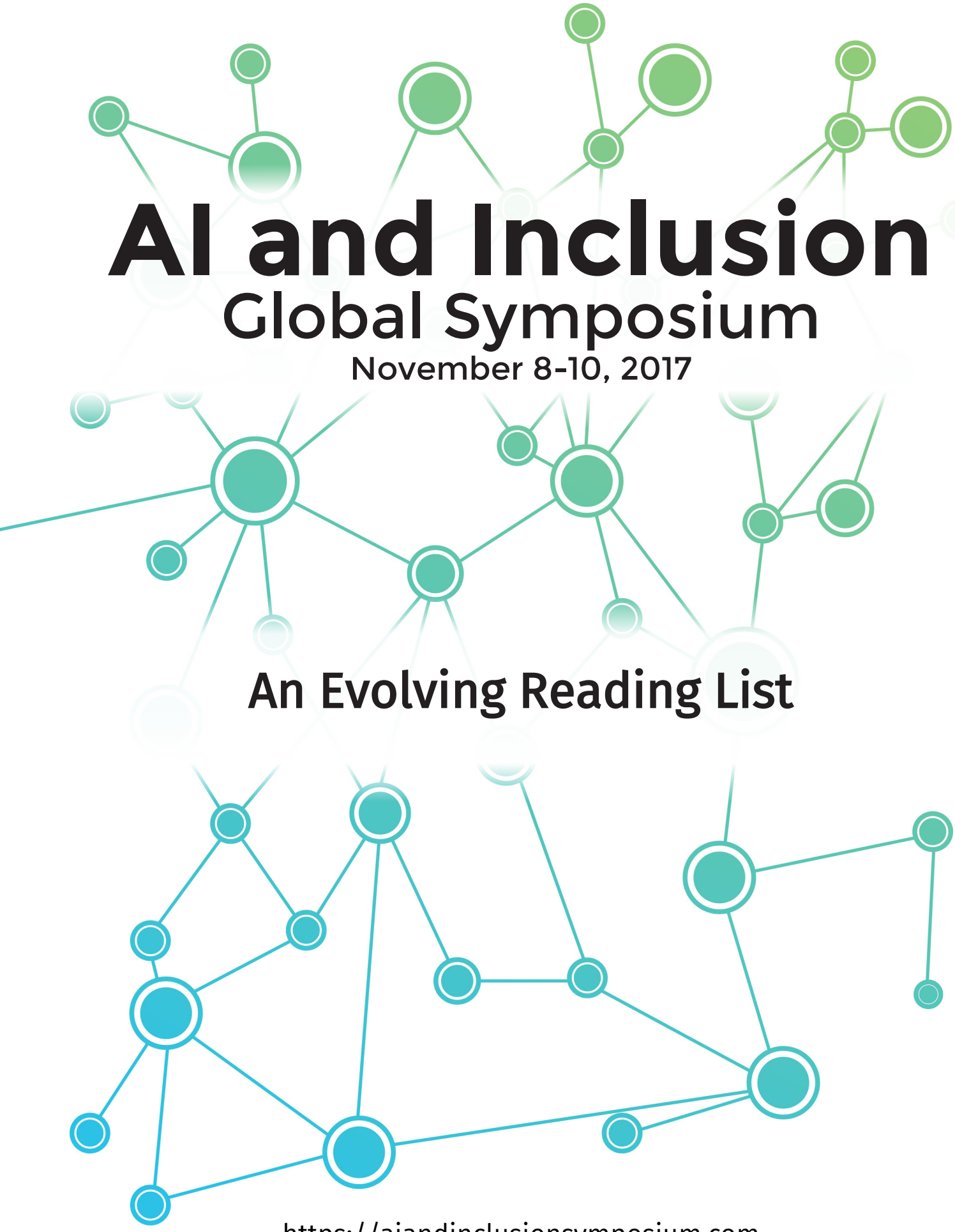
AI and Inclusion

Global Symposium

November 8-10, 2017



An Evolving Reading List

https://aiandinclusionsymposium.com

These reading suggestions are shared in preparation for the Symposium on AI and Inclusion, which will take place on November 8-10, 2017 in Rio.

ARTIFICIAL INTELLIGENCE (AI) AND INCLUSION

An Evolving Reading List

AI-based technologies and applications offer tremendous opportunities to build a better world. Advanced systems can be used to enhance educational instruction, improve the efficiency of transportation infrastructure, fight epidemics, support the elderly, invest in the stock market, and combat corruption, to name just a few examples. However, perhaps more so than ever before, these same AI-based technologies can also deepen existing divides, gaps, and inequalities, and even create new ones if not developed and deployed in thoughtful ways with appropriate safeguards and support mechanisms in place. As such, there is a growing need to critically engage with AI's potential impact on diversity and inclusiveness in our increasingly networked society.

As we look toward building effective dialogues at the Global Symposium on AI & Inclusion, this expanded reading list is meant to serve as a starting point for exploring the ways in which we may think about the very notion of social inclusion/exclusion itself in the age of AI, as well as the complex ways in which autonomous systems interface with various dimensions of inclusion and the questions that arise at this intersection. It is our hope that these resources will support thought-provoking discussions and inform future research efforts on the role of AI as we work towards a more inclusive society.

For more readings about AI, please also visit the AI Compass.

Framing Inclusion in the Context of AI

In order to better understand how AI can be conceptualized, designed, and deployed to support efforts working towards a more diverse, fair, and equitable society, it may be helpful to develop a nuanced understanding of the concept of inclusion, better enabling us to proactively identify key barriers, test new strategies for addressing these barriers, and create a shared knowledge base for action. This raises a key question — how do we begin to conceptualize "inclusion" in the age of AI? Inclusion can be framed by using different approaches and lenses, including sociology, public policy, social psychology, feminist theory, among others. An examination of relevant interdisciplinary research in this thematic context demonstrates that the concept of inclusion and social exclusion can be characterized as follows:

- Dynamic: Social exclusion is a process or a set of processes rather than a fixed condition, and individuals may be affected by many processes simultaneously.
- Intersectional: A given individual may belong to a number of identities that have been subject to exclusion, and may face social oppression that is unique from the sum of its parts.
- Relational: Individuals or groups are included or excluded from other groups, individuals, and society as a whole.
- Subjective: Individuals, groups, and societal institutions make determinations about inclusion/exclusion based on their subjective experiences.
- Contextual: Social inclusion is tied to structural changes due to globalization, and is affected by local and national dimensions.
- Institutional: Legal and cultural institutions create structures that replicate and reinforce historical inequalities.

Key Resources

- Inclusion Matters, World Bank (Chapters 1, 2)
    - A report highlighting the concept of inclusion on a global scale, and outlining the opportunities and challenges of creating a more inclusive society.
- Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color, Kimberle Crenshaw
    - A primer on the concept of intersectionality — that most marginalized individuals belong to more than one marginalized group and how the concept has remained absent from social inclusion discourses.

Initial Questions at the Intersection of AI and Inclusion

When considering the social impact of AI, conceptualizing inclusion may become even more difficult because of the nature of AI itself. AI-based technologies are driven by their ability to recognize patterns and to classify data. The processes of pattern recognition and classification also drive exclusionary social processes — that is, in order to exclude a group from participation, individual members of an excluded group must first be identified by salient characteristics then classified as members of that group. There are questions that we should be asking in order to further shape our views and understandings in terms of how to conceptualize AI.

A few examples:

Classification (inclusion vs. exclusion): Classification and categorization are at the heart of social inclusion/exclusion, which depends on defining a 'criteria' of membership. Simultaneously, AI-based technologies revolutionize the ways in which data and people are systematized, often with unforeseen consequences.

- For instance, Kosinski, Stillwell, and Graepel (2013) developed an algorithm to predict, among other things, a person's gender, sexual orientation, and race based on the content they "like" on Facebook.
- Ye et al. (2017) developed a machine learning-based classification system to identify the gender, ethnicity, and nationality of a person solely on the basis of an individual's name.
- Blodgett and O'Connor (2017) discovered that many natural language processing and sentiment analysis algorithms learn, over time, to assign and mirror societal values about speech patterns — and to disregard those of underrepresented groups.

  Possible discussion questions:

  With pattern matching and learning at the core of AI, how can we ensure safeguards against such technologies identifying, using, or creating categories of identity that may lead to social exclusion? What are some ways that these new classification systems might create new categories of identity (some of which might not even be known by those being classified)? What are the larger implications of AI systems identifying patterns across groups of individuals,

especially given their lack of ability to consciously identify and negate socially learned biases? Additionally, what are the implications of using these systems knowing that there can be error in their judgements as well as a reliance on binaries for categories that may be perceived in a more nuanced way (e.g., gender, where someone is from, wellbeing) as well as for categories that are less homogenous than possibly perceived (e.g., all women, all indigenous communities, etc.)?

Key Resources on Classification

- Why Stanford Researchers Tried to Create a 'Gaydar' Machine, Heather Murphy
- Transgender YouTubers Had Their Videos Grabbed to Train Facial Recognition Software, James Vincent
- An AI That Predicts a Neighborhood's Wealth from Space, Robbie Gonzalez

Ethics: Allowing AI to make decisions, especially high-stakes ones, introduces a new challenge for inclusion only now starting to be addressed. An unfortunate byproduct of using terms such as "artificial intelligence" and "machine learning" is that some assume these systems are intelligent and/or learn in the same ways as humans or that all AI systems are intelligent. Only recently have researchers begun to investigate how machine decision-making processes align with and may influence human views of morality.

- New moral questions arise, such as how much of our ideas about morality should be programmed into autonomous vehicles given the inevitability of moral inconsistencies in such situations.
- Using a web-based game, researchers at MIT are investigating how humans think autonomous vehicles should make decisions, especially when faced with no-win scenarios, such as the infamous trolley problem.

Possible discussion questions:

With the rise of AI, it is helpful to revisit questions such as: What is fair? What is diverse? What is true? What is good? What impact will current deployments of AI, that have not focused on these populations, have on underrepresented

communities? How can we ensure that negative impacts will not be replicated in the development of future AI-based technologies? When AI architects develop new systems, who will have the power to judge whether their applications are ethical? How do we ensure that this power is distributed equitably?

Key Resources on Ethics

- AI Research Is in Desperate Need of an Ethical Watchdog, Sophia Chen
- The Robot Dog Fetches for Whom?, Judith Donath
- Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction, M.C. Elish

Bias and Discrimination: AI is not neutral. For example, AI-based technologies are developed by people who may hold explicit or implicit biases against members of underrepresented groups. In addition to biases held by AI developers (the "White Guy problem"), many AI systems learn to make classifications by training on data sets that reflect sociocultural biases. It is unsurprising that AI-based technologies replicate inequalities when they have been taught using biased data.

- In the United States, courts are already using AI to predict whether a defendant will commit another crime in the future; these predictions are then used in sentencing and bail decisions. Research by Larson, Mattu, Kirchner and Angwin (2016) demonstrates that decisions made by these AI-based technologies are biased — for instance, black defendants are much more likely to be incorrectly identified at risk of re-offending than white defendants. Although there is evidence of bias in this system, the Wisconsin Supreme Court ruled that these algorithms can indeed be used to sentence defendants and by extension, that they cannot be challenged. The United States Supreme Court has refused to hear an appeal of the Wisconsin decision. AI-based applications are also being integrated into Chinese courts.
- Neural networks, an important component of AI systems, are trained with data that are usually collected by humans and contain implicit biases. Lum and Isaac (2016) examined bias in a predictive policing system (PredPol) that was developed to flag areas where crimes may occur. The data fed into the PredPol algorithm were already biased: police arrests for drug crimes was disproportionately located in nonwhite areas, even though drug crimes were

estimated to be distributed throughout the city in question. Lum and Isaac (2016) then show that by training the predictive algorithm on these data, the algorithm inappropriately flags people from underrepresented groups as at risk of committing a crime.

Possible discussion questions:

What would a taxonomy of bias in AI-based technologies look like, for instance, based on the sources of bias? How can each type of bias be addressed? What techniques are already available (e.g., to de-bias training data sets, or to create non-biased algorithms)? What are the limits of such techniques? Who is responsible for what? How can governance systems support or require the design of de-biased AI? Who bears the costs of bias in AI-based technologies?

Key Resources on Bias and Discrimination

- Technology is Biased Too. How Do We Fix It?, Laura Hudson
- Hiring Algorithms Are Not Neutral, Gideon Mann and Cathy O'Neil
- Machines Taught by Photos Learn a Sexist View of Women, Tom Simonite

Transparency: The lack of transparency in AI systems is highlighted in two important areas: First, most AI and other algorithm-based systems used in areas such as the criminal justice system, employment, and health care do not allow for easy inspection. This issue is compounded by the fact that few existing AI systems can explain how they arrived at a given conclusion. Second, there are large knowledge gaps between AI developers, end users, and policy makers.

- As noted, the developers of AI-based systems may have implicit biases of their own that are transferred to the systems they create. Humans are prone to having implicit biases that shape the way they view people from traditionally disadvantaged groups. Without transparency in the algorithms they build, there is no way for policy makers or end users to evaluate the bias in their algorithms.
- MIT researchers have already developed a way to train neural nets to explain how they arrived at a decision. Perhaps full transparency may not be possible or practical given the nature of neural networks; however, we can hold the

algorithms and their users accountable. Ed Felten provides a good example of how an airport screening algorithm can be made accountable using a verifiable process.

Possible discussion questions:

How can knowledge gaps among stakeholders be bridged? Can we design interventions in the realm of education or public policy to addresses these gaps? Assuming education plays a key role in narrowing the gap between developers, end users, and even policy makers, what are some effective ways to inform the different stakeholders? At the same time, what are the technical opportunities and challenges in improving the explainability of machine-made decisions?

Key Resources on Transparency

- Holding AI to Account: Will Algorithms Ever Be Free from Bias If They're Created by Humans?, Matt Burgess
- There Is a Blind Spot in AI Research, Kate Crawford and Ryan Calo
- The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues, Nuffield Council on Bioethics

Knowledge/Development Gaps: In thinking about a more inclusive society on a global scale, it may be helpful to consider the geographies in which AI technologies, as well as the supporting infrastructure in the form of computing resources and data, are developed and deployed. Without this, we not only risk deepening pre-existing digital divides, including within national boundaries, but also fail to progress towards a more inclusive society for all citizens. While AI has tremendous potential to solve some of humanity's biggest problems, these solutions can be situated in the contexts in which they will be deployed in order to take effect.

- Domestic policies that are deployed asymmetrically only foster further divides; resources may be deployed in a manner that considers equal regional and community-level access to communications resources. Current inequalities should be considered when determining the need for specific resources and further investments.
- The lack of awareness and infrastructure to gather data in Global South countries hinders progress towards AI solutions to problems endemic in the

region. It's important to keep in mind that many of the risks that autonomous systems pose are shared by both Global North and Global South countries, but their effects may also be more prominent in the Global South.

- Automated systems in the workplace may cause a major disruption in global and local economies. Some have encouraged policy makers to start to think about social safety nets for workers that will be displaced by automation while others tout the benefits of AI automation for their economies.

  Possible discussion questions:

  What are the main barriers that fundamentally contribute to the digital divide when it comes to AI? Which of these barriers are new and AI-specific, and which are more generic in nature? What are possible approaches (technical, social, policy, etc.) to close divides and gaps, and what are interventions that might work on a global scale? What are effective ways to de-centralize the discourse, development, and deployment of AI systems from the Global North, so that they may be contextualized within the Global South? How can we ensure that designers of AI systems adopt a user and context-centered approach in designing autonomous systems?

Key Resources on Knowledge/Development Gaps

- Artificial Intelligence (AI) and the Evolution of Digital Divides, Andres Lombana Bermudez
- The Global Economy Will Be $16 trillion Bigger by 2030 Thanks to AI, Ross Chainey
- Big Data from the South: The Beginning of a Conversation We Must Have, Stefania Milan and Emiliano Treré

Deep-Dive #1: AI and Youth

Over the past few years, technologies based on AI have started changing our daily lives as they are rolled out at an accelerated pace not only in professional working environments but also at home and in schools. Hello Barbie or Green Dino are just two examples of AI-enabled toys that have already made their way into some children's homes, with many more under development across the globe. AI-powered toys offer playful opportunities for

children, with some systems promoting enhanced social skills, literacy learning, and language development.

Thus far, there is less research on the beneficial impact of AI-based technologies for older children and youth. However, recent reports and studies indicate that AI systems are playing an increasingly important role in learning, whether in formal educational institutions or when engaging with interactive online platforms, advanced games, or the like in informal individual and social learning environments. Within the formal education setting, AI-powered 'ed tech' — such as digital tutors, tailored curriculum plans, and intelligent virtual reality — can enhance educational outcomes, and offer rich and engaging interactive learning experiences for youth. Over time, as the reduction in costs allow more schools to implement these and other AI-based educational technologies, the future of AI in education holds the potential for personalized learning at scale. In informal and connected learning environments, such as MIT Media Lab's Scratch platform and the Minecraft virtual world, youth have the opportunity to design and program AI-based interactive games, simulations, chatbots, and robots with great benefits for creativity, learning, and self-expression.

The promise of AI-based technologies for older youth is also being examined within the context of health and well-being. AI-driven applications are being designed and deployed to address health care concerns for young people — particularly in the context of mental and behavioral health — in the form of diagnostic tools, therapeutic chatbots, and public health interventions. These technologies open up the potential for earlier intervention for vulnerable youth, increased access to and engagement with therapeutic services, and greater awareness around public health issues.

At the same time, the complex interplay between data sets and algorithms that power these 'black box' AI systems lead to pressing questions around bias and discrimination, transparency and accountability, and — particularly when these systems are connected to the Internet — privacy and safety. Further attention may also be devoted to examining AI-based technologies' short- and long-term impacts on youth's social and behavioral development. Given that young people are increasingly growing up with AI systems — from intelligent toys to in-home voice assistants — how do these interactions shift the way youth relate to the animate and inanimate objects around them?

The full brief on AI and youth can be found here.

Working Toward Solutions

When conceptualizing solutions, it may be helpful to keep in mind Joi Ito's premise that AI technologies are one of a wide range of integrated, adaptive systems; they both receive and provide input to humans, and are not free of the cultures in which they have been developed. For this reason, we believe that solutions can and should be multifaceted, with inputs from multiple levels of social systems. While it is tempting to focus on the AI systems themselves or the developers of said systems, it is important to keep in mind that these systems are products of our cultures, which are already imbued with biases. In thinking about some of the potential agents of change at the intersection of AI and inclusion, here are a few starting points:

Code

- Because the algorithms and datasets driving AI systems are built and collected by humans, Baur et al. (2017) indicate that we have an opportunity to develop AI that will help us remove bias in our decisions. Baur et al. (2017) also believe that those companies that develop transparent and explainable AI will have an advantage over those who don't. This change, however, will not occur without public awareness that creates social and market pressures on companies.
- Rob Speer, Luminoso's Head of Science describes how he created a method to de-bias algorithms and offers de-biased word embedding vectors to developers for free so they can do the same. Researchers have provided details about how gender biases can be removed in word embedding. In fact, the Social Media Collective team at Microsoft Research suggests that it's actually easier to remove gender biases from algorithms than from people.

Education

- It is important to diversify the machine learning industry's workforce as most developers are men who are either from the majority group or majority-group signaling. The educational pipeline can be bolstered by encouraging those

from underrepresented groups to become interested in computer science early.
- Education about AI and algorithmic prediction is important for the public at large. We need to strive to ensure that citizens have the necessary knowledge in order to ask the right questions of AI developers.

Deep-Dive #2: Learn more about AI

To empower people to think more deeply, critically, and creatively about AI's current and future impact on their lives, the team at the Berkman Klein Center designed three initial learning experiences that aim to help individuals better understand some of the fundamental concepts and methods of AI. A learning experience is a semi-structured activity that one completes to gain knowledge about a particular topic. This set of activities explores the following questions:

- What Is an Algorithm?
- What Is Artificial Intelligence?
- What Is Machine Learning?

Each learning experience, which is organized around a specific goal, includes a brief overview of the subject matter at hand with supporting multimedia resources, and encourages the learner to apply their understanding of the topic to real-world contexts. Currently, the learning experiences are designed so that individuals can engage in the activity on their own, in any environment with Internet access. To view each of the three learning experiences, please visit this link.

Academia

- The development of AI has far outpaced the research examining how humans interact with these systems and how these systems influence societal institutions. Academics are uniquely positioned to collaborate with AI developers in examining these questions. Universities are in a position to help in a number of ways, including conducting research on the impact of AI systems.

- Legal scholars need to be prepared to address AI's novel legal and regulatory issues triggered, which may be specific to a given AI system or cut across the entire field of artificial intelligence.

Activism

- Activists can effectively raise awareness around specific issues in artificial intelligence, such as Brazilian human rights (with a gender focus) and racial biases that must be overcome to achieve algorithmic justice. The bottom-up approach of grassroots activism allows for people who are affected by AI systems to have more agency in the discourse around autonomous systems.

Governance/Regulation

- Governments and international organizations may expand their goals past simple access and focus on infrastructure capacity, technical education, affordability of both the Internet and connected devices, and digital inclusion that considers existing inequalities.
- Developing countries might consider existing inequalities in order to overcome disadvantages that deter wider domestic adoption, such as the gender divide, the age divide, and the income divide. Similarly, developing countries may need to focus on issues such as language differences in relation to available content, severe economic disparities, and cultural obstacles to adapting to the Internet and related technologies.

Progress towards a more inclusive future with AI will require concerted interdisciplinary efforts that span collaborations among many different stakeholders, including academia, governments, international organizations, companies, and more. Without critically engaging with the contexts in which these technologies are being developed, as well as continually examining the ramifications of our dependence on such AI systems, we will be ill-equipped to face the staggering risks that they may carry into our future. The Global Symposium on AI & Inclusion is intended to make progress towards a robust research agenda and create pathways for collaborations across disciplines, sectors, and geographies.