The Partnership on AI

Jeffrey Heer (University of Washington; jheer@uw.edu)

The Partnership on AI to Benefit People and Society (or PAI) is a consortium of industrial, non-profit, and academic partners for promoting public understanding and beneficial use of Artificial Intelligence (AI) technologies. Primary goals of the PAI are "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 membership includes representatives from over 70 partner organizations (https://www.partnershiponai.org/partners/), spanning nine countries (primarily in North America, Europe, and Asia) and including technology companies of various sizes, academic institutions, and a number of non-profits (including ACM). The Board of Directors represents a number of large tech companies (Amazon, Apple, Facebook, IBM, Google/DeepMind, Microsoft) as well as non-profits (ACLU, MacArthur Foundation, OpenAI) and academics (Arizona State University, Harvard, UC Berkeley). The founding co-chairs are Eric Horvitz (Microsoft) and Mustafa Suleyman (DeepMind), with day-to-day operations of the PAI run by a dedicated staff in San Francisco, CA.

The PAI's research and outreach efforts center around six thematic pillars:

- Safety-Critical AI
- Fair, Transparent, and Accountable AI
- AI, Labor, and the Economy
- Collaborations Between People and AI Systems
- Social and Societal Influences of AI
- AI and Social Good

For more details regarding each thematic pillar, see https://www.partnershiponai.org/about/#our-work.

The PAI kick-off meeting was held in Berlin in October 2017, with wide-ranging discussions organized around the above pillars. A primary of outcome of the meeting and subsequent membership surveys was the formation of three initial working groups focused on "Safety-Critical AI", "Fair, Transparent, and Accountable AI", and "AI, Labor, and the Economy". Each of these groups are now conducting a series of meetings to develop best practices and resources for their respective topics.

I am a member of the Fair, Transparent, and Accountable AI (or FTA) working group, co-chaired by Ed Felten (Princeton) and Verity Harding (DeepMind). The charter of the FTA working group is to study issues and opportunities for promoting fair, accountable, transparent, and explainable AI, including the development of standards and best practices as a way to avoid and minimize the risk that AI systems will undermine fairness, equality, due process, civil liberties, or human rights. The FTA group has held two in-person meetings so far: the first in London in May 2018 and the second at Princeton in August 2018. The initial meeting resulted in the formation of four specific projects.

The "Papers" project aims to produce a set of three educational primers, one each for fairness, transparency, and accountability. The goal is to provide a starting point for people to learn about fairness in AI, how it is conceived, and what people are currently doing about it. The purpose is not just to come up with definitions, but provide structure for how to think about problems of fairness, transparency and accountability in the context of the work that Partner organizations are conducting.

The "Case Studies" project focuses of identifying multiple case studies that illustrate how FTA concerns can be addressed, as well as develop a framework for the analysis of such case studies. Meanwhile, the "Grand Challenges" project is working to create challenges or endorse existing contests developed around FTA issues in practice.

The "Diverse Voices" project seeks to understand different communities, how their voices can be involved in tech policy making, and the impact of AI deployment. The project intends to engage with these groups in a meaningful manner and ensure under-represented groups' voices are heard. One goal is to provide guidance for outreach, as many companies are interested in deploying ethical AI but are unsure how to address the problem.

Each of these projects is a work-in-progress, with additional milestones and meetings planned over the coming year.

Jeffrey Heer is a Professor of Computer Science & Engineering at the University of Washington and Co-Founder of Trifacta Inc., a provider of interactive tools for scalable data transformation. His research spans human-computer interaction, visualization, data science, and interactive machine learning. He also serves as an ACM representative to the Partnership on AI.