation of AI technologies. A European approach to trustworthy AI should start from its core values which include the responsibility to prevent unnecessary harm to the natural person, society and the environment and ensure public trust and confidence in appropriate AI

technologies, public authorities and governments. For this, the Commission must enforce the highest human rights and environmental compliance standards for the investment in, development and use of AI technologies in Europe. As part of the consultation process on the European Commission's 'White Paper on Artificial Intelligence - A European Approach,' this document outlining our four key recommendations:

A) Safeguarding environmental and human rights in AI research and development funding;
B) Implement a rights-based approach, mandatory human rights impact assessment and the development of a clear risk framework;
C) Regulating AI; focus on updating existing legislation, ensuring access to justice and invest in effective enforcement;
D) Commitment to diversity and inclusion.

2. Safeguarding environmental and human rights in AI funding

Ensuring that AI technologies respect human rights, promote human agency and operate in the best interest for our societal and environmental well-being requires careful scrutiny of the entire AI supply chain, from the extractive industry for hardware components, to research and development investments, to the procurement processes, and the design, development and ongoing deployment of AI systems. In the White Paper the Commission expressed its commitment to harness the capacity of the EU to invest in next-generation technologies and infrastructure, however, it did not articulate how it would align these investments with Europe's commitment to human rights and the environment. We urge the commission to articulate an investment approach that:

- align research and development and investments with regulatory requirements for trustworthy AI;
- financing research to understand the harms of AI technologies;

- invest in the creation of a market of trustworthy AI.

Align research and development finances with regulatory requirements for trustworthy AI. Public authorities wear several hats, at times they are seed funders, regulatory bodies, enforcement authorities, and also clients of computational providers. It is important to recognize these different roles in the AI supply chain and ensure that human rights and environmental well-being are ensured throughout. It is a waste of public resources if the security research funded under Horizon 2020 and in the future Horizon Europe will not meet the proposed enforcement criteria as they negatively impact fundamental human rights and cause additional unintended harm to the environment. Take for example Horizon 2020, in recent years it has financed security projects that develop highly questionable and controversial AI technologies, like biometric identification, emotion recognition and lie detection. It is unlikely that these projects will comply with future legislation that governs AI technologies. As such regulatory measures, norms and risk assessments that apply to the use of AI technologies should also apply to all European public research and development funding.

Financing research to understand the harms of AI technologies: the White Paper proposes a risk framework to assess AI technologies, yet the added value and the harms of these systems are not well understood. What is needed is a more thorough understanding of on the one hand concrete positive use cases of AI and on the other hand the harms to the individual, collective, society and environment. As such, we ask the Commission to explicitly articulate a commitment to invest in research that explores the possible positive and negative impact of specific AI technologies:

- For environmental well-being, investments are needed to understand both the extractive and carbon footprint of AI technologies. By extractive, we mean the environmental and human costs in the mining and production process of the

hardware component needed to run AI technologies. By carbon footprint, we mean the energy consumption for the training, testing, and running of specific AI models.

- For fundamental human rights research needs to be supported that critically examines the development and deployment of these AI technologies to understand if these are not deliberately or accidentally perpetuating racial, gender and other societal and labour force inequalities. This not only requires investment in research that examines potential bias in AI technologies, but significantly more research needs to be done into the decision-making and practices of buying and deploying it, and the lived experiences of those impacted by it.

Invest in the creation of a market of trustworthy AI through clearly articulating a financial commitment to research and development of privacy-, human rights-, and environment- by design AI technologies. Take for example voice recognition technology, this research shows that training large AI models emits more than 626,000 pounds of carbon dioxide equivalent. At the moment most voice recognition engines are built on pre-trained models that are offered by big computational companies like Google. The EU should invest in European universities, research centres and companies to experiment with the creation of alternative pre-trained voice models to decrease the dependency on the large foreign computational companies while putting fundamental human rights and environmental rights at the core of the product. The deeper in the AI stack privacy-, human rights-, and environment- by design are build the bigger the potential impact. This requires specific research and development calls in Horizon Europe for the creation of trustworthy AI core infrastructures, libraries and models.

3. A rights-based approach and comprehensive risk framework

The White Paper highlights the fundamental challenge of AI, that in the end, this technology should be self-learning, ever-changing, machine-driven and sector agnostic.

As such we should recognize that our current knowledge of AI systems will merely be a snapshot of its time. While historic development of technology offers insights into the social challenges and harms that emerge with the introduction of new technologies, it is impossible to anticipate all its applications, biases, risks and harms. As such any approach needs to be holistic, problem-oriented, open, regularly reviewed and implemented through a human rights-based approach, to address current and future harms to the individual, the collective, society and the environment.

Clear articulation of the theory of risk used by the EU. While risk is the proposed mechanism to determine if AI technologies should be regulated or not, the White Paper offers no information about how the theory of risk, that supports the low and high-risk dichotomy, is constructed. There is no clear articulation if and when human rights, societal, and environmental risks are mandatory elements of a risk assessment. It also fails to address how the probability of risks is to be assessed when unknown unknowns of technology have in the past produced severe social harms, a clear example is the proliferation of fake news on social media platforms. Nor is there an explanation on how the potential harms to the individual and communities are taken into account. We, therefore, encourage the Commission to clearly articulate and publish a theory of risk which should include a risk framework, methodology, and the underlying assumptions on which it is based.

Need for a more comprehensive understanding of risk: the current binary approach to risk, it is either low or high harm, which is determined through a classification of sector and riskiness of the product, has several fundamental flaws. First, it assumes that one can make a clear distinction between high and low risk, and ignores the fact that seemingly low-risk applications in one sector can become high risk when they are used in another sector. Research by the Data Justice Lab showed how consumer credit reporting agency Experian and their geodemographic segmentation tool Mosaic is used in the public sector as a way to analyse populations, from fraud detection to law

enforcement. As such any risk framework needs to account for the tendency of technology developed in one sector to be applied in another.

The binary of low or high risk also assumes that there will be no AI technology that is considered to be 'too risky' or incompatible with the Charter of Fundamental Rights, and where mitigation of risks alone is not enough. Risk is constructed from identifying potential problems with a specific AI technology and assessing the impact and probability of this will occur. As such the EU should develop a more nuanced and comprehensive framework that explicitly formulate criteria that distinguish between 1) high harm and should be banned, 2) high harm, but only allowed under strict control, 3) medium harm, only allowed with transparency, public deliberation and oversight ex-ante and ex-post 4) low harm, allowed with ex-post oversight. This more comprehensive approach offers scope for measures and safeguards beyond mitigating risk and allows for the articulation of red lines to protect those areas where AI technologies are deemed incompatible with the Charter of Fundamental Rights. Once there is a clear theory of risk that includes red lines, measures such as bans or moratoriums can be explored.

The context in which AI technologies are deployed needs to be taken into account. Not only is technology not neutral, but it is also deployed in a specific context with specific agendas. In understanding harms it is imperative to understand the context in which AI technologies are deployed, i.e. the who, what (propose), why, where, when and directed at which group. In the context of welfare fraud detection algorithms, we have seen that these technologies are disproportionately targeting the poor and similar techniques are sub-sequentially not applied to expose tax evasion by the rich. In the context of labour targeted advertising and assessment algorithms have been known to negatively impact the opportunities of women and people of colour. As such the context of AI technologies needs to be prominently taken into account when assessing the potential risks and harms.

The need for a human rights- and environmental-based approach throughout the AI supply chain and life cycle. To ensure that Fundamental rights and environmental protections are safeguarded by all European nation-states, we support Access Now proposition that for all applications in all domains the burden of proof should be on the entity wanting to invest in, develop or deploy the AI system to demonstrate that it does not violate human rights nor put additional burden on the environment via a human rights impact assessment (HRIA), an environmental impact assessment and a mandatory disclosure scheme. The process by which an AI system is determined to be high, medium or low risk must be reliable, verifiable, trustworthy, contestable and should be reassessed throughout the system's life cycle.

Highest justification for public authorities, critical consumer products and critical infrastructure. Here we recognize that all AI technologies and their implementation are not equal, some use cases will have more impact than others. For example, the risk and harms related to the implementation of AI technologies in critical infrastructures (electricity, water management, communication and internet), essential services (welfare, police, health care, borders and transportation), public spaces (street, Facebook), crucial consumer products (access to finance and credits scoring) and (training) AI models on which third parties build applications will be far higher than of others. As such these will need high levels of scrutiny and should be subjected to a high justification threshold and independent external review for deployment.

4. Regulating AI; legislation, access to justice and enforcement

Clearly defining the problem, the Commission highlighted existing EU legislation that applies to AI technologies, i.e. European legislation on fundamental rights (e.g. data protection, privacy, non-discrimination), consumer protection, and product safety and liability rules. It remains unclear where these existing regulations fall short in the

governance of AI technologies and where updating or stricter enforcement of those existing legislation is sufficient and for which areas new AI legislation needs to be created. As such we encourage the Commission to more clearly articulate what is unique to AI technologies that require the updating of existing regulation and the creation of a new legal framework.

Access to justice. When AI technologies will become more prevalent throughout society the key question is how will Europe ensure access to justice for all. We welcome the Commission commitment to ensure effective judicial remedy and redress for parties negatively affected by AI systems. However, a fundamental problem with effective judicial redress is 1) limited knowledge that one has been subjected to and impacted by AI technologies, 2) difficult to prove harms caused by AI technologies, and 3) should be an avenue for collective action in cases where individual harms are low but systemic harm is high, think of harms to a collective group that share specific attributes, such as ethnicity, gender, religion, low-income and others. As such we urge the Commission to strengthen legal mechanisms by which individuals and collectives are informed about the AI technologies that impact their lives and provide avenues for judicial redress in the public interest (i.e. not having to demonstrate individual harm) or on the basis that they may have been subject to the system.

Publishing of datasheets and statistics to enable redress and public oversight. Information on the use of these technologies should be mandatory for both the public and private sector. Publicly accessible datasheets[1] should inform the public about the context and the purpose in which AI technologies are being used. These datasheets should provide information on the specific AI technology, the purpose of its deployment, when it started, who is providing these services, who is responsible for it, and more. Publicly available statistics should provide information about the use of these systems,

[1] Datasheets for AI systems build on the datasheets for dataset proposition of Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford http://jamiemorgenstern.com/papers/datasheet.pdf

where it is being deployed, if it is impacting a natural person (if yes, information should be provided on who is impacted, if they belong to a racialized communities, specific social-economic demographic, age group, or gender), how much carbon is being emitted while running these AI technologies and more. The nature and responsibility of public authorities require additional transparency efforts, as such, each nation-state should create a centralized national public repository of datasheets and statistics. Private actors should offer the datasheets and statistics visibly on their website.

Notification to ensure access to justice. Public authorities and those entities delivering critical consumer products and critical social infrastructures[2] have the obligation to notify an individual when a decision about them is taken by an AI system without meaningful human intervention. Meaningful human intervention, in this case, refers to the fact that a person is responsible for the decision impacting individuals, as such they need to understands how the AI technology constructs the output, feels they have the ability to challenge or contradict the decision and is held to account for when a decision goes wrong. A meaningful notification obligation will not only entail information that a person has been subjected to a decision made by an AI technology, but it should also offer access to information about when, how and why it was used, and offer a contact person to gain more information or challenge the decision.

Avenues of judicial redress. How AI technologies will be used in the future and what challenges may arise is still very much unclear, as such it is crucial that any updated or new legislation will include different avenues for judicial redress. Taking a cue from the GDPR, what is needed are judicial avenues for 1) representative actions where a group of individuals assign their rights to remedy to an organisation or body, 2) representative actions where no individual assigns their rights, but an organisation or body takes action on behalf of a group of people not related to them, and 3) class action where

[2] critical infrastructures (electricity, water management, communication and internet), essential services (welfare, police, health care, borders, and transportation), public spaces (street, Facebook), and crucial consumer products (access to finance and credits scoring).

individuals, whose claims are sufficiently similar to others, define themselves as a group and sign up to take on a court case against a specific AI implementation. As such we urge the Commission to first understand existing possibilities for judicial redress in its review of consumer rights, anti-discrimination, fundamental human rights and environmental regulation across the member states. And include the different avenues for judicial redress in the update and stricter enforcement of existing regulation, and in any new regulatory framework that will be developed to govern AI technologies.

Ban the use of biometric technologies in public spaces. Fundamentally the use of biometric technologies for untargeted mass processing of personal data in public space and communication infrastructures holds numerous issues. It facilitates mass surveillance of populations which in itself is a violation of our fundamental human rights, including privacy, data protection, equality, freedom of expression and information, freedom of assembly and association and more. For the data processing aspect, it processes highly sensitive personal data in a manner that is often unknown to the individual. These highly invasive biometric technologies are being built and tested across Europe under the guise of catching terrorist, but on closer look, these technologies are assisting police to identify pickpockets and protesters, and schools in France proposed to use it to identify students. As such these AI technologies severely threatening our fundamental rights for very mundane tasks that do not need AI technologies. We, therefore, support the call of the European digital rights movements to ban on the use of biometric technologies in public and semi-public spaces, until the impact and harms are fully understood and there are clear legal frameworks that guide them.

Enforcement authorities. Legal frameworks are only as strong as their enforcement. In the White Paper, the Commission fails to propose how it will ensure effective enforcement of existing and new legislation to govern AI technologies. We propose not to create a new and centralized body that should ensure that AI systems comply to the

EU regulatory frameworks, instead, the EU should enable current topical authorities in the field of environment, data protection, anti-discrimination and consumer protection to incorporate AI in their work processes. For example, this would mean that the existing environmental enforcement authorities will be tasked with the responsibility to ensure that the AI supply chain complies with environmental rules and regulation and has a minimum impact on the environment. And for the enforcement of data protection requirements, the national DPA's should be in the lead.

AI is a cross-cutting issue and as such is subjected to a range of legislation and enforcement authorities. With a distributive enforcement approach, there is a significant risk that certain AI technologies will lack proper governance, as none of the enforcement authority feels they are the ultimate responsible party. Here the EU should engage in a mapping of authorities and their mandate to gain a better understanding of how enforcement is currently distributed and identify ambiguous areas. In addition, enforcement authorities should articulate their organizational needs to integrate oversight on AI systems in their scope of work. For example, a current challenge for existing enforcement authorities is that there is not enough social-technical expertise to engage with AI technologies. This report highlights the lack of technical expertise even within Data Protection Authorities let alone other enforcement bodies. In their analysis, the Commission should take into account the brain-drain of critical AI expertise as a result of Brexit, at the moment the UK is leading the institutionalization of AI expertise in organizations such as the ICO, Allan Turing Institute, Ada Lovelace Institute, and the Center for Data, Ethics and Innovation. Before investing in and promoting new AI technologies, the Commission should articulate a clear plan that will ensure adequate national enforcement.

5. Commitment to diversity and inclusion

Commitment to diversity and inclusion should be reflected in its funding criteria, support to SME and regulatory approach. Developing inclusive technologies that work for everyone, no matter their ethnicity, genders, disability or class, requires diversity in the teams that build them. This is not merely a matter of upskilling the workforce with an emphasis on women like the White Paper suggests. Reports about Silicon Valley show that the male-dominated technology sectors widely marginalizes people of colour and denies them career opportunities, in addition, there have been major scandals around sexual harassment, misogyny and impunity. Despite promises to diversify their workforce none of the big technology companies has made much progress in the last 5 years, and the rate at which women and people of colour leave the technology industry is far higher than that of their white male counterparts. Diversity is key for building trustworthy AI technologies that benefit all and this requires more than skill-building, it requires a change in culture and mindset. An EU commitment to diversity and inclusion as such should be reflected in its funding criteria, support to SME and regulatory approach.

Meaningful inclusion of communities and critical experts in EU decision making around AI technologies. AI technologies are not merely technological artefacts. These technologies are integrated into social, economic and political domains. As such the investment in, development and regulation of AI technologies require more than technical scrutiny. For the development of trustworthy AI technologies the EU must commit to a more holistic understanding of technology. Decisions about how public money is invested in AI technologies should be made by a consortium of people, which include the communities impacted by them and the meaningful involvement of among other human rights, welfare, labour rights, anti-discrimination efforts.

For more information, please contact:

Fieke Jansen, PhD candidate Data Justice Lab and Mozilla public policy fellow, jansenf@cardiff.ac.uk