Fair, Transparent, and Accountable AI: Charter

The **Fair, Transparent, and Accountable AI (FTA) Working Group** aims to encourage the development of AI systems that perform effectively while respecting notions of fairness, transparency, and accountability. The group will explore interpretations of fairness, transparency, and accountability as applied to AI in order to foster discussion of these values and possible tensions with other goals of AI decision-making (for instance, tradeoffs with predictive accuracy). This work is toward the Partnership on AI's broader vision of benefiting people and society by helping to shape the development and consequences of AI through a multistakeholder perspective.

We hope to provide a practical toolkit for those engaged in AI development and deployment to engage in responsible behavior. We also hope the FTA group's work will inform discussions by the media and policymakers; we will provide input where requested and appropriate, though it is not our intention to directly recommend legislative or regulatory changes.

The FTA group's work is founded on two premises:

Our first premise is that, through techniques like identifying underlying patterns and drawing inferences from large amounts of data, AI has the potential to improve decision-making capabilities. AI may facilitate breakthroughs in fields such as safety, health, education, transportation, sustainability, public administration, and basic science.

Our second premise is that there are serious and justifiable concerns—shared both by the public and by specialists in the field—about the harms that AI may produce. As a group coming from communities shaping AI systems (such as civil society, technology, and academia), we have a responsibility to address these concerns.

One potential harm is **unfairness**. Analyses drawing on data that embody hidden biases may reinforce those biases and analyses drawing on a lack of data about particular groups may cause bias in practice; indeed, analysis techniques themselves may also be biased. In an unfair society, even accurate data may be used to support unfair treatment if practitioners single-mindedly pursue predictive accuracy without regard for broader societal consequences and human rights.

Distinct stakeholders will interpret fairness and its relationship to societal goals such as justice differently. Thus, we must foster discussion about divergent interpretations.

A second potential harm is **nontransparency**. Opaque AI may face limitations in human trust of the systems and their outputs, in the troubleshooting and debugging of systems and their applications, and in their ease of application to new areas, among other challenges. Complete transparency of all code and data will not always be possible, but degrees of transparency and methods for enabling systems to be analyzed are important to consider as AI systems become more prevalent.

A third potential harm is a **lack of accountability**. AI systems that make consequential decisions about people must be accountable to affected parties in modes akin to traditional accountability for design choices and decisions made. In all cases, AI applications should be accountable for supporting human rights. Accountability depends on a mix of technical, administrative, legal, and policy measures for which it is important to convene multistakeholder voices and ensure both responsible uses of AI and appropriate redress when due.

Researchers, government officials, business and civil society representatives, and the public should be responsive to these challenges of fairness, transparency, and accountability. Technologists should seek to develop methods that detect and correct errors and biases, promote transparency, and enable accountability, while also fostering AI efficacy.

Examples of the group's work may include:

- Consultations with stakeholders, to ensure that important interests and viewpoints are represented, especially those of groups that have historically been victims of discrimination;
- Briefing papers that summarize and introduce ideas and work in areas related to fairness, transparency, and accountability;
- Representative use cases that can serve as frameworks for discussion;
- Case studies, developed or curated by the working group;
- Codes of best practice or procedural recommendations to help organizations and technology developers put good intentions into practice; and
- Comments to governments and other bodies that solicit input from the public.

The group's work will be organized through periodic in-person meetings, audio/video conferences, and other collaboration tools among members of the Partnership on AI.