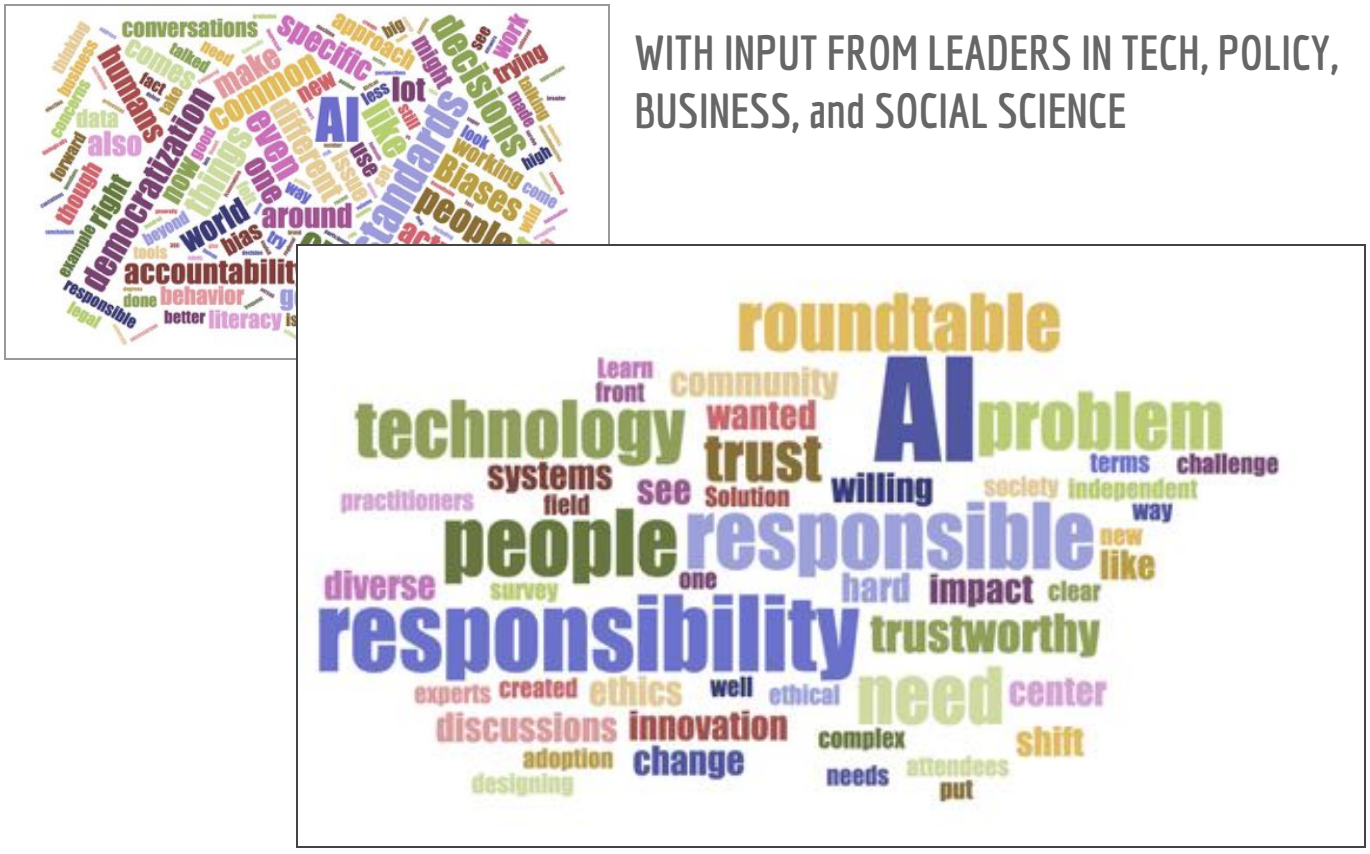
RESPONSIBLE INNOVATION PROJECT

RESPONSIBLE AI? ROUNDTABLE REPORT



WITH INPUT FROM LEADERS IN TECH, POLICY, BUSINESS, and SOCIAL SCIENCE

ALKA ROY

Founder, The Responsible Innovation Project

Released: December 2020

* Original cover image created with Collective Narrative from the Responsible AI? Roundtable using www.jasondavies.com

TABLE OF CONTENT

Responsible AI? Roundtable

Introduction

Despite increased awareness, debates and attention to technology and AI ethics, trust and responsibility, the question remains : Why are we designing technology and AI the way we are designing it? What needs to evolve?

While thoughtful concerns, alarms and ethical dilemmas have been raised by leaders in and out of the tech community, the tech and AI industries continue to accelerate the building, growth and adoption in this $13T+ industry[1] without any significant shift to privacy and ethical concerns. Even when we admit there's a problem, there is a tendency to debate principles, leverage a compliance model or push the responsibility to technology itself--leading to a call for Responsible AI, Trustworthy AI or Ethical AI.

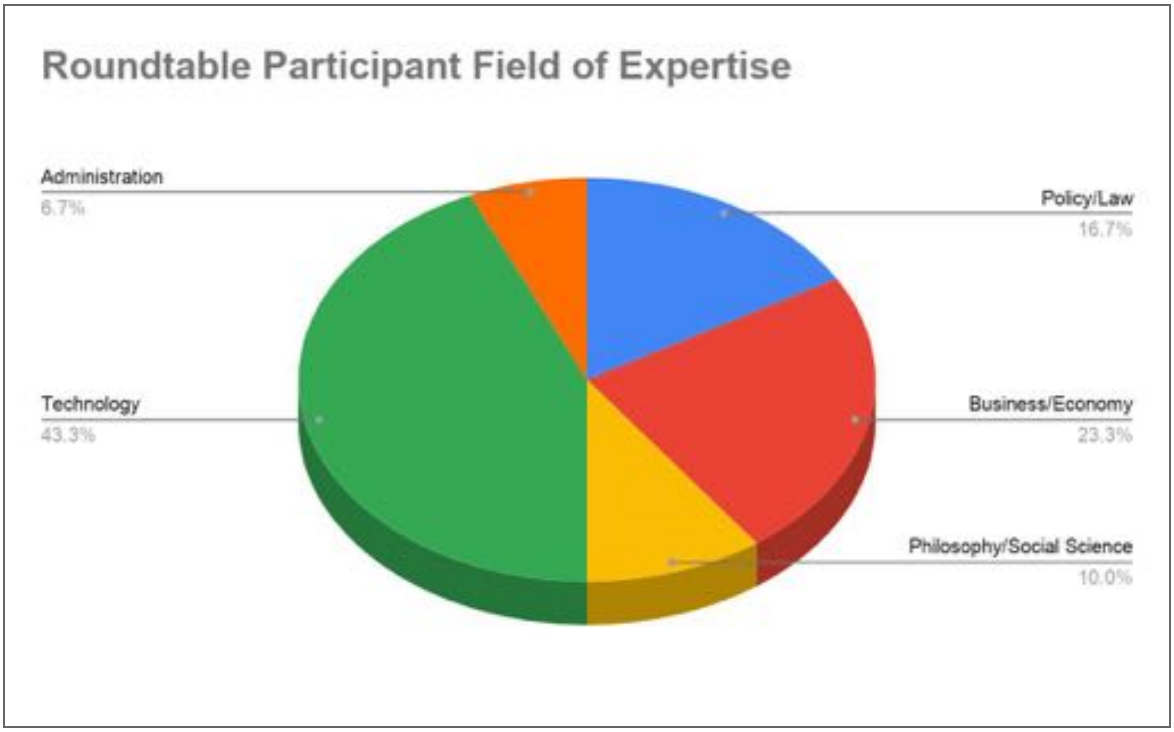
On August 20, 2020, The Responsible Innovation Project, based in the Bay Area, California (USA), designed and held an academic/industry roundtable on **Responsible AI**, raising the question: **Can AI really be responsible?** The goal was to arrive at a collective understanding of the challenges and strategies for building AI responsibly. The participatory roundtable included multi-disciplinary academic and industry leaders, practitioners and researchers working on technology and AI or at the intersection of technology, policy and humanities.

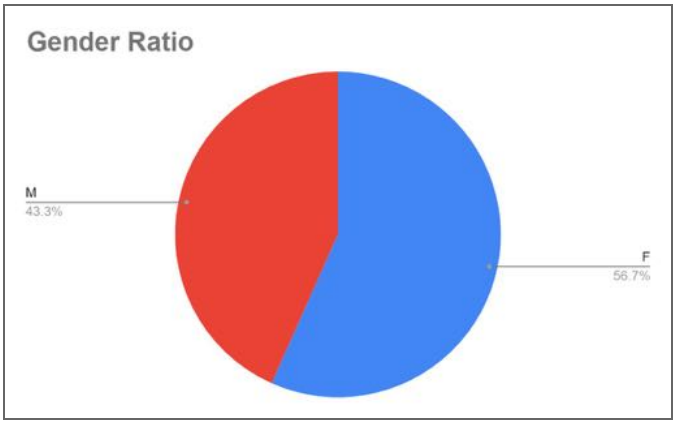
The Participants

Thirty (30) leaders and researchers with tech, policy, business and social science/philosophy expertise participated in the roundtable. They shared perspectives shaped by their experiences with several academic, industry, open source and non-profit institutions including UC Berkeley, Stanford, Harvard, Google, Dell, IBM, HPE and had affiliations with industry groups like Linux Foundation, Cognitive World, ISSIP and CITRIS Foundation. To allow for sincere and honest conversations, all delegates were asked and encouraged to share their personal perspectives rather than represent their respective organizations.

[1] Source: 2018 McKinsey Global Institute Report

Diversity of Participants: The following charts show the breakdown of the participants' primary area of expertise, while 30% of the attendees listed multiple areas of expertise. In addition, 30% of the participants also identified themselves as researchers. The roundtable attendees were 57% Female, 43% Male (the tech subgroup was 40% Female, 60% Male). Participants were primarily from North America, with 13% from Europe, Africa and Asia.





The Collective Narrative

Methodology

The two-hour online roundtable started with an overview of the current challenges and opportunities with emerging technology and AI and The Responsible Innovation Framework. The remaining 1.5 hours followed a participatory World Café process to combine the benefits of small group conversations and cross-pollination of diverse perspectives. At the end of the roundtable, the observations and strategies for building technology and AI responsibly were synthesized using the Collective Narrative Methodology and further edited and grouped into the following key areas:

1. Diversity of Perspectives
2. The Challenge of Bias
3. Will Common Standards and Frameworks Get Us There?
4. Is Democratizing AI & AI Literacy the Answer?

The Collective Narrative (shared in italics) uses the conversational style of the roundtable attendees and is a synthesis of readouts by group leads. The **we** in this section refers to the attendees and groups that were formed during the course of the roundtable. An effort has been made to preserve words and styles of the attendees to the extent feasible with permission from attendees to use the collective narrative without individual attribution. This narrative serves to highlight the importance of qualitative data and diverse points of views rather than present the position or opinion of any one individual or The Responsible Innovation Project.

The Narrative

Diversity of Perspectives

During the breakout sessions at the roundtable, we had a number of thought-provoking conversations with diverse perspectives. We were encouraged by the number of women participants (which was higher than typically present in AI and technology related discussions). We had different multidisciplinary viewpoints including those that pushed us to look beyond the EU/US/China-centered models. For example, we talked about the African context. This kind of in-buit diversity is important in these discussions and we discussed the need for this type of broader approach to AI accountability and transparency.

The Challenge of Bias

Trust is central. How do we incorporate accountability into some of the use cases for AI, particularly for customers and end users who are impacted? How can they be more informed about what decisions were made and who made the decisions? How do we put people and the satisfaction of users at the forefront of the development of Artificial Intelligence rather than an afterthought?

Every group brought up the issue of bias and what we need to do to identify biases. We did not arrive at a definite consensus to address the questions raised but rather spent time trying to understand the concerns. Our biases will continue to be exploited. Should we let AI exploit our biases or should it make us aware of our biases? Biases are such an integral part of our human condition—would we survive as a species without them? But should AI mimic or help us mitigate or better understand them? Computer scientists, researchers and engineers have been working on technical mitigation of bias. But it's not simple. Even in instances when AI systems have gone through so-called de-biasing exercises, there is still bias that creeps into anything that is being done.

When we talk about decision intelligence, is it important to consider how we define intelligence and who we think should be part of that discussion. One of the many issues we discussed was automation in the context of cascading decisions: let's say AI made a decision, then how does a human respond to that? How can we create a graph of decisions for a cascading series of human response and AI-decision scenarios? What roles should people play and what role should AI play? Part of the issue is that we do not actually capture decisions as an artifact of a business like we do processes. So how are we going to figure out how to control this machine that either appears to be moving fast with us or possibly competes with the way people process variables or make decisions? Interestingly, towards the end of the roundtable some of us commented that no one brought up automation or the rate of automation as a challenge but rather mostly focused on how to make tech. We also noted that no one raised questions about scenarios where technology or AI should or shouldn't be used.

When it comes to correlation versus causation, we discussed that in one hundred years we will still be struggling through it because humans are much better at this than machines are. Machines are good at correlations, but humans are much better at causations. Maybe it will be appropriate for AI to make high volume decisions while high risk, less frequent decisions would be made by humans. We keep making decisions about the future based on past data that is outdated or keeps changing. Several people commented on how that limits what decisions we arrive at.

Will Common Standards and Frameworks Get Us There?

We addressed the fact that from a standards perspective, even though many industry groups are working to generate it, it is the wild, wild west of standards right now. We are in the very early days of

this. There are a lot of guidelines out there as well as lots of principles and articulations of what AI should be doing and how to assess its impact. But there is a lack of consensus on how to synthesize all of those different standards and how to turn them into concrete solutions. Generally, this can lead to a sense that there's a lack of agreement or incentive to do something together.

We discussed how we can create a methodology or some common standards of behavior that we can use when it comes to AI. We discussed several parallels with the United Nations and some other organizations that use a set of common standards to monitor and to further a set of specific goals.

We found that there are specific problems with adopting common standards: not everyone wants to actually create them. One of the biggest challenges in the industry right now is that we like our existing tools and methodologies and there is an unwillingness to change. So how do we create a set of universal tools and standards of behavior? Even if we create common standards, then how can we make sure that we monitor their implementation? Is it on a voluntary basis? Even if we create a playbook for it, will everyone abide by it? Do people want to abide by it? Who would decide on the standards and who would drive accountability? When it comes to the legal side, would it even be beneficial for a company to abide by a common set standards or not?

There are a lot of issues with trying to fit all the frameworks into one set of common standards. It is also dependent on business conditions: take COVID disrupting our frameworks as an example. What adds to the problem is the challenge of translating between people who develop the principles and people who are developing the technology. There is also the issue of how teams work together. How do we create an environment where they can work together and understand each other's normative concerns?

Some of us felt that large companies are pressured to perform quickly rather than go through a deliberate process. So, even though AI has been around for a long time, the commercial interest is the new force that is driving things. And there might be potential for doing things poorly because of the pressure to do it quickly. At the same time, some pointed out that underdeveloped countries have an exasperated lag between technology and adoption of policies. Which made us look for strategies to close this gap.

Is Democratizing AI & AI Literacy the Answer?

Right now with technology and AI, there is little or no public accountability. We need to figure out how to make AI more accountable. Accountability is something that is a big challenge that we should take on and try to solve. There are too many unconnected systems and there is no 360 view of what is going on. One idea that was highlighted was that governing bodies should include people with a global approach and collect many people's input to push standardization beyond one sector or one way of thinking. Another idea was a customizable common knowledge methodology that could create a collective

knowledge framework as a next step in trying to create more transparency. In addition to that, we talked about a prevalent global data center for authentication, approval and usage. As well as leveraging global AI ethics centers and responsible AI applications as the drivers of frameworks and strategies.

We did not come up with a single strategy to ensure responsible and ethical technology and AI but the need for several different perspectives. Specifically, we identified Democratic AI as a concept to be explored more: co-creation and conscious creation are self regulating and help achieve the end goal. In that context, democratization of access to data and domain expertise and democratization of standards were consistent themes along with related challenges. It was particularly interesting to see how we could potentially bring democratization to AI, or could we even do this in a responsible and consistent way.

We talked a lot about the educational system and AI literacy along with a global approach to democratization of data collection. Some of us who came from the world of social impact and philanthropy had to catch up on terms like democratization of data collection used in technical terms in some of the conversations (it means different things to different communities). We discussed that well-rounded AI education for the younger generation was key. But perhaps we all need some AI certification at different levels. We discussed just how important AI literacy is and why academic institutions should expand AI from the Computer Science world to include the world of social impact and philanthropy. This would help bridge the gap between this "unknown world of AI" and engage the impacted people on the ground to also apply AI to create results and make impact.

Highlights and Key Takeaway
Technology for People

The roundtable narrative converged on the meta-theme: **The need and desire to put society and people front and center of Technology and AI. And the struggle to figure out how.** Some key questions that were raised included:

- How do we put people's well-being and user-centered development at the forefront of the development of AI rather than an afterthought?
- How do we define machine intelligence and who should be part of that discussion?
- Could AI be used to make high volume and low risk decisions but high risk, less frequent and specialized decisions need to be made by people?
- How can more people be more informed about the decisions that affect them and understand who is making those decisions?
- Who decides and how do we decide where technology should or shouldn't be used? Are we automating the right things at the right rate?

The four areas of strategies and approaches that the collective started to converge on to build AI responsibly come with their own set of challenges:

1. **Diversity of Perspectives:** If diversity is an embedded problem in an ecosystem, how can we begin to tackle this?
2. **The Challenge of Bias:** Defining bias itself becomes biased. Before we start buying tools to "correct" biases, how do we understand exactly what to fix and how? Who is ensuring that these tools are trustworthy?
3. **Will Common Standards and Frameworks Get Us There?:** What is the motivation for a common standard or framework? How flexible will it need to be? How accountable and to whom?
4. **Is Democratizing AI & AI Literacy the Answer?:** Sounds good, right? What could go wrong? What does this look like when we don't have a level playing field? Who gains or controls this democratization? What does AI literacy for all and the marketplace it serves look like? What is being taught? What is knowingly or inadvertently left out? And how is it taught so that we know when not to use it.

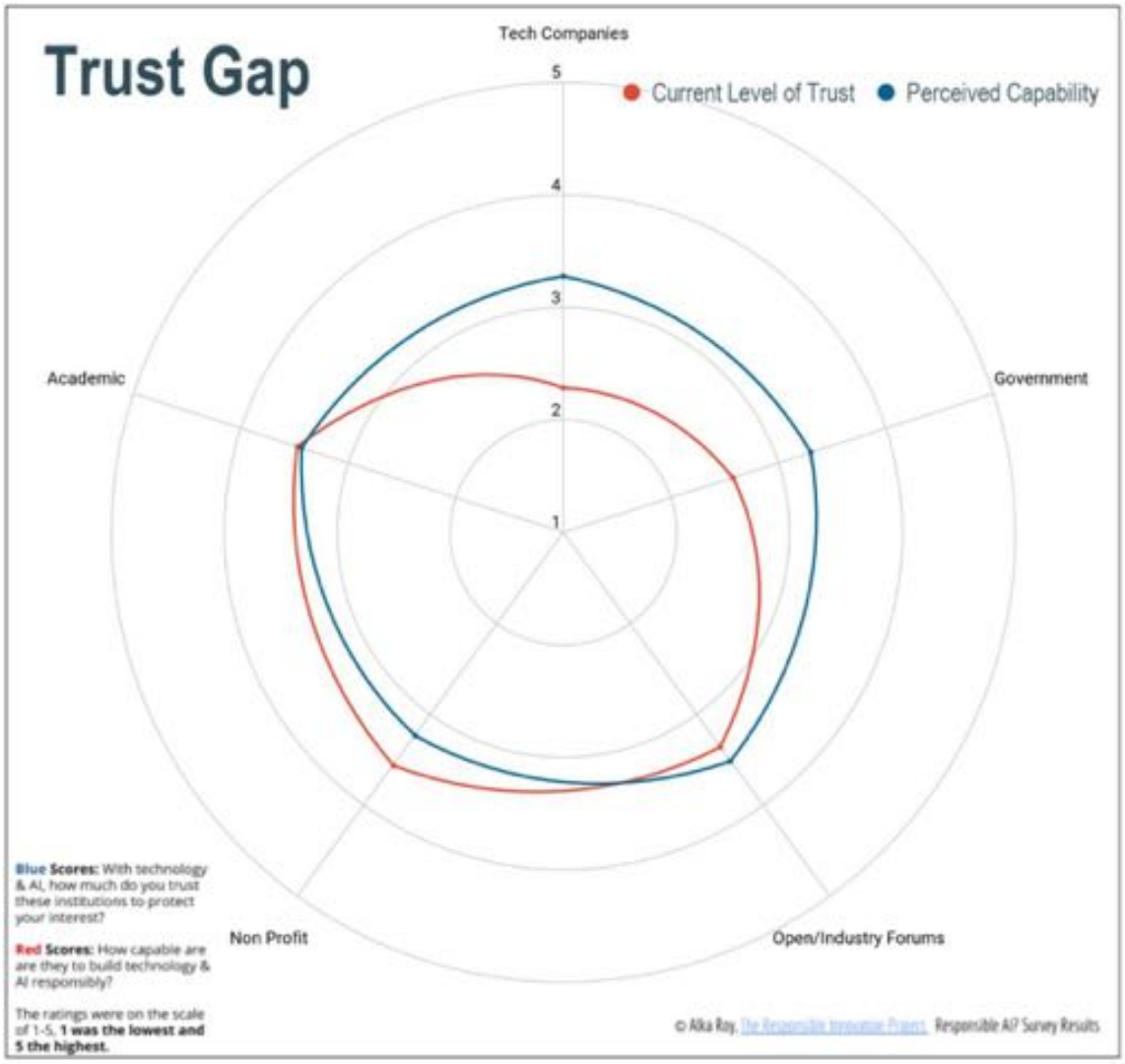
When we filter these themes through the lens of the earlier meta-theme, of putting people first, the question becomes: **How would we innovate, educate and organize differently if we put people and society's well-being first?** It comes down to how we view and practice our responsibility towards various communities and cultures.

Ultimately, how well these strategies are implemented and accepted also depends on the fundamental question of trust: **How much do we, can we, should we trust the people, the systems and institutions that we are relying on to navigate this?**

Responsible AI? Survey Results

Trust Gap

In Oct 2020, The Responsible Innovation Project released the Responsible AI? Survey Results. It showed a trust gap with both tech companies and government institutions. **How do we begin to develop effective strategies when we are starting out with a trust gap with the same institutions that fund most of the research in this area and have the highest capability to solve these issues?**



The survey participants were asked to rate how much they trust tech companies, governments, open/industry forums, and academic institutions to protect their interests. They were also asked how capable these same institutions are to drive accountability and responsibility. The scale was 1 (lowest trust/capability) to 5 (highest trust/capability). The graph shows the average scores.

Can AI Really Be Responsible?

When terms like Ethical AI, Trustworthy AI, and now Responsible AI are widely used, the language and framing shift the primary responsibility of responsibility, trust, and ethics to technology. Considering that what is defined as AI can be a complex ecosystem of systems that keeps evolving, it is not clear how it can be ethical, trustworthy, and responsible. The Responsible Innovation Project wanted to probe what Responsible AI means to those working on AI. Though the survey respondents or the roundtable attendees did not explicitly question the term, Responsible AI, their definitions and discussions put the responsibility on the makers and designers of the technology--people.

What is Responsible Innovation and AI? *

It is the responsible use of technology with considerations for upstream and downstream consequences. It is human and planet centered AI. It is people applying AI with integrity while respecting the rights of individuals and the collective.

Building or Using AI responsibly means developing technology that is both useful and accountable to the people who are meant to use it or who could be impacted by its use. It requires people and entities such as businesses, open source forums, nonprofits and governments to act responsibly when building, deploying, and using AI systems.

*Modified Collective Definition Based on The Responsible Innovation's Responsible AI? Survey Responses

Conclusion

Why hold this roundtable when the adoption of new technologies often outpaces our understanding of how they really work or impact society?

Why bother with more surveys and discussions when the Trust & AI and Ethics field is already crowded with countless principles and guidelines?

This is a hard and messy problem and there is no one size fits all solution. We can not solve hard and interconnected problems that were created over time in isolation, without a community that helps us shift the culture that created the problem.

The roundtable was held to see what would happen if we got a diverse group of multidisciplinary experts and enthusiasts together in an independent setting, outside of their departments and companies affiliations demands and the usual whitepapers, conferences and presentations. We wanted to see if that collective would converge on a common definition of the challenge and responsibility. Finally, instead of getting relegated to a separate field of responsibility, trust or ethics, we wanted to gauge the appetite for bringing the lens of responsible innovation front and center to technology (including AI) and ask: **Why are we designing technology and AI the way we are designing it? What needs to evolve? Where can we start?**

So, how did we do? What did we learn? And what next?

How Did We Do?
Taking Responsibility

The greatest success and impact of the roundtable was that a set of people with diverse experiences had a thoughtful and respectful exchange. At the roundtable, these leaders, practitioners and researchers were willing to take responsibility for figuring out what needs to change in their own domains. They were also willing to accept that they didn't have all the answers. They were willing to be vulnerable as well as simplify complex concepts for non-tech participants.

It is easier to point at others and look at what they need to change. It is harder to figure out what we should change. When the focus moves closer to us, the challenges amplify. In the post-roundtable survey and discussions, several attendees shared that they have begun to consider the impact of their **design** on others, and started asking questions about what a **diverse community** can look like. They were invigorated by the power of new connections, energy, and perspectives. The most important shift after the roundtable came from **technology practitioners who were interested in the problem but started to see responsible innovation as their responsibility**.

This shift is an important step in our personal and collective journey of figuring out what can and needs to be done.

What Did We Learn?
People Are the Problem and the Solution

There is a desire to align and the challenge is that we have to figure out how to trust the same people who may be at the center of creating the problem. We have to understand things all over again. Reframe them, which can be hard for experts. Other than the clear cases of theft and violence like cyber attacks and killer robots, no one has it figured out.

That doesn't mean that we don't do anything. **We haven't stopped tech adoption and massive digitization of our interactions for consensus and regulations. Why are we waiting for someone else to figure out responsibility and accountability? What if we start including social impact to innovation in that same iterative way? Otherwise, we'll keep amassing more and more technical and social debt and it'll get harder for us to dig our way out.** The most trustworthy solutions and people are those who admit what they don't know, design their inquiry with consideration and develop strategies that can be evolved and adapted.

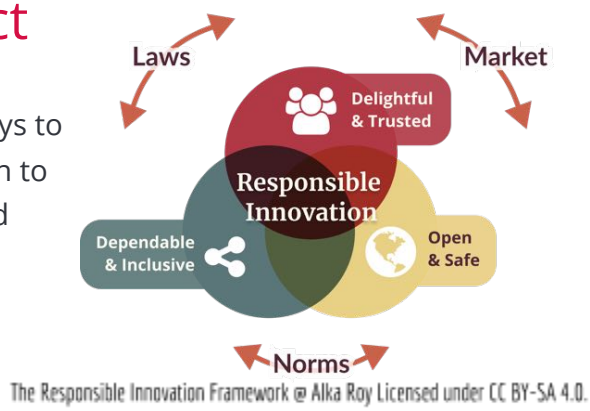
What Next?
Getting to Trust

How do we get to trust? To arrive at the key takeaways from the roundtable: diversity, bias, common frameworks and wider access and literacy to AI in a trustworthy way, here's what we need:

1. **An independent, trustworthy and accountable community.**
2. **To collaborate and put society and people front and center of Technology and AI.**
3. **Address the trust gap in the people, the systems and institutions that are both creating the problem and are capable of solving it.**
4. **Reframe how and what we teach people.**
5. **Revisit what and how we define and build systems and technology.**

We need to take responsibility and be accountable. We need to understand different industries, environments and ecosystems and collaborate. **We need a strong and collaborative community to lead us to a delightful, trustworthy, dependable, inclusive, open and safe world.**

Call to Action
The Responsible Innovation Project



The Responsible Innovation Project is exploring innovative ways to create a community and ecosystem for Responsible Innovation to become the norm. Our community is participation, inquiry and project-based. Our approach is to stay lean, trustworthy and collaborative.

The Responsible Innovation 2020 projects included:

1) **Launching The Responsible Innovation Project**
2) **Responsible AI? Survey and Report**
3) **Responsible AI? Roundtable and Collective Report**
4) **The Responsible Innovation Framework & White Paper**
5) **Responsible Innovation & AI Masterclass**
6) **Speaking, writing, teaching, engaging and collaborating with an amazing network of industry groups, global practitioners, researchers, leaders, startup founders, advisors, students & enthusiasts**

We have found that examples and applications give the most practical insights and tell the most powerful stories. To get us there, in 2021 we plan to continue working on industry & technology-specific case studies, assessment methodologies, workshops/learning tools for actionable RI as well as hosting idea development labs (RI Labs).

Whether as an individual or an organization, bring your ideas, your work and challenges to share, kick-off a new project or participate in one of our ongoing projects!

If you are working on **projects or research related to economic and behavioral incentives** or **using art and design to explore the impact of Responsible Innovation and AI**, please reach out.

Additional Information

Responsible Innovation (RI) Lab

This roundtable is part of The Responsible Innovation Project's **RI (Responsible Innovation) Lab**. RI Lab is bringing together a collective of multidisciplinary researchers, professional leaders, community members, artists, and tech practitioners to investigate the impact and possibilities of emerging ideas and technologies. The goal is to make space for thoughtful inquiry and bridge across academic, industry, startup, and non-profit communities for practice-based collaboration and projects.

Acknowledgments

First and foremost, we are grateful to all the roundtable participants for their time and candid and valuable thoughts and ideas. Without their participation, this report would not exist. We are also grateful to the breakout group leads, Angela Chatzidimitriou, Daniel Conway, Eliot Dreiband, Mehtab Khan, and Deborah Stokes for serving as breakout leads on the day of the event.

Our lean and stellar core roundtable volunteers who made this event possible included Leo Chang, Laura Hughes and Madeleine Sibert. In addition to Leo, Laura and Madeleine, other invaluable volunteer contributors and reviewers of this report include Alice Albrecht, Fyodor Ovchinnikov, Lavina Ramkissoon, Jana Thompson. As the primary author, I take sole responsibility for any oversights or shortcomings that might appear now or surface in the future.

Join Us!

Curious? Ready to Engage? Contact Us to advise, mentor, support, or work on projects at the RI Lab collective or Join The Responsible Innovation Project mailing list for reports and invitations to future projects and initiatives.

Alka Roy is the founder of The Responsible Innovation Project working on building delight, trust and inclusion into technology and AI and a guest lecturer on Responsible Innovation & AI at UC Berkeley. She is preoccupied with the questions: How are we designing our future? And how will that design, design us?

Alka is a product and technology leader who has been part of several industry firsts for AT&T and Cingular Wireless. She was instrumental in setting up the Bay Area 5G co-create lab for AT&T to spur innovation and lead the Responsible AI initiative for AT&T Innovation Center. In addition to a Computer Science & Electrical Engineering degree, Alka also holds an MFA in creative writing & literature and speaks, writes, mentors and hosts multidisciplinary and deep tech sessions on innovating with values and serves on several open source and industry ML/AI, Data Science and Trusted AI committees.