



Support for Timnit Gebru and Google’s Ethical AI Team

For the last three years [AFOG at UC Berkeley](#) has benefited from the participation of several members of the Ethical AI team at Google in our working group meetings and workshops. Recently Dr. Timnit Gebru who co-leads the team was fired from her position over her request for greater transparency and procedural fairness in the internal review process of the research produced by herself and her team members.

Members of the AFOG group do research on the role of digital technologies in society and seek ways of building better and more equitable systems. Many decades of research in the history of technology shows that technologies are inescapably *political*. They structure and mediate relationships between people in distinctive ways. While digital technologies have great potential for benefit, there are always downsides, risks, and forms of harm to consider. Those risks are not evenly distributed, but often follow the lines of existing power hierarchies.

Based on our committed research in this domain, this is what we know:

- **Technology’s capacity to do harm must be fully confronted.** “Sunlight is the best disinfectant,” as Supreme Court Justice Louis Brandeis once said. It is important to specifically support the individuals and groups that are most vulnerable to harm from the creation of new technologies. We laud the efforts of the Ethical AI team to do this hard work. Dr. Gebru’s efforts with her collaborators to publish work on large language models is fully in line with this important goal. To address harms they must be acknowledged fully, not downplayed or dismissed for opaque procedural reasons.
- **There is no “[view from nowhere](#).”** We must recognize that all technologies, systems, algorithms, or products are entangled inescapably in human values. The perspectives of researchers and designers -- life experiences, motivations, and the human, social, and material resources they have access to -- shape the questions they ask, the methods they use, and the solutions they build. No demographic group has the advantage in being more “objective.” Diverse research teams such as the one Dr. Gebru helped build at Google are essential to the identification, acknowledgement and mitigation of harms.
- **Listening to members of non-dominant groups is not enough.** There is a long history of scholarship in Academia and Industry that considers the needs and priorities of technology “users” while the power to enact those priorities remains in the hands of a select few. The field of AI research as a whole reflects a skewed demographic, white and Asian men, while AI *ethics* research draws many more women and non-white researchers. This is reflected in the composition of the Ethical AI team led by two women

including Dr. Gebru who is Black. Members of non-dominant groups are too often invited in but ultimately marginalized or removed, as Dr. Gebru was, once they push back or question the status quo. True inclusion means experts like Dr. Gebru and her team are able to [drive decision-making](#) -- about research agendas, technology designs, and institutional practices -- and [not just be consulted](#). Ultimately change requires that dominant groups cede power. Institutional commitment must be embodied in practices and processes to enact meaningful change.

- **Knowledgeable and powerful insiders have a key role to play.** Research in multiple fields underscores the important role expert insiders play in identifying harms and risks of corporate work and bringing attention to social values within the firm. Google must adopt structures that support the voice and independence of the Ethical AI team, and others doing values work within Google. Shaping Google requires the Ethical AI team to meaningfully participate in high-level policy decisions. It also requires routine internal practices that successfully build insights of the Ethical AI team's work into products. External structures that support and elevate their work and hold Google accountable for acting on the knowledge they produce are essential to the Ethical AI team's success. We recommend the establishment of an external advisory board comprised of experts from the interdisciplinary community engaged with these issues *chosen* by the Ethical AI team themselves. Ethics and values work within firms is complicated. It requires advancing the interests of society against, between, and intertwined with corporate profits. Researchers doing this work within firms will cause trouble, but it is the "good trouble, necessary trouble" required to build systems aligned with justice that Google and other technology companies ought to embrace.

We [support the call of the Ethical AI team](#) to confront power structures within their institution that reproduce the marginalization of non-dominant groups from technology design.

Members of the AFOG Group:

Dr. Jenna Burrell, Associate Professor, School of Information, UC Berkeley

Deirdre K. Mulligan, Professor, School of Information, UC Berkeley

Niloufar Salehi, Assistant Professor, School of Information, UC Berkeley

Celeste Kidd, Assistant Professor, Department of Psychology, UC Berkeley

Morgan G. Ames, Assistant Adjunct Professor, School of Information, UC Berkeley

Dr. Razvan Amironesei, Research Fellow in Applied Data Ethics, University of San Francisco
Data Institute