AI NOW

AI Now Institute, New York University

Submission to the European Commission on "White Paper on AI - A European Approach"

June 14 2020

Written comments[1] of:
Amba Kak, Director of Global Programs, AI Now
Rashida Richardson, Director of Policy Research, AI Now
Roel Dobbe, Postdoctoral Researcher, AI Now

As part of the consultation process on the European Commission's 'White Paper on Artificial Intelligence - A European Approach,' (hereafter, "the White Paper").[2] AI Now Institute submits this document highlighting our key responses and recommendations.

AI Now Institute is a research institute at New York University (NYU) dedicated to studying the social implications of artificial intelligence and algorithmic technologies (AI). Our work examines the rapid proliferation of AI systems through social domains such as criminal justice, healthcare, employment, and education. We focus on concerns in the areas of bias and inclusion, safety and critical infrastructure, rights and liberties, and labor. As we identify problems in each of these spaces, we work to address them through robust research, community engagement, and key policy interventions.

In this document, we respond to certain key elements of the White Paper including:

I. EU strategy to "promote the uptake of AI"
II. Future regulatory framework for AI
    a. Relationship to existing legal frameworks
    b. Definition, scope, and risk-based assessment
    c. Impact Assessment frameworks
    d. Documentation, transparency, and information provision
    e. Human oversight
    f. Biometric & affect recognition
    g. Product safety & liability
III. Government use of AI and public procurement guidelines

We have also attached **(in annexure)** a position paper prepared and submitted jointly with the City of Amsterdam, Helsinki, Nesta and Mozilla Foundation, on strengthening accountability through public procurement contracts.

---

[1] The authors extend gratitude to Meredith Whittaker and Jason Schultz for their feedback and insights while drafting these comments and Luke Strathmann for editorial assistance.
[2] We also respond to certain recommendations on the Commission's 'Report on the Safety and Liability implications of Artificial Intelligence, the Internet of Things and robotics'.

AINOW

Summary of recommendations:

1. European industrial strategy related to AI should be informed and guided by the social impacts of these technologies. Rather than indiscriminately promoting the uptake of "more AI," the Commission should ground its strategy in evidence of the social and environmental impacts of these technologies, whether they work, and which communities are likely to benefit more or bear the risk of bias, exploitation and other harms. It should not be presumed that any AI technology is either needed or beneficial until proven.

2. European AI Strategy should prioritize the environmental implications of AI systems and digital infrastructure. Regulation should mandate transparency about the carbon footprint of cloud services to allow any organization to calculate their digital carbon footprint and that of AI systems. In addition, the use of AI for fossil fuel exploration and production should be prohibited.

3. Rather than the risk-based approach proposed by the Commission, the scope of regulation should be determined based on the nature and impact of the AI system, irrespective of the sector in which it is used. The "legal or similarly significant effect" standard used in Article 22 of the GDPR could be under-inclusive, so we recommend including more descriptive criteria, for example that all AI systems that have "an *impact on opportunities, access to resources, preservation of liberties, legal rights, or ongoing safety of individuals, groups, or communities"* be included within the scope of regulation.

4. Algorithmic Impact Assessments (AIA) should be required before an AI system is deployed to inform any decision of whether (and how) such systems are used. Public sector use of AI that has been demonstrated to produce biased outcomes and harms including but not limited to, predictive policing, child welfare predictive analytics, pre-trial risk assessment, and public benefits decision systems, should be considered for moratoria or other restrictions.

5. In order to have meaningful and equitable public participation in AIAs, government agencies should consider proactive measures to address financial and capacity barriers, including financial remuneration to community members appointed to serve on decision-making bodies, and inclusive educational opportunities.

6. AIAs should evaluate whether the AI system creates the conditions and capacity for meaningful human oversight, which includes oversight by those who are directly impacted by these systems.

Regulation on human oversight over AI systems should not be based on a binary classification between those systems that are "solely automated" versus all others. Where such oversight cannot be meaningful because of technical features or knowledge and capacity constraints, the use of AI systems should be entirely restricted, especially in sensitive social domains.

7. Regulation should ensure that external researchers and auditors have access to AI systems in order to understand their workings, as well as the design choices and incentives that informed their development and commercialization, and to engage the public and impacted communities in the process. Meaningful access includes making software toolchains and APIs open to auditing by third parties. It should also involve a review of commercial confidentiality, IP, and access to information laws, that operate to hinder accountability and protect corporate secrecy.

8. Policymakers should impose moratoriums on all uses of facial recognition in sensitive social and political domains, including law enforcement use, education, and employment. Lawmakers must supplement moratoriums with (1) transparency requirements that allow researchers, policymakers, and communities to assess and understand the best possible approach to restricting and regulating facial recognition; and (2) protections that provide the communities on whom such technologies are used with the power to make their own evaluations and rejections of its deployment.

   Given its contested scientific foundations and evidence of amplifying racial and gender bias, affect recognition technology should be banned for all important decisions that impact people's lives and access to opportunities.

9. Product safety regulations are insufficient to address the safety risks of AI systems in critical infrastructure and sensitive domains. Regulation should focus instead on system safety, and should draw on standards from safety-critical domains such as aerospace and nuclear engineering, which have a long history of working with sophisticated automation. As current AI safety research is too narrowly focused on technological fixes, the Commission should also invest in research and development that addresses all crucial dimensions of AI system safety, both technical and empirical.

10. Government agencies should not procure or use AI that are shielded from independent validation, public review, or legal challenges because of trade-secret or confidentiality claims. In addition, procurement contracts should 1) Include specific waivers to trade secrecy; 2) provide government staff with training modules by vendors to help understand the systems and develo public-education materials; 3) restrict broad indemnity clauses; 4) require validation studies; and 5) mandate an open, competitive bidding process.

I. EU INDUSTRIAL STRATEGY TO "PROMOTE THE UPTAKE OF AI" (Section 1)

In the White Paper, the Commission commits to both an *enabling* policy framework to promote development of the European AI industry, as well as a *restrictive* regulatory framework in accordance with fundamental rights. We appreciate that these two conversations are happening alongside rather than divorced from one another as the social and political impacts of AI systems must inform AI industrial strategy.

However, the Commission's justification for introducing regulatory frameworks is that it will eventually lead to greater uptake of AI (*"will give citizens the confidence to take up AI applications and companies the legal certainty to innovate using AI")*. This view of regulation as merely instrumental to increased uptake of AI is misplaced. The creation of regulatory frameworks should determine which AI applications and technological futures are worth pursuing in the first place, which are impermissible, and to and to enact modes of democratic decision making capable of steering such decisions.

The White Paper puts particular emphasis on boosting small and medium-sized AI enterprises. The concentration of market power in the AI industry has negative implications for society and poses a major challenge for governance. An important dynamic that should be accounted for, however, is that control over data and computational infrastructure rests with a handful of large tech companies and so, in most cases, these smaller firms have no choice but to license their computational infrastructure from the large players. Without tackling this dependence, it might be the case that SMEs do not offer a real alternative to the large players, but are interconnected with them, and in some ways serve to further entrench their concentration of power.

The Commission should be wary of encouraging competition against purely quantitative metrics of "more AI" for the sake of it, a logic that is commonly referred to in narratives around the so-called global "AI arms race".[3] This apparent competition serves to ramp up AI development and deployment but is also used to push back against calls for slower, more intentional development and stronger regulatory protections. More broadly, this view of progress tends to see all calls for restraint, reflection, and regulation as a strategic disadvantage to national or regional interest. It turns accountability into a barrier to progress and suppresses calls for oversight.

For these reasons, we would urge the European Commission to demonstrate global leadership by going beyond measuring industrial progress solely in terms of AI adoption, and instead focus on

[3] AI Now 2019 Report, "China Arms Race Narrative", at page 42
https://ainowinstitute.org/AI_Now_2019_Report.pdf

an evidence-led approach that investigates the social impacts of these technologies, whether they work (and how they are improvements on existing tools or modes), and which communities are likely to bear the risk of bias and other harms they create. Given the mounting evidence of harms caused due to AI systems being applied in sensitive social contexts, these questions are urgent.

II. A FUTURE REGULATORY FRAMEWORK FOR AI (Section 5)

RELATIONSHIP TO EXISTING DATA PROTECTION AND OTHER LEGAL FRAMEWORKS (Section 5a)

As the Commission considers additional regulatory requirements for AI systems, it is important to emphasize that any new protections must be complementary to existing legal frameworks that apply to AI systems, including data protection and non-discrimination law, and which should continue to be strengthened rather than eroded. Taking the GDPR in particular, there are multiple points at which the regulation applies to data-related activities within AI systems and puts in place critical safeguards such as collection and purpose limitation principles, rights to data access and against solely automated decision making, and data protection impact assessments. These should be dynamically interpreted in the context of AI applications.[4]

In the long run, however, the data protection framework might come up against its own limitations when it comes to some of the most pernicious uses of AI, especially to the extent it proceeds from an individualistic rather than a collective understanding of privacy and harm. The personal data threshold, as discussed, routinely breaks down in the context of AI systems that often make sensitive inferences about people (or the communities they are part of) based on discrete non-personal data categories. The personal data threshold might be inappropriate when the goal is to minimize harms that emanate from algorithmic profiling, which is often on the basis of classes, aggregates, and patterns.[5] As demonstrated by the recently concluded SyRI case in

[4] *See* UK ICO, Guidance on the AI auditing framework: Draft for Consultation 2020 https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf; Reuben Binns (UK ICO), Enabling access, erasure, and rectification rights in AI systems, October 15 2019 https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/; Amba Kak and Rashida Richardson, AI Now Submission to the Office of the Privacy Commissioner of Canada, February 2020, https://ainowinstitute.org/ainow-comments-to-canadian-office-of-the-privacy-commissioner.html.

[5] *See* Kate Crawford & Jason Schultz, Big Data and Due Process: Toward a Framework to Redress Predictive Privacy, Boston College Law Review 2014 https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr; Taylor, L., Floridi, L., van der Sloot, B. eds. Group Privacy: new challenges of data technologies, 2017.

the Netherlands where algorithms were being used to detect welfare fraud[6] and recent research[7] on predictive policing tools,[8] AI systems are often being applied with harmful collective impacts on the rights to public benefits, personal autonomy, access to opportunities, and privacy of the most vulnerable communities.

DEFINITION, SCOPE, AND RISK-BASED ASSESSMENT (Section 5C)

The Commission does not clearly define what counts as "AI " in the White Paper, but proposes a risk-based approach to determining the kinds of AI applications should be subject to additional regulatory requirements (where high risk applications fall within scope of these additional requirements, whereas lower risk would be exempt).

We would caution against building a regulatory framework based on a rigid threshold between high and low risk applications. There are already lessons in the implementation of the GDPR where, as pointed out by organizations like EDRi[9] and Access Now,[10] systems based on self-assessment of risk have allowed for significant loopholes and insufficient guidance to entities. We can also draw from the experience of the New York City Automated Decision System Task Force (Task Force) which demonstrated the high levels of discretion involved in the assessment of what counts as a "risky" application. The Task Force's report suggested that spreadsheets themselves may be an application warranting exemption because of its quotidian functions;[11] but as research demonstrates, spreadsheets too can facilitate a variety of

[6] Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems (2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.html.
[7] Liberty Human Rights, Policing by Machine (2019), https://www.libertyhumanrights.org.uk/issue/policing-by-machine/ (finding that predictive policing programs entrench pre-existin discrimiation by using biased police data after reviewing fourteen police forces in the UK); Alexander Babuta & Marion Oswald, Briefing Paper: Data Analytics and Algorithmic Bias in Policing (Royal United Services Institute for Defence and Security Studies 2019), https://rusi.org/sites/default/files/20190916_data_analytics_and_algorithmic_bias_in_policing_web.pdf (finding use of predictive data analytics technologies in policing can lead to discrimination and skewing of the descision making process in ways that disproportionately affect particular groups).
[8] Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems (2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.html.
[9] EDRI, Recommendations to EC White Paper on AI, June 2020 https://edri.org/wp-content/uploads/2020/06/AI_EDRiRecommendations.pdf
[10] Access Now, (draft) Submission to the Consultation on the "White Paper on Artificial Intelligence - a European approach to excellence and trust, May 2020, available at https://www.accessnow.org/EU-white-paper-consultation
[11] New York City Automated Decision Systems Task Force Report, 26 (2019), https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.

computational functions and in some contexts have been used to automate decision-making that produced harmful results.[12]

Appreciating the AI's operational context is imperative to understanding the full scope of risks and opportunities that must be contemplated when assessing effective regulation. This is especially true given that AI systems can produce predictions, classifications, and decisions that may appear inconsequential at the time, but that can have lasting effects on people's lives, especially if they are used in assessments or means testing in the future. **For these reasons, we think it is important to first define the scope of the term "AI" for regulatory intervention where the threshold condition is determined by the nature of the impact and the particular effects of the technology, irrespective of the sector in which they are being used.** Adopting a comprehensive and precise definition for regulatory purposes is challenging, so we offer the following recommendations for consideration:

- First, **a definition of AI should emphasize and center the impact and potential effects of the technology rather than centering the underlying technical mechanisms or methodologies.** Many existing legal definitions of AI and related systems tend to lead with and emphasize the technical mechanisms or methodologies that facilitate the various capabilities or functions of the technology. This focus on the technical details is not ideal because it can reinforce automation bias and legitimize certain technical solutions over alternatives,[13] or ignore autonomous systems already in place and their present risks.[14] In this vein, as put forth in the AI Now Algorithmic Accountability Policy Toolkit,[15] we recommend drawing from the following definition: "An Automated Decision[-making/-support] System is a system that uses automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans."

- Second, Article 22 of the GDPR that deals with automated decision systems uses the threshold of whether such systems have "legal or similarly significant effect" on the individual. Such a definition could be underinclusive of the potential harms that AI systems pose and, for example, could fail to include harmful impacts that are clear and well-documented but are not specifically legally actionable. **Rather than "similarly significant" we recommend including more descriptive criteria, for example that all AI systems that have "an impact on opportunities, access to resources, preservation of**

[12] Jay Stanley, Pitfalls of Artificial Intelligence Decision Making Highlighted in Idaho ACLU Case (2017), https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case.
[13] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018); Ruha Benjamin, Race After Technology (2019).
[14] P. M. Krafft, M. Young, M. Katell, K. Huang, and G. Bugingo, Defining AI in Policy versus Practice, in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 2020, at 77,10.1145/3375627.3375835.
[15] AI Now, Algorithmic Accountability Policy Toolkit, 2018 https://ainowinstitute.org/aap-toolkit.pdf

liberties, legal rights, or ongoing safety of individuals, groups, or communities" be included.

- Third, a definition of AI must demonstrate awareness of the context in which the technology is operationalized. Since AI includes a constellation of processes and capabilities that can be related yet distinct (i.e. ranking versus classification), the definition must evaluate or speculate current and future applications to ensure these distinctions are not attenuated. Adopting a reflexive definition can enable cross-sectoral regulation because it does not ignore the fact that uses in different contexts can produce divergent and sometimes harmful outcomes. For example, the Gale-Shapley algorithm is commonly used to match recent medical school graduates with residency programs without concern. However, when it was used in a highly segregated and dense school district like New York City, it has been accused of worsening disparities between schools and further exacerbating segregation.[16]

ALGORITHMIC IMPACT ASSESSMENTS (Section 5F)

Rather than "conformity assessments" for high risk AI applications, we would instead recommend the structure of "algorithmic impact assessments" (AIA) which are designed to support democratic participation and accountability for the decisions of whether and how AI or algorithmic decision systems should be used. There are growing resources that can guide the creation of AIA frameworks:

- AI Now's detailed AIA framework[17] that public agencies can draw from when mandating AIAs.
- The Canadian government's Algorithmic Impact Assessment tool[18] is a useful template for regulatory agencies.
- ICO's draft auditing framework for AI systems[19] too has helpful guidance on how to document risks, manage inevitable trade-offs, and increase reflexivity at every stage of ADS procurement or development.

[16] Jay Cassano, *NYC Students take aim at segregation by hacking an algorithm*, Fast Company, April 16, 2019, https://www.fastcompany.com/90331368/nyc-students-take-aim-at-segregation-by-hacking-an-algorithm
[17] Dillon Reisman et al, Algorithmic Impact Assessments: A practical framework for public agency accountability, AI Now Institute (2018) https://ainowinstitute.org/aiareport2018.html
[18] Government of Canada, AIA (2019) https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html
[19] ICO, ICO consultation on the draft AI auditing framework guidance for organisations (2020) https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-consultation-on-the-draft-ai-auditing-framework-guidance-for-organisations/ (ICO draft AI auditing framework)

In addition to AIAs, there are also similar proposals, including from the Council of Europe,[20] to mandate "human rights impact assessments" that guide similar evaluations into whether uses of AI are compatible with human rights.

For any AIA to be a meaningful exercise, entities using AI should document the risks identified, strategies of mitigation, and a roadmap to implement those strategies *before development.* Critically, the AIA must not have a predetermined commitment to eventually implementing the AI system. **Where the risks identified cannot be sufficiently mitigated, or where the concerns of the affected community remain unresolved, there needs to be scope and space to stop or prevent an AI system's development and/or deployment altogether.[21]**

Recent work aiming to situate the AI development process in its sociopolitical context shows how safety risks, normative controversy, and sensitive tradeoffs may arise throughout the lifecycle of AI systems.[22] As such, robust AIAs should ensure wide public consultation before and during the early stages[23] of implementation of the AI system and establish channels for dissent for citizens to raise concerns once a system is integrated. Such consultation should be balanced to ensure robust and diverse participation among experts and individuals and communities affected by AI use.

Following the conclusion of the New York City Automated Decision Systems Task Force, a coalition of researchers, advocates and community members developed several recommendations on how to lead meaningful public engagement based on lessons learned from the Task Force process. **Recommendations included providing financial remuneration to community members appointed to serve on decision-making bodies, which can help address some of the financial and capacity barriers and engender more equitable and representative participation.[24] Other recommendations spoke to the need for more inclusive educational opportunities and information sharing, including providing printed material at libraries and other community centers for residents that may not have access to the internet or sufficient**

---

[20] Council of Europe recommendation, Unboxing artificial intelligence: 10 steps to protect human rights by the Human Rights Commissioner, May 2019, https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

[21] Cory Doctorow, "Chicago PD's Predictive Policing Tool Has Been Shut Down after 8 Years of Catastrophically Bad Results," Boing Boing, January 25, 2020, https://boingboing.net/2020/01/25/robo-racism.html.

[22] Dobbe, Roel I.J., Thomas Krendl Gilbert, and Yonatan Mintz. "Hard Choices in Artificial Intelligence: Addressing Normative Uncertainty through Sociotechnical Commitments." In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 242. AIES '20. New York, NY, USA: Association for Computing Machinery, 2020. https://doi.org/10.1145/3375627.3375861.

[23] Rashida Richardson, ed., Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force, AI Now Institute (2019) https://ainowinstitute.org/ads-shadowreport-2019.html (hereafter, Shadow Report)

[24] Shadow Report of the NYC Task Force, *supra note 20* at 47.

technical literacy.[25] We recommend including some of these approaches in the design of an AIA.

RULES RELATING TO DOCUMENTATION, INFORMATION PROVISION, AND TRANSPARENCY (Section 4d)

The Commission rightly emphasizes the importance of transparency both in terms of record-keeping about data provenance as well as proactive disclosures to those deploying and using AI, and those who are impacted by these systems. We provide the following recommendations to create a robust transparency framework:

- **Explanations for individuals should include but are not limited to:**
    - The types of decisions or situations being subjected to automated processing;
    - Factors involved in a decision relying on automated processing operations (e.g. behavioral data; socioeconomic indicators; legally defined categories of data; location data);
    - Descriptions of the types of data used in automated processing, and how these are collected;
    - A legible description of the methodology and mechanism underlying the automated processing (e.g. "this technology employs a linear regression model to predict who will succeed in the program"); and
    - Description of potential legal or other significant effects or consequences of automated processing. When the automated processing operations are run or facilitated through a government agency or authority, additional explanation requirements should exist.

- **Entities using AI should provide a comprehensive plan for giving external researchers and auditors meaningful, ongoing access to examine specific systems, to gain a fuller account of their workings, and to engage the public and affected communities in the process.** The "technical information" required to do this will differ from system to system. As we describe in our AIA report,[26] many systems may only require analysis based on inputs, outputs, and simple information about the algorithms used without needing access to the underlying source code. For others it might be critical to obtain access to training data or a record of past decisions to researchers. We believe that the best way for

[25] Shadow Report of the NYC Task Force, *supra note 20* at 24 & 47-49.
[26] Dillon Reisman, et al., Algorithmic Impact Assessments: A practical Framework For Public Agency Accountability, AI Now Institute (2018), https://ainowinstitute.org/aiareport2018.pdf. *See also*, Koene, A., Clifton, C., Hatada, Y., Webb, H., Patel, M., Machado, C., LaViolette, J., Richardson, R., & Reisman, D. *A governance framework for algorithmic accountability and transparency*, European Parliamentary Research Service, (2019) https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

entities to develop an appropriate research access process would be to work with community stakeholders and interdisciplinary researchers through the notice and comment process. **Rather than the typical linear process used to present AI system development, this would acknowledge the need for an iterative process, in which the channels for dissent and democratic deliberation are integrated throughout the lifecycle of the system to prevent harmful outcomes and respond swiftly when new issues emerge.**

- We recommend mapping the legal barriers that typically emerge when trying to access technical information from AI systems used by both private and public actors. When faced with these requests for information, vendors of AI systems often do make broad trade-secrecy or confidentiality claims. The invocation of such corporate-secrecy laws then functions as a barrier to due process, making it difficult to assess bias, contest decisions, remedy errors, or verify specifications. **Commercial confidentiality, IP, and access to information laws that operate to prevent accessibility of information should be reformed with this goal in mind.** Often it isn't the laws themselves that are in need of reform, as much as it is blanket interpretations that are put forth to claim that all aspects of a technical system have competitive commercial value or are otherwise protected.

RULES RELATING TO HUMAN OVERSIGHT (SECTION 5e)

The Commission notes that human oversight *"can ensure that an AI system does not undermine human autonomy or cause other adverse effects"* and gives a range of examples including ensuring that the "final decision" on a social security benefit is eventually taken by a human. This mirrors the intent of Article 22 of the GDPR which requires meaningful human intervention when legal rights might be impacted by an algorithmic decision.

While this intent is well-meaning, we caution against regulating based on a rigid distinction between "solely" automated decisions versus decisions that are informed, aided, or supported by algorithms. In practice, these distinctions are slippery and the fact that there is human intervention in the final decision does not address major concerns over opacity or control and should not be automatically presumed to be at a lower risk level. In fact, where AI systems are used as "decision making aids", research demonstrates that humans are often unable to accurately evaluate the quality or fairness of the predictions made. People fail to rely more heavily on accurate predictions compared to inaccurate predictions, and often respond to predictions in biased and inaccurate ways.[27] This follows from a large body of research showing that people

[27] See Ben Green & Yiling Chen, Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments, https://www.benzevgreen.com/wp-content/uploads/2019/02/19-fat.pdf; Ben Green & Yiling Chen, The Principles and Limits of Algorithm-in-the-Loop Decision Making, https://www.benzevgreen.com/wp-content/uploads/2019/09/19-cscw.pdf

struggle to effectively interpret, use, and oversee algorithms when making decisions.[28] Governance must therefore consider the full sociotechnical system of the human-algorithm collaboration, rather than consider the algorithm or human in isolation.

We instead encourage the European Commission to adopt regulatory regimes to incentivize models and human-algorithm interactions that enhance the real capacity for human oversight and restrict the use of ADS entirely where such oversight cannot be meaningful. Where governments are adopting AI systems to determine the allocation of welfare benefits or deciding criminal justice outcomes, the consequences of overestimating human oversight has serious consequences on basic civil liberties. In other high risk domains, like self-driving cars or automated pilots, research has found over-reliance of drivers[29] or pilots[30] on automated systems led to complacency and a degradation in manual skills eventually putting human life at risk.

For these reasons, we also recommend that any impact assessments of AI systems include an internal assessment of the knowledge differentials or inefficiencies contribute to people's inability to adequately assess and anticipate problems that may arise from such systems. The UK ICO's recent draft auditing framework has some useful guidance on documenting these limits on human capacity to engage with the AI systems.[31] They recommend documenting not just potential risks emanating from these systems, but also the capacity of those interacting with the system to recognize such risks. Where risks and strategies of mitigation (if they exist) are identified, they encourage creating a knowledge base that can be drawn upon by others interacting with the system. It is important that such efforts be grounded in rigorous evidence of what mechanisms improve human oversight, as some mechanisms with intuitive appeal (such as providing explanations of the model's predictions) have been found to provide little benefit.[32]

---

[28] Megan Stevenson, Assessing Risk Assessment in Action (2019) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3016088; Dietvorst Berkeley, Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err (2015) https://psycnet.apa.org/fulltext/2014-48748-001.html; Amirhossein Kiani, Impact of a deep learning assistant on the histopathologic classification of liver cancer (2020) https://www.nature.com/articles/s41746-020-0232-8

[29] John Markoff, Google's Next Phase in Driverless Cars: No Steering Wheel or Brake Pedals, https://www.nytimes.com/2014/05/28/technology/googles-next-phase-in-driverless-cars-no-brakes-or-steering-wheel.html

[30] House Committee on Transport & Infrastructure, The Boeing 737 MAX Aircraft: Costs, Consequences, and Lessons From its Design, Development, and Certification https://transportation.house.gov/imo/media/doc/TI%20Preliminary%20Investigative%20Findings%20Boeing%20737%20MAX%20March%202020.pdf.

[31] ICO draft AI auditing framework, *supra* note 17.

[32] Ben Green & Yiling Chen, The Principles and Limits of Algorithm-in-the-Loop Decision Making (2019) https://www.benzevgreen.com/wp-content/uploads/2019/09/19-cscw.pdf; Forough Poursabzi-Sangdeh, Manipulating and Measuring Model Interpretability https://arxiv.org/abs/1802.07810

BIOMETRIC & AFFECT RECOGNITION (Section 5)

As noted by the Commission, the use of biometric data for remote or passive identification has produced clear harms and violations of the fundamental rights of EU citizens. In the context of law enforcement in particular, these harms have been disproportionately borne by racial and ethnic minorities. Existing legal frameworks like the GDPR and the necessity and proportionality standards in the law enforcement data directive (Directive 2016/680) have largely failed to prevent the proliferation of these biometric surveillance tools, demonstrating that procedural safeguards are difficult to enforce, particularly in the context of law enforcement use. Courts, too, have been reluctant to limit the tools available to law enforcement as evidenced by the Court's approval of live facial recognition in use by the London Metropolitan Police, in the absence of a sanctioning law.[33]

In this context, we would urge the Commission to exercise leadership in entirely prohibiting certain applications of these technologies, rather than relying on procedural safeguards for their use. In particular, **as demanded by a coalition of more than forty European digital rights groups, the use of biometric mass surveillance, like the widespread use of facial recognition in public space, should be prohibited.**[34]

In addition, **we also recommend an urgent moratorium on all uses of facial recognition in sensitive social and political domains**—including law enforcement use, education, and employment—where facial recognition poses risks and consequences that cannot be remedied retroactively. **Lawmakers must supplement a moratorium with (1) transparency requirements that allow researchers, policymakers, and communities to assess and understand the best possible approach to restricting and regulating facial recognition; and (2) protections that provide the communities on whom such technologies are used with the power to make their own evaluations and rejections of its deployment.**

Globally, there has also been a rise in emerging technologies that claim to make evaluations about individuals based on biometric data. They go beyond determining or verifying identity, and include affect recognition systems that claim to be able to detect an individual's emotional state or interior character based on the use of computer-vision algorithms to analyze their appearance and behavior, from facial microexpressions, tone of voice, or even their gait. Such technologies are rapidly being commercialized for a wide range of purposes—from attempts to identify the perfect employee[35] to assessing patient pain[36] to tracking which students are being attentive in

[33] R(Bridges) v. CCSWP and SSHD, [2019] EWHC 2341 (Admin), Case No. CO/4085/2018, 4 September 2019, para. 78.
[34] EDRI, Ban Biometric Mass Surveillance (2020) https://edri.org/blog-ban-biometric-mass-surveillance/
[35] Drew Harwell, Rights Group Files Federal Complaint against AI-Hiring Firm HireVue, Citing 'Unfair and Deceptive' Practices," Washington Post, (2019) https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices.

class.[37] Despite these broad claims, the case for affect recognition is built on questionable research that recalls discredited race science, physiognomy, and other pseudoscientific methods that claimed to detect a person's worth and character based on their physical attributes. **We recommend an immediate moratorium on the development and deployment of affect recognition and similar systems in public and private contexts.**

PRODUCT SAFETY & LIABILITY (Section 5a)

While we applaud the Commission for its rigorous revisiting of product safety and liability regulation, the narrow frame of AI "products" as the main scope of possible interventions would be insufficient to address the safety risks of AI *systems* in critical infrastructure and sensitive domains. Decades of experience on system safety show that a focus on the product alone does not cover all dimensions needed to safeguard "systems", which rely on a broad set of aspects including people, organization, hardware and software infrastructures, and the context in which a given system is applied.[38] The harms from AI systems failures, including those in autonomous vehicles, medical prediction algorithms, and online recommendation systems, are rarely because of the physical or technical product features alone. More often they stem from a lack of a safety culture, human-machine interactions, inadequate specifications in the engineering development process, or a lack of empirically verified safety assurances.[39] While some of these aspects are mentioned in the Commission's reports, we would encourage a more holistic perspective on the safety of AI systems, which takes the total systematic nature of the technology, and its development and deployment environments, into account. This should be central to the overall legislative framework for every sector.

Vulnerable IoT devices and AI models should not be used in safety-critical applications: The AI, IoT and robotics markets in particular desperately need rules, standards and incentives to prioritize security and safety in their products, services, and infrastructures. An Ernst and Young (EY) cybersecurity expert found that for each major stage of assembly in the IoT industry, namely chip production, software specification and implementation, "the focus is on maximizing profits rather than ensuring security or integrity, [t]his short term, profit-first approach is a sure recipe for

[36] Clarice Smith, Facial Recognition Enters into Healthcare, Journal of AHIMA, (2018) https://journal.ahima.org/2018/09/04/facial-recognition-enters-into-healthcare.
[37] Jane Li, A 'Brain-Reading' Headband for Students Is Too Much Even for Chinese Parents, Quartz, (2019), https://qz.com/1742279/a-mind-reading-headband-is-facing-backlash-in-china/.
[38] Richard A.Stephans, "System Safety for the 21st Century: The Updated and Revised Edition of System Safety 2000" (2012).
[39] Inioluwa Deborah Raji and Roel Dobbe. "Concrete Problems in AI Safety, Revisited." In Workshop on Machine Learning In Real Life. Addis Abeba, Ethiopia (2020) https://drive.google.com/open?id=1Re_yQDNFuejoqjZloTgQpILosDGtt5ei .

disaster."[40] While security issues are becoming more apparent and a higher priority for the IoT industry, security is still regarded as "one of the most challenging technical problems faced by IoT implementers."

Data that faulty IoT devices generate may form the input to various safety-critical AI systems. The trends in the IoT industry are hence worrying and compounding the already inherent vulnerability of machine learning models.[41] Recent advances at the intersection of computer security and machine learning indicate that popular models, such as deep learning, are inherently brittle, and prone to errors, misuse and attacks. While the ML security research field has made some strides, and is getting lots of attention, leaders in the field question whether this research might be as applicable and ground-breaking as it is widely perceived.[42] It remains unclear and unlikely that machine learning will reach a level of robustness in the foreseeable future that supports its use in safety-critical environments. As such, the Commission should be extremely careful with allowing the uptake of any of such systems in sensitive domains.

Need For A Shift In Business Models To Incentivize Safety: The lack of investment in security during the production of IoT devices, motivates the EY expert to conclude that "[i]t is critically urgent that the organizations and institutions developing and using these technologies adopt a Security by Design perspective." However, the ability to ensure AI systems are safe or secure by design is compromised by the economic incentives and power structures in the software industry. As Gürses et al. show, more and more tools are hidden away in "services" in the form of software libraries, toolchains and modular application programming interfaces (APIs).[43] These are controlled by a handful of tech companies which effectively reign over both software tooling and the accompanied digital infrastructure. As a result, building a system may be more efficient in terms of coding. However, the same modularity makes addressing the vulnerabilities of the overall system, which is now a *composition* of modular services (consisting of different APIs, toolchains and libraries), a much more complex task, as the tech companies that provide these often don't enable access to documentation that would allow these users to validate their security. This is further intensified when certain service modules are developed by third parties

[40] Lovejoy, Kris. "How to Manage Cyber Risk with a Security by Design Approach." Ernst & Young, February 7, 2020. https://www.ey.com/en_gl/advisory/how-to-manage-cyber-risk-with-a-security-by-design-approach.
[41] Roel Dobbe, AI Vulnerabilities Report. AI Now Institute. In Preparation, (expected Summer 2020)
[42] Nicholas Carlini, Lessons Learned from Evaluating the Robustness of Defenses to Adversarial Examples. Santa Clara, CA: USENIX Association (2019) https://www.usenix.org/conference/usenixsecurity19/presentation/carlini-talk.
[43] A software service library is way to integrate pre-developed code into a newly written program. A toolchain is a set of programming tools that is used to perform a complex software development task or to create a software product, which is typically another computer program or a set of related programs. An application programming interfaces (API) is a computing interface which defines interactions between multiple software systems and can be used to extend the functionality of one program with another. An API typically asks for certain inputs and returns outputs, often without visibility of the internal workings of the code that processed the inputs into outputs.

and hidden away behind APIs. To make matters more challenging, different modules tend to be developed by different teams or companies and may change in an asynchronous manner.[44]

Taken together, these challenges render the secure-by-design of AI systems a noble but impractical goal in the current industrial landscape. They make impossible the ability to observe and validate the workings of a system *from end to end*. In addition, those parties responsible for different modules or the development of extra services for security have to be trusted. This goes against the central premise of security engineering which encourages "trust, but verify," and ultimately means that those practicing responsible security engineering must treat every module or service that is not transparent as an extra vulnerability.

In addition, AI systems may be hidden behind such service APIs. As these tend to be built on utilitarian models that are optimized over a narrow set of metrics (such as prediction accuracy or test error), implementing such code at scale without having access to the underlying code may lead to disastrous outcomes. For example, some researchers had to develop detours to show how the AI models in Facebook's advertising API discriminate on the basis of race and gender; they did not have direct access to this software.[45] Given the importance of Facebook's advertising for domains like housing or recruitment, such APIs deserve to be open to be scrutinized and audited by third parties. These discriminatory economic effects should be acknowledged as safety risks and prevented at all costs.

In order to promote AI systems that are safe and secure by design, the Commission hence has a crucial role to play to remove barriers that have been erected by the software industry. Service libraries, tool chains and APIs need to be open for auditing by third parties, especially when used in software development for government institutions and in safety-critical and sensitive domains. The Dutch government has recently committed to an 'open source by default' policy for procurement,[46] which is now experimented with in the development of a contact tracing app[47].

---

[44] Seda Gürses and Joris Van Hoboken. Privacy after the Agile Turn, (2017) https://osf.io/preprints/socarxiv/9gy73/

[45] Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. "Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes." *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (November 7, 2019): 1–30. https://doi.org/10.1145/3359301; "Facebook's Ad Delivery System Still Discriminates by Race, Gender, Age." Accessed June 12, 2020. https://news.northeastern.edu/2019/12/18/facebooks-ad-delivery-system-still-discriminates-by-race-gender-age-y/.

[46] Rikken, Nico. "Netherlands Commits to Free Software by Default." FSFE - Free Software Foundation Europe. Accessed June 12, 2020. https://fsfe.org/news/2020/news-20200424-01.html.

[47] Security.nl. "Overheid Laat Experts Nieuwe, Open Source Corona-App Ontwikkelen," April 22, 2020. https://www.security.nl/posting/653533/Overheid+laat+experts+nieuwe%2C+open+source+corona-app+ontwikkelen.

III. GOVERNMENT USE OF AI & PUBLIC PROCUREMENT GUIDELINES (Section 4)

We would strongly contest the suggestion in the White Paper that AI uptake must be prioritized in the public sector. Government agencies make many decisions that can have immediate and long-term effects on a citizen's rights and opportunities, and while the use of AI products can make decisions or forms of policy implementation more efficient, they can also produce serious harms and risks.[48] AI products used to make decisions and allocate public benefits were successfully challenged in the Netherlands for privacy and equity concerns,[49] but unregulated use of similar products by government agencies in the United States has resulted in wrongful benefits termination, unmet care needs, and in some cases death of residents that could not resolve technology errors expeditiously.[50] There have also been grave results in cases where the AI product failed to perform as expected, notably several jurisdictions in the United State stopped using predictive analytics in child protection agencies after several high profile deaths of children the technology failed to identify.[51]

For these reasons, public sector AI systems should benefit from additional scrutiny and accountability safeguards, including rigorous AIAs as described above. **Public sector use of AI that has been demonstrated to produce biased outcomes and harms including but not limited to, predictive policing, child welfare predictive analytics, pre-trial risk assessment, and public benefits decision systems, should be considered for moratoria or other restrictions.**

It is also important to note that government agencies largely contract AI technologies from third-party private vendors who provide the algorithmic systems for public services, including welfare benefits and criminal risk assessments. This unique role of AI vendors in supplying the

---

[48] Eric Corbett & Christopher A. Le Dantec, *'Removing Barriers' and 'Creating Distance': Exploring the Logics of Efficiency and Trust in Civic Technology*, Media and Communications (2019).
[49] ECLI:NL:RBDHA:2020:865, https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878 (finding the Dutch public authorities use of public benefits fraud detection ADS to be an unlawful violation of the right to privacy).
[50] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (New York: St. Martin's Press, 2018); Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems (AI Now Institute, September 2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.html; AI Now Institute, NYU Law Center on Race, Inequality, and the Law & the Electronic Frontier Foundation, Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems (AI Now Institute, September 2018), https://ainowinstitute.org/litigatingalgorithms.pdf; ECLI:NL:RBDHA:2020:865, https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878 (finding the Dutch public authorities use of public benefits fraud detection ADS to be an unlawful violation of the right to privacy).
[51] Michael Nash, "Examination of Using Structured Decision Making and Predictive Analytics in Assessing Safety and Risk in Child Welfare" (Los Angeles: County of Los Angeles Office of Child Protection, May 4, 2017), http://file.lacounty.gov/SDSInter/bos/bc/1023048_05.04.17OCPReportonRiskAssessmentTools_SDMandPredictiveAnalytics_.pdf; David Jackson & Gary Marx, *Data mining program designed to predict child abuse proves unreliable, DCFS says*, CHICAGO TRIBUNE, Dec. 6, 2019, https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html.

core logics and consequential actions of these systems calls for increased accountability and transparency in public procurement processes. **We recommend the following procedural and substantive protections that should be incorporated in vendor contracts:**

- **Waiver to trade secrecy or other barriers to information**: As recommended above, all public agencies that use AI systems should require vendors to waive any trade secrecy or other legal claim that might inhibit algorithmic accountability, including the ability to explain a decision or audit its validity.[52] Additionally, public agencies should avoid agreeing to Non-Disclosure Agreements or other confidentiality clauses that can inhibit the agency's ability to assess the technology and comply with transparency and accountability measures.
- **Requiring data provenance documentation:** We would recommend that documentation in accordance with prototypes like model cards[53] and datasheets[54] are made mandatory for all government AI vendors. We would also encourage the government to make this documentation public at the consultation stage In order to invite scrutiny from the active community of public interest researchers that work on these issues.
- **Training modules:**[55] Government agencies should require vendors to provide more training materials to help agency staff understand the system, and require the vendor to collaborate with the agency in developing public-education materials that engage the public. Vendors are often in the best position to ascertain whether public-education documents adequately describe the capabilities and potential risks of a given system, while agencies are in a better position to assess the needs of people affected by a given system, as well as the needs of the broader community.
- **Restrict broad indemnity clauses:**[56] Government agencies procuring AI systems should not enter purchase agreements of licenses that require the agency to indemnify vendors for any negative outcomes. There have been incidents where prominent vendors include such clauses, absolving them of any responsibility for negative consequences that were caused by design errors or oversights in the AI system that vendors should be accountable and responsible for.
- **Mandatory validation studies:**[57] Given the hype associated with AI systems, and the fact that government agencies may not always have the capacity to evaluate claims, it is critical to mandate comprehensive validation studies and audits. These studies (including the methodology and results) should typically audit for discriminatory impact on protected classes, accuracy, and the value of using AI as compared to existing practices. These

[52] AI Now 2018 Report, https://ainowinstitute.org/AI_Now_2018_Report.pdf
[53] M. Mitchell et al, Model Cards for Model Reporting (2019) https://arxiv.org/pdf/1810.03993.pdf
[54] Timnit Gebru et al, Datasheets for Datasets (2020) https://arxiv.org/pdf/1803.09010.pdf; see also Inioluwa Deborah Raji and Jingying Yang, ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (2019) https://arxiv.org/abs/1912.06166
[55] Shadow Report of the NYC Task Force, *supra note 20*
[56] Shadow Report of the NYC Task Force, *supra note 20*
[57] Shadow Report of the NYC Task Force, *supra note 20*

validation studies should be performed on an ongoing basis to ensure that AI use is just, accurate, and constitutes an improvement over existing practices.

- **Non-discrimination guarantees and audits:**[58] Government agencies negotiating AI system contracts should ensure the contract includes language requiring the vendor to guarantee the product or service is compliant with federal, state, and local antidiscrimination laws. Inclusion of such clauses will ensure that government agencies have standing to have the system fixed, and that vendors share liability if AI use produces discriminatory outcomes.
- **For biometric AI, like face recognition systems,**[59] where there is mounting evidence of biases on grounds of race and gender, agencies should assess whether current or prospective AI will disproportionately affect individuals or groups based on protected class. In order to verify the functionality of these systems, the agencies must demonstrate that any biometric detection system performed up to a specified standard. Because evaluations often represent a specific context (for example, NIST benchmarks tend to be skewed for age, gender and race), testing procedures should include user-representative datasets[60] which include the major intersectional demographic categories of affected users. Agencies must report the performance of the model on each demographic subgroup in order to acknowledge any performance disparities.

- **Open, competitive bidding process:**[61] In order to ensure proper scrutiny and accountability in government procurement of AI, we would recommend requiring all procurement is done through an open, competitive bidding process and should not be exempt from public-hearing requirements. Any deviations from this for a sole source contract instead should have clear justification.
- **Heightened standards for sensitive, social domains:** Given the risks to life, civil rights, and civil liberties that AI poses in sensitive social domains, the standards for an AI vendor's record should be heightened. When AI is used to assist or make life-altering decisions, like whether an individual goes to jail before trial, expectations for vendor performance appraisals should be greater than satisfactory.
- **Warranty Measures:** Government agencies should consider including warranty measures, such as staggered payments, into procurement contracts or during negotiations. Such measures can provide the agency leverage to ensure vendor accountability to correct errors or bugs, and provide a product that fits the agency's needs.

[58] Shadow Report of the NYC Task Force, *supra note 20*
[59] Shadow Report, *supra note 20*
[60] Shadow Report, *supra note 20*
[61] Shadow Report, *supra note 20*

IV. AI AND ENVIRONMENTAL IMPLICATIONS

The tech sector is a significant contributor to climate change and environmental harms, and the uptake of AI systems is likely to further accelerate these impacts.[62] As a whole, the industry's energy dependence is on an exponential trajectory, with best estimates showing that its 2020 global footprint amounts to 3.0–3.6 percent of global greenhouse emissions, which is more than double what the sector produced in 2007 and comparable to the aviation industry.[63] In the worst-case scenario, this footprint could increase to 14 percent of global emissions by 2040. The tech industry can no longer be exempt from measures to curb emissions.

A core contributor to the AI field's growing carbon footprint is a dominant belief that "bigger is better." This belief assumes that AI models that leverage massive computational resources to process large training datasets are inherently "better" and more accurate.[64] While this narrative is inherently flawed,[65] and its assumptions drive the use of increased computation in the development of AI models across the industry. In addition, AI companies are aggressively marketing their (carbon-intensive) AI services to oil and gas companies, offering to help optimize and accelerate oil production and resource extraction, meaning that both the development of AI models, and their application, are producing damaging climate consequences.[66]

These dangerous developments should be a wakeup call to the Commission. Currently, the White Paper mostly promotes AI as a *solution* to the challenge of climate change. We are happy to see that it does mention the need to address environmental implications of AI systems "throughout their lifecycle and across the entire supply chain." We ask the Commission to make these ambitions more concrete in the report. In a recent effort, we developed seven policy considerations that provide a path toward tech-aware climate policy, and climate-aware tech policy.[67]

---

[62] Roel Dobbe and Meredith Whittaker, "AI and Climate Change: How they're connected, and what we can do about it" AI Now Institute. (17 October, 2019), https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c.

[63] Belkhir, Lotfi, and Ahmed Elmeligi. "Assessing ICT Global Emissions Footprint: Trends to 2040 & Recommendations." *Journal of Cleaner Production* 177 (March 10, 2018): 448–63. https://doi.org/10.1016/j.jclepro.2017.12.239.

[64] Sutton, Richard. "The Bitter Lesson." *Incomplete Ideas* (blog), March 13, 2019. http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

[65] Welling, Max. "Do We Still Need Models or Just More Data and Compute?" University of Amsterdam, April 20, 2019. https://staff.fnwi.uva.nl/m.welling/wp-content/uploads/Model-versus-Data-AI.pdf.

[66] Greenpeace USA. "Oil in the Cloud." Accessed May 20, 2020. https://www.greenpeace.org/usa/reports/oil-in-the-cloud/.

[67] Roel Dobbe and Meredith Whittaker, "AI and Climate Change: How they're connected, and what we can do about it" AI Now Institute. (17 October, 2019),

These recommendations can inform the White Paper in various ways. A first key step is to mandate transparency about the environmental footprint of European AI efforts and the digital infrastructure that is used to run these. **At a minimum, we need insight into the greenhouse emissions related to cloud services.**[68] Reporting standards should allow any organization to understand its own digital footprints and make better informed decisions when developing AI systems. **Secondly, environmentally harmful uses of AI should be banned, including the use in fossil fuel exploration.** Beyond this, the use of AI systems to track and surveil climate refugees along national borders need to be prohibited. Such use contributes to a dynamic in which people in regions that contributed least to the climate crisis are denied resources and care from regions that are more responsible for the current crisis. Lastly, in efforts to build and extend sovereign European digital infrastructure, the Commission should clarify what it means by "high-quality digital infrastructure"; at the very least, it should set high ambitions to minimize environmental impacts across the European computing value chains, from chip manufacturing to data centers to software standards and AI applications. **It is time for technology and climate issues to be combined in new and ambitious missions,[69] which requires alignment with the European Green New Deal. Finally, given AI is integrated in a variety of policy domains, energy and climate impacts should be calculated there, as well, as a standard part of policy practice**.[70]

https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c.

[68] Cook, Gary, Jude Lee, Tamina Tsai, Ada Kong, John Deans, Brian Johnson, and Elizabeth Jardim. "Clicking Clean: Who Is Winning the Race to Build a Green Internet?" Washington, DC: Greenpeace, January 2017. http://www.clickclean.org/international/en/.

[69] Horizon: the EU Research & Innovation magazine. "Missions Could Make Europe Cool Again – Prof. Mariana Mazzucato." Accessed June 12, 2020. https://horizon-magazine.eu/article/missions-could-make-europe-cool-again-prof-mariana-mazzucato.html

[70] Cox, E., S. Royston, and J. Selby. "Impact of Non-Energy Policies on Energy Systems." UK Energy Research Centre, November 16, 2016. http://www.ukerc.ac.uk/publications/impact-of-non-energy-policies-on-energy-systems.html.

Using procurement instruments to ensure trustworthy AI

A position paper by the AI Now Institute, City of Amsterdam, City of Helsinki, Mozilla Foundation and Nesta.

1. The challenges faced by the public sector

Public sector authorities are increasingly seeking to capture the opportunities offered by AI-enabled systems to improve the provision of services to the public.

However, public sector authorities and the wider general public have justified concerns over data governance, privacy, bias, discrimination, accountability, transparency and the overal opacity of AI-enabled systems. A landmark ruling by the District Court of the Hague in the Netherlands on the use of an algorithmic risk model (SyRi) to detect social benefit fraud illustrates how public sector use of AI-enabled systems in itself can result inadvertently in new risks or harms.

Public sector authorities further often rely on the expertise, and previously developed models, of technology providers and may lack the necessary skills to fully understand or audit AI-enabled systems. In those rare cases where flaws are found, public sector authorities are often faced with 'vendor lock in'. The more training an algorithm gets using the data provided by the city and its citizens, the more valuable and useful it gets for its user. This makes the user dependent on the vendor, as they are unable to use another vendor without substantial switching costs. At present, the vendor often holds intellectual property rights in the system, and can ward off liability or requests for information using IP, trade secrecy and broad indemnity clauses.

Finally, because there are no clear rules about public oversight of tech vendor contracts, government agencies may procure and use tech that could impact large numbers of people without ever needing to notify the public. Any use of an AI-enabled system by public sector authorities could be used in such a manner that significant risks can occur. As a result, citizens should rightfully expect a high level of transparency and accountability when those systems are procured.

2. Going beyond guidelines

Some governments have taken steps to create guidelines for government agency procurement of AI-enabled systems. In the U.K., the government published a "Guide to using AI in the Public Sector" to enable public agencies to adopt AI-enabled systems in a way that benefits society. Organisations like the World Economic Forum or Data Ethics have further elaborated useful guidelines specifically focused on procurement policies.

The Cities Coalition for Digital Rights, a coalition of 39 cities in the EU and the US, are taking steps to ensure that cities use technology in an open and transparent way. In its declaration, the coalition affirms several broad principles including the transparency, accountability, and non-discrimination of algorithms. This means in practice that the public "should have access to understandable and accurate information about the technological, algorithmic, and artificial intelligence systems that impact their lives," and they should be able to "question and change unfair, biased or discriminatory systems."

However, these guidelines have largely not been implemented. There is an urgent need to go beyond mere guidelines, and provide clarity on fundamental rights safeguards, testing requirements or modelling requirements when public sector authorities decide to procure AI-enabled systems.

3. Standard Contractual Clauses for Municipalities for Fair Use of AI-enabled Systems

Procurement and contract conditions are both very powerful and practical instruments for public sector authorities to assure AI-enabled systems comply with fundamental rights and democratic values. In Amsterdam 2 billion and in Helsinki 2.5 billion euros is annually spent through procurement.

The EU's High Level Expert Group (AI HLEG) recommended to strategically use public procurement to fund innovation and ensure trustworthy AI, by introducing "clear eligibility and selection criteria in the procurement rules and processes of EU institutions, agencies and Member States that require AI systems to be trustworthy".

The next step is to begin defining what these procurement standards should be and to operationalize and standardize their use. In October 2019 the City of Amsterdam has - together with the City of Helsinki and external experts - started this process by drafting standard contractual clauses which attempt to include such criteria. They are intended for use in those situations where a municipality purchases an AI-enabled system from an external supplier. The draft contractual clauses cover all algorithmic systems that when used by the Municipality, may affect citizens of the municipality, visitors to the municipality, or companies established in the municipality to a significant extent. In such an event, the municipality wishes to implement certain safeguards, including detailed procedural transparency in all cases and technical transparency in case of a mandatory cooperation of the contractor with an audit or other type of inspection.

There is also global momentum on this issue. In December 2019, in response to the New York City Automated Decision Systems (ADS) Task Force, an NGO coalition recommended a series of protections to be included in vendor contracts. These include specific waivers to trade secrecy; provision of training modules by vendors to help government staff understand the systems and to collaborate in developing public-education materials; restricting broad indemnity clauses; mandatory validation studies and an open, competitive bidding process for these arrangements.

4. Call to action: use procurement policies to encourage trustworthy AI

These are concrete examples of how public procurement policies can be leveraged to support the development and uptake of trustworthy AI. However, we hope that other public sector authorities, including the EU, will follow suit. When public procurers represent a critical mass, they can create new standards and new demands for certain safeguards in AI-enabled systems.

That's why we ask the European Commission to

- Adopt similar contractual clauses in its own procurement and tendering processes
- Facilitate the development of common European standards and requirements for the public procurement of AI-enabled systems