THE GREENLINING INSTITUTE

The Greenlining Institute
360 14th Street, 2nd Floor
Oakland, CA 94612
www.greenlining.org

April 8th, 2025

Re: Joint California Policy Working Group on AI Frontier Models

Dr. Fei Fei Li, Dr. Mariano-Florentino Cuéllar, and Dr. Jennifer Tour Chayes,

The Greenlining Institute appreciates the opportunity to provide feedback on the Joint California Policy Working Group on AI Frontier Models' Draft Report. For over 30 years, the Greenlining Institute has worked towards a future where communities of color can build wealth, live in healthy places filled with economic opportunity, and are ready to meet the challenges posed by climate change. Inevitably, each of these efforts will be profoundly shaped by the impacts of generative artificial intelligence—likely, in ways that we can only begin to anticipate.

We believe that in order to create a more just future for all Californians, equity must be at the forefront of our technological revolution: we've worked to increase broadband access to rural and formerly redlined zip codes throughout the state; our work in financial accountability has led to millions of dollars in community reinvestments and the creation of the Department of Financial Protection and Innovation; our research in algorithmic bias has informed policymakers on the development of responsible procurement standards.

The CA AI Working Group Report offers findings that will guide policymakers in our shared goal of creating a just and equitable AI-driven future. We commend the group's dedication to a responsible, transparent, and proactive approach to AI governance. Our feedback includes recommendations to ensure that this technological transition benefits all Californians, including our most vulnerable communities.

Third-Party Assessments Include Community Consultation

While third-party assessments are indeed essential to identifying these frontier models' vulnerabilities and discriminatory biases, they are not enough. Community consultation is essential for building models that actually serve community needs.

Government agencies and policymakers should ensure that generative models are co-designed alongside the communities most affected by their implementation. This is essential for ensuring that models are properly stress-tested, downstream impacts can be anticipated *prior* to deployment, and developers are held accountable to equitable human-centered design. For

The Greenlining Institute
360 14th Street, 2nd Floor
Oakland, CA 94612
www.greenlining.org

example, a model that has been designed to support care workers ought to incorporate consistent community consultation with laborers, older adults, and care recipients of color.

Community consultation allows developers and lawmakers to anticipate potential labor impacts, discriminatory biases, and technical vulnerabilities before any harm takes place. Such was the case at DEF CON 31 in 2023, when a group of hackers, researchers, and members of civil society participated in red-teaming various generative AI models that were to be integrated into critical government or public-facing systems. They quickly found significant vulnerabilities in the model. With more deliberate and inclusive consultation from lay community members, more of these types of vulnerabilities can be identified early on. The demographic, institutional, and disciplinary diversity that results as a benefit from third-party assessments, moreover, is multiplied tenfold from intentional community consultation.

This process of community consultation should be facilitated through participatory design workshops, where agencies collaborate with community members, advocates, and experts to define success, evaluate whether the system should be implemented, and understand the tradeoffs between different outcomes. This process is essential for establishing trust between developers and consumers, facilitating responsible adoption, continuous monitoring, and proactive feedback mechanisms.

Adverse Event Reporting Should Empower Consumers of Color

Robust and accessible adverse event reporting mechanisms are essential for mitigating vulnerabilities. Users—particularly those coming from communities of color, with veteran status, living in formerly redlined zip codes, living with disabilities, young children, etc.—will be the ones experiencing these adverse events firsthand. Consumers, therefore, must serve as reporting entities.

Excluding consumers from reporting would significantly slow down developers' abilities to identify and correct vulnerabilities. While community consultations and third-party assessments do indeed offer diversity and pluralism in informing the initial development of these models, no amount of stress-testing can ever account for all the possible scenarios of the real world. We know that these models are growing at a rapid, ongoing, and exponential pace. Our accountability mechanisms, therefore, must match this pace. Direct consumer feedback is the foundation for accountable development—saving developers and regulators time and resources in the long term.

The Greenlining Institute
360 14th Street, 2nd Floor
Oakland, CA 94612
www.greenlining.org

It is essential that policymakers make incident reporting accessible and transparent. In order for a consumer to even recognize that they are being subject to an adverse event, they must know that they are interacting with an AI model. Whether that model is a generative chatbot or automated decision making system, developers ought to disclose to users when they are interacting with a model. In these disclosures, consumers should be made aware of their rights. This should include a private right of action that allows for them to opt-out of the system, in addition to instructions on how to report an adverse event.

These adverse event reports should be forwarded to an incident reporting database held by the state. This is essential for detecting patterns of discrimination, particularly as they may appear across different models, across different circumstances. While consumers are responsible for reporting these individual adverse events, the state should be responsible for aggregating the data and actually identifying these events as the result of biased or incomplete training. For example, when a Black woman interacts with a chatbot API designed to provide mental health counseling, she may receive inaccurate or even harmful advice from the chatbot. This event, on its own, does not provide the developer, regulator, or user with any significant information about how to improve the model. If these adverse events are collected and analyzed by one regulatory reporting agency, however, larger trends emerge. It may be revealed that the training data excludes culturally competent mental health data for Black women. This information, combined with the results of any disparate impact assessments submitted by each company to the state's Attorney General, may lead to the determination that the model ought to undergo further training prior to being deployed in high-impact use cases.

Discriminatory events do not occur in a vacuum. It requires zooming out in order for you to really see them for what they really are. As we shape today's regulations for tomorrow's future, policymakers and developers need to maintain this big-picture perspective. Coordinated feedback mechanisms and intentional community consultations are all a part of our larger goal: to ensure that these revolutionary technologies work for the benefit of all. We thank the CA AI Working Group for their commitment to responsible, transparent, and equitable AI governance.

Sincerely,

Angel Lin
Tech Equity Policy Fellow

Mobile: (425) 623-4720
Email: angel.lin@greenlining.org