IDRC GLOBAL SYMPOSIUM ON AI & INCLUSION OUTPUTS

Berkman Klein Center for Internet & Society at Harvard University;

;

IDRC Grant/ Subvention du CRDI: 108698-001-Addressing the Challenges of Artificial Intelligence (AI) & Inclusion

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

IDRC Global Symposium on AI & Inclusion Outputs

Table of Contents

IDRC Global Symposium on AI & Inclusion Outputs: Framing Document

I. Introduction

This complete set of materials is intended to serve as an input for the IDRC research agenda “Artificial Intelligence and Human Development." It includes this cover memo, followed by an analysis of research questions at the intersection of artificial intelligence (AI) and inclusion, and an appendices with additional outputs.

Artificial intelligence and related technologies have begun to shape important parts of the digital economy and affect core areas of our increasingly networked societies. Whether it be transportation, manufacturing, or social justice, AI has the potential to deeply impact our lives and transform our futures in ways both visible and hidden. The promise of AI-based technologies is enormous, and benefits range from efficiency gains to unprecedented improvements in quality of life. The challenges, however, are equally staggering, for instance when considering the uncertainty surrounding the future of labor or the emergence of new power structures outside the control of existing governance and accountability frameworks. More specifically, the uneven access to and impact of AI-based technologies on marginalized populations run the disturbing risk of amplifying global digital inequalities. These groups include urban and rural poor communities, women, youth, LGBTQ individuals, ethnic and racial groups, people with disabilities – and particularly those at the intersection of these identities.

A complex set of issues exist at the intersection of AI development and the application divide between the Global North and the Global South. Some of these core areas include health and wellbeing, education, and humanitarian crisis mitigation, as well as cross-cutting themes such as data and infrastructure, law and governance, and algorithms and design, among others. We are examining the core areas and cross-cutting themes through research, events, and multi-stakeholder conversations. This cover memo is informed by these efforts, including the Global Symposium on AI and Inclusion[1], which incorporated perspectives from a wide array of experts in the field.

II. Overarching Themes/Critical Issues

Grounding the conversations about AI and inclusion within a shared, contextualized understanding of the fundamental concepts at play is crucial, as doing so allows us to maintain and progress fruitful conversations and further explore connected issues. To this end, the following three overarching themes and critical issues emerged both from the research leading up to as well as the conversations that took place at the Global Symposium on AI and Inclusion.

[1] The Global Symposium, co-hosted on behalf of the Network of Centers by ITS Rio and the Berkman Klein Center for Internet & Society, involved over 170 participants from more than 40 countries around the world and took place over the course of three days (November 8-10, 2017) at the Museum of Tomorrow in Rio de Janeiro, Brazil.

Exploring the intersection of "AI and Inclusion"
It is necessary to critically engage with the ways in which we define the fundamental concepts of AI as well as inclusion. As put forth by Nishant Shah, co-founder of the Centre for Internet & Society Bangalore, we must reckon with the history and impact of AI technologies and inclusion, drawing attention to the mutually constitutive intersection between them to look beyond just computation to the actual lived realities of the computed. Simultaneously, a critical engagement with the definition of inclusion in a technological context is also necessary. From a technical development process perspective, Ansaf Salleb-Aouissi, Lecturer in Discipline at Columbia University, articulates the need to recognize a "4D Framework" — develop, de-identify, decipher, and de-bias — in developing AI to achieve better inclusivity. As we continue to examine all aspects of AI and inclusion, we must recognize the need for a new vocabulary to discuss the novel ways in which AI and inclusion interact.

Reframing "Inclusion" to embrace self-determination
It is important to understand the influence of western values within the AI discourse and how this may contribute to bias in global systems — technological and social. One important criticism of the traditional concept of inclusion is that it places the burden of "performing" inclusion on the underrepresented communities themselves. Rather than simply incorporating these communities into global conversations about AI, or "throwing money" at engineers within these communities to support inclusive development, policymakers must enable these communities to actively drive the discussion and deployment that shape their experiences. Self-determination constitutes an alternative to or amendment of the inclusion paradigm, offering agency and space for underrepresented individuals to represent themselves and guide the systems and policies that affect them.

Rewriting old AI narratives to encompass new AI capabilities
We must also reconcile the differences in speed at which the AI discourse progresses and the speed at which new AI capabilities are developed. As technological advancements are often made at a more rapid pace than developments in the discourse surrounding them, there is a tremendous potential to exacerbate existing divides and create new ones. This gap points to an urgency to collaborate quickly and meaningfully across regions and sectors in order to understand and mitigate potential risks at the intersection of AI and inclusion.

III. Research Questions and Matrix

The research questions listed in the research memo and analyzed below were drawn from each session at the Global Symposium on AI & Inclusion as well as from attendee inputs during interactive activities. They are categorized in matrix form, vertically by the stages of developing an AI system and horizontally by mechanisms for intervention. The structure of the matrix, in addition to a brief description of each category, is outlined below for reference. For more detailed information on the methodology for categorizing

research questions and where they are drawn from, please refer to the full research memo attached in the Appendix A.

	Design	Development	Deployment	Evaluation
Defining and Framing ("Back to First Principles")				
Bridge-Building (Network Building/Liaising) Infrastructure Building				
Educating				
Policy-Making				
Tool-Building				

Observations and Trends
Collected research questions range from specific and technical to higher-level and theoretical. Additionally, questions vary in the targeted timeline. Some of the lower hanging fruit that can be addressed sooner are items that can act as building blocks for even more questions and provide an overview for what a global AI landscape could look like. One example of this type of question, under the sections "Bridge Building and Infrastructure Building" and "Evaluation," is "What does the current global framework of AI production/use look like? Who are the major players, what are their initiatives, and how can we lift up more grassroots initiatives?" Global South actors can provide meaningful insights into emerging grassroots efforts, which can collaborate to address local and global challenges.

Within the matrix, the highest number of questions are contained in the cell with the horizontal of "Defining and Framing" and the vertical of "Design." This category includes more speculative and fundamental questions regarding subjects such as the definition of inclusion, AI biases in relation to human biases, and how we choose to design AI-based technologies. Keeping in mind the academic bent of the audience, this section was likely the most substantial as the topics in this category called for more significant academic and theoretical engagement. The design category more broadly houses questions that examine both the design phase of AI development as well as the phase leading up to it in which one identifies the purpose and audience of the tool.

Other major theoretical themes from the research questions include the extent to which AI can enhance or altogether replace human behavior, how cultural norms and nuances may be incorporated into AI-based tools, and the potentially inherent exclusivity of AI due

to its categorization of data. It is imperative that we consider these questions, as they serve as the foundations for how we define AI and inclusion, and can provide us a guideline for more research questions in the remainder of the matrix.

Among the less conceptual questions, a few central trends emerged. One involved the concern with the monopolization of companies, resources, and discourse concentrated in the Global North and the inequity this dynamic yields. A second encompassed the co-production of data and co-design of AI-based tools as a means of closing the gap between users and producers while fostering diverse data and more inclusive technologies. A third addressed the establishment of mechanisms for auditing AI algorithms to measure fairness and setting fairness standards. These trends represent just three among many that highlight the challenges presented by AI development and use as well as potential solutions.

IV. Opportunities and Challenges

When it comes to prioritizing research questions as well as proposing solutions derived from them, our work on AI and inclusion suggests that it is vital to distill the greatest opportunities and challenges decision-makers face while grappling with AI-based technologies, particularly in Global South contexts. Although a number of action items and constructive approaches are warranted, three primary considerations emerged as integral to addressing the opportunities and challenges at the intersection of AI and inclusion found in the research questions.

Targeting solutions

Efforts to answer these questions may differ by method, expertise, and region. Therefore, there is a need to devise specific sets of solutions for specific types of problems concerning AI-based technologies and their implications. A broad-strokes approach to formulating solutions for AI and inclusion challenges is ineffective, as the complex nature of problems at this intersection call for a targeted, multivalent approach. Meanwhile, clustering similar research questions and issues may serve as a way to harmonize discussions surrounding AI and inclusion, aligning normative goals using the levers of individual agency and/or ecosystem-driven approaches.

Adopting a multitemporal approach

The temporal aspect of identifying issues and formulating solutions was at the crux of the conversation on seizing opportunities presented by AI while identifying effective solutions to potential problems. Specifically, AI's rapid growth and multifaceted societal implications warrant a reconceptualization of the time structure of the AI research agenda. We cannot simply "act now," nor can we afford to "wait for the time it takes to conduct research;" rather, we must do both simultaneously. This approach, of course, must account for different temporalities and goals for different sets of problems and solutions, modeling itself on larger social movements that successfully build momentum and involvement on global and local levels without following a traditional chronological agenda commonplace in academia.

Breaking down silos
In terms of collaboration, all questions should ideally be explored with a diverse group of individuals from varying sectors and geographic regions in order to break down constrictive silos. However, certain questions in particular necessitate a co-research approach by nature; these types of questions include, among others: how to expand funding for Global South initiatives; the ways in which behavioral expectations vary geographically; the democratization of AI education; and the global governance of AI systems. Working towards true and organic interdisciplinarity encourages more meaningful integration of varying perspectives that may yield more inclusive AI development and a less hierarchical global AI power structure.

Devising solutions for research questions with this integrative approach may involve the democratization of nuanced and accessible education (including co-designed reading lists and syllabi), formulation of novel partnerships across regions and sectors, thoughtful allocation of funding to underrepresented developers, and engagement in sustained multi-stakeholder dialogues.

The research memo in Appendix A includes the sorted matrix, high-level key takeaways, and future considerations.

Global Symposium on AI & Inclusion Research Memo

I. Meta-level Observations

These meta-level observations address the key takeaways and points to note as the research agenda is being written.

1. Recognizing the limits of categorization (e.g. that research agendas typically categorize items by short-term or long-term goals), a research agenda should recognize the importance of working on multiple goals simultaneously and strive to resist creating temporal distinctions.
2. Recognizing that Global North institutions typically dominate the research sphere, even surrounding issues pertaining to the Global South, it is imperative to adopt a 'co-design' approach that aims to capture diverse perspectives.
3. Recognizing the academic bent of the Symposium, a research agenda must recognize and emphasize that true interdisciplinarity will require more innovative work, including building new partnerships, redistributing funding, and de-siloing research/initiatives.
4. Recognizing that the academic sphere is often restricted to knowledge gathering and that there is a distinction between research and action research methodologies, research methods utilized and/or suggested in a roadmap should be varied and set the stage for actionable recommendations based on findings that model the true interdisciplinary described above.

II. Research Questions Matrix

1) Methodology

These key research questions collected were drawn from discussion sessions at the symposium, as well as the outputs from interactive activities ('Wonderbag' activities) completed by attendees across the three days. These activities were designed to encourage collaboration as well as to collect inputs (e.g. pertinent research questions) on session topics from all participants. For more detailed information regarding the Wonderbag activities and its outputs, please refer to the 'Wonderbag Outputs' document shared previously and included in the appendices.

In order to thematically categorize the research questions, we designed a matrix in which the questions are classified both by the developmental stages of an AI system (vertical), and mechanisms for intervention (horizontal). Each cell in the matrix aligns with a segment in the AI development process (verticals), as well as a specific intervention lever that may have influence on AI and its impact (horizontals) — this categorization allows us to contextualize important research questions by sorting them according to potential areas for action. Furthermore, this schematic enables us to simultaneously identify the nature of specific questions as well as thematic trends across all questions. Though the matrix follows a more linear model with specific modes of

intervention, the questions themselves should not be thought of as confined to the boundaries of their category; rather, the categories should be used to provoke thought on potential action items coming out of the research question by considering what stage of AI is being addressed and what level of society may be affected or integral to resolving the question.

Finally, it is worth noting that there exists a myriad of approaches to sorting these questions, and that this is one approach specifically geared towards fostering research. In the appendices, we have included a copy of all research questions sorted by this matrix, as well as a copy of all research questions organized by session in chronological order. Below is the model of the matrix and a description of each category, followed by meta-level observations about the collected research questions.

	Design	**Development**	**Deployment**	**Evaluation**
Defining and Framing ("Back to First Principles")				
Bridge-Building (Network Building/Liaising) Infrastructure Building				
Educating				
Policy-Making				
Tool-Building				

Defining the Development Stages of an AI System:

- **Design:** This stage includes both the technical design of AI-based technologies (e.g. how data will be collected in the model, the target audience of an AI tool) as well as design of systems that govern AI and the implications that must be considered before moving on the the development process.

- **Development:** The development stage is the production phase of an autonomous system that follows the design process. Questions in the development category pertain to the building of more inclusive tools/methods into AI-based technologies, frameworks of AI development, and what AI tools are being developed for whom.
- **Deployment:** The deployment stage covers the distribution, utilization, ubiquitousness, and implementation of AI-based technologies within society at multiple levels including the local, national, and global ecosystems.
- **Evaluation/Impact:** The evaluation/impact phase encompasses the measurement and understanding of the impact of AI technologies, including ways to evaluate the effects of autonomous systems on different agents within society.

Defining the Mechanisms for Intervention:

- **Defining and Framing ("Back to First Principles"):** This encompasses questions examining the fundamental ways in which concepts related to AI and inclusion are defined.
- **Bridge-Building:** This mechanism contains questions related building infrastructures such as networks and liaising between multiple stakeholders.
- **Educating:** These questions highlight both the education of AI-based technologies as well as the impacts that AI may have on education.
- **Tool-Building:** This mechanism contains questions pertaining to the building and usage of tools that use AI technologies.
- **Policy-Making:** These questions pertain to the sets of principles, guidelines, laws, or regulatory frameworks that govern AI.

2) Matrix

	Design	**Development**	**Deployment**	**Evaluation**
Defining and Framing ("Back to First Principles")	**Can we study past AI evolutionary cycles to predict future AI trajectory?** **To what extent should we be looking for**	**To what extent does the development of new categories as a way to forward inclusion further entrench repressive notions? How does AI deal with this?**	**How do different communities define "Artificial Intelligence" and how does it influence varying perceptions including fear and excitement surrounding the future of AI?**	**How do we ensure that the wealth generated by using AI systems benefits society as a whole and not just a few monopolizing companies?**

	technical solutions for social problems? Is the core focus of AI on 'optimization' fundamentally at odds with the plurality and messiness of real life, which inclusion and social justice sometimes celebrates and/or fights for? How do AI biases compare to human biases? How do questions of bias and neutrality overlap? Is debiasing about being neutral, aspiring to solve a degree of neutrality? Or is it about creating a values paradigm and making that paradigm explicit? How can we design inclusive AI systems considering the strong	**How can we incorporate an international human rights framework into AI development?** How can we create AI that can help us benefit those who have the least to do with creating it? Behavioral expectations lead to "behavioral normalization": how can we study this topic without falling in the trap of "categorization" and narrowly defining what it means to be a user?	**Explaining AI behavior and decision-making is not a trivial task -- how do we trust AI without always fully understanding its functionality, and what are the implications of that "blind" trust?** Is AI condemned to be a driver of social exclusion, as it will reduce labor necessary to deliver goods and services? Can platform cooperatives, blockchain be alternatives to the present exclusion process? What does it feel like to live within an algorithmic environment — for example, at work? (think working for Uber, operating drones) How should we talk about responsible innovation? When it comes to inclusion, do we prioritize an individual or	**What are the ramifications of reproducing binary systems in code?** **Are there good proxies to look at fairness in AI? Are explainability and transparency the ultimate goal?** Do we need to know how much AI is all around us? Considering we are already cyborgs, is AI itself included when we talk about "inclusion"? Is global data mining and the power imbalance in AI production a form of digital colonialism? Is there a relationship or parallelism between the industrial revolution and AI about the challenges and opportunities of natural resource exploitation?

	and intertwined people-machine relationships AI enables? Does this relationship need less, more or the same amount of trust than a human-human relationship? How do we define inclusion? Based on groups, minorities, gender? Is there an ethical framework surrounding about what kinds of data companies can get from the user? How can data capture cultural norms in such a way that enhances inclusiveness? How do we enrich data with "data" that is unseen or slightly unnoticed such as body language, emotions, or feelings?		ecosystem-driven approach? What does it mean to be a ‘user’ or a ‘producer’ of autonomous systems? What are the ways in which users simultaneously serve as producers? How can we use this decentralization of production to achieve more equitable AI? Business is emotionally driven by exclusivity. How can this be related to inclusion in AI? What is the difference between AI informing individuals to enhance their knowledge, and AI replacing human decision-making, which yields dependency? How do the outcomes of an AI application differ when an emotional element is incorporated into algorithm design versus when it is not?	What are the costs and implications of AI "exclusion"? Why are AI ethical norms not yet explicit and what is preventing this from happening? Do we aspire for neutrality? Does the “digital native” generation suffer from the same rates and forms of mental health conditions as millennials and baby boomers? How has technology shaped mental health over the past three decades? Are there benefits of non-digital populations being affected by AI? If the future of labor primarily involves working from home/decentralized spaces, how might human interaction and psychology change?

	Are we holding algorithms to a higher standard than we are holding ourselves to? To what extent can AI replace or enhance human behavior? How do we create standardized boundaries delineating where AI should replace versus enhancing human capabilities? What are we trying to 'include' in social inclusion? What motivates us to create AI-based systems/algorithms/robots? How does this influence the types of AI that emerge? How do we include those in AI who do not want to be included in these systems?		To what extent are AI data and infrastructure challenges new? What is a public good and how is it defined? How does it intersect with corporate responsibility and sustainability? People can become data subjects by virtue of where they are born. How do we configure this in the context of social inclusion (access to digital networks) / exclusion (limited access to privacy and/or agency)? Is creating AI without the support systems of a workplace (an HR department, work benefits, a union) exploiting workers? Is the "fourth technological revolution" merely a manifestation of late phase capitalism when there is nothing to suggest fundamental shift within power	Do AI-driven applications reduce user agency? (Eg. Spotify's weekly music recommendations, tailored to every user individually which encourages reliance on algorithms). How can we avoid exacerbating existing digital divides? How does the movement for more inclusive and equitable AI mirror, or could benefit from mirroring, larger social movements such as the Civil Rights movement or the environmentalism movement? What is the public interest in having an "inclusive" AI? Is it trans/cross-national? How do we introduce opportunities and challenges together in a way that stimulates productive discussion of both, specifically in a Global South context?

	How can we rework the definition of inclusion based on how we as individuals position ourselves within the idea of inclusion? How do we denaturalize the "scripts" of AI and inclusion? How do we conceptualize the relationship between AI “and” inclusion? Inclusion often seen as secondary within AI -- maybe there is a need to flip this. How can AI aid and/or enhance inclusion?		relations or technology?	How do we define “AI success” (instead of just AI)? How is our focus on inclusion in AI giving us an excuse to not look at structural inequalities in "reality"? (race, gender, class)? And how is the normative reifying the boundaries between the online/offline, technology and the real world? How do we configure ethical standards to consider the questions we have not considered before and the ones we need to rethink in light of new technologies and circumstances? What are similar problems -- static frameworks dealing with a continually changing subjects -- in legislating as well as in history, that may shape how we navigate AI and Inclusion challenges?

Bridge-Building (Network Building/Liaising) and Infrastructure Building	**How can we align incentives between companies and activists to make profitable and beneficial AI?** **What are ways to empower software experts from the Global South? Funding?** How can we encourage international business models such as startups to capacitate local entrepreneurs? Can you imagine a homegrown AI-driven sector in your country/region, or will you be colonized? How can we make visible the bodies that do the labor of supporting AI and create a new labor framework? How do user expectations	**How can youth be stakeholders and more involved in building the AI world?** **How do we include more individuals in AI systems without erasing difference through normalization and universalization?** What are successful examples of how to engage women and minorities in AI design with the goal of helping non diverse organizations be more inclusive in their AI design? How can we empower developing countries to develop AI? What does AI for startups look like? How do we prevent AI-based health tools	**Can business conciliate their aim of collecting data with protecting the data of users?** **Is it possible for us to de-situate AI from the undeniable structure of societies based on surveillance capitalism?Give AI applications/uses a public good? Should it be publicized and insofar controlled?** Can small companies from the Global South compete with quasi-monopolies from the North? How can intellectual property regimes foster data access to promote inclusion? How do we build awareness about AI applications for the public? How can we ensure that the wealth generated by companies using AI systems benefits society as a whole	**Is it possible to set an AI Governance Forum?** **How will we make certain that we work together collaboratively in the age of AI to define and defend the public interest? How can we meaningfully collaborate?** **What does the current global framework of AI production/use look like? Who are the major players, what are their initiatives, and how can we lift up more grassroots initiatives?** If profit is the primary purpose of businesses, what can we develop to make sure AI is ethical? Who owns algorithms? Who designs algorithms? How do we empower Global South engineers? How is AI on a granular level affecting the most

	differ in the Global South from the Global North, and how can AI design accommodate those varying expectations? How can we make sure that the data going into algorithms are not only based on “Western” values as a form of colonialism? How do we get more women (in design, in these conversations)? How do we foster innovation of marginalized populations' data without also weaponizing that data? How can we move away from data protectionism to open and collaborative data sharing? Should we build our own infrastructure in the	from being focused on health issues prevalent in rich countries and not lower-income countries? Particularly as this imbalance is already the case for the health industry at large? Is there a probable/possible way to decentralize the production of AI outside of the US/China center? How will we ensure that vulnerable populations are not considered edge cases but are instead considered priority cases in AI production? What are the existing platforms and collective intelligence that we can pool to identify communities of practice in place? How do we promote	and not only a few rich people in the US and Europe? If we partner with private AI services, how can we ensure that ethical, contractual obligations are kept, when they can easily be overturned with profit seeking models? Is it possible to design business models for AI technologies that benefit data subjects economically? Should we support the proposition of a kind of public fund that would own our data (instead of digital giants)? What is your thought on the idea that Latin American is the guinea pig for AI? Are there non-exploitative AI systems people want to be a part of? Is this even possible?	marginalized individuals globally? How do we measure this benefit and/or harm? Who has the power to influence the discourse around what is fair? How can we ensure that even the way we arrive at a definition of fairness is fair? How do we make case studies more visible so that relevant AI and inclusion studies take place? What solutions are possible to address job loss in the global South resulting from AI? Is universal income one solution? How can AI increase access to quality health care in Global South if AI applications stem from Global North? Do users have a say in how they get defined as users? What are transparency and

	Global South or rely on that already built in the North? Should AI databases, particularly those produced by users, be placed - if necessary, by law - in the public domain? Cloud computing infrastructure must be a commodity, a "natural resource" of humanity—how do we get to this agreement? How could we include people with no background in computer science, poor internet connection, or no computer, to design more inclusive algorithms? What are the implications of victims of conflicts being used as data sets to train algorithms -- how	cooperation between the social sciences and developers? How does AI play into building online community, which requires a lot of data and data processing? How can development agencies ensure they aren't contributing to harms when contracting AI firms (eg. Palantir.)?	How do we open conversation to incorporate lawyers that don't work on "online" law? How do we ensure that we do not remain stuck in "AI excitement stage" in the Global South? How do we support groups and communities already working within the inclusion/social good space, but not necessarily "on AI?" How do organizations doing work localizing AI to their environments get funding to sustain their vision?	openness tactics that could help make the "user" more aware of their role in the AI ecosystem and bridge the divide between AI users and producers? Recognizing that AI tools will fail, who will they fail for and how can we redress that failure? How do we counteract technological brain drain when Global South culture does not incentivize practitioners to deliver the "tech for good" message? When using data as a proxy for research-based knowledge (especially in the social sciences), it can lack the granularity of what a social scientist would do, but we use it anyway. What are the limits and potential risks of this?

	could this data collection help or harm refugees/Internally Displaced Persons/asylum seekers? How can ground-up data collection as an AI method mitigate algorithmic bias? What are successful examples of the academic technology community bridging industry and engineers into conversations and solution creation?			
Educating	**How can we improve education systems -- especially in engineering -- to promote more inclusive and ethical AI? Should this improvement in engineering education include moving away from logic back into mathematics and will**	**Can we work with makerspaces to include AI into their offerings? Those places can host online courses in groups (marginalized youth) and can be hubs for cutting edge digital education & creativity**	**How do we democratize access to education, particularly technology-focused education, in the context of lower-income countries that speak little or no English which is typically the dominant language in widely used education tools?**	**How does privilege manifest itself in the means of education chosen by different groups/stakeholders? What is the prevailing narrative about AI and technology in the classroom, and who shapes it?**

	this curriculum require new forms of engineering? Can we revise AI courses at universities so that graduates understand inclusion implications of the tools they develop and consider inclusion - related issues, such as de-bias and interpretability, in this development? When we talk about inclusion, how can we ensure that available information & sources for all people are equal? Who will train the trainers in developing countries? How do we think critically and make decisions around data acquisition for educational data training sets?	How can we personalize technological learning resources, such as MOOCs? How do we extend the inclusion dialogue over time among developers of AI? How can law scholars contribute to the AI debate without hampering the process of development?	**How do education tools for expanding knowledge and use of AI currently reach populations globally, and how can expanding access to educational resources positively reflect back in AI systems?** How can people be educated to know how their personal data is used in AI systems when you live outside the jurisdiction where the data is processed? How does AI-enabled teaching affect resource allocation within the educational system? (i.e. is it supplemental or does it take resources away from teachers?) How does the theory of mistakes as a learning process integrate machine learning? How visible are the “mistakes” that AI technologies make in	What can AI & Inclusion do for underdeveloped countries in education when the gap is significant? It is common to take infrastructure as something there, done, without power of influence or agency. How, why should we make it explicit that infrastructure is not invisible as many try to do? What are the ways in which we could begin retraining workers before their jobs are replaced, and who would fund and initiate those trainings (government subsidies, companies themselves)? What are ways to re-skill workers of an older age to ensure they can still make a living in 5 years?

	Are education and learning based on personal relationships? If so, what role do AI-enabled teachers play? What is the role of education in a world where remembering facts is less important and AI-enabled technologies perform many functions that humans traditionally performed?		the process of learning, and what is the effect on students? How can education and related resources help define clear user expectations influenced by users themselves? Can AI help to address the challenges of scale that educational systems often face?	
Policy-Making	**Is there a need for a (legal) definition of AI? Is this definition defined at a local, national, or universal level? Is it binding? What stakeholders decide on the definition?** AI is biased, as well as the algorithms behind them. Could bias turn into a crime? Could we create an	**How can we harmonize data protection regulations and the rights to explanation with complex AI models that are not self explanatory and are nonlinear as deep neural networks? How can the right of data removal be articulated in the AI context without causing undertilting and misrepresentation of class?**	**How can we force non-democratic countries to abide by the governance systems of AI and what are the implications of a lack of global guidelines/standards?** **How do we include AI ethics into the global policy frameworks we build?** What can debates on AI and debates on Internet regulation learn from each other?	**What will labor policies look like when most of the population works in informal conditions brought on by the AI age? How can these policies best protect workers?** **As AI is going to be able to predict diseases, how do we mitigate violations of privacy and bioethics that discriminate against individuals based on existing conditions or even their**

	algorithm that respects specific country laws? What are methods of preventing ethical privacy issues in collecting data on health by establishing a governance of global health data flows? Should government data ever be merged with private company data for groups with for-profit interests? What sensitive open data does the government collect to use in algorithms and how can it be better monitored to ensure the data it's using is as inclusive as possible?	Given the dominance of Silicon Valley, will it be possible to strengthen data protection laws in the USA? Following GDPR? Considering the difficulties in designing an explainable and intelligible ai systems, as well as the embedded bias in data sets, should certain AI applications receive strict and prohibitive regulations? Should there be a tax imposed on the development of AI to address the loss of jobs in primary industries in developing countries? What mechanisms can we implement to hold the AI power-holders accountable in the private and public	What is the role of the state in the AI political economy? What (and how) should governments do to rebalance power from the private sector to individuals or users? Should we start talking about the right to be excluded from an AI controlled world? How should governments use AI to learn about the people and their necessities without identifying them? Intellectual property rights and antitrust law prevent algorithms from being scrutinized. Would AI risks of exclusion be a reason to mitigate IP rights? Consent doesn't work when systems use AI without users being aware - is this a good thing given global south issues of agency/consent & what protects users better?	**predisposition to health conditions?** How can we create constant auditing processes for AI algorithms considering that data gets renewed constantly and its amount increases each day? How will AI affect laws and governance? In formulating new legal provisions in order to adopt to AI, how should the transition be dealt with? Should an AI and human intelligence legal framework coexist for a certain amount of time? How could compliance to AI standards be enforced in countries that have a tradition on corruption? Is data protection law hampering or helping equality? More broadly, how can we use law to improve

		sectors? How do we challenge people, companies, governments controlling AI space to address challenges of AI?	How can we support consistent application of AI across contexts? e.g. background checks for drivers with ride-sharing services in many different countries If algorithms challenge the integrity of legal infrastructure that government protects, who addresses it? Who do you talk to? Do states have a role in promoting the use of AI to prevent and/or mitigate humanitarian crises? How do we ensure that corporations such as large health insurance companies do not utilize AI systems to disadvantage individuals in ways such as denying coverage? What are methods of making existing biases transparent to consumers so that they can take them into account in their	equality in the context of AI? How can/should control over data and infrastructure be governed? What are appropriate oversight mechanisms and how can they be implemented to empower people around the world? How does the regulation of AI help or hinder companies' existing business models, while being sensitive to the needs and rights of innovators and users? Who are appropriate independent algorithm auditors and what are their guidelines? What are appropriate remuneration mechanisms? What (multi-level) AI differentiations impact rights

			actions and decisions surrounding AI-based products?	and how can we address that from a legal POV? What might hacks and security breaches mean for the health sector and what are cases we can anticipate and work to mitigate? How can we equitably regulate cross-border health data flows so that they promote impactful outcomes without sacrificing privacy? How do you regulate how inclusion and bias are expressed in AI educational systems/modalities? How do we make sure that accountability and oversight mechanisms remain meaningful? What do optimal AI related legal protections/government systems look like?

				How do we get policy makers focused on the AI challenges and opportunities surrounding jobs?
Tool-Building	**AI is a hard design material due to its dynamic nature. How can we design something that is constantly changing? How do we explore "what-if" scenarios of unknown?** **Who is "the reasonable" man we should configure AI systems around? Does such a person exist? If so, what does this person look and act like?** How do we democratize the building of datasets? How can we build a coop to train systems/data well? How do we deal with	**What are ways that we can use algorithms, particularly in search engines, to disseminate legitimate helpful health information that promotes positive health behavior?** How can AI contribute to developing better participation and accessibility in society for people with disabilities? How can AI include transgender people? How can we add analytics to include multiple gender options to current AI panel systems? How can we use deep	**In what respect can businesses from the Global South provide AI applications that businesses from the Global North cannot?** **When and how does AI lead to price discrimination or "personalized pricing?"** How are big organizations like Google, Facebook, Microsoft and Amazon trying to debiase their results? How can we deal with the huge discrepancy in the market that comes along with the use of AI by a few companies at the expense of small and medium businesses, especially in the legal services business? Is it technically possible to	**What are the kinds of human-auditable control mechanisms that can be built into decision making algorithms?** **Can we develop technical models (test cases or databases) to ensure that the AI & Inclusion discourse progresses?** How to evaluate different practices of "ethics by design" Are there examples of best practice in design that can influence AI? How can the "perfect" predictability of algorithms result in greater risk or potential harm to population

	biased data due to copyright (not all content is accessible and allowed to be heard)? How can people who are currently building models start to deal with biased data and models in a practical way? How can we change the constructs of AI technology to be more inclusive? Which characteristics are relevant (depending on the domain) to be included in AI models to reduce discrimination? How do we design algorithms for vulnerable populations without imposing asymmetrical power dynamics? How can we design AI that works well with	learning in order to include youth and improve education statistics? How can we support the trust and privacy process in producing and consuming information online? Is it conceivable in the near future that commercial tools and methodologies for software development include, by default, features that address inclusion concerns? What are the most significant potential opportunities and risks at the intersection of AI and health? How do we represent 'code switching' through code?	have the advantages of AI (which are based on big data and scale) without concentrating data in big corporations? How can AI technologies contribute to the consolidation of digital-born media in developing countries? Can these technologies create a consistent business model for these small media? Could AI deployments help to create a robust environment for anonymity online? How does AI change the application/expectation of customer service? How can AI increase (or decrease) quality medicine accessibility by implementing differential pricing at an individual level? Could AI help build a universal insurance system?	health where there is not human context? What are the humanitarian risks of algorithmic discrimination in conflict settings? How can we ensure that the new HR mechanisms based on AI systems won't exacerbate exclusivity but promote a more diverse working environment? As AI is driving the improvement of human health which expands the longevity of human life, how can we balance the potential increase in global population size with the limitation in natural resources to sustain it? How do we ensure that oversight mechanisms are actually implemented in ways that empower people around the world, enabling them to use AI in ways that are good

	limited access to data (due to lack of information, literacy, and connectivity in the Global South)? Can we implement design to make algorithms more inclusive and accessible for knowledge-sharing and utilization? How do we include “point of views” into data? In other words, when there are alternative opinions how do we decide what to program in algorithms? Would it be possible to design an algorithm that monitors bias? How do we create human-driven design for AI? Is it possible to code ethics or empathy in machines?	What preventative systems utilizing AI tools could be put in place by states? How can AI-based technologies help our global society cope with and/or mitigate climate change? Plain text interpretation can be difficult -- is there a best method of making an AI to correctly interpret public texts posted on places like Twitter? What does it mean to focus on de-biasing models rather than de-biasing data? What are the best methods of configuring AI models to be the least biased regardless of the data?	How can AI help women from underdeveloped countries (in security, health, etc.)? Land property is a significant issue of this century. Rights over land of First Nations around the world are under attack. How can AI help to protect those rights for Indigenous populations? In developing countries the media industry tends to be monopolistic, which reduces pluralism. How could AI technologies be used in this industry to contribute to democracy in these countries? How can AI be used to mitigate the migration crisis at large? What potential pitfalls should we be mindful of as AI is deployed in times of humanitarian crises?	for society and do not lead to a negative human rights impact?

	What methods could be implemented to identify datasets that could train models to predict onsets of violence? AI has the potential to make very accurate diagnostics, but how is AI able to identify nuanced mental health cases that often require human evaluation? How do we create AI tools that are expansive and inclusive with culture-specific understandings of wellbeing and lifestyle? Will algorithms be able to "see" the social inequalities that underlie many health problems (eg. caste in India) and make decisions accordingly?	How can AI be developed to be inclusive of elderly individuals? Unbiased data sets don't exist, but how can we be aware in order to not introduce bias into the group of dataset constructors?	How can new business models originating from AI help excluded groups? Is there space for inclusive entrepreneurship? How might requiring AI developers to sign an “ethical oath” of conduct promote inclusivity, transparency, and in turn equity, in algorithms? Are there large distributed power systems that could be created to enable technological work and remove the cost barriers?	

	How can we build AI tools that are not easily repurposed as autocratic agents of control? How can AI explore the "data" that is not explicit, such as emotions, feelings, etc.? How do we define a “decision boundary” in AI systems that we could use in the future to make non-discriminatory predictions? What mechanism could improve transparency in the design of AI? How can we identify the limits of algorithms before or as they are being designed? How do we incentivize the exploration / use of alternative to data-intensive AI			

	approaches (learning without data, etc.)? How can we standardize algorithms while keeping them culture-specific?			

III. Future Considerations

These are examples of topics that could warrant greater investigation in the research agenda in order to yield a more comprehensive agenda.

1. Themes that may warrant more research
    - The intersection of AI and culture/religion
    - The possible impact of AI on media production and consumption (e.g. journalism, film, music)
    - The environmental impacts of the AI age
    - Research vs. actionable research methodologies

2. Underrepresented perspectives / representations that may warrant more research
    - People with physical disabilities
    - Industry (Google, Twitter, Microsoft Research, Alibaba)
    - China, India and other non-Western major players

Global Symposium on AI & Inclusion: Day 1 Research Questions

Opening Remarks + Interactive Opening Session

How do we incorporate those who cannot participate online/in AI conversations in AI decision-making? How do we ensure these individuals have a role not just as consumers?

How can we empower developing countries to benefit from the design and deployment of AI? How do we support their contributions to relevant discussions?

How can we avoid exacerbating existing digital divides?

How do we engage youth in conversations about AI?

How do we include AI ethics into the global policy frameworks we build?

How will we ensure that vulnerable populations are not considered edge cases but are instead considered priority cases in AI production?

How will we make certain that we work together collaboratively in the age of AI to define and defend the public interest? How can we meaningfully collaborate?

How do we configure ethical standards to consider the questions we have not considered before and the ones we need to rethink in light of new technologies and circumstances?

Keynote 1: AI and the Building of a More Inclusive Society

How do we define a “decision boundary” in AI systems that we could use in the future to make non-discriminatory predictions?

How do education tools for expanding knowledge and use of AI currently reach populations globally, and how can expanding access to educational resources positively reflect back in AI systems?

What does it mean to focus on de-biasing models rather than de-biasing data?

How can ground-up data collection as an AI method mitigate algorithmic bias?

What are the best methods of configuring AI models to be the least biased regardless of the data?

How does the movement for more inclusive and equitable AI mirror, or could benefit from mirroring, larger social movements such as the Civil Rights movement or the environmentalism movement?

How can we rework the definition of inclusion based on how we as individuals position ourselves within the idea of inclusion?

Keynote 2: Inclusion in the Age of AI

What is the public interest in having an "inclusive" AI? Is it trans/cross-national?

What mechanisms can we implement to hold the AI power-holders accountable in the private and public sectors? How do we challenge people, companies, governments controlling AI space to address challenges of AI?

How do we foster concrete models to ensure that the AI & Inclusion discourse progresses?

How does the human rights field stay relevant in an AI context? How do we make sure that accountability and oversight mechanisms remain meaningful?

How do we ensure that oversight mechanisms are actually implemented in ways that empower people around the world, enabling them to use AI in ways that are good for society and do not lead to a negative human rights impact?

What do optimal AI related legal protections/government systems look like?

How do we denaturalize the "scripts" of AI and inclusion?

How can we improve education systems -- especially in engineering -- to promote more inclusive and ethical AI? Should this be away from logic and back into mathematics? Will this require new forms of engineering?

How do we conceptualize the relationship between AI “and” inclusion?

Does AI become a critique of inclusion politics?

Deep Dive: Advancing Equality in the Global South

Recognizing that AI tools will fail, who will they fail for and how can we redress that failure?

Why are AI ethical norms not yet explicit and what is preventing this from happening?

How do we bridge the gap between academia and industry, especially when industry is also looking for knowledge? How do we foster interdisciplinary dialogue? How do we foster society-wide discussion of ethics?

How can we leverage the transformative power of AI tools to do good - on individual, national, and international level?

Should we be looking for technical solutions for social problems, especially regarding fairness?

How do we empower talented engineers from the Global South to develop applications that amplify cultural specific values instead of being sucked into companies that develop platforms and tailor them to specific countries?

How do we ensure that we do not remain stuck in “AI excitement stage” in the Global South?

How do we introduce opportunities and challenges together in a way that stimulates productive discussion of both, specifically in a Global South context?

Should government data ever be merged with private company data for groups with for-profit interests?

When it comes to inclusion, should we prioritize an individual- or ecosystem-driven approach?

How do we counteract technological brain drain when Global South culture does not tell or incentivize practitioners to deliver the "tech for good" message?

Inclusion often seen as secondary within AI -- maybe there is a need to flip this. How can AI aid and/or enhance inclusion?

How do we support groups and communities already working within the inclusion/social good space, but not necessarily "on AI?”

Global Symposium on AI & Inclusion: Day 2 Research Questions

Welcome Back and Introduction to Day 2 - Bridging AI & Inclusion

From Slides:

How do we incorporate the perspectives of those who cannot participate in AI development and dialogues?

What are appropriate oversight mechanisms and how can they be implemented to empower people around the world?

To what extent should we be looking for technical solutions for social problems?

When it comes to inclusion, do we prioritize an individual or ecosystem-driven approach?

Deep Dive: Data and Economic Inclusion

What does it mean to be a ‘user’ or a ‘producer’ of autonomous systems?

How can thinking about data markets in a broader way help us rethink the relationship between data and economic in/exclusion?

When using data as a proxy for research-based knowledge (especially in the social sciences), it can lack the granularity of what a social scientist would do, but we use it anyway. What are the limits and potential risks of this?

How does AI play into building online community, which requires a lot of data and data processing?

Breakouts Set 1: Drivers and Forces at Play

User / Behavioral Expectations

- How can education and related resources help define clear user expectations influenced by users themselves?
- What are the ways in which users simultaneously serve as producers? How can we use this decentralization of production to achieve more equitable AI?
- What are the largest issues pertaining to AI and inclusion globally, and how can we build towards more inclusivity from there?

- How is AI on a granular level affecting the most marginalized individuals globally? How do we measure this benefit and/or harm?
- What is the difference between AI informing individuals to enhance their knowledge, and AI replacing human decision-making, which yields dependency?
- How can we identify the limits of algorithms before or as they are being designed?
- How do the outcomes of an AI application differ when an emotional element is incorporated into algorithm design versus when it is not?
- What are methods we can use to reduce uncertainty in the algorithms we create to understand where there are gaps?

Algorithms and Design

- What can be done at the design level of an AI system to promote accountability and transparency?
- How do users influence the development of algorithms?
- Who has the power to influence the discourse around what is fair? How can we ensure that even the way we arrive at a definition of fairness is fair?
- Is it possible to code ethics or empathy in machines?
- Are we trying to solve inherently social issues with technology? Is this wise?
- Are we holding algorithms to a higher standard than we are holding ourselves to?

Data and Infrastructure

- Are there non-exploitative AI systems people want to be a part of? Is this even possible?
- To whom do we owe representation and representative algorithms?
- To what extent are AI data and infrastructure challenges new?
- Who is “the reasonable” man we should configure AI systems around? Does such a person exist? If so, what does this person look and act like?

Business Models

- How do we define inclusion in the context of AI and business models? How do we imagine the distinct roles of stakeholders involved?
- How do we ensure transparency (and inclusion) across companies and their consumers?
- How can we support consistent application of AI across contexts? e.g. background checks for drivers with ride-sharing services in many different countries
- How does the application of AI change companies’ existing business models?
- How does the regulation of AI help or hinder companies’ existing business models, while being sensitive to the needs and rights of innovators and users?
- How does AI change the application/expectation of customer service?
- What is a public good and how is it defined? How does it intersect with corporate responsibility and sustainability?
- When and how does AI lead to price discrimination or “personalized pricing?”

Law and Governance

- Is there a need for a (legal) definition of AI?
- How do standards need to be negotiated differently?

- How do we ensure ability and agency – even when it isn't "personal" data? - Bastard Data
- Who are appropriate independent algorithm auditors? What are the guidelines? Do we need new intergovernmental orgs?
- How do we audit - input, output, examination of the algorithm itself?
- What are appropriate remuneration mechanisms?
- Do we want the "recipe" or do we want to know the "health effects?"
- How do we open conversation to incorporate lawyers that don't work on "online" law?
- What (multi-level) AI differentiations impact rights and how can we address that from a legal POV?
- If algorithms challenge the integrity of legal infrastructure that government protects, who addresses it? Who do you talk to?

Breakout Set 1 Report Back

How do we define our values in the AI context?

Are there good proxies to look at fairness? Are explainability/transparency even what we want?

If all are socially/politically constructed and technology has agency, do we look for something neutral? Do we aspire for neutrality?

What is an AI-driven business model? What can we learn from corporate accountability and sustainability?

How do we incorporate social justice into this discourse?

What is the role of the Network of Centers in promoting ethical AI?

What is an ethical business model for "big tech"?

What can we learn from and movement building, especially re. long term growth and momentum?

Breakouts Set 2: Application and Impact Areas

Shifting of Industries and Workplaces

- What does it mean to have a career trajectory in this world of emerging AI? How does this change the notion of pride in one's professional identity?
- How do we get policy makers focused on the AI challenges and opportunities surrounding jobs?
- How can we make AI help ensure job security, ie. make it part of the solution and move away from its destructive capabilities?

Health and Wellbeing

- To what extent can AI replace or enhance human behavior? How do we create standardized boundaries delineating where AI should replace versus enhance? Is this based on the severity of the health condition, the risk involved, etc.?
    - e.g. is the control of AI in preventing suicide more human-dependent because of the risk and sensitivity involved than, for instance, weight control or migraine tracking?
- Does the “digital native” generation suffer from the same rates and forms of mental health conditions as millennials and baby boomers? How has technology shaped mental health over the past three decades?
- In which cases might potential harms of AI outweigh potential benefits? What are the most significant potential opportunities and risks at the intersection of AI and health?
- How can the “perfect” predictability of algorithms result in greater risk or potential harm to population health where there is not human context?
- What are ways that we can use algorithms, particularly in search engines, to disseminate legitimate helpful health information that promotes positive health behavior?
- How can AI increase (or decrease) quality medicine accessibility by implementing differential pricing at an individual level?
- What might hacks and security breaches mean for the health sector and what are cases we can anticipate and work to mitigate?
- How can we equitably regulate cross-border health data flows so that they promote impactful outcomes without sacrificing privacy?

Education and Learning

- How can we personalize technological learning resources, such as MOOCs?
- How do we think critically and make decisions around data acquisition for educational data training sets?
- Are education and learning based on personal relationships? If so, what role do AI-enabled teachers play?
- How does privilege manifest itself in the means of education chosen by different groups/stakeholders? What is the prevailing narrative about AI and technology in the classroom, and who shapes it?
- Can AI help to address the challenges of scale that educational systems often face?
- How does AI-enabled teaching affect resource allocation within the educational system? (i.e. is it supplemental or does it take resources away from teachers?)
- How does the theory of mistakes as a learning process integrate machine learning? How visible are the “mistakes” that AI technologies make in the process of learning, and what is the effect on students?
- What is the role of education in a world where remembering facts is less important and AI-enabled technologies perform many functions that humans traditionally performed?

- How do you regulate how inclusion and bias are expressed in AI educational systems/modalities?

Social Inclusion

- What are the ramifications of reproducing binary systems in code?
- How do we represent 'code switching' through code?
- What are we trying to 'include' in social inclusion?
- People can become data subjects by virtue of where they are born. How do we configure this in the context of social inclusion (access to digital networks) / exclusion (limited access to privacy and/or agency)?
- Should social inclusion extend to AI systems?
- What impact does the gendering of autonomous beings have? (ex) assistants mainly given a 'female' identity
- What are similar problems -- static frameworks dealing with a continually changing subjects -- in legislating as well as in history, that may shape how we navigate AI and Inclusion challenges?

Humanitarian Crisis Prevention and Mitigation

- What preventative systems utilizing AI tools could be put in place by states?
- Humanitarian crises encompasses a wide range of issues — how do we identify which issues to focus on in an inclusive manner?
- What potential pitfalls should we be mindful of as AI is deployed in times of humanitarian crises?
- What methods could be implemented to identify datasets that could train models to predict onsets of violence?
- Do states have a role in promoting the use of AI to prevent and/or mitigate humanitarian crises?
- What are the humanitarian risks of algorithmic discrimination in conflict settings?

Possible Approaches and Solutions: Ideas, Case Studies, and Prototypes

What motivates us to create AI-based systems/algorithms/robots? How does this influence the types of AI that emerge?

Are there large distributed power systems that could be created to enable technological work and remove the cost barriers?

How do we identify the salient issues related to AI in a given context? (i.e. AI as partner in some geographic contexts).

How do we make case studies more visible so that relevant AI and inclusion studies take place?

Should robots have rights? What should their ‘inclusion’ into human society look like?

What do context-specific robot/AI ecosystems look like, and how do/should they interact?

Global Symposium on AI & Inclusion: Day 3 Research Questions

Intervention Points and Opportunities for Collaboration

Is there a need for a (legal) definition of AI?
Local, national, universal? Binding? Who decides?

How do we define our values in the AI context? Technologies embed norms and values, not just informed by them. Rather than de-bias make inevitable biases explicit?

What is an ethical business model for "big tech?" What are appropriate remedy mechanisms?

How do we audit algorithms and who are appropriate auditors? Do we want the "recipe" or do we want to know the "health effects?" Or both?

How do we define "AI success" (instead of just AI)?

How do we sustain the inclusive conversation we are having in Rio over time?

What is new and what is not new in terms of AI opportunities and challenges?

Is the "fourth technological revolution" merely a manifestation of late phase capitalism when there is nothing to suggest fundamental shift within power relations or technology?

Report Back from Cluster Meeting #3

Is data protection law hampering or helping equality? More broadly, how can we use law to improve equality?

What sensitive open data does the government pull for algorithms and how can it be better monitored to ensure the data it's using is as inclusive as possible?

What are the existing platforms and collective intelligence that we can pool to identify communities of practice in place?

How do different communities define "Artificial Intelligence" and how does it influence varying perceptions including fear and excitement surrounding the future of AI?

What are successful examples of the academic technology community bridging industry and engineers into conversations and solution creation?

What mechanism could improve transparency in the design of AI?

How might requiring engineers to sign an “ethical oath” of conduct promote inclusivity, transparency, and in turn equity, in algorithms?

What does the current global framework of AI production/use look like? Who are the major players, what are their initiatives, and how can we lift up more grassroots initiatives?

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

GLOBAL NETWORK OF INTERNET AND SOCIETY RESEARCH CENTERS

ITS Instituto de Tecnologia & Sociedade do Rio

Global Symposium on AI & Inclusion
Notes Synthesis

November 8-10, 2017 | Museum of Tomorrow | Rio de Janeiro

Session Contents

- **Session Summary** - overview of session and key takeaways
- **Conceptual Insights** - ideas that help us understand what inclusion means in the age of AI and how AI and inclusion intersect [note: does not cover basic intro to AI, generalities like "AI will impact jobs"]
- **Research Questions** - self-explanatory, also collected and compiled separately
- **Action Items**: suggestions for what can be done to pursue a more inclusive society with AI, ie. code-based solutions, coalitions, etc.
- **Examples/Case Studies**: anecdotes or examples related to AI and inclusion raised by session participants
- **Key Perspectives - Global South & Underserved Voices** - statements and data reflecting or amplifying Global South perspectives/voices from underserved communities

Day 1: Opening Remarks + Interactive Opening Session

Wednesday, November 8, 2017 | 1:00pm - 1:45pm

Session Summary:
In his opening remarks, Carlos Affonso Souza (Institute for Technology and Society of Rio de Janeiro) set the stage for a day of foundation-building, introducing **inclusion as a unique lens** to explore AI's ethics and governance, and urging participants to consider the **opportunities** overshadowed by challenges and the **developing countries** upstaged by wealthy established players. Luis Alberto Oliveira (Museum of Tomorrow) briefly spoke on behalf of the venue, an "experiential applied science museum" whose facilitation of journeys of exploration through future scenarios rendered it a natural fit for the forward-thinking Symposium.

Sandra Cortesi (Berkman Klein Center for Internet & Society at Harvard University) and Ronaldo Lemos (ITS Rio) kicked off the first session with an interactive opening exercise using the Berkman Klein Center's DotPlot data visualization tool to share participants' pre-meeting survey inputs. In addition to highlighting regions and stakeholders present, and participants' opinions on the most important challenges and opportunities presented by AI technologies, Sandra and Ronaldo noted that the three most important things participants hoped to accomplish all involved establishing **meaningful collaboration and collective learning mechanisms**. They proceeded to call on contributors from the audience representing different sectors and regions to share their goals for the Symposium and broader AI & Inclusion discourse. Felipe Estefan (Omidyar Network) closed the session on behalf of the Ethics and Governance of Artificial Intelligence Fund with a series of provocative open questions and a call for creative global collaboration cognizant of geographic and cultural context.

Conceptual Insights:

- Conference framed as opportunity to use inclusion as lens to explore **AI opportunities, not just challenges**
    - Note: we later heard that in Global South often AI challenges remain underrepresented
- According to DotPlot, the "three most important things participants hope to do at the symposium" all involved establishing **meaningful global collaboration and learning mechanisms**
    1) Explore opportunities for collaboration
    2) Learn about innovative projects focused on AI and inclusion
    3) Learn about the current state of AI and inclusion research from a global perspective

Research Questions:

- **How do we incorporate those who cannot participate online/in AI conversations in AI decision-making? How do we ensure these individuals have a role not just as consumers?**
- How can we empower developing countries to benefit from the design and deployment of AI? How do we support their contributions to relevant discussions?
- How can we avoid exacerbating existing digital divides?
- How do we engage youth in conversations about AI?
- How do we include AI ethics into the global policy frameworks we build?
- **How will we ensure that vulnerable populations are not considered edge cases but are instead considered priority cases?**
- **How will we make certain that we work together collaboratively to define and defend the public interest? How can we meaningfully collaborate?**
- How do we configure ethical standards to consider the questions we have not considered before and the ones we need to rethink in light of new technologies and circumstances?

Action Items:

- N/A (opening session - people weren't quite there yet!)

Examples/Case Studies:

- The Institute of Electrical and Electronics Engineers (IEEE)'s Global Outreach was originally "global" but only had representatives from the European Union. Now, IEEE is expanding its reach to be more inclusive leading to partnerships with China, Japan, India, Iran, Israel, South America, Ethiopia, Brazil, Mexico, and more countries are involved (Danit Gal, Chair of the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems Outreach Committee).
- There are still only 3.9 billion people not online. A lack of universal and affordable access to broadband is one result of the digital divide and a clear example of the obstacles we face as a global community going into the age of AI (Ursula Wynhoven, International Telecommunications Union).
- Faro Digital hosted AI workshops with youth in Chile and Argentina and will be presenting results at the next Conectados al Sur Symposium to take place in Costa Rica in January (Ezequiel Passeron, Faro Digital).

Key Perspectives - Global South & Underserved Voices:

- Need for proactive, **concerted outreach** effort to Global South organizations
    - ie. IEEE Global Outreach - "global" originally only had reps from EU: now China, Japan, India, Iran, Israel, South America, Ethiopia, Brazil, Mexico, and more after reaching out
- Important to recall that the promise of AI and mass digitization cannot be realized without universal and affordable access to broadband ➜ much of global south does not have

- [Felipe from Omidyar] “For governance and citizen engagement, it can be a challenge to seize an opportunity when we see it, especially for Latin America” ➔ opportunity to explore **LatAm-specific** challenges and opportunities with global applications and implications

- Helpful to see **SDG** targets and indicators as starting points - see how they can drive conversation for underserved voices, esp. re. gender equality
- **Youth voices remain underrepresented** in conversations about tech and AI - need to factor this into dialogues and decision making

Day 1: Keynote #1 - AI and the Building of a More Inclusive Society

Wednesday, November 8, 2017 | 1:45pm - 2:30pm

Session Summary:
Ansaf Salleb-Aouissi (Columbia University) bridged the technological and academic spheres by providing a useful overview of the history of AI, underscoring how **"old AI" remains at the forefront of new applications** and how we can use lessons learned decades ago to promote more inclusive AI today. Artificial intelligence, as first defined in the 1950s by way of the **Turing Test**, is a system that perceives its environment and makes decisions to maximize its chances of success, in turn serving as an extension of humans to improve their efficiency. Ansaf mentioned bank account numbers, GPS routing, video games, and autonomous vehicles as examples of AI applications old and new that passed the Turing Test and demonstrated the technology's' pervasiveness. She proceeded to discuss how AI systems are designed, differentiating different types of machine learning and demonstrating how their variance can create both **discriminative algorithms** and deficient understandings about how these systems discriminate. She concluded with an insightful "4D framework" for achieving inclusive AI:

1. **Develop knowledge** – promote proficiency in AI through education to close the digital divide
2. **Decipher models** – create rigorous framework/science on interpretability for short and long-term
3. **De-identify data** – anonymize data feeding into algorithms to preserve privacy
4. **De-bias data/algorithms** – identify and treat bias in data, or account for bias in the ML algorithms themselves, while reducing human bias in data collection and analysis

Conceptual Insights:

- AI is an extension of humans, intended to enhance ourselves and society
- We need to switch the focus of de-biasing data to de-biasing the models themselves
- Earlier forms of AI ("Old school AI") first developed in the early 1950's, are still present in AI systems today and influence how AI is designed and impacts society -- we can look at these past applications to better understand how AI has evolved and how it might progress
- AI has the potential to free up humankind from mental drudgery
- Positioning ourselves within the idea of inclusion will aid in the necessary reframing the definition of inclusion
- The current political economy of information is that there is a small number of geographies that design the tech for the rest of the world

Research Questions:

- How do we define a "decision boundary" in AI systems that we could use in the future to make non-discriminatory predictions?

- How do education tools for expanding knowledge and use of AI currently reach populations globally, and how can expanding access to educational resources positively reflect back in AI systems?
- **What does it mean to focus on de-biasing models rather than de-biasing data?**
- **How can ground-up data collection as an AI method mitigate algorithmic bias?**
- **What are the best methods of configuring AI models to be the least biased regardless of the data?**
- **How does the movement for more inclusive and equitable AI mirror, or could benefit from mirroring, larger social movements such as the Civil Rights movement or the environmentalism movement?**
- How can we rework the definition of inclusion based on how we as individuals position ourselves within the idea of inclusion?

Action Items:

- **Create a rigorous framework or science on interpretability**
- **Create connected co-ops of ML training data** that's representatives of populations
- **Build a massive social movement** on the scale of the environmental movement / CR movement in order to achieve equitability
- **Create globally accessible AI educational tools to improve AI literacy and build sustainable learning and use of AI**
- In an effort to mitigate algorithmic bias, **collect the stories of individuals around the world who stem from underserved and marginalized populations to troubleshoot current and potential problems with AI,** as well as to identify potential opportunities

Examples/Case Studies:

- Ansaf Salleb-Aouissi, Columbia University, described an open access AI course at Columbia with global representation as a model of democratized, international AI education.
- Current political and economic ties to Silicon Valley represents where power is centralized, which creates a power imbalance in the technology world in which a few monopolies in the Global North define the digital world (Mark Surman, Mozilla Foundation).
- Health therapist bots as a potential positive application of AI in developing countries (Nagla Rizk, Access to Knowledge for Development (A2K4D)).
- Using AI tools to track jaywalking incidents as a way to target interventions is an additional potential positive application of AI in developing countries (Nagla Rizk, Access to Knowledge for Development (A2K4D)).

Key Perspectives - Global South & Underserved Voices:

- We can debias data/AI outcomes by paying attention to people, uplifting their narratives/perspectives

- Uses and interventions of AI must be context-specific and include ground-up data collection as a method to mitigate algorithmic bias

- It is necessary to include youth in the learning, design, and dialogue of AI in everyday life, as they will be the ones most impacted

Day 1: Keynote #2 - Inclusion in the Age of AI

Wednesday, November 8, 2017 | 2:40pm - 3:30pm

Session Summary:

Nishant Shah (Leuphana University of Lüneburg) suggested that when framed as **AI *"and"* inclusion**, there is a reinforcement of the binary between the human (inclusion) and the technological (AI) that reproduces three harmful tropes — 1) AI and inclusion "working together" and a corresponding erasure of their not-so-innocent histories, 2) "in AI we trust" with AI as the trainable savior offering an inherent critique of inclusion politics, and 3) AI as an evil entity easily circumventing human actors under threat. Rejecting these reductive accounts, Nishant called for an embrace of the lived realities of computed, cyborg beings and for a pursuit of an **artificially inclusive, naturally intelligent** future through exploring shared tensions between AI and inclusion. The first tension is one between **representation and simulation**; while current rights-based inclusion politics revolve around a representational construct of the human being, technology is operating towards a simulation (and transcendence) of human intelligence. The representational framing of inclusion leaves little room for intersectionality, whereas AI's focus on simulations "renders the idea of a 'people's problem' into 'people ***as*** a problem.'" The second is between **possibility and probability**; both AI and inclusion are based within the realm of probability, but there is a growing need to shift towards the computation and implementation of possibility, which asks us to think beyond what is already known and the bounds of our recognized histories. In moving away from AI ***and*** inclusion towards the **framework of inclusive intelligence**, he suggests we will be able to achieve the following:

1. **rethink and denaturalize the units of measurement**, showing how lived realities and processes of humans are constructed on multiple scales
2. **situate both AI and inclusion within a larger ecosystem**, moving from living ***with*** to living ***within*** AI and conceptualizing inclusion as a continuously negotiated entity instead of a static referent
3. conceptualize **AI and inclusion as mutually constitutive**

Conceptual Insights:

- How we get down to analyzing relationship between AI and inclusion?
    - Trope 1: AI and Inclusion working together - "human frameworks problem"
        - Machine interventions → who is responsible, who needs to take blame?
        - Create bridges between AI politics and Inclusion politics → reinforce idea that emergent AI practices are "innocent"
        - Giving up position of critique once you say they can be neutral AI, good AI, etc.
    - Trope 2: In AI we trust
        - If we train algorithm to never make mistake, it can be a solution
        - "Better than a human, hope for future"
        - Here what produced problem is also what provides great hope

    - Trope 3: AI as essentially evil
        - AI can circumvent, no human actors, humans under threat and AI is cruel
            - Ie. human labor under threat, human relationships companion robots
    - Rather, we are all cyborgs (Donna Haraway) here is an invitation to stop thinking of yourself as human → part reality, part fantasy, etc.
        - We have to be cyborgs because otherwise, every time a technology changes, we will have to force ourselves to recalibrate ourselves as human

- Removing the '**and**' from AI and inclusion - 3 goals:
    1. Shifting away from the rhetoric of 'bringing' inclusion politics to AI (inclusion politics is not perfect)
    2. Prevents the naturalization of separation between AI and inclusion
    3. Stops the framing of AI as a huge disruptor that calls for an inherent shift in thinking about inclusion

  ⇒ Instead, we need to think of both as separate complex problems, each with its own sets of resolutions and reconciliations that may shape our **artificially inclusive** and **naturally intelligent** future

- Suggesting **inclusive intelligence** as a new model of thinking about AI and inclusion together:
    1. Denaturalize units of measurement (for both AI as well as inclusion)
    2. Placement of AI as well as inclusion within a larger ecosystem
    3. AI and inclusion as mutually constitutive (inclusion is not only a human problem)
       ⇒ bringing forth a new lived reality of the computed, of cyborgs

Research Questions:

- What is the public interest in having an "inclusive" AI? Is it trans/cross-national?
- What mechanisms can we implement to hold the AI power-holders accountable in the private and public sectors? How do we challenge people, companies, governments controlling AI space to address challenges articulated in Rio?
- How do we foster concrete models to ensure that the AI & Inclusion discourse progresses? (There isn't time to discuss for five years!)
- How does the human rights field stay relevant in an AI context? How do we make sure that accountability and oversight mechanisms remain meaningful?
- **How do we ensure that oversight mechanisms are actually implemented in ways that empower people around the world, enabling them to use AI in ways that are good for society and do not lead to a negative human rights impact?**
- What do optimal AI related legal protections/government systems look like?
- How do we denaturalize the "scripts" of AI and inclusion?
- How can we improve education systems - especially in engineering - to promote more inclusive and ethical AI? Should this be away from logic and back into mathematics? Will this require new forms of engineering?

- **How do we conceptualize the relationship between AI "and" inclusion?**
- Does AI become a critique of inclusion politics?

Action Items:

- **Concrete international standard** on different ethical, legal, political, social topics on which we have intl definition impossible to change
- **Ongoing intersectional audits** necessary for algorithms: distribution of false positives, false negatives (Joy Buolamwini's intersectional benchmarks for facial recognition)

Examples/Case Studies:

- Upper caste women in India were trained to be midwives, whose refusal to treat lower caste women created new modes of exclusion. This shift is an example, Nishant Shah (ArtEZ University of the Arts) argues, of Wendy Chun's statement that "technologies that promise us freedom are eventually going to recursively become the technologies by which we will be controlled."
- Women from the 1940's - 1960's WERE women, not just operated by them. It was only later that computers became a masculine object erasing the history. Inclusion of women in STEM fields, therefore, is not a new concept -- and this and other forms of inclusion need to be placed in the historical context of computing.
- Amazon advertised a t-shirt that said "Keep Calm and Rape a Lot." The startlingly offensive shirt was created entirely by an automated algorithm designed to add verbs to riff on "Keep Calm and Carry On" and offer them for sale without human involvement. This is clear example of what can go wrong when systems are not designed and overseen well and critically, and points to a fundamental problem rooted in a hybrid of human and machine interactions (Nishant Shah, ArtEZ University of the Arts).
- Digressions from statistical gender binaries in TSA millimeter scans trigger unnecessary pat-downs at airport security because a person does not "fit" into the strict gender boxes engrained in the system (Sasha Costanza-Chock, Civic Media at MIT).
- Ongoing intersectional audits of distribution of false positives and false negatives: Joy B. with Algorithmic Justice League is testing the most widely used facial recognition systems.

Key Perspectives - Global South & Underserved Voices:

- Not just about increasing efficacy of model - what are the moments when people who are going to be most impacted get to decide whether they participate or not? **Inclusion to self-determination.**
    - "Is what I'm asking for intersex/gender conforming people in database so that TSA can better deal with it? Better training for TSA agents? I don't know! Maybe we shouldn't have the system at all. Inclusion/exclusion/ \right to not be included in a system is important question."
- Must figure out to how to design AI systems so that harms and benefits do not reproduce structural inequalities that are constituted by matrix of domination + **in addition to looking at distributive outcomes that are just, look at *procedural***

- Feminist queer and postcolonial theories and activism - language is a tool of domination and control - naturalizing language is already a problem
    - **How do we denaturalize the scripts of AI and society?** What touchstones emerge because of this?

Day 1: Deep Dive - Advancing Equality in the Global South

Wednesday, November 8, 2017 | 4:00pm - 5:15pm

Session Summary:

Kathleen Siminyu (Africa's Talking) kicked off the Symposium's first deep dive with an anecdote about technology education for women in Kenya, noting that **inclusion needs to start much earlier** in the Global South to empower women in engineering classes and careers. She stressed that **more funding will be necessary to help local practitioners innovate**, as there is not a lack of creative ideas but rather a lack of capital; during the Q&A, some participants pushed back on this point, prioritizing cultural rather than economic stimuli. Next, Kyung-Sin Park (Korea University Law School) emphasized AI's role as an **enforcer, rather than controller, of human thinking,** providing "two pieces of good news": (1) human beings are able to (re)direct AI's outputs by way of **hard coded instructions** and constraints for fairness and nondiscrimination, opening up opportunities for affirmative action equivalents, and (2) legal systems, including data protection law, are capable of reigning in AI's monopolistic aspects. Rehema Baguma (Makerere University) then drew attention to the importance of **participatory design** in advancing AI research and development from within the Global South, as well as integrating co-design principles into the training of AI experts globally. Jennie Bernstein (UNICEF Innovation) closed the panel with a call for **real-time data collection** to identify new problems and novel ways of understanding them, as well as **tailored interventions** to address real-time needs.

During the discussion that followed, participants debated the existence and transcendence of **AI's "good/bad binary,"** the technology's ability to provide valuable solutions in specific localized situations while wreaking havoc in a broader geopolitical context; the questionable use of data from refugee camps illustrated this tension. Several participants from the Global South also shared stories about their **countries' current "AI excitement phases"** and overemphasis of autonomous systems' promise. This stood in stark contrast to the overemphasis of challenges mentioned at the start of the Symposium, and raised questions regarding how the Global South can simultaneously address AI opportunities and challenges in a way that stimulates productive discussions of both.

Conceptual Insights:

- Social and cultural beliefs/ecosystems and funding are perceived as impediments to inclusive AI development and discourse for Global South and global underrepresented populations
- It is helpful to view AI as an enforcer (rather than controler) of human thinking as long as instructions are clear and not conflicting
- Data protection law is important and reworkable
- Worth looking toward possible solutions ***beyond*** do no harm
- In the Global South, many see AI as seductive and focus more on opportunities than challenges.

- Good/bad binary - need to remember history/geopolitical scale, not just project-to-project basis

Research Questions:

- **Recognizing that AI tools will fail, who will they fail for and how can we redress that failure?**
- Why are AI ethical norms not yet explicit and what is preventing this from happening?
- **How do we bridge the gap between academia and industry, especially when industry is also looking for knowledge? How do we foster interdisciplinary dialogue? How do we foster society-wide discussion of ethics?**
- How can we leverage the transformative power of AI tools to do good - on individual, national, and international level?
- **Should we be looking for technical solutions for social problems, esp. re. fairness?**
- How do we empower talented engineers from the Global South to develop applications that amplify cultural specific values instead of being sucked into companies that develop platforms and tailor them to specific countries?
- **How do we ensure that we don't remain stuck in "AI excitement stage" in the Global South? How do we introduce opportunities and challenges together in a way that stimulates productive discussion of both? How do we foster a greater awareness about Global South-specific AI risks and opportunities?**
- Should government data ever be merged with private company data for groups with for-profit interests?
- **When it comes to inclusion, should we prioritize an individual- or ecosystem-driven approach?**
- How do we counteract technological brain drain when Global South culture does not tell or incentivize practitioners to deliver the "tech for good" message?
- Inclusion often seen as secondary within AI - maybe a need to flip it. How can AI help inclusion?
- **How do we support groups and communities already working within the inclusion/social good space, but not necessarily "on AI?"**

Action Items:

- Hard coded constraints re. discrimination and justice - AI "Affirmative Action" equivalents
- Data protection law - updated and enforced
- Industry-wide standards (to evaluate AI applications around Ansaf's 4D of AI and inclusion?)
- Participatory design/co-design in training of AI experts
- Funding more R&D based in Global South to increase research/apps created locally
- AI Panic Button/Kill Switch
- Global South engineering training to foster research in local context
- Direct Action Organizing
- Real-time data collection & intervention

Examples/Case Studies:

- Kathleen Siminyu (Africa's Talking) Found that in her community designed for women in Machine Learning and Data Science Kenya, there are still more men. One reason for this is that In Kenya, girls go through high school that don't offer physics, and they cannot enroll in engineering because of pre-requisites they often don't have. Girls are also told to concentrate on languages because traditionally better at that.
- In Kenya, an individual was working on NLP in Swahili, and though it has great potential and would be an inclusive technological develop stemming from the Global South, the project is just a pet project as there is not enough funding (common among developers in the Global South).
- Partnership with Iraqi telco - see how usage corresponds with realities (Jennie Bernstein, UNICEF Innovation).
- UNICEF is using Image recognition systems to interpret someone's image and give feedback on malnutrition (Jennie Bernstein, UNICEF Innovation).
- Cameras at intersections in Iran monitoring traffic and who is wearing a hijab (Sedighe Hematpoori Farokhy, Applied Science University).
- Palantir's use of risk systems in active combat zones: partnering with Palantir on more benign uses is still supporting the "nomination of people to be murdered" - which is a questionable if not problematic use of data from refugee camps. This is an example of how there isn't a good/bad binary -- you might do partnership with firms like Palantir, but you need to remember history on a geopolitical scale, not just on project-to-project basis (Sasha Costanza-Chock, Civic Media at MIT).

Key Perspectives - Global South & Underserved Voices:

- More funding is necessary to help local practitioners innovate ➔ AI remains a "pet project" for many Global South engineers who must pursue other work to pay the bills
    - This will also ensure more research and applications come out of developing countries
- Local educational training from the Global South itself can foster research with local context
- Global south narratives of "seductive AI" - countries stuck in "AI excitement stage"
    - Crucial to foster greater awareness within Global South - everyone should know that free services/that data can be used against you

- Inclusion needs to start much earlier ➔ women often discouraged from studying STEM at early age and then lack necessary prerequisites to advance in school
- Ensuring diversity in development teams of AI applications will enable individuals to better address differences in culture, gender, ethnicity
- There may be an opportunity for affirmative action-style AI analogous solutions as a corrective for "merit"-based automations in US
- Participatory design (co-design) in training of AI experts would help ensure underserved voices that may not have the technical acumen to push development are still being accounted for

Day 2: Introduction to Day 2 - Bridging AI & Inclusion

Thursday, November 9, 2017 | 9:00am - 9:30am

Session Summary:
This session shifted the dialogue from conceptual foundations to actionable insights and case studies. Three key points from day 1 were highlighted to guide upcoming discussions: (1) deconstructing the relationship between AI "and" inclusion, (2) configuring the idea of **self-determination** in relation to AI, and (3) moving past old AI narratives in light of new AI capabilities. A key question emerged in regard to AI development on a global scale: how can we harmonize the discussions and align normative goals using the levers of **individual agency and/or ecosystem-driven approaches**? Specifically for the Global South, navigating the good/bad binary and striking a balance between challenges and opportunities to produce effective dialogue were emphasized as crucial questions to address moving forward. Action items focused on refining and building governance mechanisms under which responsibility is shared in an equitable manner by diverse stakeholders with adequate accountability and power-balancing measures in place.

Note - content below draws directly from Day 1 summary slides

Conceptual Insights:

- **Deconstructing "AI and Inclusion"**
    - Ansaf's 4D of AI: develop, de-identify, decipher, de-bias
    - Nishant's benchmark: beyond computer and computation to lived realities of the computed ➔ mutually constitutive intersection
    - ***Do we need new vocabulary/education/policy?***

- **From Inclusion to Self-Determination?**
    - Power; the role of western imperialist values in AI bias + discourse
    - Burden placed on underrepresented communities to "perform" inclusion
    - Self-determination as an alternative?

- **Old AI Narratives, New AI Capabilities**
    - Tremendous potential to equalize and exacerbate/create divides
    - Urgency to collaborate quickly and meaningfully

Research Questions:

- How do we incorporate the **perspectives** of those who cannot participate in AI development and dialogues?
- What are appropriate **oversight mechanisms** and how can they be implemented to empower people around the world?
- To what extent should we be looking for **technical solutions for social problems**?

- When it comes to inclusion, do we prioritize an **individual or ecosystem-driven approach**?

Action Items:

- Industry-wide standards
- Real-time data collection & intervention
- Data protection law
- Connected co-ops for ML training data
- Intersectional audits
- Revamping STEM education
- Globally accessible AI educational tools
- Participatory (co)-design
- Direct action organizing
- Storytelling + ground up data collection
- Affirmative action equivalents
- AI interpretability framework
- Public algorithm banks

Examples/Case Studies:

- Within “Key Perspectives” below

Key Perspectives - Global South & Underserved Voices:

- Global South stuck in “**AI excitement stage**?”
    - ie. Kenya, India (4 critical AI articles, 220 on potential)

  How do we introduce opportunities and challenges together in a way that stimulates productive discussion of both?

- Navigating and transcending the “**good/bad binary**”
    - ie. merging of datasets - health data/insurance (South Korea)
    - ie. questionable use of refugee data (Palantir)

  What mechanisms can we implement to hold AI power-holders accountable?

Day 2: Deep Dive - Data and Economic Inclusion

Thursday, November 9, 2017 | 9:30am - 10:30am

Session Summary:

The Data and Economic Inclusion Deep Dive began by **outlining the scope of economic inclusion,** including its relationship with social inclusion. An emphasis was placed on treating data as a resource as well as an output, since disparities in both accessing and producing data contribute to economic exclusion. Stefanie Felsberger (Access to Knowledge for Development Center) called for a reexamination of data markets in a more nuanced way that accounts for the **labor that users (and content producers) of platforms undertake for algorithms**. Bruno Magrani (Facebook) pointed out some of the data-for-social-good initiatives that Facebook has undertaken in recent years,including advancing alternate text, disaster mapping, public health campaigns. Malavika Jayaram (Digital Asia Hub) followed up by drawing attention to the problems that arise when corporate data – which is not as granular as research undertaken by social scientists – is continually used as a **proxy for research-based knowledge**. Florent Thouvenin (Center for Information Technology, Society, and Law at University of Zurich) argued for a broader view of economic inclusion mindful of incomplete data sets stemming from not only different scales of use due to the digital divide, but also different populations actively choosing not to use AI-driven services out of privacy concerns. He drew attention to the **the market landscape of AI services' domination by a few major players** and called for open data initiatives to facilitate the sharing of data and lessen the barriers to entry for smaller companies. Felipe Estefan (Omidyar Network) called for **examining the inequalities of power** as a way of understanding where economic exclusion may be happening, putting forth a list of "10 things that we need to talk about" (see above). Towards the end of the session, he advocated for moving away from the binary construction of for-profit companies as bad and non-profit organizations as good, suggesting **profitability should not be the only factor in judging the soundness and impact of business models** when evaluating economic inclusion.

Conceptual Insights:

- Economic Inclusion isn't something we just think about in Global South contexts → much broader (ex. Privacy becoming a privilege)
- Raises a need to rethink users' roles as a more active agent in AI systems, taking on a burden of labor of training the algorithms
- For-profit business models CAN HAVE SOCIAL IMPACT; moving away from binary construction of for-profit as bad, non-profit as good people
- Profitability cannot be the sole factor for judging the soundness of the business model

Research Questions:

- What does it mean to be a 'user' or a 'producer' of autonomous systems?
- How can thinking about data markets in a broader way help us rethink the relationship between data and economic in/exclusion?

- When using data as a proxy for research-based knowledge (especially in the social sciences), it can lack the granularity of what a social scientist would do, but we use it anyway. What are the limits and potential risks of this?
- How does AI play into building online community, which requires a lot of data and data processing?

- **From Felipe Estefan's 10 list of things we need to talk about** (Omidyar)
    - The economic impact of data and AI revolution is not equally distributed
        - Different contexts (human rights, labor laws, extc)
        - **How can we collaboratively understand where the areas of vulnerabilities are across geographies and contexts?**
    - Unpopular LatAm president leaders (crisis of representation)
        - Making people believe that the current systems that have been put in place are wrong ➜ that democracy is wrong ➜ people losing faith in political safeguards
        - **How do we use AI to instill faith in working political systems?**
    - Proactive collaborative approaches (between academics and philanthropy)
        - **How can we create proactive collaboration strategies?**

Action Items:

- Provision of more privacy-friendly online services in order to be able to gather more data that will enable better online services
- Open data initiatives, sharing of data, access rights of 3rd parties?

Examples/Case Studies:

- One of Facebook's biggest platforms is used in Egypt (it's so ubiquitous that police can ask you to show your Facebook). Users 'Tagging' of photos at a high rate lessens the amount of labor that the algorithm has to do. This raises a need to rethink users' roles as a more active "producing" agent, as they take on the burden of labor that fuels the design (Stefanie Felsberger, A2K4D).
- An automatic alternative text tool -- or a tool to recognize objects within images -- would be useful to allow people with visual impairments to have the page read to them (Bruno Magrani, Facebook Brazil).
- Using AI for disaster mapping and humanitarian response strategy: after catastrophic flooding in Peru in March of 2017, aggregated, anonymised data on the effects of the flood were available which helped organizers know where and how to send aid (Bruno Magrani, Facebook Brazil).
- During the Zika epidemic, UNICEF was targeting their awareness campaign for Zika prevention at women. Facebook gathered aggregated anonymous data about how people were talking about Zika in Brazil during the epidemic, however, and found that it was mostly men who were talking about Zika and decided they wanted to empower men to think about what prevention steps they could take and used the data to build a more effective/targeted campaign (Bruno Magrani, Facebook Brazil).

Key Perspectives - Global South & Underserved Voices:

- Unpopular LatAm president leaders (crisis of representation)
    - Making people believe that the current systems that have been put in place are wrong ➜ that democracy is wrong ➜ people losing faith in political safeguards
    - How do we use AI to instill faith in working political systems?

Day 2: Breakout 1 - Drivers and Forces at Play

Thursday, November 9, 2017 | 11:00am - 12:15pm

Breakout 1A - User / Behavioral Expectations

Session Summary:

The focus of this breakout session was the **need to create clearer user/behavioral expectations** that promote user choice and discourage dependency on AI. Because the global power dynamic of producers and users of AI is imbalanced, **user expectations vary significantly based on context**, which disadvantages vulnerable populations and discourages algorithmic accountability and transparency. Populations in the Global South, therefore, have fewer – or in some cases a complete lack of – clear user expectations. This expectations gap leaves users reliant on large producers of AI, discouraging them from pursuing a meaningful understanding of AI technologies and their personal implications, and dissuading them from embracing their roles as users in influencing global AI policy. First steps towards achieving more useful user/behavioral expectations include (1) providing accessible education that demystifies AI, (2) hiring engineers from diverse backgrounds, (3) incorporating data that promotes greater "emotional" intelligence into algorithms, and (4) identifying potential uncertainties present within existing algorithms. To supplement these actions and foster more equitable producer/user dynamics and expectations, **large producers of AI must partner with marginalized populations globally** to gain a better awareness of how AI impacts or may impact them.

Conceptual Insights:

- Idea of choices is critical in defining clear and individualized user/behavioral expectation
- Default expectation that AI will make decisions for us ➔ we need to focus AI on enhancing human decisions
- There is a need for more awareness of AI limits and how we can work towards a world in which AI helps inform users
- We need to incorporate emotional intelligence into algorithms as well as more context-specific data to help algorithms make more holistic inferences

Research Questions:

- How can education and related resources help define clear user expectations influenced by users themselves?
- What are the ways in which users simultaneously serve as producers? How can we use this decentralization of production to achieve more equitable AI?
- How is AI on a granular level affecting the most marginalized individuals globally? How do we measure this benefit and/or harm?
- What is the difference between AI informing individuals to enhance their knowledge, and AI replacing human decision-making, which yields dependency?
- How can we identify the limits of algorithms before or as they are being designed?

- How do the outcomes of an AI application differ when holistic design is incorporated into algorithm design versus when it is not?
- What are methods we can use to reduce uncertainty in the algorithms we create to understand where there are gaps?

Action Items:

- Use education and other resources to reach a point of clear user expectations
- Bridge the gap between engineers and academics, connecting them in conferences, meetings, and research projects
- Build a formalized network of diverse engineers and encourage employers to hire engineers from more diverse backgrounds using evidence of this success
- Demystify AI through informal and pervasive education channels, including YouTube tutorials, information dissemination from community leaders, and open-source online learning courses.

Examples/Case Studies:

- Credit score Study: at Yale a group solution efficiency increased significantly with aid of chat bots, demonstrating the power of chat bots and other AI-based tools in maximizing our computational abilities.
- In Brazil users suffered by blindly and fully trusting an AI-based transportation app. The app didn't have an understanding of dangerous neighborhoods as that training data wasn't incorporated and the passengers ended up in a neighborhood they shouldn't have been in and were killed. This examples demonstrates the need to holistically train data, understand algorithms, and maintain a human element in the process.
- In Colombia it takes years, and at times even decades, to pass new policy; the process of developing AI policy, therefore, needs to begin now as this technology is developing too quickly to wait.

Key Perspectives - Global South & Underserved Voices:

- In Kenya there are no or very little user expectations when it comes to AI and other emerging technologies ➔ how do we get to this point and promote the idea of choice in the process?
- Public policy needs to start now: In Colombia it takes years and sometimes decades to pass a policy, but in this technology is developing too quickly to wait on policy
- We need to move away from thinking of AI as a possible replacement of humans towards an enhancing mechanism for human decision-making

- We need to speak with the most marginalized populations globally to determine this technology on a granular level affecting them and obtain their perspectives. How do we measure/quantify who AI benefits/harms most and how?

Breakout 1B - Algorithms and Design

Session Summary:

This breakout session raised two main questions about algorithms and design: (1) How do we deal with the **multiplicity of definitions regarding fairness**? and (2) Is it possible to **design and code ethics/empathy** into machines? The power dynamics inherent in conceptualizing fairness were emphasized, with participants questioning who has the privilege to make decisions about fairness and drawing attention to the lack of (accessible) case studies from the Global South that serves as a barrier to framing fairness from a Global South perspective. During the session, participants also put forth two concrete possibilities for action items to explore. The first was designing an **independent auditing system** that combines both code-based interventions (creating and disseminating "fair" data sets, deploying test cases for algorithms,) and regulatory measures (instituting a "fairness certification"). The second was improving the integration of **ethics education in computer science curricula**, especially at the college level, with a specific focus on bridging the gap between the fundamental differences in conceptualizing the very ideas of bias and fairness.

Conceptual Insights:

- Fairness is a normative concept: would we want to enforce & legislate for fairness - how do you correct for specific kinds of unfairness?
- Algorithms also have social and political agency → a need to shift our thinking towards AI as a partner, not a subservient being. Are we holding algorithms to a higher standard than we are holding ourselves to?
- Notion of 'un-biasing' data sets or algorithms makes no sense; bias always exists, even within humans
- Engineers are not just implementers; must keep in mind that there already exists a set of engineering-specific ethics and regulations. The question should be how we merge this with the social sciences perspective, rather than how social sciences should be introduced into the engineering curriculum.

Research Questions:

- What can be done at the design level of an AI system to promote accountability and transparency?
- How do users influence the development of algorithms?
- Who has the power to influence the discourse around what is fair? How can we ensure that even the way we arrive at a definition of fairness is fair?
- Is it possible to code ethics or empathy in machines?
- Are we trying to solve inherently social issues with technology? Is this wise?
- Are we holding algorithms to a higher standard than we are holding ourselves to?

Action Items:

- Code-Based
    - Can we develop however many different algorithms for specific definitions of ‘fairness’ in specific contexts, and blending them as needed?
    - Creating and sharing ‘fair’ data sets that new ones are able to be tested against
    - Developing solid ‘test cases’ for fairness in algorithms
- Regulatory
    - Need for independent auditing systems: outside experts evaluating data sets and algorithms, especially in ‘high-risk’ contexts
    - Instituting a ‘fairness certification’?
- Education
    - Bridging the gap between how ‘bias’ is thought of in CS (issue of accuracy) and social sciences (complex ideas) through education at the university level

Examples/Case Studies:

- COMPAS, the risk and needs assessment tool designed primarily to assess an offender’s risk of recidivism is an AI tool that is already shaping our justice system. The discourse around COMPAS holds promising of providing multiple possible answers to the question of what fairness is in the criminal justice system.
- Racism is outlawed in Brazil: could this law be used as a basic set of guidelines that could encode ‘anti-racism’ into autonomous systems?

Key Perspectives - Global South & Underserved Voices:

- Lack of cases, especially from the Global South; we keep coming back to the same cases about ‘fairness,’ and we don’t know what ‘fairness’ can look like in a Global South context because there aren’t enough cases

Breakout 1C - Data & Infrastructure

Session Summary:

According to participants in this breakout session, AI technologies should be reconfigured to **emphasize representative algorithms that can restructure socio-political relations**; there is no such thing as "the reasonable person" when developing and deploying such technologies. Meanwhile, no one is unbiased and neutral, as neutrality assumes an absence of evaluative choices and values that inform bias. In addressing data and infrastructure, many participants felt stuck between two bad options: **digital colonialism** (exploitation with service provision) and **exclusion** (data invisibility, lack of access). Cognizant of the fact that structural and access inequalities will further AI challenges, they agreed consideration of the colonial gaps within and across country lines must be examined more closely and with greater specificity. Non-exploitative AI systems may only be possible through **new access modes of participation/public-private partnerships** that aim to address such starting points in new and emerging technologies in developing countries.

Conceptual Insights:

- There is no such thing as neutrality -- everyone has bias.
- You cannot know what you're missing unless you're part of the process, and able to participate.
- Emphasizing representative algorithms helps restructure socio-political structures

Research Questions:

- Are there non-exploitative AI systems people want to be a part of? Is this even possible?
- To whom do we owe representation and representative algorithms?
- To what extent are AI data and infrastructure challenges new?
- Who is "the reasonable" man we should configure AI systems around? Does such a person exist? If so, what does this person look and act like?

Action Items:

- Non-exploitative AI systems may only be possible through new access modes of participation / public-private partnerships that aim to address such starting points in new and emerging technologies in developing countries
- Gather more data / do more research in underdeveloped countries
- Emphasize representative algorithms

Examples/Case Studies:

- Facebook using India as testing subjects for digital engagement practices/expanding internet use through their application "Free Basics," which India ultimately rejected as a form of digital colonialism. Is it better to be included in conversations but exploited by AI algorithms, or better to be unexploited but also excluded? (Pranesh Prakash, The Centre for Internet & Society)

- Poland's talented doctors go abroad to Sweden to learn new technologies and possibilities in medicine with AI -- some of these doctors return but many don't; is this sort of "brain drain" good or bad for Poland, given that skilled physicians are leaving but simultaneously are gaining knowledge and expanding the country's international reach? (Alek Tarkowski, Centrum Cyfrowe)
- AI in weather forecasting and disease prognosis in Uganda has already been hugely helpful at the social level in Uganda (Julianne Sansa-Otim, Makerere University).
- The environmental movement can serve as a model for particularly Brazilians who care about issues surrounding AI including the success, failures, and need for due process for participation in a movement.
- The lack of a participatory framework surrounding AI and internet governance in China -- which is primarily dictated by norms that the states impose -- is one example of how participation varies in different contexts. These differing contexts have implications for AI (Marcelo Thompson, University of Hong Kong).

Key Perspectives - Global South & Underserved Voices:

- Better to be "data colonized" -- you get access and you're part of the debate.
- Data colonialism vs. data exclusion

- Colonialism gaps exist within and across countries.
- Data colonialism vs. data exclusion

Breakout 1D - Business Models

Session Summary:

The two liveliest conversations during this breakout session revolved around two questions: (1) Is it possible for a business model that relies on AI technologies to move beyond zero-sum and benefit multiple categories of stakeholders equitably, fairly, and "inclusively" – and what does this even mean/who decides?; and (2) How does **consumer understanding of the nature of a marketplace**, formal or informal, change the **normative perception of AI-enabled pricing**? Participants felt that the first question had more to do with traditional business and was not new to AI, drawing from corporate responsibility and sustainability debates. What was new involved the second point: **what role does/should AI-fueled decision-making play, and how does/should it change given a consumer's perspective and awareness of it?** Participants considered the ethics of "personalized pricing" and how consumer behavior changes when they perceive that a context is formal or informal (i.e. a physical marketplace or a search on Google Flights). This sparked commentary on the difficulty of **establishing trust across contexts** (e.g. background checks for ride-sharing services in different countries, verification of identity), which will likely be further complicated by AI-based black box technologies when not even engineers are able to explain outcomes. The breakout concluded with discussion of another critical question: to what degree do AI-enabled technologies actually **change a business paradigm as opposed to consumer perception**?

Conceptual Insights:

- There is a different in the production and the consumption of technology, both of which form part of a given business model. AI may be used in the production of technology, and also in the consumption of it. What practical implications does this have? How do you think about the application of AI sequentially, or staggered in time?
- Moving away from a zero-sum game was a common goal expressed by participants; this, however, is common to most discussions around the relationship between business and society. The insight is: ***what is explicitly new about AI***?

Research Questions:

- How do we define inclusion in the context of AI and business models? How do we imagine the distinct roles of stakeholders involved?
- How do we ensure transparency (and inclusion) across companies and their consumers?
- How can we support consistent application of AI across contexts? e.g. background checks for drivers with ride-sharing services in many different countries
- How does the application of AI change companies' existing business models?
- How does the regulation of AI help or hinder companies' existing business models, while being sensitive to the needs and rights of innovators and users?
- How does AI change the application/expectation of customer service?
- What is a public good and how is it defined? How does it intersect with corporate responsibility and sustainability?
- When and how does AI lead to price discrimination or "personalized pricing?"

Action Items:

- Formally investigate the notion of price discrimination / personalized pricing in the context of online markets. In what way is AI acting as a 'Robin Hood' in helping to 'distribute wealth' according to the judgments of algorithms?
- Codify types of cases / business models found within AI. What does an ethical business model look like for 'Big Tech'? How can we build a typology of the different kinds of business models?

Examples/Case Studies:

- Iran ride-share application "Snapp": women have been raped because drivers passed the background check as everyone is easily able to obtain a Snapp license, which the company itself was not aware of. This is an example of how business models need variation by context but require certain ubiquitous standards. Additionally, where is the human component of deciding whether to trust a driver if one is just trusting the technology?
- Cambridge Analytica is an organization that profiles voters and uses their data to create personalized political messaging that affects behaviors and voting patterns. This is an example of a "big tech"/"big data" business model -- what are the ethics that they impose within this model?

Key Perspectives - Global South & Underserved Voices:

- Whether a market is informal or formal influences how a consumer behaves. In a physical market (an informal context), customers expect to negotiate for goods. A similar case can be made for an online market for goods, such as eBay. However, when a customer searches a given flight that they would like to purchase on Google, and the algorithm adjusts the price higher due to this particular demand, the customer may believe they are operating in a formal market, while in reality it is more like an informal market.
- Engendering trust across contexts is not trivial. How do you do so when key decisions are made using AI-enabled technology, rather than humans?

- Technology companies using AI are currently primarily concentrated in one physical location (Silicon Valley), where the business models are being developed by a select group of people, thus excluding other perspectives/prerogatives.

Breakout 1E - Law and Governance

Session Summary:

The breakout began on the topic of AI's use in the judiciary, with participants noting that even as Supreme Court justices readily incorporate AI-based tools into judicial decisionmaking in countries ranging from Kenya to the United States, many remain ignorant of how these systems are designed and developed. This has increased the sense of urgency for stakeholders to explore challenges and opportunities at the intersection of AI and inclusion, and law and governance. Participants arrived at three core conclusions: (1) there is **no standardized legal definition** for AI, (2) there is **no attractive methodology** for addressing AI issues, and (3) there is **no clear understanding of how AI technologies function** among most lawyers. (They also observed that they had articulated their findings in the negative, and that this itself would send a message.) The group proceeded to discuss possible legal safeguards, including mandatory compliance with UDHR and open-sourced AI for public interest systems. It ultimately established that the **right of transparency/explanation may not be enough** to exercise self-determination and collective agency. A few individuals highlighted data ownership and access questions, and suggested we pursue the latter for more inclusive AI. Among the most important takeaways from this session was the need to consider **different types of solutions** for different clusters of issues; specifically, should we address them on the basis of **substance** (transparency, bias, access,) or **technology** (different algorithms that use distinct types of data)?

Conceptual Insights:

- Judiciaries increasingly using AI tools but rarely understand how they are used and how to use them. (U.S. and Kenya)
- Lack of legal definition presents problems when such definitions enable us to interpret and administer justice
- Right of transparency/explanation are not enough to be able to able to self-determinate/collectively exercise agency
- Address problems via substantive baskets (transparency, access) or technological clusters ie. algorithms that use specific types of data?

Research Questions:

- Is there a need for a (legal) definition of AI?
- How do standards need to be negotiated differently?
- How do we ensure ability and agency – even when it isn't "personal" data? - Bastard Data
- Who are appropriate independent algorithm auditors? What are the guidelines? Do we need new intergovernmental orgs?
- How do we audit - input, output, examination of the algorithm itself?
- What are appropriate remuneration mechanisms?
- Do we want the "recipe" or do we want to know the "health effects?"
- How do we open conversation to incorporate lawyers that don't work on "online" law?
- What (multi-level) AI differentiations impact rights and how can we address that from a legal POV?

- If algorithms challenge the integrity of legal infrastructure that government protects, who addresses it? Who do you talk to?

Action Items:

- Ensure every AI is compelled to follow UDHR
- Open source requirement for public interest facing AI
- Policy guidelines to access versus ownership models that don't work
- New legislation
- Copyleft for AI

Examples/Case Studies:

- Platform Cooperativism is an example of disintermediated business and governance model. (Melanie Dulong de Rosnay, Institute of Communication Sciences)
- The Creative Commons International Community does not agree on everything yet still achieved meaningful discussion/consensus. (Melanie Dulong de Rosnay, Institute of Communication Sciences)
- Google in Canada linked a real name that should have been anonymized to a corporate document outing the victim of a sex crime. Who is accountable? This is an example of how it can be problematic when states need to protect citizens but encounter an "algorithmic nuisance." (Marketa Geislerova, Global Affairs Canada)

Key Perspectives - Global South & Underserved Voices:

- Algorithm do not take into account cultural differences? (Could they?)
    - ie. travel to India – honking as form of politeness

Day 2: Breakouts #1 Report Back

Thursday, November 9, 2017 | 12:15pm - 1:00pm

Session Summary:
In the report back from the first set of breakouts, several salient trends emerged including: (1) **definitional challenges**, specifically with regard to what we mean by business models, liability, fairness, and AI itself; (2) the **relevance of lessons learned** from past related technology policy challenges, including internet governance, corporate accountability and sustainability, and broadband access; (3) **neutrality** as a difficult and likely doomed aspiration; and (4) AI's reinforcement and enactment of **norms and values**. Participants suggested **learning from local and global social movements** to foster long term solutions and momentum, and one flagged an **interdisciplinary "reading list"** to get diverse collaborators up to speed on each other's vocabularies and contributions.

Conceptual Insights:
- Definitional piece - what are we really talking about
- **Bringing in time scale to bridge divide**
- **AI Alternatives**

Research Questions:
- **How do we define our values in the AI context?**
- Are there good proxies to look at fairness? Are explainability/transparency even what we want?
- If all are socially/politically constructed and technology has agency, do we look for something neutral? Do we aspire for neutrality?
- What is an AI-driven business model? What can we learn from corporate accountability and sustainability?
- How do we incorporate social justice into this discourse?
- What is the role of the Network of Centers in promoting ethical AI?
- **What is an ethical business model for "big tech"?**
- **What can we learn from and movement building, especially re. long term growth and momentum**?

Action Items:
- Legal (binding?) definition for AI
- Build upon corporate accountability and sustainability frameworks, expanding them for AI
- Advocacy campaigns to engage the public in algorithm design
- Identify and formalize proxies for fairness
- Storytelling for co-design - talking to the populations that would benefit or suffer most from these technologies in order to understand and reframe user expectations

- Comprehensive strategy for development for collective inclusion ecosystem and collaboration
- Engaging NoC network on AI and Inclusion
- Interdisciplinary list of resources to bridge knowledge gaps

Examples/Case Studies:
N/A

Key Perspectives - Global South & Underserved Voices:

- Cannot address many data and infrastructure issues until we address access, poorly distributed
- Often a question of data-based exclusion vs. data-based colonialism; surveillance capitalism vs. excluding poor
- Cannot separate human rights and development goals - need to address AI with everything else, complement human right instruments with development agenda

- Need to talk to the populations that would benefit or suffer most from these technologies in order to understand user expectations
- What can we learn from and **movement building** in general? esp. in keeping eye on long haul -- black lives matter, women's movement, etc.

Day 2: Breakouts #2 - Application and Impact Areas

Thursday, November 9, 2017 | 2:00pm - 3:15pm

Breakout 2A - Shifting of Industries and Workplaces

Session Summary:
In this breakout, participants established that it is difficult to predict the future of work; we are at a starting point since much of the evidence is indirect and assumptive in a mature, industrialized world. The group agreed that progress can be made moving forward if we carry out the following three tasks: (1) make **distinctions about context** (state, region, culture, etc.) when discussing work and labor; (2) move past the narratives of fear surrounding AI and look at how new emerging technologies can be part of the **solution for job protection**; and (3) examine and identify the role of educators, employers, and, most importantly, policy makers (on account of their formalized responsibility to their citizens) to ensure that they uphold their duties, which will be critical to securing jobs in the context of AI.

Conceptual Insights:

- It is difficult to predict the future of work -- we are currently at a starting point, since much of our evidence is indirect and assumptive in the mature, industrialized world. And we have even less information about impact / resulting consequences (potential and actualized) on informal economies that penetrate developing countries, as such case studies / examples are incidental in nature.

Research Questions:

- What does it mean to have a career trajectory in this world of emerging AI? How does this change the notion of pride in one's professional identity?
- How do we get policy makers focused on the AI challenges and opportunities surrounding jobs?
- How can we make AI help ensure job security, ie. make it part of the solution and move away from its destructive capabilities?

Action Items:

- Possible coalition opportunity for Alison Gillwald (Research ICT Africa - looking at Rwanda and Tanzania right now) and German woman (name?) for examining vulnerable / remote communities with little data accumulation

Examples/Case Studies:

- Uber is a case study for a technology yielding different work dynamics/labor markets across contexts: In Egypt it empowers workers; in the in US it is primarily used for temporary work / an in between job; vs. in Brazil it has revealed corruption of the

formalized taxi industry -- giving more power, agency, and money to people through Uber driving. Though the application is the same in each locale, the outcomes are not, demonstrating the context matters and should be considered when examining the shifting of industries and workplaces during the rise of AI. (Nagla Rizk, A2K4D, US and Fabro Steibel, ITS Rio).

- World Bank study highlights how new technologies have the potential to mitigate job misfortune and measure country impact (Silja Baller, World Economic Forum (WEF)).
- Twitter has an increasing contingent workforce, who could be more easily undermined and vulnerable without a full time contract (Susan Ariel Aaronson, Elliott School of International Affairs, George Washington University).
- Research ICT Africa is looking into Rwanda and Tanzania as case studies of structural inequality globally and how we can address it with AI-based technologies or in the context of rising AI-basd technologies (Alison Gillwald, Research ICT Africa).

Key Perspectives - Global South & Underserved Voices:

- Center focus on state / regional context -- AI will have different effects on areas with variant communities, environments, industrialization, etc.
- Cannot discuss work without getting specific. (Otherwise, conversation gets too general and becomes irrelevant and ineffective when determining solutions for underserved voices.)

Breakout 2B - Health and Wellbeing

Session Summary:

During this breakout, the intersection of health and AI was framed as both the **potential of AI to enhance and/or pose a risk to health care** (e.g. EHR, genomics, improved health infrastructure), and **the influence that the ubiquitousness of AI will have directly on the health and wellbeing of society**. The group highlighted concrete case studies demonstrating how AI might influence health, staying away from the help/harm binary and focusing rather on existing uncertainties. By **first identifying uncertainties in how AI might be utilized to aid global health**, the group argued, we will be able to best address those gaps now, which is imperative since complications in AI health tools can prove fatal. For example, health devices such as pacemakers have the potential to use AI to adapt to your body over time without human interference, and we must anticipate possible cybersecurity risks preemptively to ensure this tool constitutes a health benefit rather than a peril. Four main areas of AI and health discussed that warrant further examination include: the distinction between prevention and on-site response; computerized versus human response (identifying what parts of care warrant full/partial human oversight); the role of pharmaceuticals and AI in aiding vulnerable communities (e.g. automated differential pricing); and patient protection with respect to privacy, autonomy, and knowledge of rights.

Conceptual Insights:

- There is really nothing in AI that can replace that human component that ensures the people on the other end feel safe and cared for -- particularly when it comes to mental health care
- There could be potential gaps in the “perfect” predictability of AI where there is a lack of human context
- How will integrating humans and AI influence human behavior and mental health -- Will we want to be in the care of technology and other AI? How will we as humans relate to one another if we are constantly interacting with AI?
- AI care-giving efficiency will be significantly greater than human capacities, which can yield greater care. Eg. one doctor can see maybe a few hundred or 1,000 patients per year, whereas an AI’s capacity is nearly limitless

Research Questions:

- To what extent can AI replace or enhance human behavior? How do we create standardized boundaries delineating where AI should replace versus enhance? Is this based on the severity of the health condition, the risk involved, etc.?
    - e.g. is the control of AI in preventing suicide more human-dependent because of the risk and sensitivity involved than, for instance, weight control or migraine tracking?
- Does the “digital native” generation suffer from the same rates and forms of mental health conditions as millennials and baby boomers? How has technology shaped mental health over the past three decades?

- In which cases might potential harms of AI outweigh potential benefits? What are the most significant potential opportunities and risks at the intersection of AI and health?
- How can the “perfect” predictability of algorithms result in greater risk or potential harm to population health where there is not human context?
- What are ways that we can use algorithms, particularly in search engines, to disseminate legitimate helpful health information that promotes positive health behavior?
- How can AI increase (or decrease) quality medicine accessibility by implementing differential pricing at an individual level?
- What might hacks and security breaches mean for the health sector and what are cases we can anticipate and work to mitigate?
- How can we equitably regulate cross-border health data flows so that they promote impactful outcomes without sacrificing privacy?

Action Items:

- Formally identify potential risks of intersecting health and AI (eg. mis-diagnoses, loss of human element in care). Identifying these uncertainties will allow us to work towards mitigating them before moving forward full speed in developing AI for health applications.
- Identify components of doctors’ practice that, if replaced by AI, would not benefit patients in order to create a standard for where AI can be applied in healthcare and where it cannot.
- Research the difference in how AI might influence physical health versus mental/overall well being.

Examples/Case Studies:

- Toda Mejora is a foundation that addresses LGBTQ teen suicide with trained therapists available to teens online 24-hours. The foundation is considering using chatbot therapists to reach more youth. Could this program be enhanced by chatbots while acknowledging the high risk and that they could not serve as a complete human replacement?
- AI could significantly aid public health through rapid diagnostics and prognoses in ambulances that could determine your condition, the urgency, the specialist you need, and what hospital you should go to
- If an AI was trying to determine why young girls miss school is rural Nepal, it likely based solely on reporting data would determine they had higher health issues; in reality, however, it would be because they were menstruating which is taboo -- this is an issue an that AI could not understand solely based off of reporting data.
- What happens when individuals can maliciously hack or control machines that are sustaining our health or even keeping us alive, such as a smart pacemaker monitor? How can we anticipate these potential cybersecurity risks and best mitigate them?
- There is a possibility of encouraging differential pricing of pharmaceutical drugs implemented case-by-case based on algorithms: when one is diagnosed, they are prescribed a drug based on factors such as their income-level and pre-existing health information, which would increase medicine accessibility overall.

- In Chile in years past, when a person google searched “how to kill myself,” it came up with sites that gave instructions on different ways of killing yourself. After Toda Mejora worked with Google, Google changed the algorithm to first bring up Toda Mejora’s page which offers help and counseling when a person Google searched that phrase. Programming algorithms in this way can be a method of utilizing AI to promote access to sound and helpful health information that promote positive health behavior.

Key Perspectives - Global South & Underserved Voices:

- Cybersecurity of AI and health needs to be considered
- AI has the power to disseminate helpful/legitimate health information to those who need it, and this can be context-specific. We need to identify what the largest challenges are affecting those in specific countries and adjust the algorithms accordingly

- How will health AI’s impact the work/work ethic and knowledge set of human caregivers?
    - Needs to be considered particularly when considering that there are already fewer trained caregivers in low-income countries/regions, and AI entirely replacing caregiving can pose a risk there where it wouldn’t in more economically developed countries

Breakout 2C - Education and Wellbeing

Session Summary:

There are many issues within education and learning that continue to pose difficulties as AI is applied in this field, such as challenges with **privilege, perspective, colonialism, and language**. According to participants, the primary dimension along which AI changes the paradigm is the shift in power and/or resources from exclusively human-powered institutions and capital to AI-based technologies. An insightful debate during this breakout session centered on the **fundamentals of teaching and learning separate from the application of AI**: the mechanisms by which people learn, the actors involved and their relationships, and education's normative societal purpose. In the aftermath of the industrial age, participants wondered what the role of education should be and where students and teachers alike belong. Some posited that learning occurs in relationship and questioned what effect an AI-enabled non-human teacher could have on the quality of learning, and to what extent a human teacher could be replaced with an autonomous system. **Many agreed that societal resources dedicated to teaching and learning will likely shift with the advent of AI, but the degree to which this will happen remained unclear.** Will resources currently assigned to human teachers be removed? How does privilege factor into these decisions, and what are downstream effects?

Conceptual Insights:

- There are many issues within education/learning that continue to pose difficulties as AI is applied to this field, such as challenges with privilege, perspective, colonialism, language, and so forth.
- **The primary dimension along which AI changes the paradigm is the shift in power and/or resources from exclusively human-powered institutions to AI-based technologies.**
- If the student-teacher relationship is important, what changes when the teacher moves from human to AI? What effect does it have in terms of the rate of learning, the absorption of information, learning of new skills, development of passion/interest, etc.?
- Societies must make a decision around if, how, when, and to what extent they want to incorporate AI-based technologies into the teaching process. Particularly important is the decision on allocation of resources, and the relative roles that human teachers and AI technologies have relative to the students, and each other.

Research Questions:

- How can we personalize technological learning resources, such as MOOCs?
- How do we think critically and make decisions around data acquisition for educational data training sets?
- Are education and learning based on personal relationships? If so, what role do AI-enabled teachers play?
- How does privilege manifest itself in the means of education chosen by different groups/stakeholders? What is the prevailing narrative about AI and technology in the classroom, and who shapes it?

- Can AI help to address the challenges of scale that educational systems often face?
- How does AI-enabled teaching affect resource allocation within the educational system? (i.e. is it supplemental or does it take resources away from teachers?)
- How does the theory of mistakes as a learning process integrate machine learning? How visible are the "mistakes" that AI technologies make in the process of learning, and what is the effect on students?
- What is the role of education in a world where remembering facts is less important and AI-enabled technologies perform many functions that humans traditionally performed?
- How do you regulate how inclusion and bias are expressed in AI educational systems/modalities?

Action Items:

- N/A

Examples/Case Studies:

- Teachers unions in Liberia are against technological education solutions as they see them as a fill-in for education rather than building up the human capital and education systems. Will these AI-based educational tools be siloed to the underprivileged while the privileged maintain more robust human-centered educational structures?

Key Perspectives - Global South & Underserved Voices:

- Perspective is critical when it comes to choosing course material: what facts do you learn; how is the story told? It is a delicate and difficult process to recognize one's own perspective and biases, without having direct experience of any other.

- The inclusion of a diversity of teaching/learning modalities is important when considering this overall question (e.g. apprenticeships). The predominant modality is learning at the moment colors our view of traditional education's normative role and impact.

Breakout 2D - Social Inclusion

Session Summary:

The majority of this session focused on **defining social inclusion** and the key issues it encompasses. Contributors agreed that participation in the discourse around inclusion is a privilege in itself; in many cases, academics and NGOs shaping the conversation may not be held directly accountable to the underrepresented populations of which they speak. Issues of **identity and privacy** emerged as the key dimensions of social inclusion, with privacy framed as a human right and an ability to exercise agency over one's personal data. Participants discussed whether identity's **non-binary and contextual nature could be represented in static code**, and law and history were introduced as possible areas in which similar problems arising from static frameworks and continually evolving subjects had been dealt with. Participants also looked toward the changing technical landscape of modular construction (using smaller, independent, and often third-party apps to build larger apps and services,) as a complicating factor for privacy-related actors in the networked ecosystem. Finally, the suggestion of radically expanding our notion of **social inclusion to include autonomous systems** (as opposed to just humans) drew attention to possible limitations in our current thinking about social inclusion as an inherently human domain, and the potential of a more fluid understanding of man and machine.

Conceptual Insights:

- Defining social exclusion in the digital age as: lacking the information technology and capacity needed for full participation in society
- Inclusion (and the ability to think about inclusion) is a privilege that involves capacity, consent, infrastructure, even the way you're born
- Identity is contextual, based on who it is being presented to; software can't inherently account for this, as it assumes that things are static

Research Questions:

- What are the ramifications of reproducing binary systems in code?
- How do we represent 'code switching' through code?
- What are we trying to 'include' in social inclusion?
- People can become data subjects by virtue of where they are born. How do we configure this in the context of social inclusion (access to digital networks) / exclusion (limited access to privacy and/or agency)?
- Should social inclusion extend to AI systems?
- What impact does the gendering of autonomous beings have? (ex) assistants mainly given a 'female' identity
- What are similar problems -- static frameworks dealing with a continually changing subjects -- in legislating as well as in history, that may shape how we navigate AI and Inclusion challenges?

Action Items:

- N/A

Examples/Case Studies:

- Databases produced by individuals on social media are rife with racism, sexism, homophobia, etc. Examples include the Tay chatbot, which used Twitter data to train Microsoft's AI chatbot which projected bigoted tweets in less than 24 hours, as well as facial recognition tools on Facebook, which crowdsourced tagging data that yielded racial biases.
- Current gender classifiers in databases and the algorithms they feed are mostly binary systems, which doesn't fit a more socially inclusive understanding of gender identity. What are the ramifications of reproducing binary systems in code?
- Free Basics, an application created by Facebook and other social network companies that hand-picks internet services to be used for free is an issue of people becoming data subjects without any consent -- serving as an example of how individuals and populations can become data subjects by virtue of where they are born. How do we configure this dynamic in the context of social inclusion (access to digital networks) / exclusion (limited access to privacy and/or agency)?
- Sophia is a robot that got citizenship in Saudi Arabia -- the first robot to get citizenship in the world. Should social inclusion be extended to AI systems themselves?

Key Perspectives - Global South & Underserved Voices:

- N/A

Breakout 2E - Humanitarian Crisis Prevention and Mitigation

Session Summary:

AI technologies have the potential to significantly **improve mechanisms for preventing humanitarian crises and decrease intervention response times**. While the discourse regarding AI's ability to mitigate the severity or altogether avert humanitarian crises is in its early stages, the potential impact is significant given the billions of individuals worldwide regularly affected by natural disasters, forced migration, and conflict. Examples of where AI could be utilized as a tool for preventing and alleviating global crises include (1) identifying trends of crop plight to prepare for and avoid famines; (2) tracking satellite images in real-time after a disaster to quantify the impacts and identify the most urgently needed interventions; (3) using chatbots to communicate with people affected, and (4) recommending peacekeeping tools post-conflict based on contextualized data. **Development of these tools for humanitarian applications requires careful design and oversight** due to the potential for abuse and human rights violations during and after crises, which could yield damaging consequences. Yet particularly as climate change continues to worsen, and as the aftermath of the 2016 refugee crisis continues to transform the geopolitical landscape, **AI's application in this sphere presents unprecedented opportunities not only to promote inclusivity but also to directly save the lives of the most vulnerable populations worldwide.**

Conceptual Insights:

- TBA

Research Questions:

- What preventative systems utilizing AI tools could be put in place by states?
- Humanitarian crises encompasses a wide range of issues — how do we identify which issues to focus on in an inclusive manner?
- What potential pitfalls should we be mindful of as AI is deployed in times of humanitarian crises?
- What methods could be implemented to identify datasets that could train models to predict onsets of violence?
- Do states have a role in promoting the use of AI to prevent and/or mitigate humanitarian crises?
- What are the humanitarian risks of algorithmic discrimination in conflict settings?

Action Items:

- N/A

Examples/Case Studies:

- Machine learning and satellite imaging designed to predicting crop yields and/or plight in an effort to control for food and prevent famines.
- Unmanned aerial vehicles used in conflicts/emergencies to identify most urgent areas in need of intervention.

- AI-based technologies tracking natural disaster patterns, such as earthquakes, to predict and best prevent future disasters.
- The Harvard Humanitarian Initiative using AI tools to analyze satellite images to differentiate between the burning of towns by way of arson and wildfires.

Key Perspectives - Global South & Underserved Voices:

- N/A

Day 2: Possible Approaches and Solutions - Ideas, Case Studies, and Prototypes

Thursday, November 9, 2017 | 4:30pm - 5:45pm

Session Summary:
The stories shared during this session made clear the urgent need for case studies in the domain of AI and inclusion, as **the application and perception of AI-based technologies are highly contextual**. How a given culture or subculture conceptualizes AI has an enormous effect on inclusion: in Japan, for instance, robots are often viewed as companions and even partners, which presents questions regarding whether they should be formally included within society and whether they should be afforded related rights and/or citizenship. In other contexts, the primary association or impact of AI-based technologies may be algorithmic discrimination, such as in Brazil, where stock image databases systematically discriminate against Afro-Brazilians or anyone possessing a darker skin tone. **Targeted interventions need to be developed alongside broader strategies that address cross-cutting issues to equalize the AI playing field.** Some examples discussed during this session include mitigating the cost of accessing training data sets by establishing a "data commons" open to all, and borrowing from advocacy movements, such as global and local environmental campaigns. AI-based technologies, created by humans, mirror humans' abiding complexities; identifying salient, specific case studies is essential to creating a common understanding of the ways that negative externalities of AI applications can manifest, and the ways that they can be addressed to promote AI for the social good.

Conceptual Insights:

- The way that different cultures (generally) perceive AI/robots can be vastly different. In Japan, the notion of an AI-enabled robot as a friend, companion, and partner is salient, while in other countries such as the U.S., robots elicit images of Terminator.
- What is needed is building a movement of citizens who are in various different ways challenging the power structures, empower people to understand how these systems are working, hold them all accountable + transparent. Combination of long-term/evidence-based research and short-term campaigns on particular companies and issues. Testing legal limits and frameworks, finding where there are gaps that require new policies to resolve.

Research Questions:

- What motivates us to create AI-based systems/algorithms/robots? How does this influence the types of AI that emerge?
- Are there large distributed power systems that could be created to enable technological work and remove the cost barriers?
- How do we identify the salient issues related to AI in a given context? (i.e. AI as partner in some geographic contexts).

- How do we make case studies more visible so that relevant AI and inclusion studies take place?
- Should robots have rights? What should their 'inclusion' into human society look like?
- What do context-specific robot/AI ecosystems look like, and how do/should they interact?

Action Items:

- Social movements to create a healthier digital environment [Mark Surman, Mozilla]:
    - (1) Citizen action to create a training data set, **data commons**, for machine learning available to all.
    - (2) **Push for ethical products** that citizens can trust (verifiability, identifying bias in an algorithm, knowing how data is used.
    - (3) Customers **embracing market power**, being vocal and demanding.
    - (4) **Mind bombs** that shift the narratives, written and told by people who can imagine different futures; embracing the **value of art as a political strategy**, a citizen movement.
- Create multistakeholder avenues for collaboration to break down siloes and develop ethical guidelines for R&D and usage [Arisa Ema].
- Continue to discover and pursue concrete and relatable ways in which algorithmic bias is manifested (a la the 'family' / 'black family' Google search) [Lucas Santana].

Examples/Case Studies:

- In an effort to create and promote data commons as a means for encouraging co-design that rival monopolies, Mozilla developed "Common Voice" which was a coop training data set for voice recognition. The benefits were that the platform was public commons and that in this open source approach we can can start to have languages in AI-based tools that wouldn't be represented by larger corporations (Mark Surman, Mozilla Foundation).
- Mozilla co-wrote a paper on "Building a Trustmark for IoT" modeled off of the movement for organic and ethically-sourced foods to encourage the labeling of ethical, diverse, and transparent technologies. In building on this idea we can either encourage companies to label their products, and/or can eventually regulate and enforce the practice (Mark Surman, Mozilla Foundation).
- Mozilla created a holiday shopping guide to IoT in order to create an accessible document on items that are surveilling you. This list is part of an effort to build a lexicon of ethical items to help individuals know what they're buying (Mark Surman, Mozilla Foundation).
- Mozilla designed and developed a film called "Our Friends Electric" with Superflux as a means of creating fiction where to prototype a different, preferred reality. This prototyping is an approach to expand the AI narrative beyond the confines of how it's currently being defined in Silicon Valley (Mark Surman, Mozilla Foundation).
- The Journal of the Japanese society of AI changed its cover from a technical drawing to a gendered/attractive cover of a female robot cleaning the room, plugged into the wall with hollow eyes. The picture was chosen from within the community and reflects a gendered issue within AI that was picked up by the BBC and other global news outlets that

portrayed the image as an “AI servant girl” playing into sexism. This incident sparked a broader conversation and led to the convening of a community now being established to unpack the ethics and social responsibility surrounding AI (Arisa Ema, Unversity of Tokyo).

- The Japanese government has created a set of AI R&D principles as well as AI usage guidelines (Arisa Ema, Unversity of Tokyo).
- The Journal of the Japanese society of AI held an AI & Society Symposium hosting mostly colleagues from the West to discuss how we can collectively build beneficial AI and treat AI more like partners (Arisa Ema, Unversity of Tokyo).
- Shutterstock image search engine algorithms are skewed towards highlighting white people and families as the standard. For example, when one searches for ‘black family’ images of black families come up vs. when you just search ‘family’ only white families come up (Lucas Santana, Desabafo Social). Desabafo Social created videos demonstrating this discrimination which generated cross-sectional dialogue that spurred some companies to change their algorithms and/or communication strategies. Desabafo argues that algorithms need to be built with diverse data by diverse people (Arisa Ema, Unversity of Tokyo).
- According to Harvard Business School, Black people have 16% greater Airbnb cancellations than other users (Lucas Santana, Desabafo Social).
- On the University of Illinois Champlain campus the supercomputer is so expensive that students are only able to use 8% of the supercomputer’s power. The sheer cost of technologies such as this makes it difficult for particularly small organizations in the global south to utilize.
- It was formerly the case that when you Google searched ‘female scientists’ the algorithm would asked ‘did you mean male scientists’?

Key Perspectives - Global South & Underserved Voices:

- We must consider how we make the few case studies more visible so that the AI and inclusion studies can happen in the Global South.
- Nothing is going to change if the black boxes are made by white men in Silicon Valley.

- We cannot look at binary or single race/gender categories when it comes to algorithmic discrimination. We need to look at the intersection of other cross-cutting factors such as disability, etc. Otherwise even an audit would not have an equalizing impact.

Day 3: Intervention Points and Opportunities for Collaboration

Friday, November 10, 2017 | 10:00am - 11:00am

Session Summary:

Juan Carlos De Martin (Nexa Center for Internet & Society) opened up the final day of the Symposium by shifting the focus from the identification of opportunities and challenges to their translation into explicit policy plans and actions. He invited participants to consider (1) different **temporalities** for addressing different types of AI challenges; (2) different **ways of clustering AI solutions** (by concept, technology, and stage of development); and (3) different ways of fostering sincere and sustained **interdisciplinarity**. Alison Gillwald (Research ICT Africa) underscored the need for genuine transdisciplinary research and proper critical engagement with ICT disciplines, SDGs, and multilateral agencies to create an informed evidence base for policy. She concluded with a recommendation of **robust regulation** that she felt was not present in the current open internet domain. Chinmayi Arun (Center for Communication Governance), however, felt that contemporary internet governance served as a helpful regulation model, yet emphasized that we would need to **expand what is meant by investing in the Global South**, incorporating not only money but people and points of view for the long term. Maria Paz Canales (Derechos Digitales) agreed with the call for multidisciplinary work bridging academics and activists, but noted that many civil society organizations lack in-house researchers and that such efforts would require innovative sustainable collaborative models. Vera Franz (Open Society Foundations) stressed the need to **extract concrete policy recommendations from case studies and prior technology policy work**, drawing from the A2K movement and antitrust law, among other examples. Victor Akinwande (IBM Research Africa) drew attention to the role of **ethics education in engineering curricula** as a concrete next step to building a shared understanding for interdisciplinary solutions. The session concluded with proposals from the audience, one of which involved a **platform or strategic map of countries, conferences, and organizations** working in the AI and inclusion space.

Note - *blue content* below drawn directly from Day 2 summary slides*

Conceptual Insights:

- **Different Types of Solutions to Different Types of Problems**
  - Clustering by concept - transparency, access, bias, etc.
  - Clustering by technology - specific types of data and algorithms
  - Clustering by stage in process - input/output (beyond outcomes)

- **Time Scales for Solutions**
  - Can't "act now" or "wait for time to conduct research" → both simultaneously
  - Different temporalities/goals for different problems/solutions
  - Social movement approach - building momentum and sustaining it long term
    - Global/local level - ie. Greenpeace vs. Brazil's environmental movement

- **True Interdisciplinarity**
    - More than interdisciplinary perspectives ➔ interdisciplinary education
        - ie. reading lists
    - Also in the sense of importance of inter-stakeholder and inter- professional dialogue

- **Expanding what we mean by "investment in the Global South"**
    - Not just money but in POVs + people
    - Long term solutions, short term solutions

- **Ties to Internet Governance?**
    - Disagreement in room ➔ Gillwald vs. Arun

Research Questions:

- Is there a need for a (legal) **definition** of AI?
    - Local, national, universal? Binding? Who decides?
- How do we define our **values** in the AI context?
    - Technologies embed norms and values, not just informed by them
    - Rather than de-bias make inevitable biases explicit?
- What is an **ethical business model** for "big tech?"
    - What are appropriate remedy mechanisms?
- How do we **audit** algorithms and who are appropriate auditors?
    - Do we want the "recipe" or do we want to know the "health effects?" Or both?

- How do we define "AI success" (instead of just AI)?
- How do we sustain the inclusive conversation we are having in Rio over time?
- What is new and what is not new in terms of AI opportunities and challenges?
- Is the "fourth technological revolution" merely a manifestation of late phase capitalism when there is nothing to suggest fundamental shift within power relations or technology?

Action Items:

- Interdisciplinary Coalitions (ie. Partnership on AI)
- Data Access Agreements/Policy Guidelines (vs. ownership)
- Legal Definition of AI
- Context-Specific User Expectations/Terms
- Network of Centers Commitment
- "Copyleft" AI Equivalents
- Interdisciplinary "Reading List"
- Accountability Audits
- UDHR Requirement
- Open Source Requirement

- Data Commons

- **Map of countries, conferences, regulation, and organizations in the AI & Inclusion space**
- Comprehensive human rights assessment on AI decision-making, design, & implementation
- **Extract from explicit case studies concrete policy recommendations**

Examples/Case Studies:

- Patent pools could be a potential step forward for the access to knowledge and data commons movement by encouraging data sharing and getting a consortium of players to agree to proselytize patents and data (Mark Surman, Mozilla Foundation).
- Countries such as Costa Rica are already accepting technologies and not considering the impact, representing the need for greater discussions about impact/influence on AI and society short term and long term.
- Studies have shown that Black men are made to wait longer for rides and have more cancelled rides in ride sharing apps when the driver can identify the race of the user.
- AI is being implemented to forecast weather and provide disease prognosis in Uganda.

Key Perspectives - Global South & Underserved Voices:

- Desabafo Social (Brazil)
    - “not speaking to AI in its language”
- Background check applications (Iran)
    - Balancing transparency and sensitive information
- AI in weather forecasting and disease prognosis (Uganda)
    - Implications for human and data security
- “4th revolution” less relevant in developed economies because no balance; difference in underdeveloped economies because unindustrialized (won’t lose those jobs)
- Investment in Global South is not just money but in POVs + people

- Disparities in service based on race/ethnicity (Uber/Airbnb)
    - Black men waiting longer for a car

Day 3: Report Back from Cluster Meeting #3

Friday, November 10, 2017 | 12:30pm - 1:30pm

Session Summary:
This report back session provided a forum for cluster groups to share tangible action items that would harness momentum cultivated at the Symposium to work towards short and long-term AI and Inclusion goals. Broadly, the action items fell into three groups: (1) **network building**, (2) **project and research initiatives**, and (3) **policy-making/governance**. For network building, short-term next steps focused on **strengthening the network advanced in Rio** through creating a Symposium mailing list, developing a repository for AI and inclusion case studies, and appointing regional leaders immersed in the field to stay connected, among others. There was also an expressed interest in **identifying who is not at the table** through an AI area mapping, and expanding our network on a grassroots level by hosting parallel symposiums locally, encouraging decentralized engagement using a shared AI and inclusion hashtag, and initiating formalized partnerships across sectors and disciplines worldwide.

Meanwhile, participants suggested that project and research initiatives could provide **opportunities to work toward middle- and long-term goals** to more deeply understand the potential societal implications of AI and how we can promote greater inclusivity. Proposed points of collaboration included creating a living lexicon of AI terms across contexts, researching the open data movement to inform private and public sector data use, and designing a technical methodology to quickly assess the ethical standing of an algorithm as it is created. Lastly, participants felt that **policy making/governance would demand both short-term and long-term engagement**, ranging from bringing evidence-based research to UN agencies in-person to sow the seeds for future policy initiatives, to identifying a **standardized "ethical oath"** for engineers to guide inclusive AI development.

Conceptual Insights:

- Creating strong communities that are intentional about inclusivity and bridging wide arrays of backgrounds is vital to the success of this movement

Research Questions:

- Is data protection law hampering or helping equality? More broadly, how can we use law to improve equality?
- What sensitive open data does the government pull for algorithms and how can it be better monitored to ensure the data it's using is as inclusive as possible?
- What are the existing platforms and collective intelligence that we can pool to identify communities of practice in place?
- How do different communities define "Artificial Intelligence" and how does it influence varying perceptions including fear and excitement surrounding the future of AI?

- What are successful examples of the academic technology community bridging industry and engineers into conversations and solution creation?
- What mechanism could improve transparency in the design of AI?
    - How might requiring engineers to sign an “ethical oath” of conduct promote inclusivity, transparency, and in turn equity, in algorithms?
- What does the current global framework of AI production/use look like? Who are the major players, what are their initiatives, and how can we lift up more grassroots initiatives?

Action Items:

- Conduct in-depth research on data protection law as it pertains to AI; in particular, is AI hampering or helping equality in the eyes of the law, and how can we utilize data protection law to promote more inclusive algorithms and their impacts?
- Conduct in-depth research on the open data movement; the government pulls a significant amount of sensitive data for building AI, and we need to decipher how the government uses this data as it will influence the inclusiveness of the AI framework (eg. governments pulling selective data for predictive policing).
- Each attendee go back to their home/region of work and hold workshops and knowledge exchanges connected thematically to the symposium, potentially using the same name (similar to TED).
    - Outputs from these localized events can feed back into the NoC trajectory and shape the larger, global events to come.
- The organizers of the symposium should help organize a repository of case studies and ongoing work that is open access with links to resources.
- Create a collaborative reading list that could be heavily curated and focused expanding collective knowledge to reach our larger goals.
    - One particular subset of this list should be on law and governance issues.
- Design a protocol for quickly assessing the risks and opportunities of AI and emerging technologies at large.
- Identifying communities of practice and creating a map of the current AI framework with the users/producers of AI -- what are the existing platforms and collective intelligence that we can pool?
- Influence policy-making by going to the relevant UN agencies and bring our evidence on AI in person, so that we are making in-person connections and can start to shape the policy based on evidence.
- Identify legal obstacles and solutions to principles proposed throughout the symposium (eg. how would implementing a requirement for using tools to inspect bias in algorithms be enforced in practice -- through policy, the law? What are the major obstacles for achieving this goal and how can we overcome them?).
- Define which coalition we could revive in order to define our ideal vision of AI in the world
- Create a living lexicon to identify individually and collectively define terms of AI and understand definitions across disciplines.
- Create an organization for creating new datasets -- the more datasets we have, the more work that can be done.

- Design a methodology to identify the measure the ethics of an algorithm as it's created.
    - Including developing tools to test for bias in machine learning algorithms.
- Bridge the industry side into the conversation more effectively by striving for more equal sector balances during conferences and building cross-sector coalitions at local levels (eg. BKC creating an AI & Inclusion coalition with CRCS).
- Write some sort of ethical oath for AI developers.
- Map inclusion -- who's here, who is not, who can we include and how? What were the successes of this conference and where else can we plug in?
- Appoint regional leaders to build these networks locally.
- Choose a hashtag that we all attendees and folks who plug into the movement can use when publishing information.
- Creating a mailing list on which all attendees and folks interested in the themes of the symposium can communicate on.

Examples/Case Studies:

- Canada has ongoing projects investigating the reaches of AI, as well as a current initiative examining the use of sex bots to disseminate misogynistic messages.
- KS (Kyung-Sin) Park (Korea University Law Schol) has an ongoing project on how to abate the polarizing effects of AI, aimed at encouraging more critical thinking about how we can make data available more equally.

Key Perspectives - Global South & Underserved Voices:

- There are great successes from this event, but we need to define who is still not at the table and how we can truly yield inclusivity -- not just for those who are working in AI, but those who will be most impacted

- Creating a map of the current users/producers of AI will make evident the imbalanced power dynamics that are inherently disadvantaging marginalized/underserved populations

Links to Relevant Public Resources

The following materials are relevant inputs and outputs from the Global Symposium on AI & Inclusion that have been shared publicly. For more information about the event as well as a complete list of public resources, please refer to the Symposium website at www.aiandinclusionsymposium.com.

- Pre-Symposium Briefing Materials
    - Evolving Reading List
    - Analysis of Pre-event Survey Responses
- Wonderbag Output Highlights
- Global Symposium on AI & Inclusion Event write-up on Berkman Klein Center's Medium