

# WORKSHOP REPORT

Towards the promise of  
open source AI:  
Co-creating a vision for  
responsibility and  
research roadmap

---

Berkeley AI Research Lab, UC Berkeley  
August 2025



# Table of Contents

---

- I. Executive Summary
- II. About the Workshop
- III. Co-Created Vision
- IV. Actions to Move Towards the Vision & Research Roadmap
- V. Conclusion

# I. Executive Summary

---

On August 12 and 13, 2025, the Responsible AI Initiative of the Berkeley AI Research Lab at the University of California, Berkeley (UCB) hosted a workshop, “Towards the promise of open source AI: Co-creating a vision for responsibility and research roadmap”. The workshop, funded by the National Science Foundation’s Responsible Design, Development and Deployment of Technologies Program (ReDDDoT) with support and in partnership from Mozilla, brought together researchers, practitioners, policymakers, and community leaders to explore what responsible openness in AI means and how it can be realized and sustained. The overarching goal was to co-create a shared vision for “responsible” openness in AI – with a focus on open source foundation models – and chart a research roadmap towards that vision.

This report brings together the outcomes of three interconnected sessions of the workshop resulting in: a co-created vision for responsible open AI, an action plan to move towards the co-created vision, and a research roadmap. Together, the vision, actions, and research priorities create a layered agenda: a north star, practical steps to get there, and a roadmap for research and learning.

## II. About the workshop

### **Why was this workshop held?**

The future of AI is at a critical juncture, with growing debate over whether – and to what extent – foundation models should be open. Open source AI models offer significant potential to advance research, innovation, transparency, and equity. They can help democratize AI by enabling broader access and participation in model development and application. However, open models also present risks and openness alone is not enough. Bad actors may more easily use and exploit them in ways that harm individuals, communities, and society. And whether models are open or closed, key design decisions – such as data sources, training processes, and transparency – remain shaped by the values and market priorities of their largely for-profit driven creators and managers, with limited public input.

Now is the time to ask not just whether models should be open, but how open models can be made responsible – and what openness in AI should mean more broadly – in ways that are democratically informed, transparent, and aligned with the public interest.

## **What is open source and open AI?**

There is no single definition of open source AI, though we follow the Open Source Initiative to define open source AI as “a system made available under terms that grant users the freedoms to use, study, modify and share.” This includes fully open weights, data, and code available for download with no use restriction. Openness is a spectrum, from fully closed to open source. Many models today are open in that they release model weights, but not fully open source as they do not share training data.

In addition to openness being a spectrum for models, openness extends across the entire AI stack: data, code, documentation, standards, and more. The workshop and this report considers actions and research for openness beyond only the model to consider openness across the AI stack.

## **What is the current state of open source and openness in AI?**

The landscape of open source AI is complex and evolving. Frameworks like Stanford’s Foundation Model Transparency Index and the Linux Foundation’s Model Openness Framework reveal that most widely used models sit on a spectrum of openness rather than meeting full open-source standards. For example, industry releases like Meta’s LLaMA-2 and DeepSeek’s R1 remain only partially open whereas academic and non-profit efforts such as MIT’s Boltz-1 and the Allen Institute’s OLMo and Tülu models have fully open-sourced their architecture.

Open models have fueled innovation and collaboration, lowering barriers to entry and creating a more inclusive AI ecosystem. They allow researchers and developers to build on shared work, accelerating experimentation and knowledge-sharing beyond well-resourced corporations. This culture of openness has supported broader access and collective progress in AI. However, once released, open models can be repurposed in harmful ways – from disinformation and fraud to deepfake abuse – with little possibility of recall or control. Closed systems, on the other hand, risk concentrating power among a handful of firms that control the largest models and infrastructure. These dynamics underscore both the promise and complexity of open source AI today, setting the stage for a collective discussion on how openness can be made responsible and aligned with the public interest. For a deeper dive into the current state of open source AI, see the full literature review [here](#).

*Day 1 of the workshop included presentations on the state of openness and open source AI, as well as a panel discussion (see Agenda in Appendix).*

---

## What was the workshop approach?

This invite-only workshop – hosted by UC Berkeley’s Responsible AI Initiative in partnership with Mozilla and funded by the NSF’s ReDDDoT program – brought together researchers, practitioners, policymakers, and community leaders to explore what responsible openness in AI means and how it can be realized and sustained. The overarching goal is to co-create a shared vision for “responsible” openness in AI – with a focus on open source foundation models – and chart a research roadmap towards that vision.

The first day of the workshop included academic presentations and a panel to discuss the current state of openness in AI, tensions, and opportunities. The second day began with welcome remarks by Jennifer Chayes, Dean of the UCB College of Computing, Data Science, and Society; followed by introductory keynotes from Mitchell Baker, Co-Founder of Mozilla, and Vera Zakem, Chief Innovation Technology Officer of California. It then utilized participatory design methods to inform a co-created vision, action plan towards the co-created vision, and research roadmap.

## Who attended the workshop?

Participants include researchers, industry practitioners, policymakers, and community leaders from the following organizations: Brown University, California Department of Technology, California Office of Data & Innovation, Centre for AI & Digital Humanism, City of San Jose, Columbia University, Cornell University (*Digital Life Initiative, Cornell Tech, AI4Abolition*), Creative Commons, Data & Society, Dimagi, Digital Public Goods Alliance, EleutherAI, GitHub, Google, Harvard University (*Institutional Data Initiative, Berkman Klein Center*), Heising-Simons Foundation, Hugging Face, Institute for Technology & Society of Rio de Janeiro, Linux Foundation, Meta, Microsoft, MIT, Mozilla, National Fair Housing Alliance, Open Source Initiative, Partnership on AI, Queen Mary University of London, Rutgers University, Stanford University (*Center for Research on Foundation Models, Institute for Human-Centered AI*), Tech Governance Project, Telepath, The Brookings Institution, United Kingdom AI Security Institute, United Nations Development Programme, University of California-Berkeley (*Berkeley AI Research Lab, Berkeley Institute for Data Science, Center for Responsible Business, Human Technology Futures, CITRIS and the Banatao Institute, College of Computing, Data Science, and Society, School of Information, AI Security Initiative*), University of California-Los Angeles, and University of Texas at Austin.

---

## Who informed the workshop approach?

This workshop approach, agenda, and list of participants was informed by a Steering Committee led by Genevieve Smith (PI, UCB) that includes: David Widder (Cornell University), Ezinne Nwanko (UCB), Jake Metcalf (Data & Society), Matthew Zhou (California Office of Data & Innovation), Meredith Lee (UCB) & Maximilian Gahntz (Mozilla). It is supported by research assistants, Erik Cedarblade & Hiral Patel (UCB). Participants and institutions were identified based on their experience and backgrounds in contributing to discussions, research and practice in open-source and openness in AI. In addition to developing an invitation list of target participants, we also had a Google Form where people could self-nominate to be invited or where other registered or invited participants could nominate attendees. The names shared on this form were then reviewed for alignment and suitability for the workshop, along with space, before invitations were extended.

# III. Co-Created Vision

---

The debate around open source AI often centers on the binary question: open or closed? This workshop moved beyond that dichotomy by first inviting participants to take a step back and “Imagine it’s 10 years from now”. Small groups envisioned the world they wanted to live in, which established a north star for technology development by grounding the discussion of what values should define our future and how AI could contribute to realizing that vision. The collective visions can be categorized into the following pillars:

<b>Pillar</b>	<b>About</b>
<b>Openness and transparency</b>	AI systems and the broader AI stack are developed and shared as part of a commons, with transparency and interoperability at their core. Openness is treated as a spectrum – from fully open source releases partial or conditional openness of models and datasets.
<b>Economic and social justice</b>	There is equitable access to jobs, healthcare, education, and government services – a world where every person has the resources to live a dignified life. AI is a tool that bridges gaps in opportunity, rather than exacerbating unequal distributions of wealth.
<b>Community and human connection</b>	Technology is designed to foster authentic human connections and strengthen communities. Digital tools serve to deepen our bonds, rather than replace or distort our reality of genuine human connection.
<b>Demographic governance</b>	Power is decentralized and decision-making is participatory, ensuring everyone has a stake in stewardship of AI through public oversight and accountability. Trusted institutions ensure technology serves the public interest. Governance that is by the people, for the people.
<b>Agency</b>	A digital world rooted in continuous and meaningful consent, where individuals can choose their level of interaction with technology. People are fully informed about how and when their data is used, where, and have the power to control their personal data on their own terms.

## Pillar (cont'd)

## About (cont'd)

---

### Sustainability

Technology operates in harmony with a thriving environmental ecosystem. We have moved beyond exploitation and extraction of natural resources to develop a green infrastructure that powers the AI supply chain end to end.

---

**In short, we envision a world where people and communities can fully participate in education, work, and civic life, with access to the resources and opportunities they need to thrive.**

**AI technologies – and the broader AI stack – serve to support people (to the degree of their choice) in navigating information and making decisions, while honoring our values, autonomy, and well-being. Openness and transparency are core, though the degree of openness or open source may vary in different models, applications or contexts. The technologies are governed in ways that are democratic, environmentally sustainable, and value-aligned.**

*“We must tap the well of our own collective imaginations, that we do what earlier generations have done: dream. Without new visions we don’t know what to build, only what to knock down.”*

*– Robin D. G. Kelley*

## IV. Actions and research roadmap towards the co-created vision

The workshop surfaced actions and research priorities to move towards the co-created vision, clustered into nine cross-cutting themes (Table 1). The nine themes are presented in a sequence that begins with clarifying definitions of openness and how to measure them, then moves into evaluation and safeguards, before addressing governance, resources, and applications within broader social contexts.

Table 1. Actions and research priorities towards the co-created vision

Theme	Actions	Research priorities
<b>Openness definitions and gradients</b>	Not explicitly surfaced in actions, but foundational across other categories	<ul style="list-style-type: none"><li>• Define measures of openness, tradeoffs, and perceptions across stakeholders</li></ul>
<b>Evaluation and standards</b>	Advance standards, evaluation practices, interoperability, and environmental safeguards	<ul style="list-style-type: none"><li>• Develop benchmarks, open evaluation ecosystems, and frameworks for safety and accountability</li></ul>
<b>Safety, security, and risk</b>	Strengthen safety and security practices, as well as address risks	<ul style="list-style-type: none"><li>• Assess safety frameworks, guardrails, and methods to capture social harms and misuse</li></ul>
<b>Governance and institutions</b>	Establish participatory governance, embed data rights, and strengthen oversight	<ul style="list-style-type: none"><li>• Study effective governance models, roles of public and private actors, and regulatory approaches.</li></ul>
<b>Capacity building and communities</b>	Create platforms and literacy programs to broaden participation and fair recognition	<ul style="list-style-type: none"><li>• Explore training, labor models, and interdisciplinary collaborations that support diverse contributors</li></ul>

Table 1. Actions and research priorities towards the co-created vision (*cont'd*)

Theme	Actions	Research priorities
<b>Infrastructure and resources</b>	Build public infrastructure and “public option” AI systems	<ul style="list-style-type: none"> <li>Investigate models for public compute, open datasets, and resource-sharing across regions</li> </ul>
<b>Funding, economics, and incentives</b>	Develop sustainable funding and fair attribution systems	<ul style="list-style-type: none"> <li>Design public goods funding, contributor compensation, and incentive structures</li> </ul>
<b>Applications and public good</b>	No standalone actions, but implied in public services, health, and climate discussions	<ul style="list-style-type: none"> <li>Assess impacts of open models in local, multilingual, and decentralized contexts</li> </ul>
<b>Social, political, and historical contexts</b>	Broader social and historical issues came through mainly in research needs	<ul style="list-style-type: none"> <li>Identify lessons from past technologies; examine openness in relation to inequality, labor, sustainability</li> </ul>

Throughout the discussions, participants emphasized that while “open source” and “openness” are often used interchangeably, they are not the same. Open source implies full release of code, data, or models. Openness, as discussed here, includes conditional forms that may limit full release in order to protect sovereignty, safety, or consent. Several of the actions and research priorities therefore focus on open source models directly [OS], while many others address how openness as a gradient can be designed and governed to serve people and communities [O], and others work for either. These are denoted accordingly in the actions section and the research roadmap, which the next sections explore in further detail.

**Legend:**

- **OS** = Specific to *open source* (full release of model weights, code, and data).
- **O** = Specific to *openness as a spectrum*
- *Unmarked* = Relevant to both open source and openness more broadly.

## A. Action plan towards the co-created vision

---

In a participatory design session, small groups identified actions to move from today's status quo toward the co-created vision. Importantly, actions extend beyond the "usual" open source playbook by focusing on how openness can serve people, communities, and society. While not exhaustive, these actions are illustrative of how and where progress can be made across the key themes.

### 1. Evaluation and Standards

**Build standards, benchmarks, and sustainability requirements so that open AI systems are interoperable, reproducible, and environmentally responsible.**

*Examples/how:*

- Fund and support pre-standardization work and benchmark development for models.
- Require reproducibility guidelines by formalizing good open source repository standards for model releases. [OS]
- Develop transparency and consent standards for dataset sharing.
- Encourage cross-institutional dataset benchmarks and governance frameworks.
- Develop transparency and consent standards for dataset sharing.
- Embed sustainability criteria in funding, procurement, and oversight.
- Mandate climate impact reporting for training and deployment.

### 2. Governance and institutions

**Move beyond surface-level "community input" by giving communities real power over how open AI systems are governed, owned, and safeguarded.**

*Examples / how:*

- Involve communities directly in model design.
- Allow conditional openness (e.g., limiting access by nation, tribe, or group where sovereignty/safety warrants it). [O]
- Establish permanent multi-stakeholder working groups to escalate issues if/when harms occur and coordinate remediation.

**Ensure strong data rights and community oversight within AI ecosystems, so contributors have meaningful power, transparency, and remedies over how their data is used.**

*Examples / how:*

- Shift from individual, one-off consent to collective data governance, such as by enabling community or worker bargaining over data rights, ensuring fair terms, compensation, and protections.

- 
- Establish the right of withdrawal by making model unlearning or retraining a standard remedy [O].
  - Support collective dataset structures – such as data cooperatives or trusts – that can license, audit, and veto downstream uses of their data

**Strengthen procurement, licensing, and enforcement frameworks so that openness, accountability, and safeguards are built into the way AI systems are funded, licensed, and released.**

*Examples / how:*

- Publish procurement templates requiring open data standards, model documentation, and environmental disclosures.
- Tie government funding to openness commitments (e.g., releasing code, models, and/or data), whether as fully open or under conditional openness with clear reasoning.
- Make openness the norm for publicly funded AI systems [OS/O], while allowing carefully defined exceptions when openness would put vulnerable groups at risk [O].

**Strengthen accountability and oversight in open AI systems by introducing “good friction” – legal, technical, and informational safeguards that preserve responsibility, transparency, and safe use.**

*Examples / how:*

- Establish liability rules and whistleblower protections so insiders can safely disclose risks in training or deployment (e.g., a contributor raising concerns about copyrighted books in the training data)
- Mandate explainability through standardized documentation (e.g., model cards) and impact reporting

### 3. Capacity building

**Expand participation and literacy so people and institutions can meaningfully contribute to and shape the open AI ecosystem.**

*Examples / how:*

- Build participation platforms (e.g., online spaces) that teach AI literacy and support contributors.
- Use universities as hubs that both test participatory approaches (e.g., model auditing approaches, governance simulations) in the classroom and implement public AI education, while contributing to open AI ecosystems.

## 4. Infrastructure and resources

**Develop publicly governed AI infrastructure and services that provide safe, equitable, and non-proprietary alternatives to private platforms.**

*Examples / how:*

- Offer a government-backed, open source AI option with interoperability standards so agencies aren't locked into proprietary corporate systems. [OS]
- Advance open hardware and shared compute resources.
- Provide public sandboxes where organizations can safely experiment with open models and datasets.

## 5. Funding, economics, and incentives

**Align funding and economic incentives so open AI systems are sustainable, fairly compensated, and designed for long-term public benefit rather than short-term engagement.**

*Examples / how:*

- Pool public funds and private contributions to support collectively governed models.
- Compensate all contributors – including testers, translators, documentation writers.
- Publicly finance public open models for high-value applications like climate science, translation for marginalized languages, and civic AI tools.

**Ensure open AI models are designed to be sustainable, transparent, and user-controlled, with clear lifecycle planning and safeguards tied to data rights.**

*Examples/how:*

- Favor small, purpose-built models that are openly shareable, secure, locally runnable, and informed by domain experts.
- Support personalizable, self-trainable, on-device agents that are built on open code and weights so users can adapt and retrain systems under their own control, versus centralized platforms.
- Require published plans for updates, deprecation, and model unlearning to account for changes in data rights or consent.

## B. Research Roadmap

A final session of the workshop explored a research roadmap. Research priorities emerged across the nine themes. Participants emphasized not only what research is most needed, but also how that research should be pursued. Six good practices emerged to guide how such research should be conducted. The sections that follow expand on each priority area (the “what”) with research goals, topics, and guiding questions, before turning to the practices (the “how”).

Some research can and should begin immediately, such as developing new evaluation frameworks and piloting participatory governance models. Other priorities require medium-term institution building, such as sustainable funding mechanisms, standards, and public infrastructure. Still others represent longer-term horizons, including exploring alternative economic and labor models.

*Areas emphasized repeatedly across breakout groups, the most common themes are marked with a \*.*

### 1. Openness definitions and gradients

**Goal:** Clarify what openness means, how to measure it, and when different levels of openness are appropriate

#### **Research topics:**

- Frameworks for measuring gradients of openness across the AI stack
- Conditions in which openness is beneficial versus when restrictions are valuable or necessary [O]
- Public and governmental perceptions of openness and sources of resistance
- Epistemic openness that incorporates diverse forms of knowledge and expertise
- Comparative analysis of open source software and open source AI governance [OS]

#### **Example research questions:**

- *How can we measure “degrees of openness” across data and models?*
- *In what circumstances is partial openness (e.g., open interfaces but closed data) appropriate or necessary?*
- *How do perceptions of openness differ between governments, corporations, and civil society groups?*
- *How can epistemic openness be broadened to include non-technical knowledge systems in shaping open AI?*
- *What can the governance of open source software teach us, and where do those lessons fail, in the context of AI?*

## 2. Evaluation and standards

**Goal:** Develop rigorous, transparent, and socially relevant evaluations, frameworks, and standards to guide responsible open AI development and deployment.

### Research topics:

- Domain- and task-specific benchmarks beyond traditional leaderboards\*
- Standardized frameworks for safety, robustness, and accountability
- Research into marginal costs/benefits of open vs. closed AI, including environmental impacts
- Open evaluation ecosystems that move away from black-box testing\*

### Example research questions:

- *How do evaluation needs differ for open-source models that can be widely modified, fine-tuned, or deployed in different contexts?*
- *How can open evaluation ecosystems balance transparency (e.g., public test suites) with security against misuse?*
- *How do we assess the marginal costs and benefits of open vs closed systems?*
- *How do the environmental costs of maintaining many distributed open models compare with centralized closed systems?*

## 3. Safety, security, and risk

**Goal:** Develop safeguards and frameworks to manage risks specific to open AI models without undermining accessibility.

### Research topics:

- Technical safety and security frameworks
- Guardrail effectiveness and limits against harmful or malicious content [OS/O], including what kinds of guardrails if any are resilient in open source contexts and governance backstops [OS]
- Risk frameworks inclusive of human rights, disability, and civil liberties
- Safeguards against malicious use, bias, and manipulative design
- Evaluation methodologies that capture social harms and benefits

### Example research questions:

- *What unique security vulnerabilities emerge for models at different levels of openness?*
- *How might safety guardrails adapt and evolve, and what are the safety limitations, regarding open-source AI models?*
- *What strategies can reduce the misuse of open models while preserving openness?*
- *How can risk frameworks for open-source AI be more inclusive, such as including disability rights and civil liberties?*
- *How can open evaluation methods capture and address social harms in distributed deployments?*

## 4. Governance and institutions

**Goal:** Identify and test governance models that balance openness with accountability across public, corporate, and community actors.

### Research topics:

- Participatory and collective governance models for open AI (e.g., co-ops, councils, biobanks)
- Mechanisms for honoring a “right to withdraw”, including technical unlearning, licensing constraints, and governance models that define remedies before or after release [OS/O]
- Roles for governments, regulators, and public institutions in stewarding openness
- Corporate governance practices and diversity of thought shaping openness
- Non-regulatory levers for accountability, including consumer pressure and framing strategies
- Comparative studies of regulations, antitrust measures, and legal frameworks across countries

### Example research questions:

- *What models of governance are effective for overseeing open source AI?*
- *How do corporate practices and incentives affect the degree of openness in industry-led AI projects?*
- *How do regulatory approaches to openness vary across countries, and what lessons can be drawn?*

## 5. Capacity building and communities

**Goal:** Build the skills, structures, and inclusion needed for diverse communities to meaningfully participate in open AI.

### Research topics:

- Training and education programs that equip diverse communities to contribute to open AI
- Labor models that ensure compensation and recognition for contributors
- Interdisciplinary collaboration frameworks bridging technical, social, and policy expertise
- Incentive structures that support participation from historically marginalized groups
- Universities as hubs for AI literacy, safe practice, and public engagement\*

### Example research questions:

- *How can labor and compensation models be structured to sustain long-term contributions to open AI ecosystems?*
- *What role should universities and public institutions play in the use and governance of open models and datasets?*
- *How can marginalized groups be supported to meaningfully participate in open AI design and governance?*

## 6. Infrastructure and resources

**Goal:** Identify and build the infrastructures required to make openness feasible, equitable, and sustainable.

### Research topics:

- Public compute infrastructure models and cost-sharing approaches\*
- Open datasets, transparency requirements, and data portability standards
- Roadmaps for smaller, efficient models (quantization, distillation, federated learning)
- Community resource needs for meaningful participation in open AI
- Cross-regional data sharing to support government and public services

### Example research questions:

- *How can public compute infrastructure be designed and governed to make large-scale open AI feasible for diverse actors – including researchers, communities, and governments – in equitable and sustainable ways?*
- *How can datasets support openness while also maintaining consent and allowing for community governance? What types of models are effective (e.g., data cooperatives) and under what conditions?*
- *What resources are most critical for communities to meaningfully participate in open AI ecosystems?*

## 7. Funding, economics, and incentives

**Goal:** Design sustainable economic systems that support openness in AI while ensuring fairness and alignment with public benefit.

### Research topics:

- Public goods funding models and hybrid public–private mechanisms for open AI
- Attribution and compensation systems for data contributors and development labor\*
- Incentive structures that align open AI investments with public benefit
- Lessons from economic models of past open source ecosystems [OS]
- Assessing the economic geography of open AI contributions and benefits

### Example research questions:

- *How might public–private partnerships balance openness with accountability and public benefit?*
- *What lessons from the history of open source software economics are transferable, and what breaks down, in the AI context?*
- *Where and in what ways do benefits of open-source AI accrue?*

## 8. Applications and public good

**Goal:** Assess / demonstrate if and how open-source and openness in AI can deliver concrete public benefits across domains and communities.

### Research topics:

- Compare open source and closed AI approaches in delivering outcomes for public services, health, and climate modeling [OS]
- Context-specific, localized, multilingual, and multimodal open systems
- Privacy-preserving, decentralized, and personalizable AI models [O]
- Open AI applications that meet community needs relative to closed systems
- Publicly financed AI infrastructure to enhance equitable access

### Example research questions:

- *How do open models improve accessibility for multilingual and marginalized communities compared to closed models?*
- *How do open-source applications compare with closed ones in meeting local or community-specific needs?*
- *What funding models best support open AI infrastructure for public benefit?*

## 9. Social, political, and historical contexts

**Goal:** Understand openness in AI as a social and political phenomenon shaped by history, culture, and power.

### Research topics:

- Messaging and media approaches that shape public opinion on openness
- Historical lessons from governance of nuclear, telecom, and other technologies
- Intersections of openness with inequality, colonialism, and justice
- Roles of journalists, watchdogs, and whistleblowers in open AI accountability
- Labor and job impacts of openness, including programming and teaching

### Example research questions:

- *How does media framing influence public support for or resistance to open AI?*
- *What lessons from past technologies (e.g., nuclear, telecom) can guide the governance of open AI?*
- *How does openness intersect with global inequalities, colonial legacies, and justice concerns?*

## Guiding practices for research on open source and openness in AI

In addition to highlighting research priorities, participants also emphasized *how* research on open source and openness in AI should be conducted. Good practices identified include:

- **Foster interdisciplinary collaboration:** Research should bring together computer scientists, social scientists, legal scholars, ethicists, and policymakers to capture the full range of technical and societal challenges.
- **Utilize participatory approaches:** Communities most impacted by AI should play an active role in shaping research, whether through partnerships with community-based organizations, participatory design, or citizen science methods.
- **Create pathways for meaningful participation:** Research should create pathways for historically marginalized groups, underrepresented regions, and early-career researchers to participate meaningfully in shaping openness.
- **Ensure transparency and accountability:** Research processes should themselves embody openness, including reproducible methods, transparent reporting, and (where appropriate) open data and code sharing with safeguards for consent and privacy.
- **Support capacity building:** Universities, training programs, and public institutions should support community AI literacy and develop skills pipelines that allow diverse contributors to shape and govern open models.
- **Practice reflexivity:** Researchers should remain mindful of historical lessons from past technologies (e.g., nuclear, telecom, biotechnology) and of how power dynamics can shape outcomes, ensuring that openness does not inadvertently reproduce inequities.

## V. Conclusion

---

This workshop report highlighted a co-created vision for responsible open source AI, as well as actions to move towards the co-created vision and a research roadmap with nine cross-cutting themes. The research agenda highlights a holistic approach for future research on open source and openness in AI. This includes research priorities for each theme, as well as how research conducted can take shape. Taken together, these priorities and practices provide a roadmap for advancing research that is rigorous, socially relevant, and impactful.

Realizing this research roadmap, moving ahead with the actions identified, and realizing the co-created vision will require cross-sector collaboration, sustained investment, and inclusive participation. Researchers are called to pursue these questions with openness, interdisciplinarity, and reflexivity. Funders and policymakers are called to resource the necessary infrastructures, communities, and safeguards that make this work possible. Together, we can work towards our collective imagination of the world and AI future that could be.

*“Look around: Humanity is in the eye of multiple storms. Will we continue shutting off the power of the masses so that a minority can stay warm, or will we build the necessary infrastructure so that everyone can thrive? Like Arundhati Roy, I believe ‘another world is not only possible, she is on her way... On a quiet day, if I listen very carefully, I can hear her breathing.’”*

*– Ruha Benjamin*

---

*This report was developed by Genevieve Smith (PI, UC Berkeley) with Hiral Patel (UC Berkeley) and Steven Luo (UC Berkeley). We would like to acknowledge and thank all the participants of the workshop who contributed to the ideas presented in this report.*

*With any questions, including information on the agenda and workshop approach, contact Genevieve Smith: [genevieve.smith@berkeley.edu](mailto:genevieve.smith@berkeley.edu). To follow the work and forthcoming white paper, please see the UC Berkeley Responsible AI Initiative: [re-ai.berkeley.edu](http://re-ai.berkeley.edu)*

# Appendix. Agenda

## Day 1: August 12, 2025

Time	About	Speakers
2:30-3:00PM	<b>Introduction</b>	<b>Genevieve Smith</b> , Founding Director of the Responsible AI Initiative, BAIR & Faculty at Haas School of Business, UC Berkeley
3:00-4:00PM	<b>Research presentations:</b> Key topics related to openness in AI	<ul style="list-style-type: none"> <li>• <b>Rishi Bommasani</b>, Society Lead, Stanford Center for Research on Foundation Models; PhD researcher at Stanford University – <i>What is the state of openness in AI?</i></li> <li>• <b>Matt White</b>, General Manager of AI, Linux Foundation – <i>The Model Openness Framework</i></li> <li>• <b>Dr. David Gray Widder</b>, Postdoc Fellow, Cornell Tech – <i>Power in open AI systems</i></li> </ul>
4:00-4:15PM	<b>Reflections and discussion at tables</b>	
4:15-4:30PM	<b>Break</b>	
4:30-5:30PM	<b>Panel discussion:</b> Leveraging open foundation models & democratizing AI	<ul style="list-style-type: none"> <li>• <b>Chloé Bakalar</b>, Chief Ethicist, Superintelligence Labs, Meta</li> <li>• <b>Jacob Metcalf</b>, Program Director of AI on the Ground, Data &amp; Society</li> <li>• <b>Matthew Zhou</b>, Assistant Chief Data Officer, California Office of Data and Innovation</li> <li>• <b>Maximilian Gahntz</b>, AI Policy Lead, Mozilla</li> </ul> <p><i>Moderated by Genevieve Smith, UC Berkeley</i></p>

## Day 2: August 13, 2025

Time	About	Speakers and additional information
9-9:05AM	<b>Welcome</b>	<b>Dean Jennifer Chayes</b> , College of Computing, Data Science, and Society (CDSS), UC Berkeley
9:05-9:10AM	<b>Introduction</b>	<b>Genevieve Smith</b> , UC Berkeley
9:10- 9:30AM	<b>Opening remarks</b>	<ul style="list-style-type: none"> <li>• <b>Mitchell Baker</b>, Co-Founder, Mozilla</li> <li>• <b>Vera Zakem</b>, Chief Technology Innovation Officer, State of California</li> </ul>
9:30 - 11:15AM	<b>Group visioning &amp; brainstorming</b>	Part A: Envisioning the world we want Part B: Responsible open source models Part C: Reimagining openness Part D: Sharing & synthesis
11:15 - 11:30AM	<b>Break</b>	
11:30-12:20PM	<b>From Vision to Action</b>	Part A: Discussion & action mapping
12:20- 1:20PM	<b>Lunch</b>	
1:20-1:50PM	<b>From Vision to Action – (Cont'd)</b>	Part B: Group share-out & synthesis
1:50-2:40PM	<b>Building a research roadmap</b>	Part A: Small group work Part B: Group share-out & discussion
2:40-3:15PM	<b>Closing Discussion:</b> Reflecting forward – Challenges & catalysts	
3:15-3:30PM	<b>Wrap-up</b>	<b>Genevieve Smith</b> , UC Berkeley
3:30PM	<b>Closing</b>	