PARTNERSHIP ON AI

June 13, 2020

Dear Commission President von der Leyen,
Dear Executive Vice President Vestager,
Dear Commissioner Breton,

The Partnership on AI (PAI, The Partnership) is a U.S.-based global nonprofit organization dedicated to research, action, and education on the responsible development and deployment of artificial intelligence (AI) technologies. Its mission is to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society. PAI convenes more than 100 partner organizations from locations around the globe, including nonprofit and civil society organizations, commercial entities, and academic institutions.

PAI staff are pleased to offer this response to the draft White Paper, "On Artificial Intelligence - A European approach to excellence and trust." This document should not be taken as stating the view of any particular PAI member organization, nor as representing a consensus view of our partner organizations.

Prior to the pandemic and global protests, the Commission outlined bold timelines for putting in place a regulatory regime to promote the uptake of AI and address the risks associated with uses of this technology. The tensions are real; AI is being deployed rapidly, and at times negatively impacting people's lives. PAI believes, however, that it will take years to develop practices that result in trustworthy AI, and those practices will need to be tested and revised over time to reflect the rapid pace of innovation and changes in society. We are taking a deliberate approach, in cooperation with our partners, to develop inclusive practices that we hope will reduce inequities for people.

The emergence of the global pandemic and protests prompted PAI to re-evaluate the white paper. The pandemic has further exposed the depths of long-standing inequities across the globe and between and among regions, states, and cities. Inequality touches all aspects of society, and in the technology sector the lack of educational opportunities, and the financial means to access technologies and broadband, are particularly acute. Today's new reality for people across the globe places a special sense of urgency to address inequality. And addressing these questions - through budgetary and other actions - will understandably dominate the Commission's attention for the coming months.

While AI may play a role in helping to reduce these inequities, the deployment of some systems can increase inequality. Rapidly incorporating AI systems in public or other sectors, as noted in the white paper, without carefully considering the risks of doing so, risks magnifying inequities. Careful deliberation regarding who is involved in designing and deploying these products, who they are designed to benefit and whether they might directly or indirectly restrict access to economic opportunities for entire groups of society, is absolutely essential to ensure inequities are not further exacerbated by technology. Those who design and deploy these systems are generally not those who are negatively impacted by this technology.

PAI appreciates the Commission's commitment to consulting stakeholders at several stages in the process of implementing its AI strategy. PAI, too, is focused on identifying and including multidisciplinary stakeholders in our work, and emphasizes the importance of including those most impacted by technology. Today's new reality reinforces how vitally important it is to meaningfully include representatives of affected communities - citizens (also migrants, refugees, and permanent residents, incarcerated individuals), and including racial and ethnic groups; women and non-binary people; the LGBTQI community; children; indigenous people; low-income people; and those with disabilities - in essence, people whose life experiences have been marginalized or ignored by technology and technology policy, and who have not been consulted.

PAI has much to learn in this area. We are hiring a research fellow to help us investigate how to meaningfully and nonextractively include people who are most affected by technology in these discussions, and to ensure they realize value from these technologies. PAI aims to produce methodologies and protocols for inclusion that are replicable so that inclusion becomes a systematic and sustainable part of the way AI is built. This research, and the resulting deliberations, are essential to ensure this expertise and needs are not overlooked when developing the next generations of technologies and infrastructures, including AI.

PAI looks forward to sharing our work with the Commission, including research, best practices, and informational materials developed in concert with our partners, to fulfill our shared goals of ensuring AI benefits people and society. We have provided feedback and additional information in the attachment below, and look forward to an opportunity to discuss our response with the Commission at your convenience.

Regards,

Lisa Dyer
Director of Policy, on Behalf of Partnership on AI Staff

Attachment: Partnership on AI Specific Comments to "On Artificial Intelligence - A European approach to excellence and trust."

Specific Comments to
"On Artificial Intelligence - A European approach to excellence and trust."

Overall Comments

Europe is a global leader in digital technology and policy. Key industries and technological advancements are emerging from and within Europe's borders. The white paper notes the depths of Europe's talent pool, and PAI is fortunate to say that many of those experts are represented in our partner organizations. From a regulatory standpoint, Europe's values and views on fundamental rights influence discussions among and within governments, academia, civil society, industry, and the media. Notably, the General Data Protection Regulation (GDPR) elevated discussions regarding privacy and the handling of personal data to a global level. The white paper further notes Europe's involvement in global AI activities at organizations such as the Organization for Economic Cooperation and Development and in the G-20, and the recommendations from the High-Level Expert Group (HLEG) on Artificial Intelligence. PAI believes Europe is, and can remain, a global leader in artificial intelligence.

The Partnership shares the Commission's desire to bring an ethical approach to artificial intelligence. Four representatives of PAI partner organizations served on the HLEG. The Ethical Guidelines published by the HLEG, and the seven requirements for trustworthy AI systems identified by the Group, are consistent with the aspirations expressed by many PAI partners. The Partnership's research, programs, and policy initiatives to develop frameworks, best practices, and informational materials are intended to advance the ethical and responsible development and deployment of AI systems.

Much of our work centers on defining what it means in practice to ensure human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability. *We have found that it is not only essential, but also critical, to explore the nuances of putting into practice such requirements to ensure outcomes that can benefit people and society*. Very often, the last mile between requirements and practice is a unique terrain that needs to be traversed in conjunction with the people actually impacted by these requirements. It is also essential to recognize the complexity of developing these practices, and to involve stakeholders

representing research, engineering, product, policy, civil society, nonprofits, communications and media, academic institutes, and the public.

It is with these insights that we question whether the pursuit of European data pools is relevant to Europe's goals to lead *globally*, and purposefully, in the ethical, responsible, and human approach to artificial intelligence. *Access to data is not enough to ensure trustworthy AI. It is more, if not equally, important to know how the data was collected, if consent was obtained, what data is found in the datasets, whether that data is representative and fit for purpose, rather than focusing on the data's geographic provenance.* The HLEG's seven requirements for trustworthy AI, which focus on related issues such as reducing bias and increasing transparency, can represent important components of the Commission's strategy.

PAI recommends the Commission offer additional details and discussion in its strategy:

- How will the Commission continue to engage citizens and external stakeholders to obtain feedback regarding the white paper? Many countries expect people to *react* to proposed rulemaking through online sites. If governments continue to ask people to react, an elite group of developers, civil society organizations, policymakers, and others, ends up talking to each other, rather than the government truly engaging its citizens. In addition to the Commission's traditional practices of collecting feedback, additional *proactive outreach* will be necessary to reach all stakeholders. Furthermore, policymakers should provide accessible, easy-to-understand information to the public as to how these systems work for these engagements to be productive.
- Implementation, compliance, and enforcement is unclear. Rulemaking in and of itself is very difficult. Far more difficult is actually *implementing* the rules. How many people will you need to create a quality and timely regulatory environment? What expertise will they require, namely from oversight institutions? Will their process be documented so they can be examined and audited if questions or concerns arise? Answers to these questions can help to answer questions surrounding whether it is possible to create an environment of "legal certainty for AI innovation."
- What informs the Commission's belief that it is possible to establish a regulatory framework that builds trust *and* speeds up the uptake of technology? Surveys indicate people are seeking less paperwork, less complex processes, to work effectively with their governments. Why will this regulatory infrastructure be different?

- The white paper did not address the uncertainties associated with AI and the future of work, and PAI believes this is an important component to consider, particularly when contemplating any revisions to the Coordinated Plan. What will the future of work, particularly AI-enabled or complemented work, look like, especially in light of COVID-19? What measures should be put in place to ensure AI advancement acts as an economic "tide that lifts all boats" instead of widening and entrenching existing inequalities in access to quality jobs and income sources? What government actions and societal structures (i.e. portable benefits schemes) are needed in order to enable decent working conditions in the growing online platform economy? What do future work prospects look like for those with limited access to the internet and technology, if remote work and distance learning become the standard? Robotics have emerged as tools to support the pandemic response; what does this mean for future employment opportunities? How will people displaced be measured? What support can be provided for them?

Comments: An Ecosystem of Excellence

The research and innovation investments outlined in the white paper represent important opportunities to advance the field of artificial intelligence, in Europe and across the globe as a spillover effect. Research indicates government and private investments in innovation impact the long term economic growth and productivity of countries. As of this writing, it remains to be seen how the financial interventions undertaken by countries around the world in light of COVID-19 will impact budgetary priorities, including in research and development. The proposed recovery plan's continued emphasis on the digital economy is encouraging.

The white paper's proposal to continue public sector investments in healthcare, to take advantage of Europe's leadership in this area and where technology is poised for large-scale deployment, is particularly relevant in light of COVID-19. A newsletter from deeplearning.ai notes the challenge of "...building AI systems that are robust to real-world conditions." The question posed: Why do AI systems diagnose medical images with accuracy rates commensurate or higher than those of radiologists, but they are not yet widely deployed? The author, Andrew Ng, posits it is necessary to address robustness in a more structured way, to improve the performance of AI systems on new datasets and environments. A recent paper finds issues with instability in image processing when using machine learning methods, for instance. PAI notes also the importance of interpretability and explainability when considering AI's role in

healthcare, which can enable robustness, carefully examining whether these systems are producing biased outcomes. So, too, does involving patients and their loved ones in healthcare decisions. Together, they are a powerful combination: one needs to create processes for people to understand and work well with AI systems. An AI system that is understandable to people can make AI more robust by making sure that people benefit from the AI's recommendations and decisions. PAI highlights these concerns in light of the interest in advancing AI in healthcare in Europe, the potential for AI in healthcare to be considered a high-risk application, and the AI community's efforts to contribute solutions to address the global pandemic.

Comments: An Ecosystem of Trust

PAI also shares the Commission's goal of creating an ecosystem of trust. "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims" resulted from a workshop co-organized by OpenAI, the Leverhulme Center for the Future of Intelligence, and PAI. The paper analyzes and recommends improvements to ten institutional processes, hardware, and software mechanisms that can improve trust between AI developers and other stakeholders such as users, policymakers, and civil society. We believe this document can offer ideas that can be further explored by the Commission.

PAI has embarked on a number of research projects that align with the seven requirements outlined by the HLEG for trustworthy AI. We hope these projects, developed in conjunction with our expert partners in civil society, industry, and academia, can inform the Commission's vision for trustworthy AI:

- PAI's experts in collaborations between people and AI systems note, "AI systems promise to augment human perception, cognition, and problem-solving abilities. They also pose risks of manipulation, abuse, and other negative consequences, both foreseen and unintended." PAI and its expert partners created a framework consisting of 36 key questions to examine when thinking about "human-AI collaboration." This team also published seven case studies to demonstrate the variety of these collaborations in the real world. Together the framework and case studies can help policymakers identify the nuances associated with different AI technologies, and the key elements of responsible AI development and deployment.
- PAI is leading a multi-year, multi-stakeholder effort to scale documentation practices for machine learning systems. ABOUT ML is designed to increase transparency in these systems by convening multilateral discussions about what

information stakeholders need to know about these systems, how to design internal processes to create this documentation, and ultimately driving adoption and use of documentation to improve transparency and accountability in machine learning development and deployment. Documentation questions can focus on many priority topics for the practice of responsible AI, such as interpretability, explainability, fairness, robustness, etc, which align well with regulatory priorities. One eventual goal is to help inform regulatory processes, including those being contemplated by the Commission, with a rough consensus view on what documentation questions are both important and feasible to answer. This process is open for public comment, and we invite the Commission and other parties to share their insights to improve this initiative.

- "Explainable AI" is being cited as a potential solution for addressing a range of AI ethics concerns, ranging from reducing bias to improving transparency by providing insights into how an algorithmic decision was made. PAI authored a first-of-its-kind research paper that determined most explainability solutions currently deployed are designed for machine learning engineers to determine how to debug the model, rather than for end users to better understand how automated systems arrived at a decision. The findings of this paper can inform the Commission's approach to information provision, as the paper highlights shortcomings with current explainable AI approaches that might prevent these techniques from providing greater transparency into how an AI system works. In addition to highlighting specific technical limitations, the paper recommends that organizations consider what policymakers, end users, and other stakeholders need to know, and why, to better understand and evaluate these automated decisions. PAI looks forward to working with the Commission and other policymakers to determine the role of explainability in information provision requirements.
- PAI is leading a research project that seeks to answer the question, "When and how should demographic data be collected and used in service of algorithmic bias detection and mitigation?" At a convening PAI hosted with experts from industry, civil society, and academia, PAI partners cited lack of clarity, across geographies and jurisdictions, around the accessible uses of demographic data to reduce bias.

  In particular, many partners emphasized the inherent tension between bias mitigation and privacy considerations. Detecting or mitigating bias will often require collection and usage of sensitive personal data. Indeed, this tension is reflected in the white paper itself, which instructs AI system developers to "use

data sets that are sufficiently representative" while also adhering to GDPR's protection of privacy and personal data.

Moreover, proxy variables were not seen as acceptable alternatives for demographic data; for instance, the GDPR definition of "sensitive personal data" and the processing of sensitive personal data used to predict demographics raises questions concerning the legal viability of doing so. We would appreciate the opportunity to interview Commission members to inform our study.

- PAI works closely with global actors on the frontlines of information integrity challenges with those building and deploying technology through its AI and Media Integrity priority area. PAI works to mitigate the impact of mis- and disinformation, promote the integrity of information and content online, and ultimately, ensure high quality public discourse. The AI and Media Integrity Steering Committee is a formal body consisting of nine PAI partner representatives developing and advising projects that strengthen mis- and disinformation solutions, including detection of manipulated and synthetic content. Through the group's oversight of the Deepfake Detection Challenge, a machine learning challenge to incentivize development of technical solutions that detect AI-generated videos, the group advocated to expand access to synthetic and manipulated media detection tools to journalists, civil society organizations, and fact-checkers working to verify media globally. The group also emphasized the need for research into trusted explanations and how audiences make sense of labels and other mechanisms for disclosing content manipulations online.

Comments: High Risk Systems and Remote Biometric Identification Systems

PAI notes the Commission's two-pronged consideration of what constitutes a high-risk system, but questions whether it is possible to achieve the Commission's goal: "The determination of what is a high-risk AI application should be clear and easily understandable and applicable for all parties concerned." Many interpretations of risk vary, depending upon the organization evaluating the intended uses of the system. Further, we recommend evaluating other potential risk frameworks, such as that put forward by Germany, to see if they offer specificity beyond the Commission's two criteria. We would appreciate the opportunity to contribute to discussions that will further specify which entity will decide which systems are "high-risk," offer a greater

level of prescriptiveness in the requirements, how such requirements will be applied, how consistency with regards to GDPR will be obtained, and other considerations.

PAI has found that first understanding how these technologies work can bridge difficult conversations between policymakers, the public, and AI developers regarding the appropriate role of such systems in society. Taking the time to make sure all stakeholders understand how these systems work is an important first step, and is fundamental to enable domain experts to identify second-order effects and long-term systemic risks. Without this information, the discussions surrounding the sectors and their applications can be fraught; not all may agree whether a particular sector or application is considered risky.

The white paper notes in particular the risks associated with remote biometric identification systems. The recent protests highlighted concerns about the uses of technology to monitor people peacefully assembling in cities around the world. One such technology, facial recognition systems, are frequently cited by the media and by civil society organizations as one of the most pervasive remote biometric identification systems. Cameras are everywhere; it is often difficult to discern when images or video are being used in by officials to monitor people in public places.

PAI's Facial Recognition Project illustrates the complexity associated with evaluating the appropriateness of intended uses of high risk systems. PAI and its partners produced an interactive graphic and a paper that is designed to demystify facial recognition systems and provide a common language for policymakers and other stakeholders to use when discussing and evaluating their capabilities as they consider regulating the uses of these systems. *The paper additionally includes a list of questions that policymakers can use to elicit additional technical and related information about facial recognition systems. Many of these questions are not unique to facial recognition systems, and PAI hopes that they can inform the Commission's deliberations concerning the potential regulation of high-risk systems.* As of this writing, and subsequent to the protests that emerged after the murders of Ahmaud Arbery, Breonna Taylor, and George Floyd - and far too many others - three PAI partner organizations have decided to end or put in place a time-bound moratorium on their sales of this technology, including to high-risk sectors such as law enforcement, pending regulations governing its use.