

Global AI Dialogue Series

Observations from the China-US Workshop in Beijing (December 2, 2017)

Berkman Klein Center for Internet & Society at Harvard University
in collaboration with the School of Public Policy and Management at Tsinghua University
and the MIT Media Lab's Ethics Initiative

China and the United States are home to leading players in the research and development of artificial intelligence (AI) and autonomous systems, which promise enormous benefits for the social good and pose significant risks. Investment in startups to apply and commercialize AI technologies is rapidly advancing in both countries, while in parallel different branches of the Chinese and American governments are preparing strategic policy plans for the future of AI. AI's social impact, however, remains insufficiently examined, and many probable and prospective national and international decision points have yet to be clearly identified owing to differing political, economic, and cultural contexts.

In order to establish a cross-cultural dialogue about specific AI issues and build a learning network for investigating approaches to address these issues within and across domestic and global contexts, the Berkman Klein Center for Internet & Society at Harvard University in collaboration with the School of Public Policy and Management at Tsinghua University and the MIT Media Lab's Ethics Initiative hosted a China-US AI Workshop to bring together experts and practitioners from both countries. The meeting was designed to strengthen and build interfaces for bilateral learning and information sharing on research questions surrounding AI of mutual interest, while fostering trust between the AI research communities in China and the United States.

The purpose of this write-up is to share observations from this initial discussion-based workshop, highlight overarching themes that emerged, and extract insights on next steps for sustaining the cross-cultural, global dialogue.

I. OBSERVATIONS FROM THE SESSIONS

The event discussion evolved from a conversation on both baselines for AI metrics research and the methodological issues broadly defined, to the tools, approaches, and case studies available to address the societal impact of AI. Designed to effectively engage the approximately 30 participants drawing from a range of sectors and regions, the workshop yielded a number of key themes from the sessions, particularly with regard to measuring the broader social impact of AI as a product of not just AI technologies alone but rather a complex interplay of variables including policies, laws, and educational strategies. These themes are outlined below as a method of organizing the most salient observations from the workshop.

1. Overarching Themes

Identifying Metrics for Impact

Across the sessions and topics discussed, the need to measure the development of AI served as a common thread, particularly as there is uncertainty concerning impact that requires a level of technological foresight and risk analysis. While there was a consensus that AI differs from previous technologies, there was also agreement that, in order to understand AI, stakeholders would need to observe how previous technology revolutions have shaped society. Policymakers recognize that AI's social impacts extend far beyond replacing human manpower, yet they still do not know which criteria to set, metrics to formulate, and signs to observe. Thus, participants suggested that the first critical step in AI policy making is the baseline setting of AI metrics research, thereby creating a cornerstone to tackle the "unknown unknowns."

Moving from Specialized Cases to Broader Applications

Given the complexity and fluidity of the scope of AI's social impact, participants proposed that policy-makers must be pragmatic and find concrete examples with which to begin and upon which to build. By narrowing down on one index in one industry, and distilling the implications of AI from those vantage points, one can begin to measure and understand the broader social impact. Although impact will differ across industries and at different junctures in time, an initial breakthrough made in any field can be spread horizontally and applied in other fields as well as developed vertically as time progresses. From one point, one can gain a multi-dimensional understanding of AI's social impact.

Anticipating Ripple Effects

The societal impact of AI does not rest in any a particular sector or country and will have ripple effects that affect industries and individuals throughout the world. Participants suggested that as boundaries between industries begin to blur and sectors grow more connected and globalized, it is not only vital to examine the impact of AI on the surface level but also equally important to consider the secondary and tertiary effects of AI that affect populations regardless of their involvement or knowledge of the development process. For example, as manufacturers of pretrial risk assessment tools integrate AI into their products sold to jurisdictions within the US criminal justice system, they run the risk of perpetuating race and gender bias; though the tool, and tools such as this, may be efficient and beneficial for the judiciary, social implications for individuals who engage the criminal justice system and their families must be considered.

Maneuvering Situational Differences

As China and the United States claim binary leadership in AI development, there remain substantial contextual differences that participants felt must be addressed. With varying cultural backgrounds, political environments, economic structures, and decision-making processes, AI and its social impact have to be examined differently in the two countries while establishing common ground for implementing solutions to mitigate potential

challenges. In a number of specific cases brought up by participants, similar issues pertaining to AI had profound but divergent impacts on China and the US respectively.

2. *Role of the Government*

In the effort to minimize the negative impacts and maximize AI's potential enabling power, government can play a role as regulator, educator, and facilitator. Participants agreed that the public sector will undeniably become a critical stakeholder in the development of AI.

As AI development progresses, there must be a mindful balance between regulating, guiding, and preserving the vigor of the tech companies and capital pushing the advanced technologies forward. Participants repeatedly underscored the importance of taking a more laissez-faire approach, as premature regulations could kill a burgeoning sector with great potential or major subsections thereof. At the same time, however, participants also sounded the alarms about data abuse, privacy infringement, and cybercrime, which are areas that presently demand government regulatory attention. Since the AI sector is rapidly evolving, negative and potentially irreversible externalities could emerge if the sector is want of regulation.

Furthermore, participants suggested that governments should begin to retrain those who are likely to be replaced by AI-based systems. As the popularization and mass-application of AI technology grows imminent, a substantial amount of the workforce across industries—from stock brokers to construction workers—will have their posts amended or altogether eliminated to accommodate capable autonomous systems. The government should prepare for this massive shift in the employment landscape at the systemic level — using the full spectrum of approaches, including educational policy, taxation, and other interventions — as it is unfair for the individual to bear the responsibility for this radical departure.

3. *Metrics Nexus*

In the second half of the workshop, participants focused on three specific areas for metrics to assess AI's societal impact: employment, inclusion, and education. To identify appropriate impact metrics, participants worked through specific case studies, examining new opportunities and needs by drawing comparisons to previously developed technologies.

Metrics for Impact on Employment

- Internet and Digital Economy Case Study

Participants agreed on the need to examine the social impact of the internet and digital economy—the backbone of the recent tech landscape shift—as a means shedding light

on the trajectory of AI and labor. Since 2008, the internet and digital economy have been gradually replacing traditional accounting systems, and thus generating a plethora of new employment opportunities in the tech sector. The UN has held multiple dialogues on this matter. However, there has been no corresponding increase in employment opportunities in global economy. In fact, there seems to have been a transfer of employment opportunities across sectors, with people losing jobs in traditional posts as new jobs are created. The internet and digital economy have been causing job migration rather than creation, and it worth considering implications for AI development and deployment.

- Autonomous Vehicles Case Study

An autonomous truck will reduce the trip from Boston to LA from three days to one day. In the US, there are currently 3.5 million truck drivers who are among the highest-paid blue collar workers at risk of losing their jobs to autonomous vehicles, and many more individuals will be affected by the ripple effects of this shift. As AI reduces the travel times and replaces driver-operated vehicles, there will not only be fewer truck drivers but fewer trucks. AI will in turn have a secondary effects on the diesel and automobile industries, as well as the service stations and fast food chains that line interstate highways, creating a domino effect. Truck transportation and industries related to or reliant on this industry are one of the economic pillars of many midwestern US states. Bearing this in mind, it is not only the 3.5 million truck drivers being impacted; it is ultimately tens of millions of people.

However, as discussed by participants, this specific case study is more applicable in the US. China is less dependent on trucks and highway transportation systems and has a significantly lower cost of labor. It is predicted that autonomous vehicles will have a marginal impact on the Chinese transportation market and will not replace the current railroad-based system. Because of these country-specific concerns, while truck unions are blocking autonomous vehicle research in the US, China is less limited in this capacity and is able to readily progress in research and development. Examining cases such as these demonstrates how similar circumstances in different countries yield distinct, context-specific outcomes and help us anticipate context-specific challenges and opportunities with global implications.

- AI Employment Takeaways

How might the global economy respond to rapid changes in one country leading the future of AI, such as China, while another is in a different stage? Using specific metrics such as employment to measure existing AI impacts while anticipating future impacts allows us to see to answering this question and respond accordingly as a global community.

Metrics for Impact on Inclusion

From an application standpoint, capital has been a driving force behind rapid technology development. Participants raised concerns about the role in society of the old and elderly in China and the US, who do not possess the capital to guide investments in technology to benefit themselves. This group is just one example of many who have the potential to be further excluded from society in the age of AI if government does not implement thoughtful prevention measures. Similar issues of inclusion will also reverberate on a global scale given the technological dominance of AI leaders such as China and the US in relation to less economically developed countries.

Moreover, we as humans encode our values into the technologies that we produce, and these values do not always take into account geopolitical, economic, and cultural diversity, and divergent applications of AI technologies that result from these differences.

There is a need to reverse the developmental paradigm from conceptualizing how to bake inclusion into existing technologies to conceptualizing how technologies can be created and utilized in order to build a more inclusive world. One approach mentioned by some participants to achieve positive AI development and deployment involves better understanding the public perceptions and attitudes toward AI-based technologies and their uses in order to develop an inclusive framework for impact measurement. To achieve such a framework and gain public trust, appropriate baseline questionnaires, studies, and discussions must be pursued that carefully gauge what different communities believe and want. In order to take such measurements and settle on meaningful social impact metrics, we must determine how to negotiate and assess public opinion and priorities.

Metrics for Impact on Education

The role of education in shaping how people address the opportunities and challenges posed by AI is crucial and often neglected. According to some participants, one must educate both the oldest and the youngest generations effectively so that AI does not follow the tracks of Genetically Modified Organisms (GMOs), which garnered negative public perceptions among the elderly as being “unhealthy” despite maintaining the same nutritional value as non-GMO foods. Democratizing AI education will help to both mitigate the risk of this “fear-induced” perception of AI, as well as encourage more equitable representation in the AI economy.

In order to achieve this education paradigm, governments must take measures to publicize and popularize the lived effects of AI among the senior generation and simultaneously integrate ethics and technology into the pedagogy of the next generation. Education, in this way, needs to be revitalized, and government should play a significant role in spearheading that revitalization.

II. SPECIFIC SUGGESTIONS BY PARTICIPANTS

Throughout the workshop, participants identified several concrete next steps for anticipating and measuring potential societal impacts of AI and working transnationally to

build a more inclusive future. Suggestions included identifying specific impact metrics, retraining workers, maintaining healthy competition internationally and nationally within AI development, supporting non-obvious entrepreneurial resources particularly in the Global South, implementing more egalitarian AI education as well as adaptable educational tools, and encouraging government integration in the AI development and deployment agenda in order to promote formalized policies and oversight mechanisms.

Designed to periodically assess the impact of AI on every facet of global society, [Stanford University's 2017 AI Index](#), released this past November as part of their One Hundred Year Study on Artificial Intelligence (AI100), is one apt example of an ongoing initiative that integrates a number of the above suggestions. The study serves as both a model for effective impact measurement, as well as a point of collaboration from the workshop given the need for more Chinese data to complement the AI Index.

When it comes to pursuing potential suggestions and solutions, encouraging information sharing through joint initiatives such as these is critical, as countries differ in their approaches to, and desired uses for, AI technologies. Over the course of the past several decades, China has sought to catch up developmentally in the global market, leading them to maintain a more positive perception of new AI technologies and less skepticism towards a “rapid development” approach. Facial recognition startups in China, for instance, outnumber those in the US because China has the larger data sets, facing fewer privacy-driven challenges to aggregating the information. This zeal, however, is a double-edged sword, as the challenges are as staggering as the opportunities and may be overlooked in the race for development; one key challenge, as discussed above, is a loss of (or drastic shift in) workforces that currently drive the US and Chinese economies.

Developing and modifying policy at the domestic and international levels was perceived as a vital next step for addressing such challenges. Participants reiterated throughout the workshop that as the technological sphere is changing with the rise of AI-based technologies, policy in turn must evolve, states must be proactive, and global governance must take on an adaptive approach. As government does not work independently and is driven to action by other sectors like civil society and academia, sustained transnational dialogues that convene a variety of stakeholders and consider societal impact, such as this workshop, must continue as a means for devising solutions and pressuring governments to adopt them as policy.

III. CONCLUSIONS

Throughout the workshop, as dialogue moved from establishing the need for baseline social impact metrics to exploring the array of fields in which they would prove helpful, participants reached a consensus on the need to conceptualize analytical projects and quantifiable paradigms, and identify specific ways in which to build upon this meeting and move forward with actionable next steps through networks and targeted partnerships.

Meanwhile, participants agreed that it was crucial to continue fostering a cross-cultural dialogue between China and the US. As of now, many informational asymmetries

remain, and such dialogues can serve as a means of developing meaningful modes of collaboration in conjunction with robust AI infrastructure. Participants felt that by continuing these conversations, it would be possible to create an inclusive roadmap for developing a metrics system to meaningfully measure AI's social impact, as well as to inform an AI global governance framework for international policy-makers.

Finally, in encouraging these collaborations, participants emphasized the importance of measuring and addressing the impact of AI on society holistically, noting that there can't be a discussion about AI benefits while limiting the scope to economic gains, and that likewise there can't be a discussion of "AI for Good" without environmentalism serving as a central theme. Each issue is integrated with another, and it is necessary to consider them all in order to make progress in closing the gap between the expected, often negative implications of AI-based technologies and its real capabilities.