CENTER FOR DEMOCRACY & TECHNOLOGY

Center for Democracy & Technology's Response to the European Commission Consultation on the "White Paper on Artificial Intelligence - a European Approach to Excellence and Trust"

Supporting Document

CDT's consultation response is focused on Section 2: "An ecosystem of trust". This is because the contents of this section involve the intersection of AI and fundamental rights, and this area of policy is central to CDT's mission and the expertise we can bring to bear.

1. Artificial Intelligence in online content moderation

One of CDT's core areas of focus has been the promotion and protection of free expression and access to information online. Online content moderation has been a dominant theme in policy debates over the past few years, and technical tools for automated content moderation play an increasingly important role in shaping today's information environment. The forthcoming Digital Services Act process will include discussion about the role of AI-powered content moderation technologies, and likely bring forward proposals for creating transparency, accountability and regulatory oversight. CDT has written extensively[1] on the role of automation in online content moderation systems[2].

We urge policy makers to refrain from mandating the use of filtering and other technologies, and we caution that reliance on automated tools involve serious risks for free expression and access to information.

2. Risk Assessment: towards a more nuanced approach

We believe prioritizing regulatory action according to risk is an efficient way to govern public intervention because it focuses on regulatory action where it is likely most needed. This approach demands at least some level of risk assessment, which can help inform policy discussions and make acknowledgement of risk a part of the public record. However, we

[1] Natasha Duarte, Emma Llansó, *Mixed Messages? The Limits of Automated Social Media Content Analysis* (November 28, 2017) https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/.
[2] Emma Llansó, Joris van Hoboken, Paddy Leerssen, Jaron Harambam, Transatlantic Working Group: *Artificial Intelligence, Content Moderation, and Freedom of Expression, (*February 26, 2020), https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf.

consider that the bifurcation of AI applications into high- and low-risk could be too rigid an approach to AI regulation, failing to capture the nuanced scale of AI risks and the highly contextual nature of the potential uses of AI applications. Under a binary approach, the only options are no regulation or heavy regulation[3], making it likely that many moderately risky AI systems will end up falling into the high-risk section and being subjected to disproportionate requirements while other, similarly risky applications could face no regulation.

Further, the Commission may wish to consider whether existing legal concepts addressing harms are capable of accounting for all of the possible harms an AI application could produce. For example, in the United States, courts have had difficulty reconciling the various harms to individual privacy caused by certain uses of data with the kinds of harms traditionally recognized in law.[4] If there are similar gaps in European legal frameworks and concepts, people may have trouble defending their own interests against some types of harm created by AI systems. This is an area of active research.[5]

Rather than a binary approach to risk assessment, CDT suggests a more scaled approach. First, any future regulation should follow a differentiated risk-based approach, based on an application's potential harm to individuals or fundamental rights. Second, the specific use context of the system or application should be considered, but risk assessments should also consider whether the same technology could be deployed in other contexts and take those potential risks into account. Third, legal obligations should gradually increase with the identified risk level. In the lowest risk category (e.g. music recommendation systems) there may be little need for regulation, but as risks to people or their rights increase, so should the level of regulatory oversight and control. The highest risk applications, such as those likely to violate fundamental rights, should receive the highest levels of scrutiny and control. In between those two extremes, there will likely be many applications that warrant varying levels of oversight and regulation.

Additionally, the Commission should consider articulating the differences between risks (likelihood) and harms (impact), and how regulators should consider those two independent factors. For example, some applications may present a relatively low likelihood of causing significant harm, such as an automated driving application causing a traffic accident, while

---

[3] William Crumpler, *Europe's Strategy for AI Regulation,* CSIS, (February 21, 2020), https://www.csis.org/blogs/technology-policy-blog/europes-strategy-ai-regulation.
[4] *See*, eg. *Spokeo, Inc. v. Robins*, 578 US _ (2016).
[5] Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI* (May 2020), https://www.researchgate.net/publication/341340407_Why_Fairness_Cannot_Be_Automated_Bridging_the_Gap_Between_EU_Non-Discrimination_Law_and_AI.

other applications may present a higher likelihood of causing lesser harms, such as a recommender system that promotes misinformation or conspiracy videos. Clearly these two applications should receive different levels of scrutiny and control, yet regulators will need guidance to help them structure and prioritize their approaches to each.

In some cases, a person may have little choice but to use or be subject to algorithmic systems. These systems should be treated as having a higher risk of harm. For example, the Whitepaper mentions the use of algorithms in recruitment and hiring contexts as an example of a high-risk application. Rather than relying on an "exceptional instance" to classify such uses as high-risk, we suggest that the Commission should articulate and consider additional factors, such as the degree of choice people have in their use. In the context of recruitment and hiring, people have little choice but to accept the use of such applications even though they will have a significant impact on their lives. In general, a clearer articulation of the factors influencing an application's potential for harm would improve the quality of risk assessments in the future.

CDT recognizes that machine learning applications hold the potential to improve the outcomes and fairness of many decision processes.[6] When used appropriately, well-designed systems can improve the speed, consistency, and accuracy of many decisions compared to "human-only" processes. However, these technologies are far from mature and should be subject to at least the same level of regulation and oversight as humans in similar contexts. Likewise, people whose lives are impacted by machine-made or machine-supported decisions should enjoy at least the same level of legal protection against erroneous or discriminatory outcomes. As discussed above, a lack of choice heightens the importance of oversight and legal protections, no matter how decisions are made.

Regardless of the Commission's final approach to assessing risk, it should follow clear and transparent rules. For applications that are most likely to cause harm, or that have the potential to cause significant harm to individuals or fundamental rights, it is appropriate to ensure oversight of EU-level or Member State-level authorities.

When regulators determine that a system poses risks to the rights and freedoms of natural persons, it must be subject to a human rights impact assessment. This assessment should consider whether and how to modify or control the uses of such systems to prevent the infringement of rights.

[6] Frederik J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, The International Journal of Human Rights (2020) https://www.tandfonline.com/doi/full/10.1080/13642987.2020.1743976.

In addition, we urge the Commission to clearly define its 'exceptional instances' clause, according to which the use of AI applications for certain purposes is to be considered as high-risk as such. As it is currently written, the clause does not allow sufficient clarity for developers and companies. Without clarity, companies' incentives to develop and deploy new applications will be chilled due to the possibility of having their offerings classified as "high risk," resulting in barriers to the growth of a European AI market.

3. The use of remote biometric identification systems in public spaces

In CDTs view, several legislative instruments, if properly enforced, already effectively prohibit the use of many biometric surveillance applications in public spaces: the European Convention on Human Rights, the EU Charter of Fundamental Rights and the General Data Protection Regulation. The general principle underlying these instruments is that any interference with fundamental rights must be necessary and proportionate and serve a legitimate aim. It is difficult to imagine scenarios in which deployment of remote biometric identification in public spaces would meet this threshold. We urge the European Commission to issue guidance that demonstrates the drastic interference with fundamental rights its use would cause, and to actively discourage biometric surveillance performed by Member States. A recent report[7] from the EU Fundamental Rights Agency provides a comprehensive treatment of this subject.

Under the ECHR (Arts 8-11), the Charter (Art. 52(1)), and the GDPR, biometric processing for uniquely identifying people in public spaces cannot be considered necessary or proportionate because it negatively impacts citizens' rights and freedoms. Although there may be some benefits to the use of these systems from a resource efficiency perspective, those benefits are outweighed by these rights violations. Likewise, compared to that of existing non-automated solutions, the scope and sensitivity of the data collected and processed by biometric surveillance systems is dramatic. Therefore, it is difficult to imagine a scenario in which authorities would be able to justify the use of biometric surveillance applications.

[7] Fundamental Rights Agency, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, (November 27, 2019) https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law.

Article 8(2) of the EU Charter of Fundamental Rights states that everyone has a right to access and rectify data that has been collected about them.[8] Therefore, any biometric surveillance system must offer public access to both the data set used for development and training as well as any data collected through use of the system. Even if biometric surveillance could be legally justified, this requirement would make such systems impractical to administer.

Moreover, the ECHR requires that any interference with Article 10 (right to free expression) or Article 11 (right to free association), is in accordance with law, and both necessary and proportionate. In this regard, the use of facial recognition technology can be highly intimidating. In order to protect their anonymity citizens may decide not to attend public meetings, or to change their everyday social behaviour in public spaces. This will undermine citizens' willingness to express opinions, communicate with others and engage in democratic processes.

Article 9 of the GDPR prohibits the processing of special categories of personal data, including biometric data, to uniquely identify a natural person.[9] Legitimate exceptions are possible, for example, based on consent.[10] However, the deployment of biometric monitoring in public spaces precludes the ability for people to give informed freely-given consent, thereby violating their rights to data protection. Even where notice of surveillance is given, the decision to venture into a public space should not be equated with consent to biometric surveillance. Therefore, consent cannot be the basis for an exception under Art. 9.[11]

According to GDPR Art 9(2)(g), national and EU legislators have the discretion to decide the cases where the use of this technology guarantees a proportionate and necessary interference with human rights. Following the same approach, article 8 ECHR requires that any interference with the right to a private life is under the law and is both a necessary and proportionate means for achieving a legitimate aim. In CDT's view, the use of biometric

[8] EU Charter of Fundamental Rights, Art. 8(2).
[9] GDPR Art. 9(1) "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."
[10] GDPR Art. 9(2) "Paragraph 1 shall not apply if one of the following applies": (a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject".
[11] An additional exception in Art. 9(2)(e) allows processing of data that the subject has "manifestly made public," but as with the consent exception, simply venturing into a public space cannot equate to manifestly making public a person's biometric data.

surveillance systems capable of identity matching in public spaces fails to meet these thresholds because it involves the indiscriminate scanning and checking of the identity of every person within the camera's range.

The US city of San Francisco[12] has banned the use of facial recognition software by the police and other agencies. Two other US cities (Somerville, Massachusetts; Oakland, California[13]), have done the same. These cities have arrived at the conclusion that biometric surveillance cannot presently be reconciled with civil rights and legal protections.

Given that the current regulatory and enforcement framework has not been successful in preventing the Member States from deploying[14] what appear to be unlawful biometric mass surveillance systems, we urge the Commission to issue clear guidelines and clarify that remote biometric identification in public spaces is incompatible with European law,

4. The unclear added value of a voluntary labelling system

As discussed in paragraph 3, CDT finds that categorizing systems into a high-risk/low-risk dichotomy for the purpose of regulation offers insufficient nuance and flexibility to address the spectrum of risks and harms posed by different AI systems. In our view, it is unclear what value a voluntary labelling system would bring for European citizens. If the Commission wishes to pursue such a system, we encourage it to consider what meaning labels would have for consumers, especially in situations where consumers have little or no choice among products or services. Even where consumers may have meaningful choices, the Commission should consider to what extent consumers would factor a voluntarily applied label into their decisions.

Further, the Commission should consider the implications of instituting a labelling system. For example, it would require the establishment of industry standards, the creation of an independent body to certify labels, and continual updates to both standards and the certification process. These would require significant investments of time and money, yet may not yield the same level of benefit to the public. The Commission should also consider

[12] Kate Conger, Richard Fausset, Serge F. Kovaleski, *San Francisco Bans Facial Recognition Technology*, The New York Times, (May 14, 2019), https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.
[13] Rachel Metz, *Beyond San Francisco, more cities are saying no to facial recognition*, CNN Business, (July 17, 2019), https://edition.cnn.com/2019/07/17/tech/cities-ban-facial-recognition/index.html.
[14] Daniel Leufer, Fieke Jansen, *The EU is funding dystopian Artificial Intelligence projects*, Euractiv, (January 28, 2020), https://www.euractiv.com/section/digital/opinion/the-eu-is-funding-dystopian-artificial-intelligence-projects/.

how adoption of a labelling system would impact smaller businesses and startups, which would likely face greater resource constraints than larger, more established companies.

5. CDT's recommendations for moving toward trustworthy AI

CDT believes that a new AI regulatory framework should integrate the current European legislative framework and, in particular, the General Data Protection Regulation.
In this regard, the new law should set precise language, explicit and well-defined rights and safeguards[15] to individual subjects to an automated decision making process. It should establish transparency requirements for citizens to know when they are subject to automated decision making and the level of accuracy of the tool in achieving the purpose. Furthermore, more precise descriptions of different kinds of AI systems will be important in the new framework.

Any new legislation should ensure that human rights impact assessments related to potentially impactful algorithmic systems are regularly carried out and submitted for independent expert review and inspection. In circumstances where the human rights impact assessment identifies significant human rights risks that cannot be mitigated, the algorithmic system should not be implemented or otherwise used by any public authority. When deployed algorithmic systems pose risks to fundamental rights, their use should be discontinued at least until substantial risk mitigation measures are in place. Human rights impact assessments conducted by or for States should be publicly accessible, reviewed by subject matter experts, and followed by additional assessments after implementing risk mitigation measures. Companies deploying AI systems should make available an application programming interface (API) or other technical capability to enable "legitimate, independent and reasonable tests" for "accuracy and unfair performance differences across distinct subpopulations."[16] In addition, developers and deployers of AI systems should disclose regular reports about bias regarding any service with the potential to generate discriminatory effects.

---

[15] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, International Data Privacy Law, (December 2016), https://www.researchgate.net/publication/312597416_Why_a_Right_to_Explanation_of_Automated_Decision-Making_Does_Not_Exist_in_the_General_Data_Protection_Regulation.

[16] Washington State, 66th Legislature, SB 6280, Passed by the Senate March 12 2020, http://lawfilesext.leg.wa.gov/biennium/2019-20/Pdf/Bills/Senate%20Passed%20Legislature/6280-S.PL.pdf?q=20200331083729.

Developers and deployers of AI systems should be using a human in the loop approach. Human in the loop means that humans are directly involved in training, tuning, and verifying the data used in a machine learning algorithm. This allows groups of experts with specialized knowledge to correct or fix errors in machine predictions as the process develops. In this way, humans are more actively involved in making normative judgements about the output of an AI system, rather than offloading decisions to the model.

While developers should regularly audit their AI systems, they can also build those systems in a way that facilitates third-party audits. One way companies do this is through open data archives. We are already seeing progress in this area with political ads, platforms like Facebook,[17] and Google[18] have developed open political ad libraries that provide information about who paid for an ad, how it was targeted, the size of the audience that saw it, and other information. Although even a high degree of transparency cannot substitute for some kinds of regulatory protections, CDT supports transparent communications from competent agencies and companies as one element of good public policy.

Legislation should require companies that offer AI-supported services to provide documentation that explains the overall capabilities and limitations of the technology in terms that consumers can understand. Moreover, AI developers should share more information about known or potential flaws in their systems, including anonymously through collaborative networks, to benefit the overall progress of the technology. To improve the consistency of such reporting, the Commission should consider steps to standardize reporting metrics. Standardized reporting metrics, such as “data sheets” for datasets and model cards, would help create a baseline against which authorities and other reviewers can make more informed and accurate comparisons among different systems.[19] The Commission may wish to consider additional mechanisms in pursuit of verifying claims for

[17] *See,* Facebook, Facebook Ad Library, https://www.facebook.com/ads/library/.
[18] *See*, Google, Google Transparency Report, https://transparencyreport.google.com/political-ads/region/US.
[19] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford, *Datasheets for Datasets*, (2018), available at: https://www.fatml.org/media/documents/datasheets_for_datasets.pdf; While model cards could help to understand some ML systems, they might not be applicable to other types of systems, such as neural nets. See Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T., *Model cards for model reporting*, (January 2019), Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229). ACM, available at: https://arxiv.org/abs/1810.03993.

compliance purposes, such as those described in the paper "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims."[20]

In line with the importance given in the White Paper to the key concept of transparency, we believe that If private entities are contracted to deploy AI systems for the public sector, they must be subjected to the highest levels of scrutiny and transparency. The protection of trade secrets must not undermine public sector transparency and citizens' trust.[21] Moreover, to build trust, AI applications should be designed so that people can understand how decisions are generated. Scholars and researchers have proposed different ways to create 'explainable AI,'[22] such as decision trees[23] or adding a layer of interpretability on a complex model. Not all systems need the same levels of transparency. For example, product recommendation systems require a lower level of explainability than applications such as medical diagnosis systems.[24]

While much can be done during the development process to address potential sources of bias, inaccuracy, or inconsistency in automated systems, no AI can be assumed to remain free of these issues once put into use. These systems require regular auditing to ensure that they are accurate, fair, predictable, and in compliance with legal and ethical standards. Internal and independent evaluators should be able to examine components of an automated system, such as its training data, and inputs and outputs. Robust auditing is necessary to foster trust and confidence in AI technologies and protect civil liberties and fundamental rights.[25] In addition to scrutinizing inputs and outputs from an automated system, auditing techniques may also examine the datasets used to train an AI system and

---

[20] Miles Brundage, et al., *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims*, (April 2020), https://arxiv.org/pdf/2004.07213.pdf.

[21] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies), https://www.statewatch.org/news/2020/apr/coe-recommendation-algorithms-automation-human-rights-4-20.pdf

[22] Hamon, R., Junklewitz, H. and Sanchez Martin, J., *Robustness and Explainability of Artificial Intelligence*, Publications Office of the European Union, Luxembourg, (2020), https://ec.europa.eu/jrc/en/publication/robustness-and-explainability-artificial-intelligence.

[23] Jaime Zornoza, *Explainable Artificial Intelligence*, Medium, Towards Data Science, (April 15, 2020) https://towardsdatascience.com/explainable-artificial-intelligence-14944563cc79.

[24] Ron Schmelzer, *Understanding Explainable AI*, Forbes, (Jul 23, 2019), https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/#347e8ee57c9e.

[25] The White House Office of Science and Technology Policy, *American Artificial Intelligence Initiative: year one annual report* (February 2020), https://www.whitehouse.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf.

the weights given to distinct variables in a system's statistical models. For additional recommendations to standardize the assessment of AI systems, please see our comments to the United States National Institute for Standards and Technology.[26]

Although no single element of a system can provide a perfect picture of its performance or its impacts, more visibility into AI systems enables more forms of independent evaluation which yield more insights into how systems make decisions and the potential biases or inaccuracies that may creep into the decision making process. Ultimately, auditing, testing, and refining systems will improve the overall quality of the technology and should reduce the number and degree of flaws. CDT believes that a European board can play a vital role in this process by identifying how national agencies are overseeing the development, deployment, and audits of systems and highlighting those lessons for broader regulation and engagement on AI.

For more information, please contact:

Pasquale Esposito, European Affairs Associate, Center for Democracy and Technology, pesposito@cdt.org

Stan Adams, Deputy General Counsel & Open Internet Counsel, Center for Democracy and Technology, sadams@cdt.org

[26] Comments of the Center for Democracy & Technology in response to The National Institute of Standards and Technology Request for Information: Developing a Federal AI Standards Engagement Plan (June 10, 2019), https://cdt.org/insights/comments-to-nists-request-for-information-on-developing-a-federal-ai-standards-engagement-plan/