UC Berkeley's Responsible AI Workshop 2024

BAIR Responsible AI Initiative

4 min read · Jan 20, 2025

Written by Koby Tesfay (College of Computing, Data Science & Society (CDSS) at UC Berkeley, FA24)

In November 2024, the first annual Responsible AI Workshop was held at the Berkeley AI Research Lab (BAIR) and hosted by BAIR's Responsible & Equitable AI (RE-AI) Initiative. Participants had the opportunity to engage with a refreshing & diverse array of speakers involved in advancing this space.

Genevieve Smith kicking off the workshop

Topics of **Fairness, Accountability & Power** were discussed in the first round of the afternoon via 15-minute research talks, 3-minute lightning rounds, and panel discussions. There were great presentations like Professor Lily Irani's breakdown on Amazon's Mechanical Turk Platform, in which she

highlighted the exploitation of data workers and her efforts to help worker organizing. In the same breath, we got an insightful introduction to Linguistic Bias in ChatGPT where Eve Fleisig took us through her research, which was conducted with other Berkeley collaborators including Genevieve Smith who leads BAIR's Responsible & Equitable AI Initiative. This research examines how different populations of English speakers experience and are (mis)represented in conversations with LLMs. In the same amount of time it takes me to finish a cup of coffee (3 minutes), Janna Huang informed us about the booming increase in water consumption by data centers as a direct response to the spark in generative AI use globally. And, a concise yet

model that consistently recommends male candidates over their female counterparts might be a direct result of training data that 'reflects the companies past hiring practices, which favored men for leadership roles'. All of the presentations during this section did a great job of bringing alternative questions and introspection to those lucky enough to listen.

The second part of the workshop followed a similar format, with an emphasis on **Security/Safety, Transparency & Privacy.** Professor Dawn Song spoke of how AI will change the landscape of cybersecurity, and the safety mechanisms needed to be resilient against adversarial attacks. Dynamic duo Cedric Whitney & Justin Norman showed how synthetic data is being used for face reidentification accuracy testing, and the risks that come with it, like diversity washing & consent circumvention. Are AI companies using synthetic data generation as a loophole that circumvents existing data consent and protection mechanisms that rely on tracking the origin and history of data? Maybe so. On the topic of Data Provenance, Shayne Longpre details his work with dataprovenance.org and on the Foundation Model Transparency Index, exploring questions about what data should be used as training data. If you're wondering if ByteDance can use GPT to build a rival AI product, they can't. OpenAI will (and has) suspended their account. Again, all of the speakers in this section had forward thinking presentations.

Prof. Dawn Song presenting on the spectrum of AI risks

The last portion of the workshop focused on an industry panel with responsible AI

leaders from Salesforce, Mozilla, Google, and Anthropic. There were a ton of great ideas floating around during this hour, with Jamila Smith-Loud setting expectations of Responsibility for AI at Google and emphasizing the importance of understanding its impacts and risks on the environment and

industry to develop effective mitigation strategies. Dr. Rachel Gillum addresses a question about AI's impact on jobs by highlighting how Salesforce is tackling these challenges through extensive AI training for employees and funding startups and nonprofits to drive innovation and foster new growth. Stephen Hood advocates for an Open Source future, and Kevin Lin speaks of AI regulation for everyone. This panel's effortless conversation was facilitated by Genevieve Smith, who did a great job navigating the differences and similarities between the vast expertise the speakers had to offer.

Industry panel

The workshop concludes with remarks from Reena Jana on various approaches to RAI, including technical methods like Model Cards and hands-on workshops like this one. Genevieve Smith then wrapped up the session by debuting the Responsible Use of Gen AI research and playbook,

which explores how models are integrated into new products and day-to-day work by product managers and others.

Overall, the workshop had a philosophical feel, with buzzwords like safety and responsibility resonating throughout. I thoroughly enjoyed listening to the discussions and insights shared.

*I used Claude & ChatGPT to help with sentence structure and checking grammar.

Responsible Ai | Ethical Ai

Written by BAIR Responsible AI Initiative

6 followers · 2 following

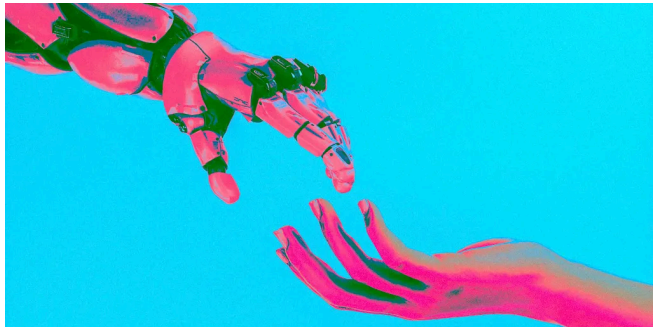
The BAIR Responsible AI Initiative (at UC Berkeley) is driving critical research, innovation and collaboration towards responsible and equitable AI.

No responses yet

Write a response

What are your thoughts?

More from BAIR Responsible AI Initiative



BAIR Responsible AI Initiative · Apr 25, 2025

Useful but Untrustworthy: Student's Relationship with AI

BAIR Responsible AI Initiative · Feb 5, 2025

How do we use generative AI responsibly?

College students have a love-hate relationship with AI that reveals a fascinatin...

What does it mean and look like to use generative AI “responsibly” in the workplace...

See all from BAIR Responsible AI Initiative