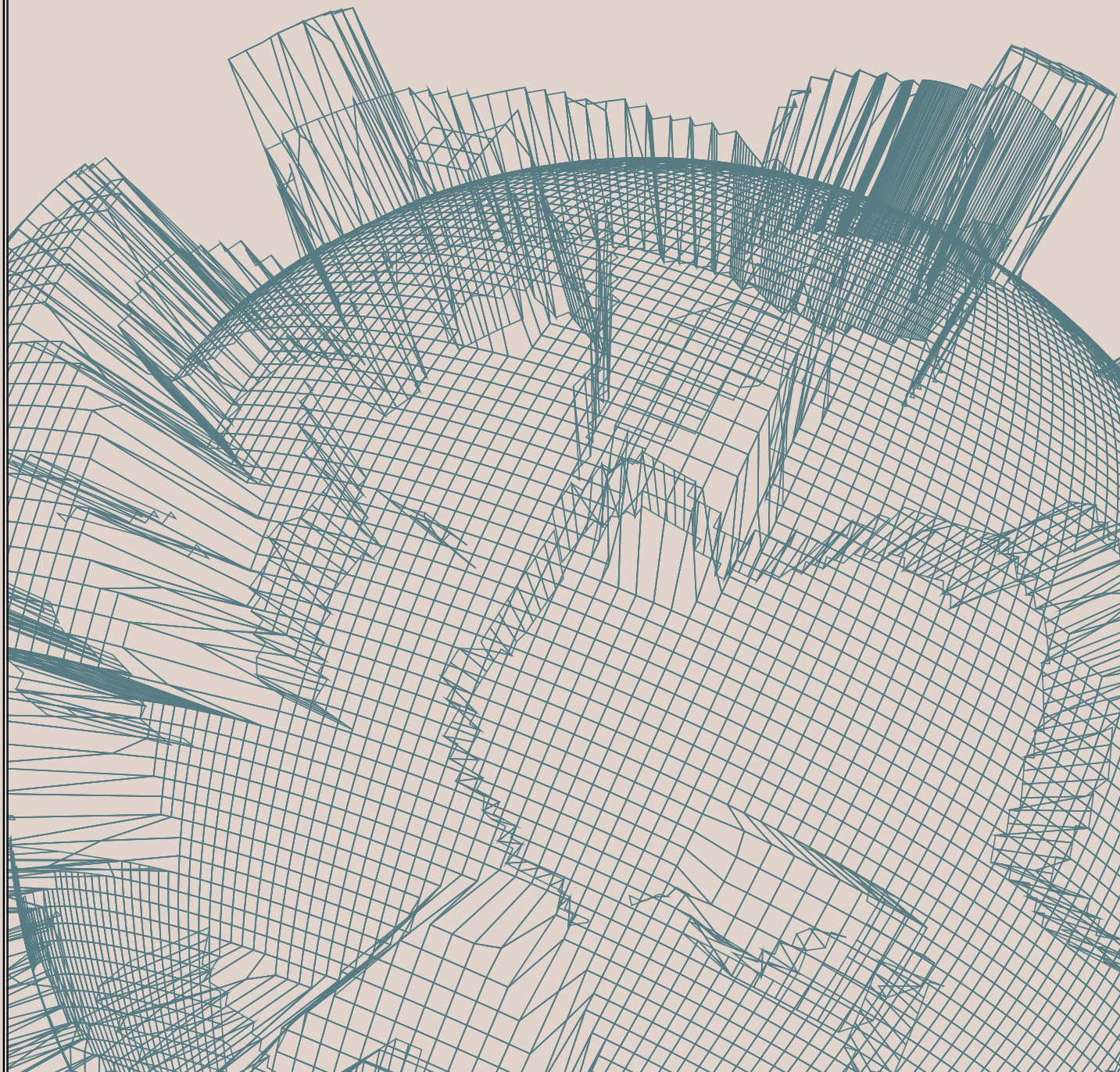
AI and Implicit Bias Policy Lab
Spring 2022

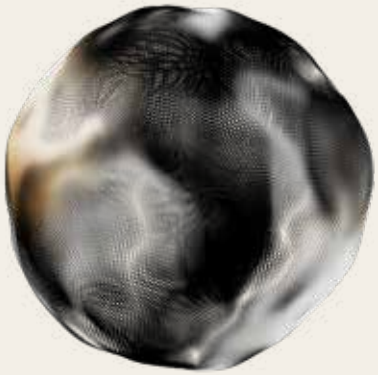
ARTIFICIAL INVISIBILITY

EXPOSING THE INTERSECTIONAL HARMS OF ALGORITHMIC BIAS



University of Pennsylvania
Carey Law School

FOREWORD



I started out my career as an attorney in an all-women's technology law practice in India. My typical day seldom went by without an exasperated sigh or two from my colleagues as we laboured over convincing our clients to comply with the numerous strange regulatory requirements that their businesses had to adhere to. We regularly advised tech start-ups and giants alike in structuring and launching their digital products and services in the Indian market.

In the course of our counsel, the innovators who designed technological offerings, often remained possessive of their creations, and sceptical of our advice. Any suggestions to tweak their digital products to ensure better regulatory compliance were usually looked upon doubtfully. On the other hand, I am sure our pedanticism didn't help win the technologists over either!

It is with this limited experience as a technology law practitioner that I joined the select group of nearly 40 students in the Policy Lab on Artificial Intelligence and Implicit Bias at Penn Law. However, after the very first lecture, I was amazed to have all my preconceived notions about the friction between technologists and attorneys dispelled! Over the course of this Spring, the Lab brought together experts from diverse fields and countries to share with the class their experiences of working with AI systems. I was pleasantly surprised to see computer scientists, engineers, lawyers, policy-makers and business-persons join hands to engage in a discourse on systemic biases in AI and how these can be mitigated. Here lay the biggest revelation from this Lab for me- the articulation of a shared goal for building fairer and equitable AI systems, acted as a bridge not just between professionals, but also between academic institutions, private sector enterprises and regulatory bodies.

The collaborative nature of the Lab next demonstrated the value of interdisciplinary co-operation in reaching policy solutions. Our guest speakers displayed how multistakeholder engagement is vital for problematizing the societal impacts of AI tools, and how solutions to such impacts can only emerge through a partnership across various interest-groups. For instance, Prof. Sandra Wachter-one of the guest speakers in the Lab explained how counterfactual explanations to automated decision-making can be a step towards bringing

about explainability, transparency and accountability in such determinations. Through the lens of data scientists, she analyzed how the data sets upon which these decisions are made, may be re-designed to enable generation of counterfactuals. At the same time, she also discussed in her capacity as a legal scholar, the inadequacy of prevailing regulatory frameworks in guaranteeing a meaningful right to explanation in AI. Thereby, Prof. Wachter and her colleagues have taken a multidisciplinary approach towards the opacity in AI. We were also privileged to witness, a European data scientist- Ms. Alexa Pavliuc speak of the disastrous impacts of AI-run disinformation campaigns in developing countries. On the other hand, leading policy-maker Dr. Virgillo Almeida, the former Secretary of Technology and Innovation in Brazil emphasized the need to include underrepresented states and communities from the Global South in the conversation on AI-governance.

Through our conversations with technologists such as Ms. Deborah Raji, we further learnt that implicit biases in technology can run across gender, race and class divides. Any solution for removal of such biases has to take into account the multiple facets of individual identity and how biases may be intersectional.

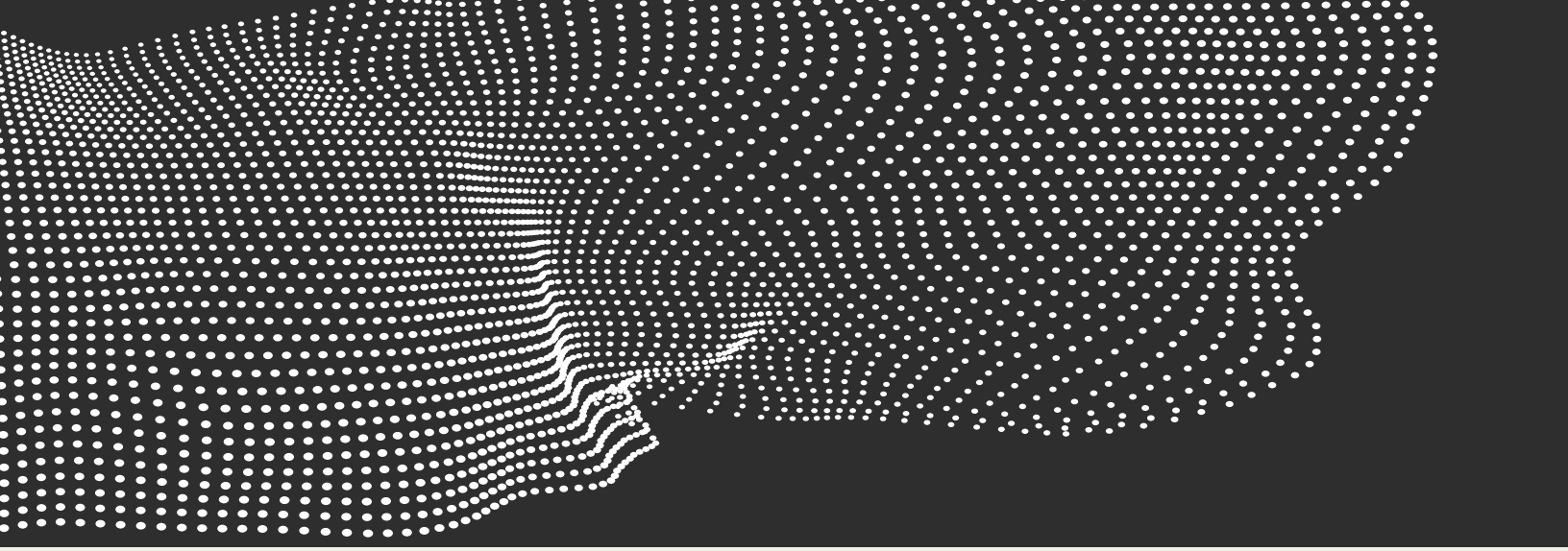
With such thought leaders paving the way for us, my fellow students and I at the Lab collaborated for understanding how AI systems contribute to prevailing gender inequalities in the work force. The use of automated filtration tools for screening job-applicants during recruitment process is increasing by the day. At the Lab, we set out to understand whether and how the pre-existing biases of the persons designing such systems seep into the technologies being used in the hiring process. Using survey-based research tools, the Lab engaged with young professionals between the ages of 20-35 to collect empirical data on their experiences with online recruitment portals and job-search websites and applications. This Report sets out our key findings from the data we collected and tries to identify the algorithmic biases that have bled into AI-driven recruitment mechanisms.

Further, the Lab demanded each one of us to prepare a policy brief addressing the impending question of implicit biases in AI and asked us to suggest innovative policy measures for resolution of this issue. With a focus on the Global South and gender-equality, we worked towards identifying new generation of biases such as immigration status, zip code and address, etc., creeping across AI tools. We further discovered several contexts, as varied as housing, social media platforms and primary education, where such biases →

have trickled. This Report will also touch upon our human rights-centric recommendations for unmasking biases found across such diverse use-cases of automated systems.

This Report is the fruit of our labour over the last semester and our effort in assisting AI developers and businesses in freeing algorithmic systems of structural biases. It is with great pride that we present our findings to the Vice-President and Deputy General Counsel, Human Rights, Microsoft Corporation- Mr. Steve Crown. He has championed the cause of developing and following human rights-based policies in Microsoft's conduct of its business. We feel fortunate to have found in him a partner who can encourage the deployment of emerging technologies while in a manner which is cognizant of the impact of these tools on individuals and society, and for the empowerment of communities. Using the identification of biases in algorithmic tools used in job applications and hiring as a starting point, through our research findings and policy recommendations, we hope to contribute to ongoing endeavours towards making AI systems and their applications more inclusive and pluralistic.

Now, keeping this conversation alive is a responsibility that each one of us at the Lab bears. I feel honoured to have been a part of this vibrant cohort of students. With their varied educational and social backgrounds and skill-sets, my fellow classmates have exemplified the learnings that our distinguished guest lecturers brought to the class. Throughout the semester, we have worked together to utilize our global perspectives and wide-ranging experiences in law, technology and humanities to critically engage with ideas on AI design, application and regulation. This journey of ours has culminated in the formation of a global fraternity of future policy-makers. We recognize the duty that we bear in giving voice to all communities and bringing about positive impact in the discourse on AI governance.



We spring forth from the shoulders of our beloved Professor- Dr. Rangita De Alwis De Silva, to whom we are grateful for the wonderful opportunities that the Lab has presented. The brains behind this mammoth experiment, Prof. De Alwis De Silva has evolved a new pedagogical tool for law schools where the classroom serves as a space for incubating new policy solutions and goes beyond traditional legal training. By connecting us with pioneering thinkers, she has facilitated dialogue as the primary means of instruction, and encouraged us to build from the experiences of our guest speakers. As the architect of this Lab, she has further provided us with a dynamic illustration of meaningful multi-stakeholder engagement in policy-making.

Over the course of this Spring, my eyes have opened up to the possibility of a marriage between information technologists and attorneys (and made me more empathetic towards my former clients!) I hope that my classmates and I can continue to advance this relationship as we assume our responsibilities as policy-makers. I wish for us to carry our learnings from the Lab forward in making AI-governance a more participatory and representative process across the world. In this quest, I remain committed to taking experiential labs to law schools back to our home countries and across the world. This teaching methodology can be a road-map for re-imagining legal education, and preparing law students better for their responsibilities as future lawmakers.

Veda Handa

Dean's Merit Scholar, LLM Candidate,
Class of 2022

ABOUT THE LAB

The University of Pennsylvania Law School Policy Lab on AI and Implicit Bias incubates ideas for an intersectional approach to inclusive artificial intelligence. Primarily through a series of multilateral conversations with international stakeholders, including leaders in technology, technologists, lawyers, researchers, and designers, we will seek to understand whether and how gender and intersectional bias, including implicit and unconscious biases are being baked into technological design and algorithms, and whether and how these biases are being reproduced in new technologies.

Currently, there is gender and intersectional asymmetry in the AI workforce. Those designing, coding, engineering and programming AI technologies do not represent a diverse demographic. Our theoretical explorations included the human rights framework, gender equality theory, post-colonial theory, implicit bias, in group favoritism, and affinity bias to explore subtle barriers to equality that bleed into the design of AI technologies. The lab engaged in survey-based research and data collection on a new generation of algorithmic bias and the human rights tools that could address them. Our work will help tech leaders, designers, and technologists in their efforts to embed pluralism and inclusion into AI systems and the future of work.

STUDENT LEADERS

BIANCA NACHMANI

Teaching Assistant, Report design

- JD, University of Pennsylvania Law School, Weitzman School of Art and Design, 2022
- BA, University of Florida, 2019
- Hometown: Penn Valley, PA

ASHLEY MARCUS

Teaching Assistant, Report research

- JD, University of Pennsylvania Law School, 2022
- MSc Policy Studies, University of Edinburgh, 2019
- BA, University of Pennsylvania, 2018
- Hometown: Golden Beach, FL

DANA DYER

Research Assistant

- JD, University of Pennsylvania Law School, 2022
- BA, University of Pittsburgh, 2018
- Hometown, Norristown, PA

BROOKE BERNSTEIN

Data Specialist

- MS, Social Policy and Data Analytics, University of Pennsylvania, 2022
- BA, University of Florida, 2021
- Hometown: Ponte Vedra Beach, FL.

CLASS TAUGHT BY: PROFESSOR. RANGITA DE SILVA DE ALWIS

STUDENTS OF THE SPRING 2022 LAB

Yusef Ahmad
Fajar Arafah
Brooke Bernstein
Shahrani Bhatti
Carolina Brandão
Carolina Bueno
Olivia Chowdry
Tatiana Corcoran
Ashlynn Dobbs
Dana Dyer
Ellen Yang Gao
Victoria Gross
Veda Handa
Ting-Hao Hsu
Anna Huebner
Karis Jackson
Mary Loftus
Jordan Oh
Alezeh Rauf
Carolina Romero
Keshara Senanayake
Maitreya Shah
Falak Shahid
Manan Shishodia
Tianyu Song
Alexandra Speiss
Michael Stephenson
Andrea Stoller
Sharanya Sukumar
Laurel Sutherland
Aubrey Thompson
Melanie Toledo
Rith Tshimanga
Jane Wang
Miriam Weinstock
Justin Wong
Angela Wu
Junyi Zhu
Carlos Rendon

CONTENTS

INTRODUCTION

A CONVERSATION ACROSS DISCIPLINARY BOUNDARIES

In 1950, Alan Turing anticipated that by the year 2000, "One will be able to speak of machines thinking without expecting to be contradicted." It is that risk that Machine Learning and AI is unassailable that poses the greatest threat to the 21st century.

Artificial intelligence is not any longer an engineering discipline but requires engagement across disciplinary boundaries. Our Policy Lab on AI and Bias was not only an incubator for new ideas, but it was marked by a methodological approach that focused on disciplinary and cultural diversity that brought together a heterogenous group of experts drawn from technology, business, law, public policy, venture capital and the humanities. The lab provided a space for debate and discursive thought where a new generation of scholars could challenge and question traditional views on AI related bias with established leaders in technology and the digital humanities.

Global South Approach

A Global South approach was unique to our Policy Lab.

The term "South" traditionally has referred to the "Third World" and has covered countries that share a postcolonial history. However, a more plural understanding of the Global South includes underrepresented and disenfranchised populations. The AI-related risks for the Global South heightens the concerns of discrimination, bias, oppression, and invisibility of those who are most left behind by economic and social development.
A perspective from the Global South examines a way in which AI risks the subordination of people who most need to share in its achievements. At the same time, AI has great potential for good in the Global South. For example, in rural Rwanda, a partnership between its health ministry and Zipline, a Silicon Valley-based tech startup, is giving

doctors in hard-to-reach clinics the ability to order blood by text message, and then have it arrive by parachute in a matter of minutes.

A perspective from the Global South examines a way in which AI risks the subordination of people who most need to share in its achievements. At the same time, AI has great potential for good in the Global South. For example, in rural Rwanda, a partnership between its health ministry and Zipline, a Silicon Valley-based tech start-up, is giving doctors in hard-to-reach clinics the ability to order blood by text message, and then have it arrive by parachute in a matter of minutes.

Our Lab convened leaders from the Global South such as Dr. Virgilio Almeida the former Secretary of Technology and Innovation in Brazil who chaired NetMundial, a global conference that sought a shift from a multilateral to a multistakeholder approach to AI.

On the last day of class, the Lab hosted Ambassador Ammo Aziza Baroud, Chad's Ambassador to the United Nations. AI can play a role in addressing access to water in the Sahel region where severe climate change has dried up 80 percent of Lake Chad and in turn impacted migration, food security and farmer herder conflict. At the same time, Ambassador Baroud's vision for an African green belt calls for the engagement of African women in the development of AI.

Human Rights Approach

Our Lab is grounded in a human right–centric, inclusive, and intersectional approach to addressing algorithmic bias and exclusion.

AI governance anchored in human rights as enshrined in the Universal Declaration of Human Rights (UDHR)—provides an established, global construct that is acknowledged by governments, businesses, and civil society. Secondly, this framework offers a legal basis for more specific regulation in a way that ethical frameworks do not. Thirdly, there are

established procedures and norms for assessing the human rights impact of business operations and remediating harm.

The Universal Declaration on Human Rights drafted in December 1948 is a shared standard of achievement for all people and nations. Article 27 of the UDHR establishes that "Everyone has the right to share in scientific advancement and its benefits." At the same time, Article 12 establishes that no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, and everyone has the right to the protection of the law against such interference or attacks. Moreover, the "due diligence" standards of the "Guiding Principles on Business" and Human Rights provide an agreed set of norms, a shared language, and institutional infrastructure to hold the business ecosystem accountable to AI-related discrimination and bias.

Intersectional Approach

Despite progress, most AI systems are designed with an understanding that gender or race is only one axis of difference. However, gender-based discrimination is often compounded by its intersection with socioeconomic status, race, age, disability, and other categories of marginality.

Deborah Raji, Joy Buolamwini of the MIT Media Lab and Timnit Gebru, then of Microsoft Research, have examined how algorithms have different outcomes on different intersectional groups. By examining algorithmic performance at the intersection of race and gender, Raji discussed in our Lab the significance of conducting intersectional audits for these types of AI systems. An example of bias is when a chatbot assumes that "doctor" indicates "man" and "nurse" indicates "woman." In our readings when Latayna Sweeney, the first Black woman to receive a Ph.D. in computer science at MIT and current Harvard Professor, googles her own name, she comes across ads like: "Latanya Sweeney, Arrested? 1) Enter name and state 2) Access full background. Checks instantly....." In situations of women of color and disability status, algorithmic bias is further pronounced.

Data-Driven Approach

My co-authored paper to be published by the Michigan Law Journal of Technology, a special edition on the intersections of race, gender, technology, and the law examines the relationship between investor diversity and diversity in AI. Our research with Crunchbase revealed the following data: Investors in AI: 79.8% (male) 19.8% (female) 80.0% (white) 16.0% (Asian) 1.7% (Hispanic)1.3% (Black). This brings into question diversity across the entire ecosystem and how gender gaps in venture capital can lead to algorithmic bias?

In the final analysis, the Lab engaged in preliminary data collection as a prelude to an argument for a new policy regime to address algorithmic and classification bias. These new categories of bias address the classification schemes in hiring platforms that sort employees in ways that amplify inequality or disadvantage according to race, sex, disability, and other protected categories. When decision- making algorithms produce biased outcomes, we need a new human rights- based policy regime to redress those disparate outcomes.

BRIDGE BUILDING

An overall theme of the Policy Lab was the importance of bridge building across and between academic institutions, multilaterals, and the private sector. Our collaborators were founders, CEOs, investors, technologists, entrepreneurs, lawyers, policymakers, writers, designers, and academic leaders. We collaborated with the Berkman Klein Center for Internet and Society at Harvard University and thank John Palfrey the former head of the Berkman Klein Center and the current president of MacArthur Foundation for his inspiration as a thought leader on emerging media and increased online transparency and accountability.

- Professor Rangita De Silva De Alwis

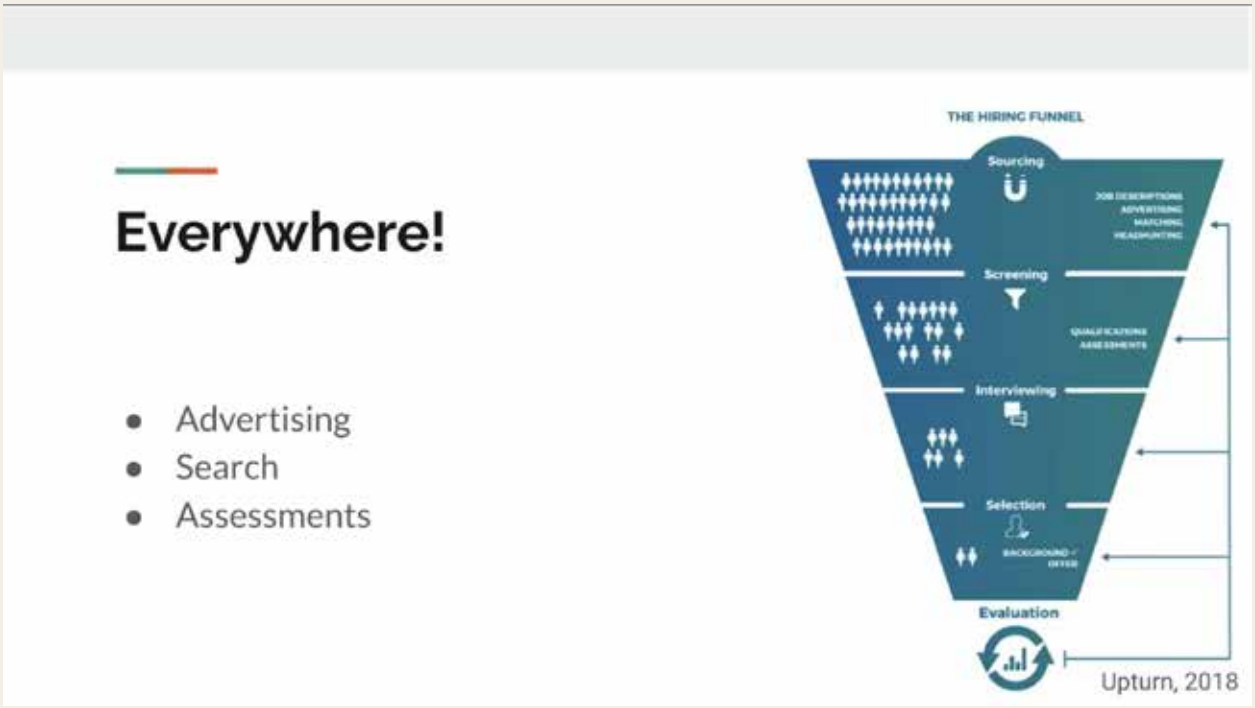
INTRODUCTION

In leading Penn's Spring 2022 AI and Bias Policy Lab, Professor Rangita de Silva de Alwis brought together scholars and industry and government leaders as guest speakers to uncloak the realities of artificial intelligence and its impacts on our world. Their discussions with the students centered around the policy lab's five core pillars: a global south approach, a human rights based approach, an intersectional approach, a data-driven approach, and bridge-building. To help convey their expertise, many speakers offered PowerPoint presentations with insightful explanations. The following section contains pieces of those presentations.

Five Pillars:

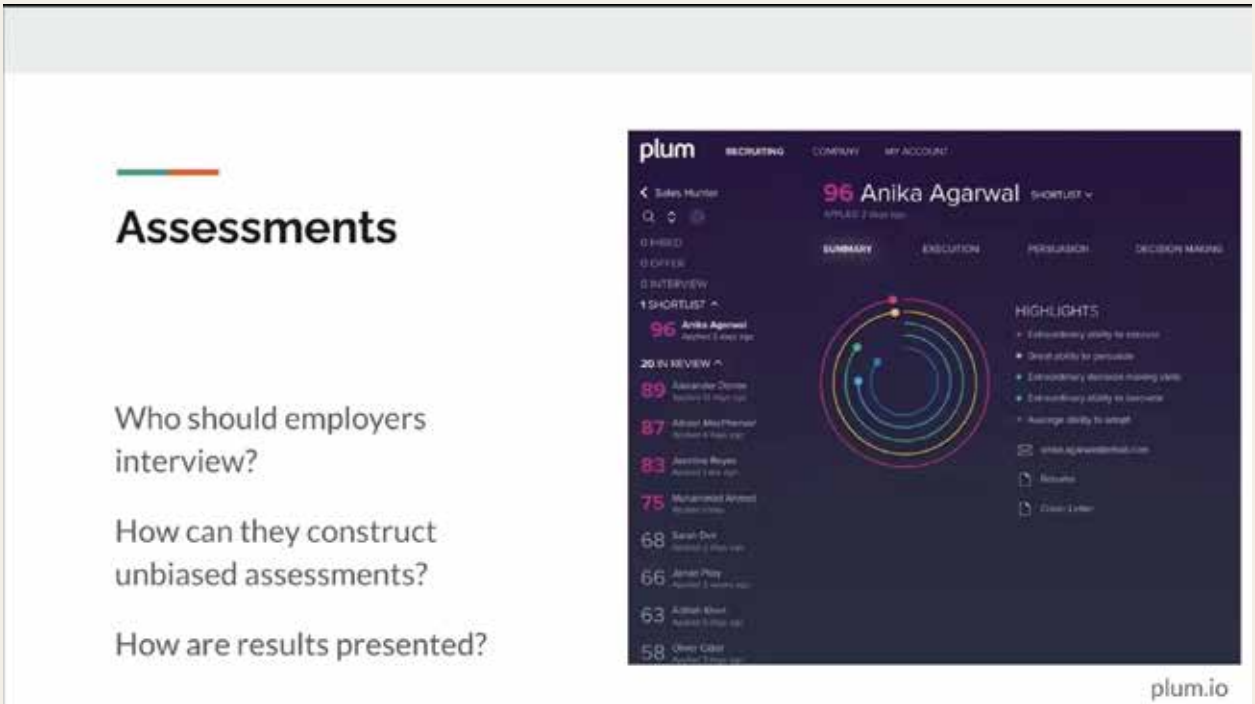
A. *Global South Approach*
B. *Human Rights Approach*
C. *Intersectional Approach*
D. *Data-Driven Approach*
E. *Bridge Building*

"[AI is] actually used in a wide variety of places. So from everywhere from sourcing, screening, interviewing, selection, all of these stages in the hiring pipeline, there are various applications for AI. For instance, when advertising a job. If you, let's say, buy advertisements from Facebook and that gets distributed to some amount of people, there's an algorithm determining who actually sees your job ad. Similarly, if you go to LinkedIn and you search for 'I want a software engineer' in let's say Atlanta, Georgia, you run that search on LinkedIn. Some people will show up in the results of that search. LinkedIn has an algorithm somewhere that is going to determine who gets shown to you in response to your search. The example that we're going to focus on today is assessments, where you apply and an algorithm somewhere evaluates you as a candidate. So for instance, you might submit your resume and then your resume gets screened by some algorithm and whether or not you pass the resume screening filter depends on that algorithm. Similarly, you may go through these more complex assessment procedures, where you record a video interview and then that video is algorithmically scored and again, that determines whether you were invited to apply for the job or whether you get an interview. So there's all these places where algorithms seem to be mediating people's access to employment opportunities." - *Manish Rhagavan*



Slide Courtesy of
Manish Rhagavan

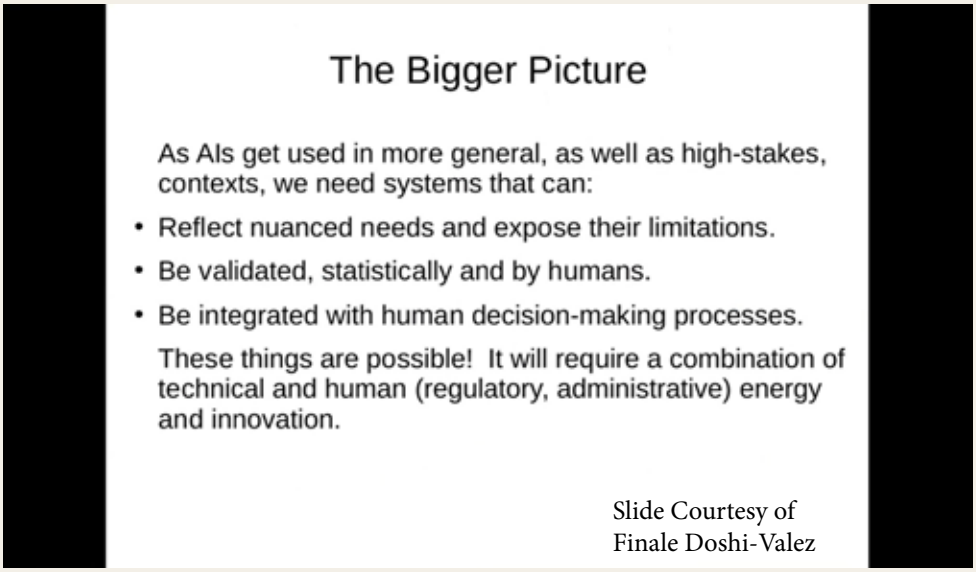
"We're going to focus on assessments today and in particular, on auditing those assessments because I find them a particularly instructive way to try to better understand what the algorithms are doing. So here's an example of what one of these algorithms might show you as a recruiter. It shows you here all the candidates who applied and here's the scores that they all got, right? Now these scores are entirely computer generated. They went through some tests and somehow, that number pops out. Now where does this come from? Ultimately, what's happening on the back end here is the employer is training an algorithm based on people who have appeared in the data set as good candidates in the past, right? Or people who've been hired in the past or whatever that might be…these are going to determine who employers should interview." - *Manish Rhagavan*



Slide Courtesy of
Manish Rhagavan

Sometimes the law falls short of adequate protection

"When we say don't include things like gender [or other protected characteristics] in your models you might think that this is going to prevent any sorts of outcome differences from arising. Because you say if the algorithm doesn't know about gender, how can it discriminate on the basis of gender? And of course, what ends up happening is that there are all sorts of proxies in the data that we observe that can lead to outcome differences down the line. The most common example that you hear of this, not in the context of hiring, but in the context of lending, is zip code as a proxy for race." -*Manish Rhagavan*

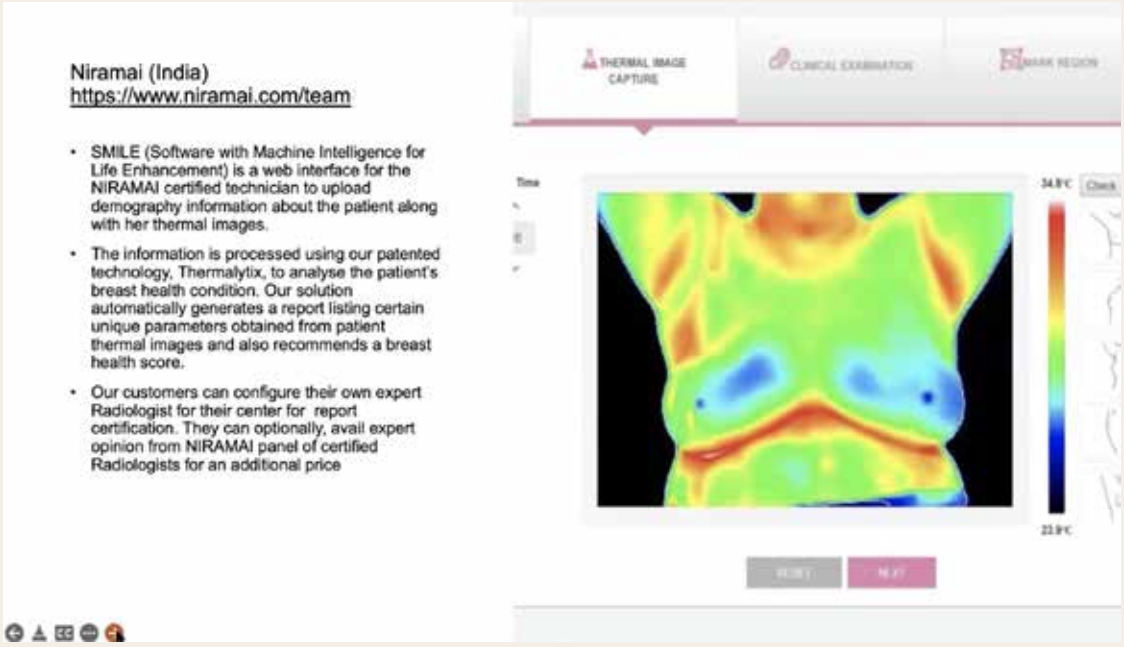


"We really need systems that can reflect nuanced needs and expose their limitations. I hope in this room we're kind of beyond this setting that, oh AI can solve all of our problems. But still, a lot of the world kind of thinks that AIs will remove biases because humans are biased, right? Well, it is true that humans are biased, but I have a different set of biases than they bring. And what we really need are ways in which humans and AI can reduce each other's warts and weaknesses and build upon their strengths. We need systems that can be validated statistically and by humans, because both again have different strengths. There are things that numbers can tell us that human inspection cannot and there are things that human inspection can tell us that numbers cannot. And then in terms of, as I've been saying over and over, we just really need to be able to integrate these eyes with human decision making processes. And that last thing that I want to mention from a technical perspective is that these things are possible." -*Finale Doshi-Valez*

A. GLOBAL SOUTH APPROACH

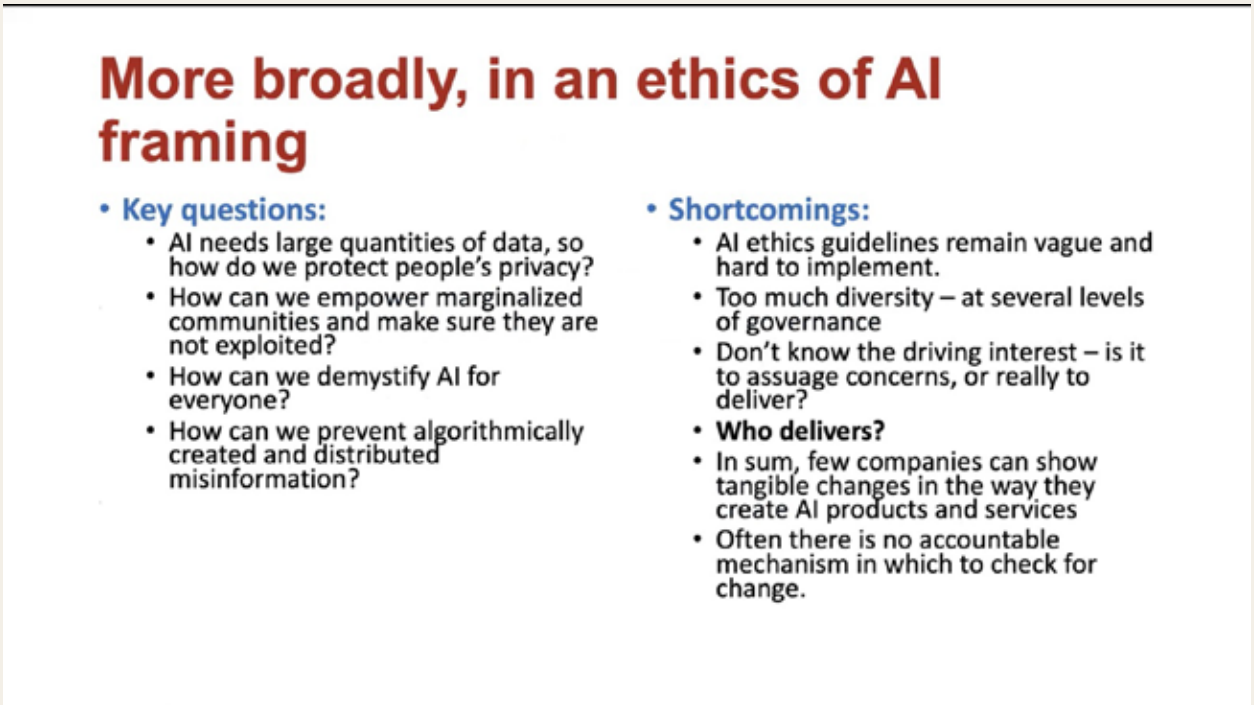
"The whole ethics of the debate more or less of Fair AI, as you might want to call it, is trying to tweak shortcomings. Now, I'm going to discuss this from a global south perspective or a global perspective because we know how the debate looks in the US–where it's relatively advanced and heavily debated about and you have everything out in the open…from a comparative perspective…When you try to move that to the global south you see how much less things are in the open…I want to say two things right at the start before I explain what I'm going to explain: (1) I'm not against the use of AI for social good, not at all, because I think it can contribute extensively to a lot of things; (2) I'm not against the use of including people's traditional ID program or furthering the SGD's program. What I'm against is a siloed narrow approach to implementing this, which is leading to much more exclusion than inclusion. One of the reasons this happens is because we are fixated on establishing the correct relationship between cause and effect, but we are not fixated on trying to see what impact it has on society." - *Padmashree Gehl Sampat*

Slide Courtesy of Padmashree Gehl Sampath



Slide Courtesy of Padmashree Gehl Sampath

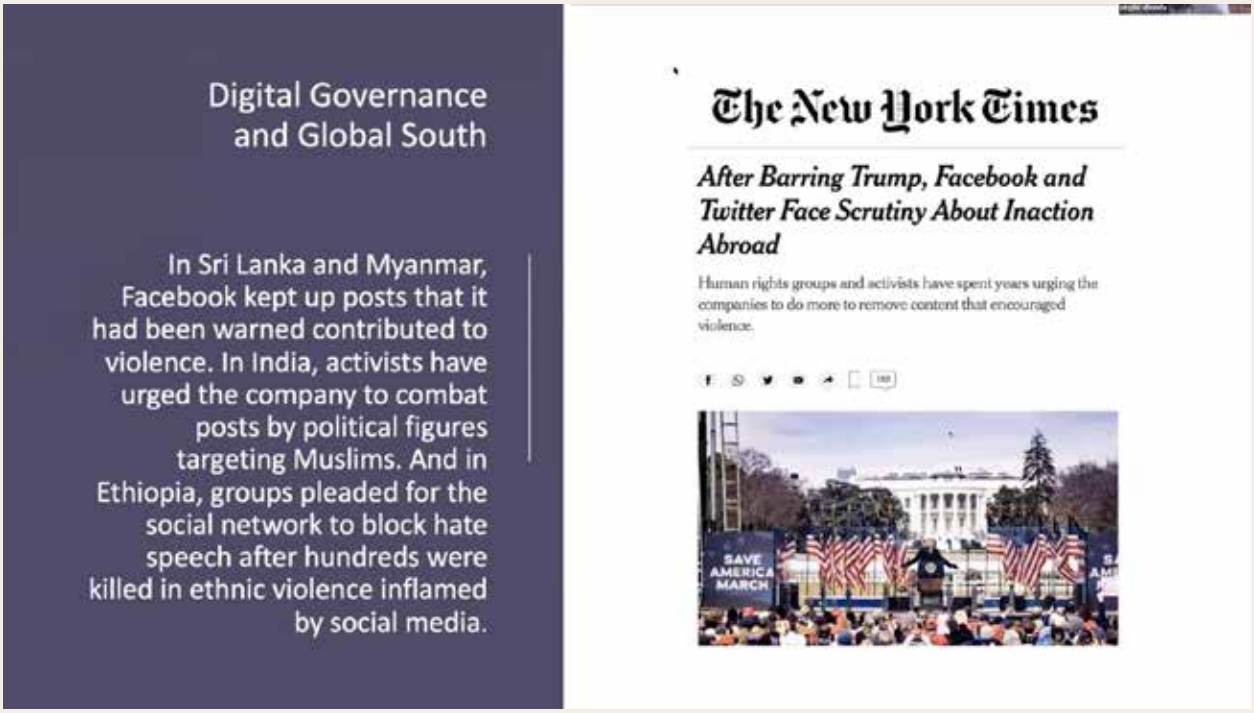
"AI and health, simply AI can help diagnose, for example, early onset of several diseases, and they can use scalable technology. So what you see is that in a large number of countries where I've done fieldwork over the last four years, you see a lot of companies mushrooming to use AI for health care products, okay, for healthcare outcomes. So here is a company called India. It's in India. It's actually using AI to predict certain kinds of clinical pathology outcomes. This is Niramai. This is also an Indian company, and this is basically using some sort of machine learning software to tell you if you're going to develop breast health conditionalities by using certain parameters in your body…If I was an Indian policymaker and I've spoken to a lot of Indian policymakers, they will tell you that AI is doing fantastically well in India. We have so many software for small scale companies that are adapting. They are thriving in different sectors. As a consumer in these instances, you might opt for early diagnosis. But the benefit of that early diagnosis may offset the cost, because many people whom you–beyond who you authorize may have access to the data that might not entirely be anonymous, and it risks that you might use that in detrimental ways…The question that I ask in many of my papers and work is whether an ethics framework is sufficient to resolve this. In an ethics framework, you will ask objectively is this the right thing to do, but how do I know that my ethics are the same as yours, and in many instances, our ethics framework tends to externalize this risk onto you." -*Padmashhree Gehl Sampat*



Slide Courtesy of Padmashree Gehl Sampath

"There's very little investigation into the impacts of AI for facial recognition in the global south. For instance, I have a lot of case studies that I've collected and I know colleagues that have collected case studies where AI is being used for facial recognition in a very bad way in the global south. AI is also being used to actually segment certain parts of the city for the police force in the way that we know is wrong from the US in the global south in countries like India, for instance, which marginalizes certain communities much more than others. It's the same thing in the case of South Africa, for instance, or the same thing is true in parts of Latin America. So when we talk about the failures in AI, some are intentional in design, which we can actually fix. We can improve the quality of data. We can actually do something about it because we are conscious of it. We are conscious of it from the way that we just automatically react to it. But there are many things that we as people just do, which is unintentional, and it's subconscious. ANd that doesn't make it right because it's subconscious. So when it's not structured to take into account the social reality or a historical artifact, which becomes a social reality for some people, then AI reproduces patterns of inequality and bias." -*Padmashree Gehl Sampat*

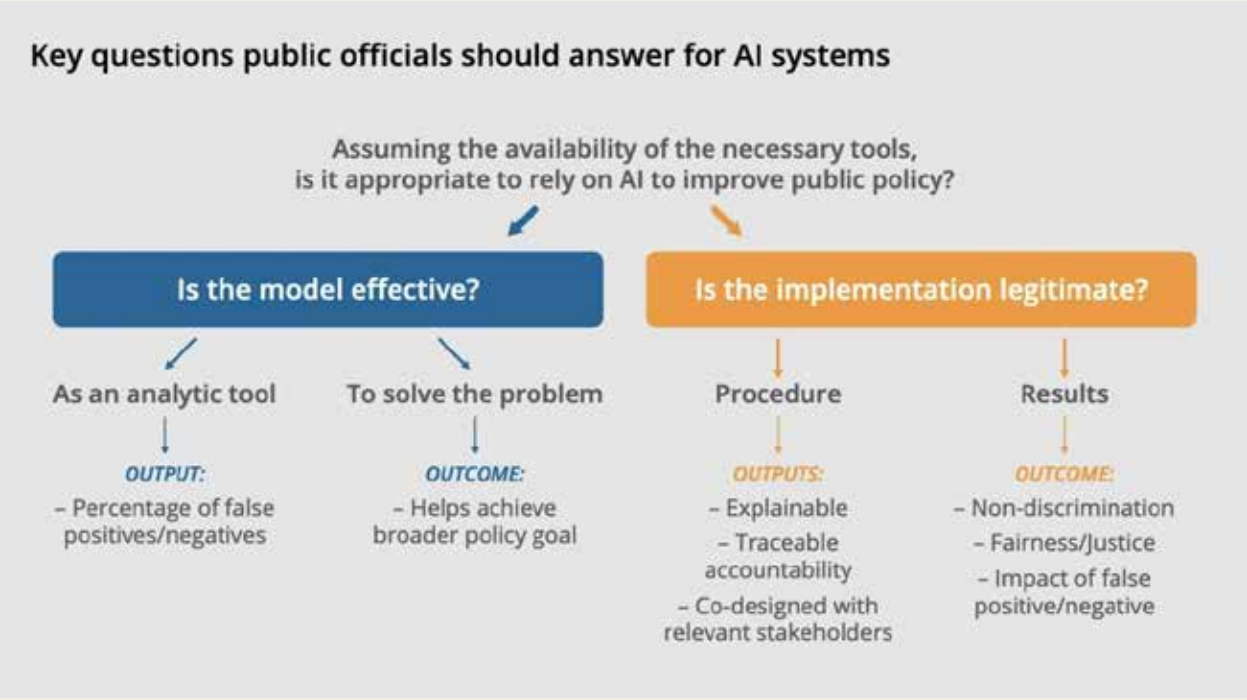
"Digital governance has been a development developed by the global north and they do not represent the needs and problems of the global south. You see here that in the case of Sri Lanka and Myanmar, Facebook kept up posts that it had been warned contributed to violence and that was a big problem for activists in media and in particular for Muslim people because most of the content moderation teams were dedicated to content moderation in the global north."- *Virgilio Almeida*



Slide Courtesy of
Virgilio Almeida

Slide Courtesy of Virgilio Almeida

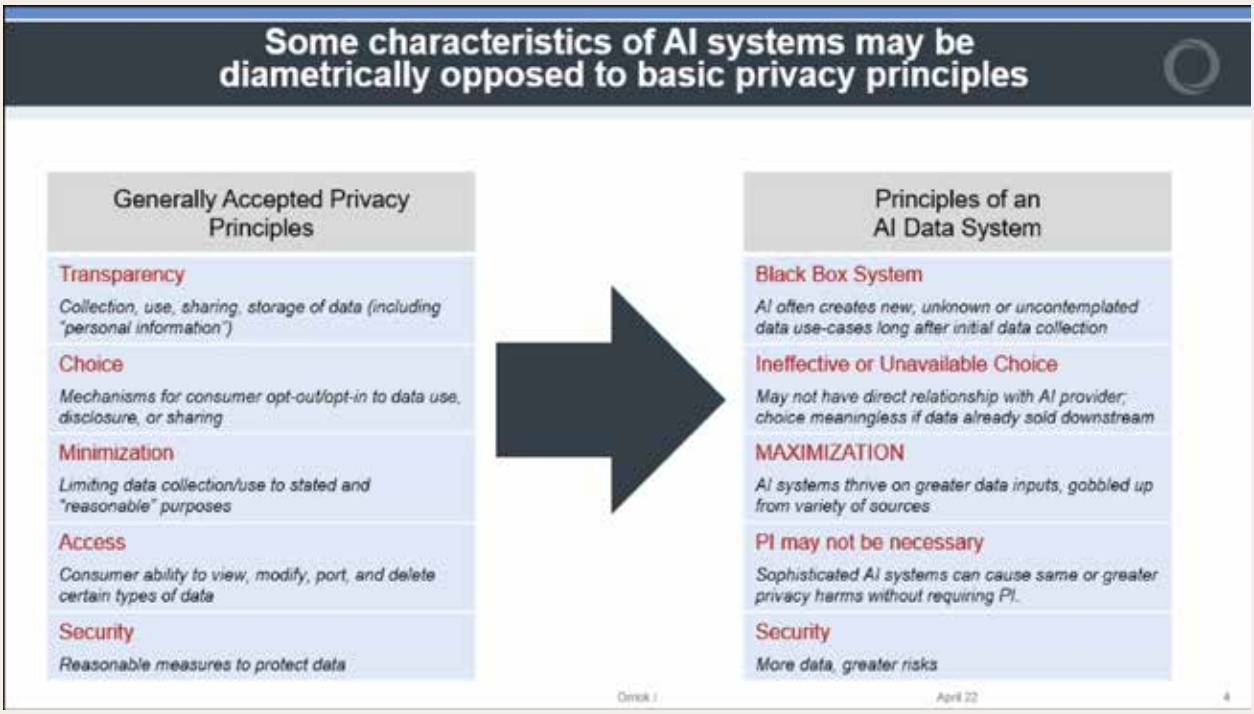
"It's important to include the global south to have new experience and to look for innovation…We proposed the creation of a globally interconnected space for interdisciplinary as a nation and building as a necessary condition for systematic and more frequent civilization. We need to bring together people, computer science, [and] political science in order to address the challenges brought by the digital world and we need to invest not only in technological innovation but also in these governance innovations." - *Virgilio Almeida*



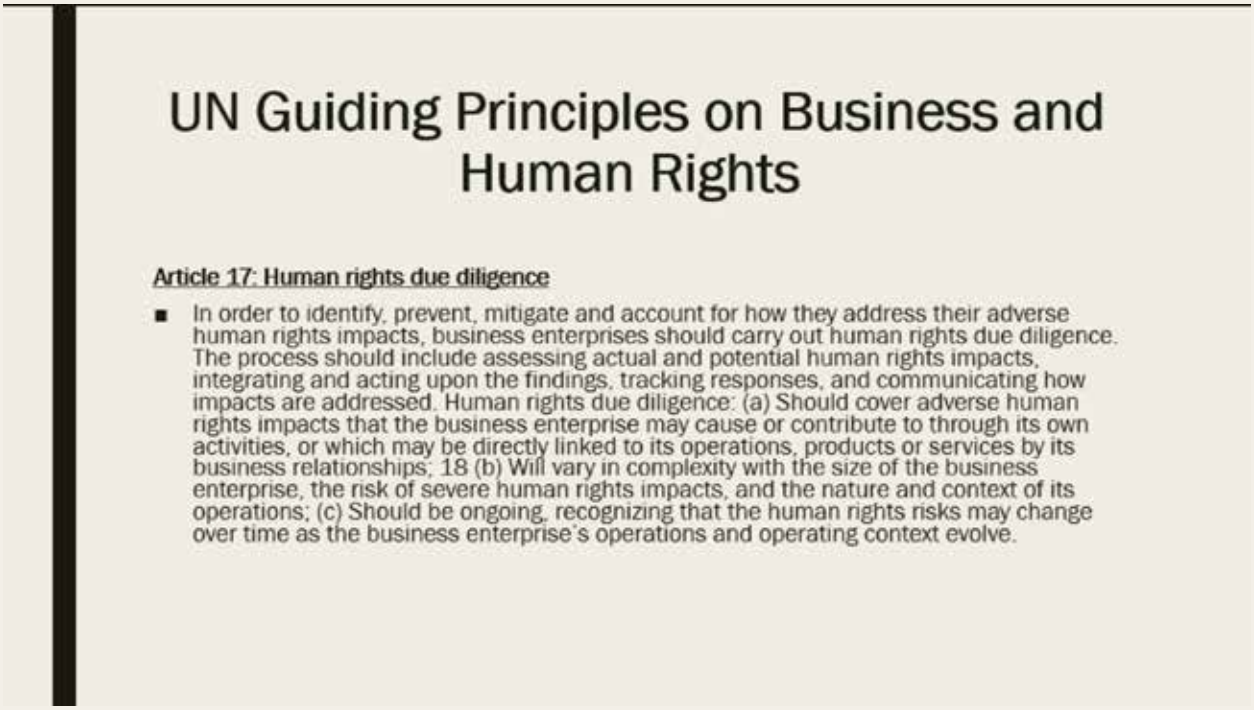
Slide Courtesy of Juan Ortiz Fueler

B. HUMAN RIGHTS APPROACH

"Some characteristics and age systems are potentially diametrically opposed to basic privacy principles. So sometimes called the Fair Information Practice Principles the FIPS gap. And so we're going to, of course, leave you with these slides for distribution. I don't need to run through each of these, but it's also this concept of data maximization on the one hand versus data minimization on the other. And the larger the datasets that we're working with, the greater the potential for privacy law implications, right? Privacy laws are generally grounded in personal information, personal data surveillance. So it's not just about whether you're collecting data about an individual name, identification number, location or an online identifier, but it's also this privacy is this concept of intrusion upon seclusion, the right to be left alone." -*Heather Egan Sussman*



Slide Courtesy of Heather Egan Sussman and Ryan McKenney



Slide Courtesy of AI Policy Lab

Slide Courtesy of AI Policy Lab

Article 9 – Accessibility (cont.)

2. **States Parties shall also take appropriate measures:**

- a) To develop, promulgate and monitor the implementation of minimum standards and guidelines for the accessibility of facilities and services open or provided to the public;
- b) To ensure that private entities that offer facilities and services which are open or provided to the public take into account all aspects of accessibility for persons with disabilities;
- c) To provide training for stakeholders on accessibility issues facing persons with disabilities;
- d) To provide in buildings and other facilities open to the public signage in Braille and in easy to read and understand forms;
- e) To provide forms of live assistance and intermediaries, including guides, readers and professional sign language interpreters, to facilitate accessibility to buildings and other facilities open to the public;
- f) To promote other appropriate forms of assistance and support to persons with disabilities to ensure their access to information;
- g) To promote access for persons with disabilities to new information and communications technologies and systems, including the Internet;
- h) **To promote the design, development, production and distribution of accessible information and communications technologies and systems at an early stage, so that these technologies and systems become accessible at minimum cost.**

Slide Courtesy of AI Policy Lab

UN Sustainability Development Goals

9

Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

17

Strengthen the means of implementation and revitalize the global partnership for sustainable development

Slide Courtesy of AI Policy Lab

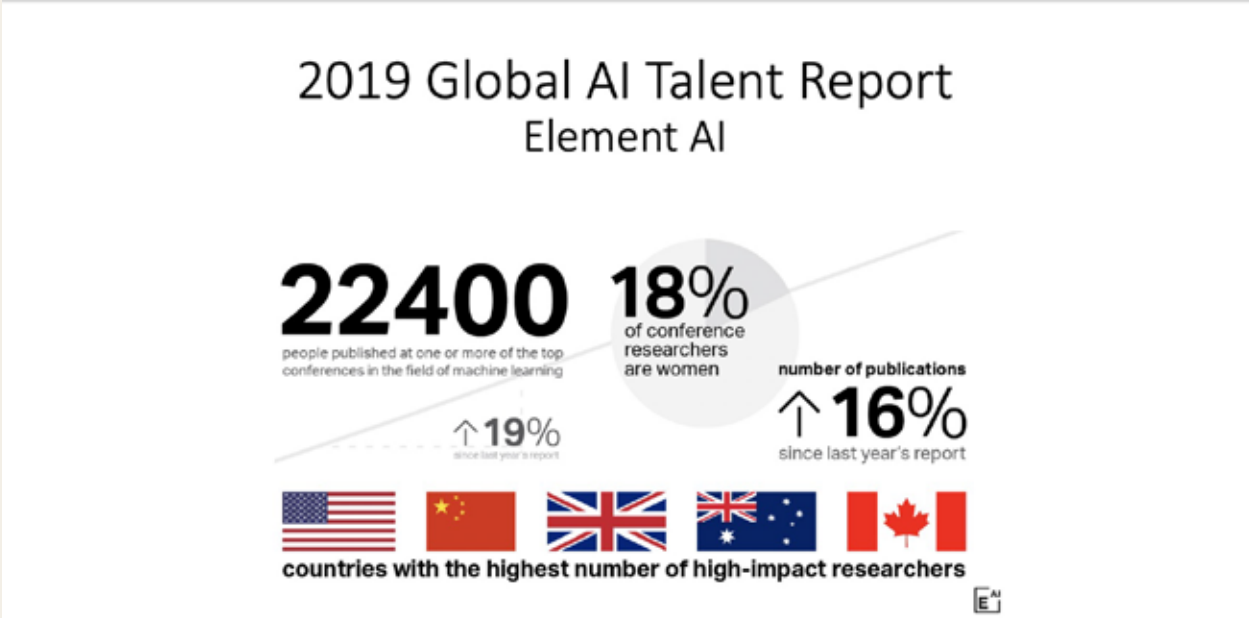
Convention on the Rights of Persons with Disabilities

Goal 9: Industry, Innovation and Infrastructure

Build resilient infrastructure, promote inclusive and sustainable industrialization: Access to financial services including affordable credit is ensured by equal recognition before the law (CRPD article 12); ensuring personal mobility with the greatest possible independence for persons with disabilities can be guaranteed by implementing article 20 of the CRPD; freedom to access information on an equal basis with others and through all forms of communication is part of article 21 of the CRPD; promoting opportunities for self-employment, entrepreneurship, the development of cooperatives and starting one's own business are recognized rights under article 27 pf the CRPD; articles 5, 6, 7, 9, 31 and 32 are also applicable.

Slide Courtesy of AI Policy Lab

Slide Courtesy of AI Policy Lab



Slide Courtesy of AI Policy Lab

Women are still underrepresented

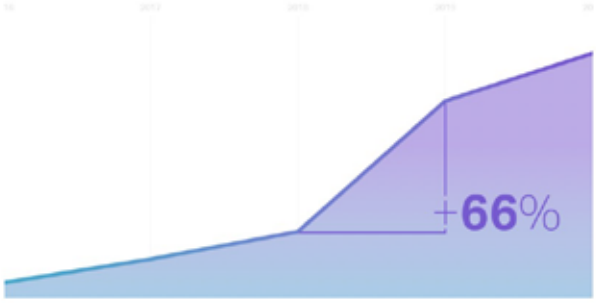
Last year, Element AI looked at three leading AI conferences and found that only 12 per cent of the listed authors for accepted papers were women. This year, we broadened the survey out to 21 conferences in the field and found that 18 per cent of authors were women. If a global talent shortage exists, one solution seems obvious: increasing the number of women in the field to something approaching equality. Yet this remains a larger problem for the AI research community.

The top 10 countries for women authors (research institution):

Spain (26 per cent)
Taiwan and Singapore (23 per cent)
China and Australia (each 22 per cent)
United States (20 per cent)
Switzerland (19 per cent)
UK and Italy (18 per cent)
India (17 per cent)

Slide Courtesy of AI Policy Lab

A handful of countries dominate AI research
Within the field of AI, there is a select group — around 18 per cent of the authors who published papers at the 21 conferences in 2018 — that contribute research with a major impact on the field, as measured by citation counts over the past two years. These 4,000 researchers are located mainly in the United States, China, the United Kingdom, Australia and Canada.



Rapid growth in self-proclaimed AI expertise
With the increase in popularity and excitement around AI, many people appear to be augmenting their own expertise with knowledge of how machine intelligence will change their particular industries or roles. The number of AI training options has also increased, including self-directed options such as those on Coursera or Udacity.

Slide Courtesy of AI Policy Lab

Artificial Intelligence & Human Rights: Opportunities & Risks
Berkman Klein Center for Internet & Society at Harvard University

A Human Rights-based Approach to AI's Impacts

The ongoing dialogue regarding the ethics of artificial intelligence (AI) should expand to consider the human rights implications of these technologies.

International human rights law provides a universally accepted framework for considering, evaluating, and ultimately redressing the impacts of artificial intelligence on individuals and society.

Since businesses are at the forefront of developing and implementing AI, the United Nations Guiding Principles on Business and Human Rights are especially salient in ensuring that AI is deployed in a rights-respecting manner.

Determining Impacts

We propose that the best way to understand the impact of AI on human rights is by examining the difference, both positive and negative, that the introduction of AI into a given social institution makes to its human rights impacts. We take this view for two reasons:

1. Determining the human rights impacts of AI is no easy feat, for these technologies are being introduced and incorporated into existing social institutions, which are not rights-neutral.
2. Each application of AI impacts a multitude of rights in complicated and, occasionally, contradictory ways. Exploring these relationships within use cases allows for more nuanced analysis.

Slide Courtesy of AI Policy Lab

Measuring Impacts

Current implementations of AI impact the full range of human rights guaranteed by international human rights instruments, including civil and political rights, as well as economic, cultural, and social rights.

Privacy is the single right that is most impacted by current implementations of AI. Other rights that are also significantly impacted by current AI implementations include the rights to equality, free expression, association, assembly, and work. Regrettably, the impact of AI on these rights has been more negative than positive to date.

The positive and negative impacts of AI on human rights are not distributed equally throughout society. Some individuals and groups are affected more strongly than others, whether negatively or positively. And at times, certain AI implementations can positively impact the enjoyment of a human right by some while adversely impacting it for others.

Addressing Impacts

Addressing the human rights impacts of AI is challenging because these systems can be accurate and unfair at the same time. Accurate data can embed deep-seated injustices that, when fed into AI systems, produce unfair results. This problem can only be addressed through the conscious efforts of AI systems designers, end users, and ultimately of governments, too.

Many of the existing formal and informal institutions that govern various fields of social endeavor are ill-suited to addressing the challenges posed by AI. Institutional innovation is needed to ensure the appropriate governance of these technologies and to provide accountability for their inevitable adverse effects.

Slide Courtesy of AI Policy Lab

The Path Forward

Human rights due diligence by businesses can help avoid many of the adverse human rights impacts of AI.

Non-state grievance and remedy mechanisms can provide effective redress for some, but by no means all, of the inevitable adverse impacts that AI will produce.

Governments have an important role to play in creating effective mechanisms to remedy the adverse human rights impacts of AI.

The role of government is essential to addressing the distributive consequences of AI by means of the democratic process.

Slide Courtesy of AI Policy Lab

Why and How Investors Should Act on Human Rights
Principles of Responsible Investment

How to respect human rights in investment activities

Institutional investors have a three-part responsibility to respect human rights:

1. policy commitment;
2. due diligence processes;
3. enabling or providing access to remedy.

POLICY	DUE DILIGENCE PROCESSES				ACCESS TO REMEDY
Adopt a policy commitment to respect internationally recognised human rights	Identify actual and potential negative outcomes for people, arising from investees	Prevent and mitigate the actual and potential negative outcomes identified	Track ongoing management of human rights outcomes	Communicate to clients, beneficiaries, affected stakeholders and publicly about outcomes, and the actions take	Enable or provide access to remedy

Slide Courtesy of AI Policy Lab

The PRI will:

- support institutional investors with their implementation of the UNGPs through knowledge-sharing, examples and other practical materials;
- increase accountability among signatories, by introducing human rights questions into the PRI Reporting Framework – initially on a voluntary basis;
- facilitate investor collaboration to address industry challenges to implementing respect for human rights;
- promote policy measures that enable investors and investees to manage human rights issues;
- drive meaningful data that allows investors to manage risks to people.

Slide Courtesy of AI Policy Lab

Taking Stock of Investor Implementation of the UN Guiding Principles on Business and Human Rights

UN Human Rights Special Procedures

- The efforts to promote implementation of the UNGPs to date have enabled broader levels of participation from a wider range of stakeholders, challenging them, but also bringing them together to learn from each other and to generate the diversity of responses that the complex nature of business and human rights requires.

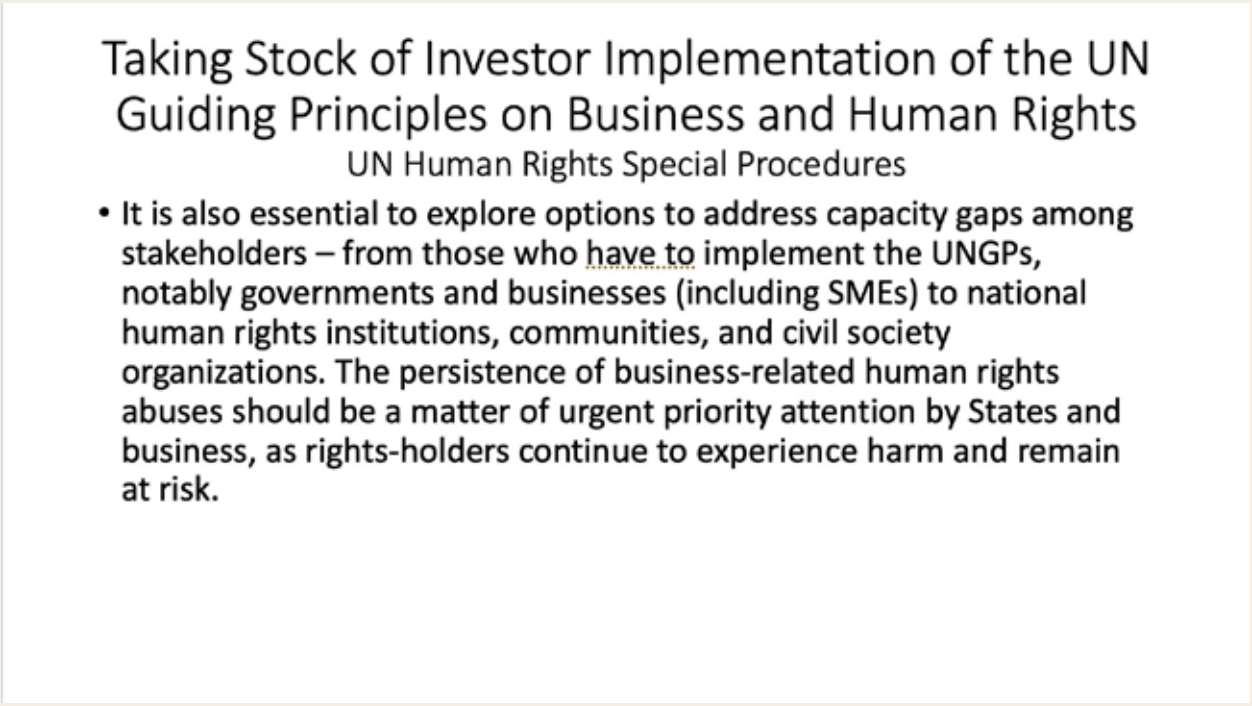
- Yet, more action by more States, including making full use of the available tools beyond voluntary measures, and more action by more businesses, including among small and medium-sized enterprises (SMEs) that are challenged by limited resources and few practical tailored tools, is still a matter of urgency.

Slide Courtesy of AI Policy Lab

- Looking ahead, the mandatory human rights due diligence wave and the increasing focus on effective regulation offer opportunities and drivers. The onus is on States to develop effective laws and regulations, but also to use the wider range of policy tools – a "smart mix" – to incentivize responsible business and due diligence. In parallel, there is a need to leverage the financial sector and the momentum of the increasing focus of investors on environmental, social, and governance (ESG) factors.

- Implement business models that do not undermine human rights, for example in the context of supply chains, purchasing practices, and in digital technology. This includes the development of more systematic peer learning and accountability platforms for State implementation, supported by reinforcing efforts at regional level, as well as better tracking, including through enhancing the role of the Universal Periodic Review and the UN Forum on Business and Human Rights.

Slide Courtesy of AI Policy Lab



Slide Courtesy of AI Policy Lab

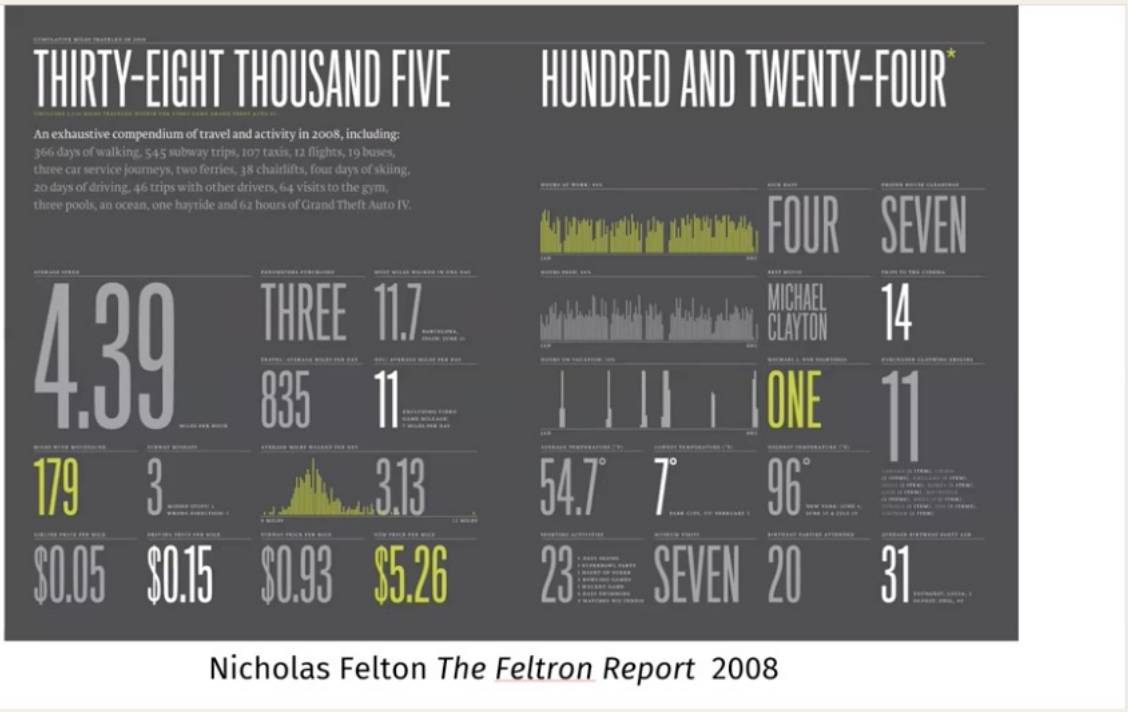
C. INTERSECTIONAL APPROACH

Portrait of a Man Dieric Bouts 1462

Portrait of Isabel Goldsmith Steve Miller 1993

Slide Courtesy of
Judith Donath

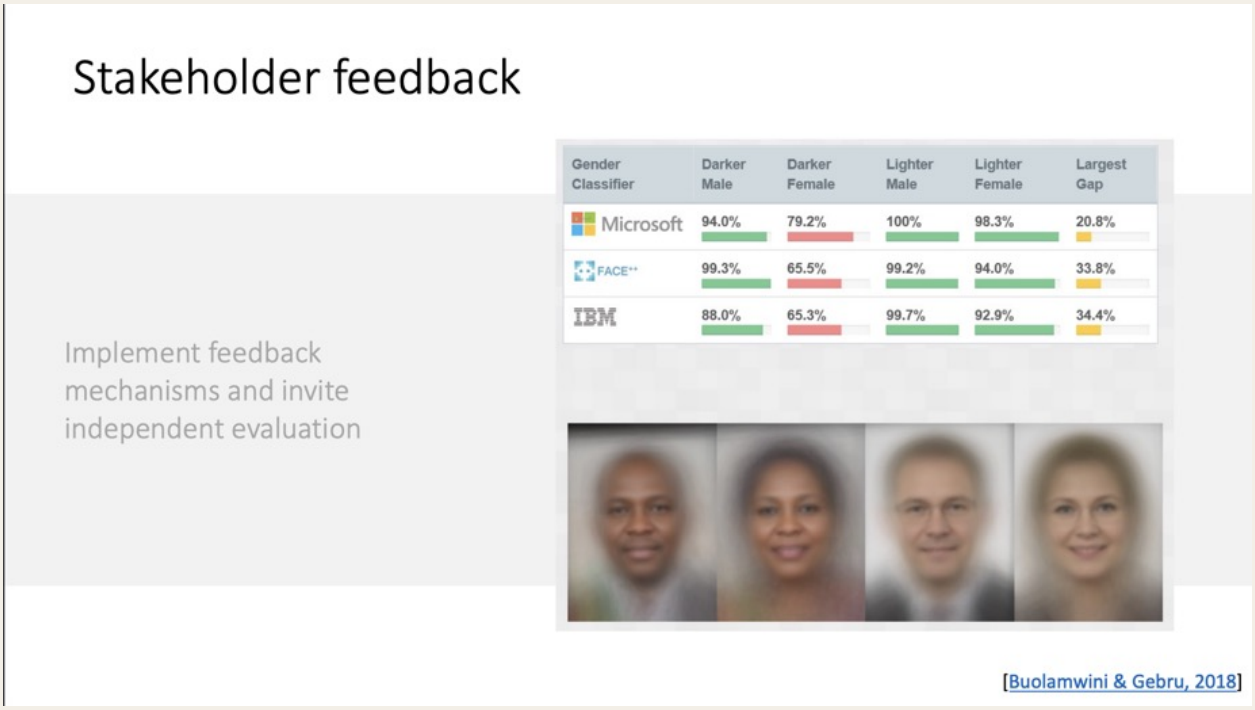
"We think of a portrait as being the light glancing off your facial structure, and we glean a tremendous amount of information from that. But here we can have anything. DNA, in theory, should be the most expressive portrait because it's all the instructions that how to make you. But this painting, assuming there you know, there is some correspondence in the painting to the DNA data, it's completely opaque to us. We can't make any sense of it. So that's the other big tension that comes in with thinking about how you portray people based on data is how do you translate the data into the kind of cue that the audience is going to use to take, but are basically their existing stereotypes and ideas of what people are and expand it in their mind into an impression of that person?" - *Judith Donath*



Nicholas Felton *The Feltron Report* 2008

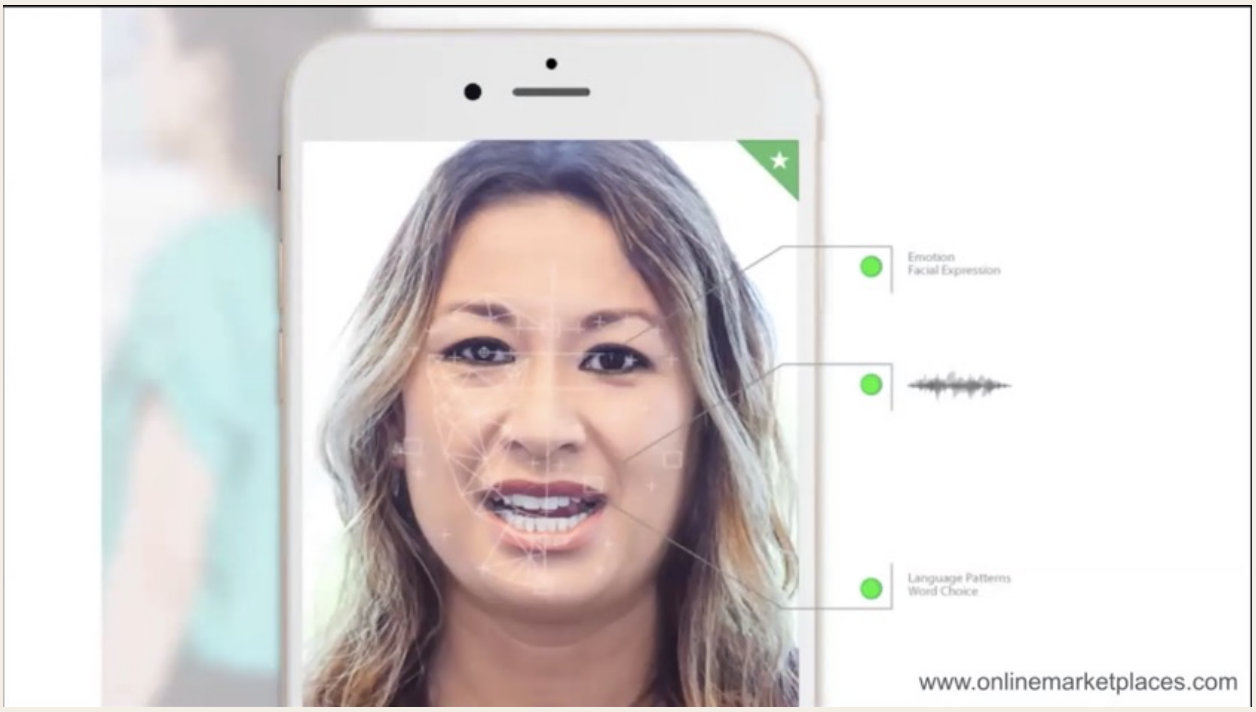
Slide Courtesy of Judith Donath

"We started doing this in 2008. [The Felton report portraits] were kind of a parody of a corporate annual report, so it's called the Felton Annual Report, but he hand-picked statistics over a period of a year in his life…And then he moved into something a little bit more like a quantified self version where he automated this because by 2012 or so, the devices that we have had the ability to record so much of what you were doing, he could basically spend a year recording the weather every day and having all the photos he took be analyzed and how much time he spent doing different things. And then the later reports are actually computer generated and they have all these kind of details and they're just boring because they are kind of like, OK, he's my heart rate over a year. Here's my weight over a year. Here's all the number of hours I slept every night. So by offloading that sort of personal picking out of the data, it's much more algorithmic. But it also stops actually giving us any sense of him as a person. And so I think it's an important lesson for anyone thinking about how we think about data and representing others because right now there's such a temptation to do everything algorithmically without recognizing the expressiveness of that role of even if you're looking at a huge amount of of computer generated data of human being as artist as opposed to the algorithm as the artist in choose what data to represent." -*Judith Donath*



Slide Courtesy of
Mehrnoosh Sameki

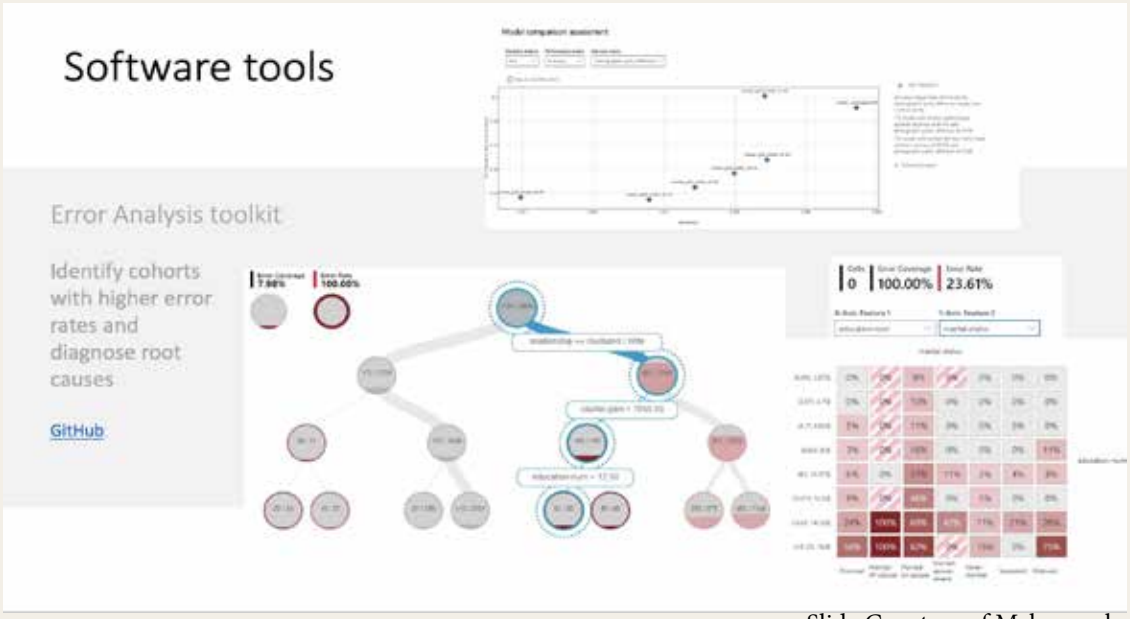
"Nowadays, all the companies, including Microsoft are quite passionate about diversity and bringing diverse voices…In conversations that stakeholders have around the ideation and design and implementation and deployment of an AI system, it's quite important to make sure that you are receiving and responding to that diverse group of stakeholder's feedback including their complaints, including their concerns. If they're only a certain ethnicity, gender, making all these decisions for any AI system, I won't be shocked that AI will just build on the unfairness and bias that exist in the society or even make it worse."- *Mehrnoosh Sameki*



Slide Courtesy of
Manish Rhaghavan

"You may have seen examples of things the purport to be hiring algorithms that will do things like analyze your emotion, your facial expression, the words that you use, the cadence of your speech, so on, and produce some evaluation of you as a candidate…[along with] the eye-catching headlines you may have seen…come a number of stories that talk about the potential of bias that these algorithms have. So hiring algorithms are going to replicate all the biases that humans have. That's the perception and that's what we're worried about. On the flip side, you also have people who claim to have strategies to actually remove those biases and to make recruitment more fair and so on." - *Manish Rhagavan*

D. DATA-DRIVEN APPROACH



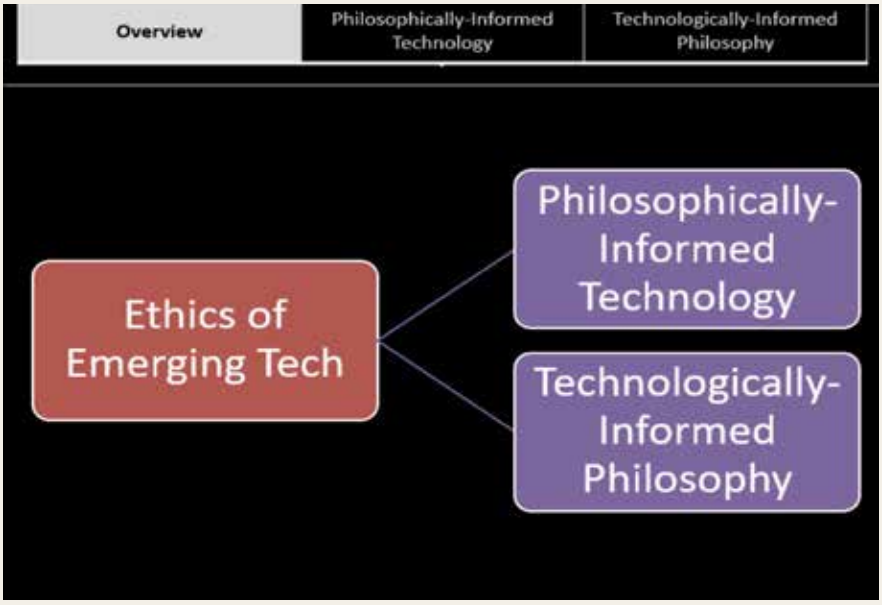
Slide Courtesy of Mehrnoosh Sameki

"One of the algorithms that we have is an algorithm that can allow you to observe how the model has made its predictions, then accept a fairness criteria in mind. For instance, in this particular case, I might say this is the same loan scenario across my females and males. I can see that 18.8% more men are getting the loan compared to females. So out of all the men, 25% of them are getting the loan. For females, 7% of them are getting the loan. And there's a disparity of 18.8% which is the difference between these two numbers. Imagine you cannot tolerate that…I [can] call one of the algorithms of this unfairness mitigation algorithm that Fairlean supports…This is the machine learning model that I have trained–but also take a fairness constraint as another input." -*Mehrnoosh Sameki*

	Biomedical Ethics	AI Ethics
Specific guidelines for practitioners	✓	✗
Robust Oversight Infrastructure	✓	✗
Accountability Frameworks	✓	✗
Research Programs and Info Exchange	Robust	Growing
Common Language	✓	✗
Interdisciplinary Specialists	✓	✗

Slide Courtesy of John Basl

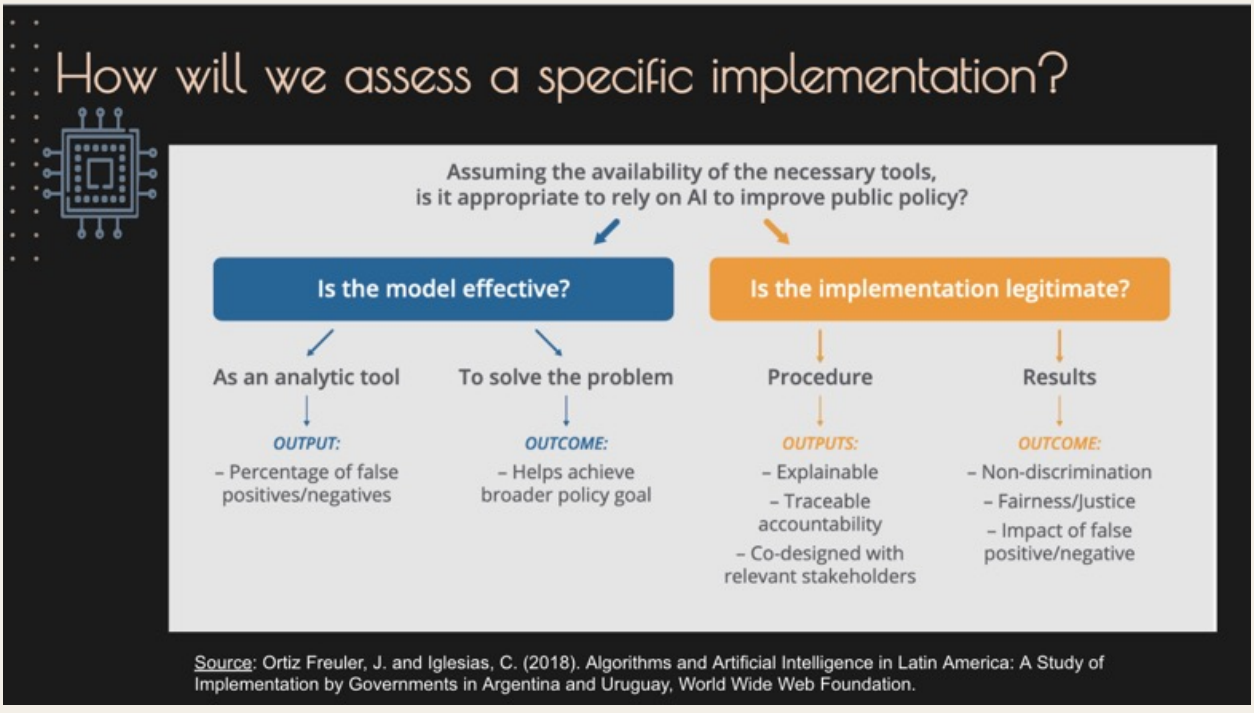
"There is a significant lack of what I'm going to call ethnical capacity or a lack of an ethics ecosystem for AI. If we compare this to, for example, biomedical ethics, we can see a very stark contrast. On this slide, I have a table comparing that state of biomedical ethics to AI ethics. We have a robust oversight infrastructure. We have IRB's, we have animal care and use committees in the case of animal research. We have accountability frameworks. We have research programs, there's a whole field of research in bioethics, and libraries of case studies and tools for exchanging information. There's a shared language–a common language that practitioners and ethicists speak to one another so they can understand and translate between one another. And we have interdisciplinary specialists, not just the four people in this room, or the 10 people in the world that are interdisciplinary specialists in AI ethics, but a whole field that's just trained up to manage this and cultivate this and deploy this ethics ecosystem. This is not to say biomedical ethics is perfect. And Kendra alluded to some of the problems embedded in biomedical practice that aren't adequately addressed by the current ecosystem. But compared to our current ethics ecosystem for AI, it is mature, it's robust, it's action guiding, and there are significant resources for translating across disciplines to grapple with problems. Now the lack of an ethics ecosystem wouldn't be so bad in the case of AI if the resources from other areas could serve as off the shelf tools for addressing challenges in AI. And this does happen. For example, people have grabbed onto the tool of informed consent from bioethics and applied that in the tech sphere and you have this notice and consent procedure. But notice and consent is an extremely poor analog. It completely fails to serve the purposes that informed consent procedures are designed to achieve…Even though we lack a robust ecosystem in the case of AI, there's still a wide recognition that applications of AI face deep challenges–especially concerns about fairness and bias. That's the most prominent concern in the public and in the literature. There are attempts to address these challenges; however the solution space, especially with computer science, is oriented towards, for example, fairness metrics. So the standard picture of the course of, and solution to issues of fairness is that, bias comes from the data, and is to be addressed in model creation."
-John Basl



Slide Courtesy of John Basl

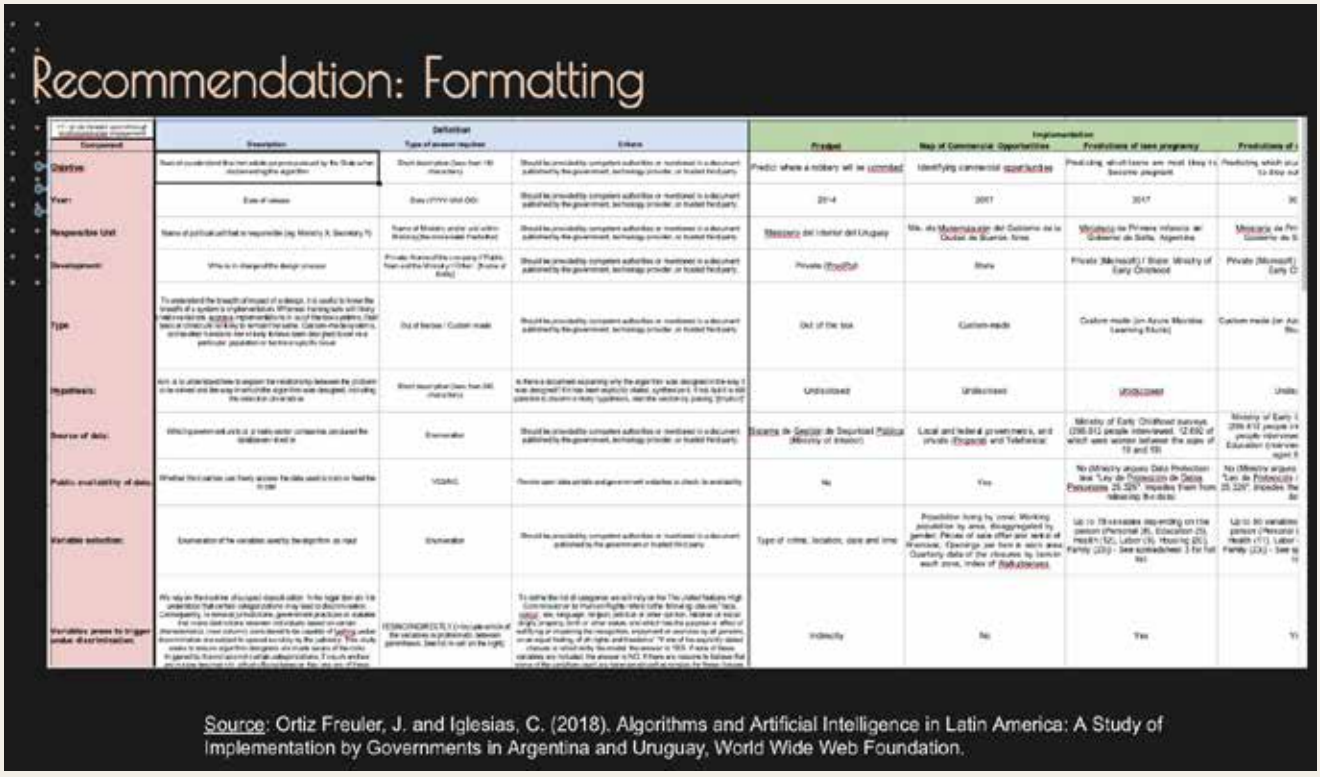
"Philosophers can draw on technology to engage in new research programs to solve fundamental problems to then put back in the toolkit to then do philosophically informed technology…On the philosophically informed technology side, much of my work has been building scaffolding that helps other to draw on the expertise from moral philosophy and other disciplines." -*John Basl*

E. BRIDGE BUILDING



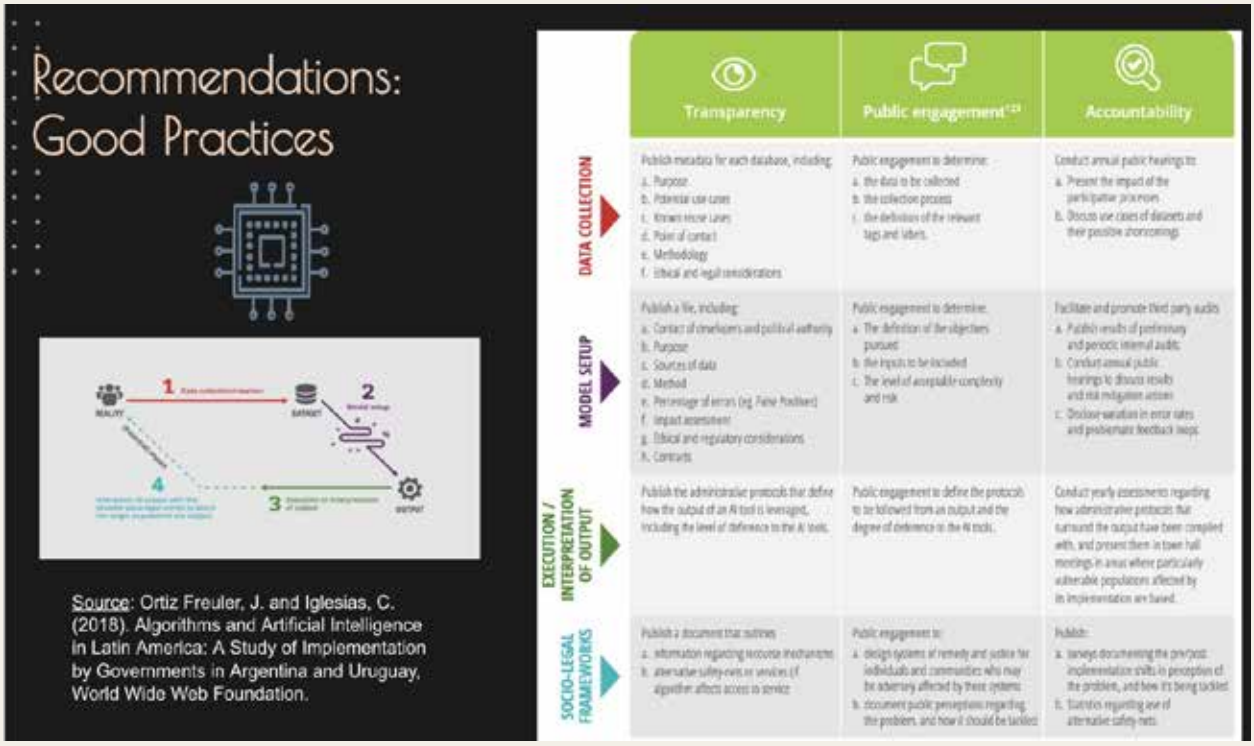
Slide Courtesy of Juan Ortiz Fueler

"So how will we assess a specific implementation? This is not something or which I think anyone would have a complete answer, but one of the things we are suggesting that government officials take into account is splitting this into two parts. First, are these models effective either as an analytic tool basically to help them define what types of issues they're facing within their government, of what types of issues their constituents are facing? And for this, we might want to look at the percentage of false positives and false negatives to understand, is it actually telling me something about the communities I am trying to engage with but also to solve a problem they might want to try to understand whether or not it might help achieve a broader policy goal that was defined by the political leadership and separate. We have to understand whether it's an implementation that is legitimate." -*Juan Ortiz Fueler*



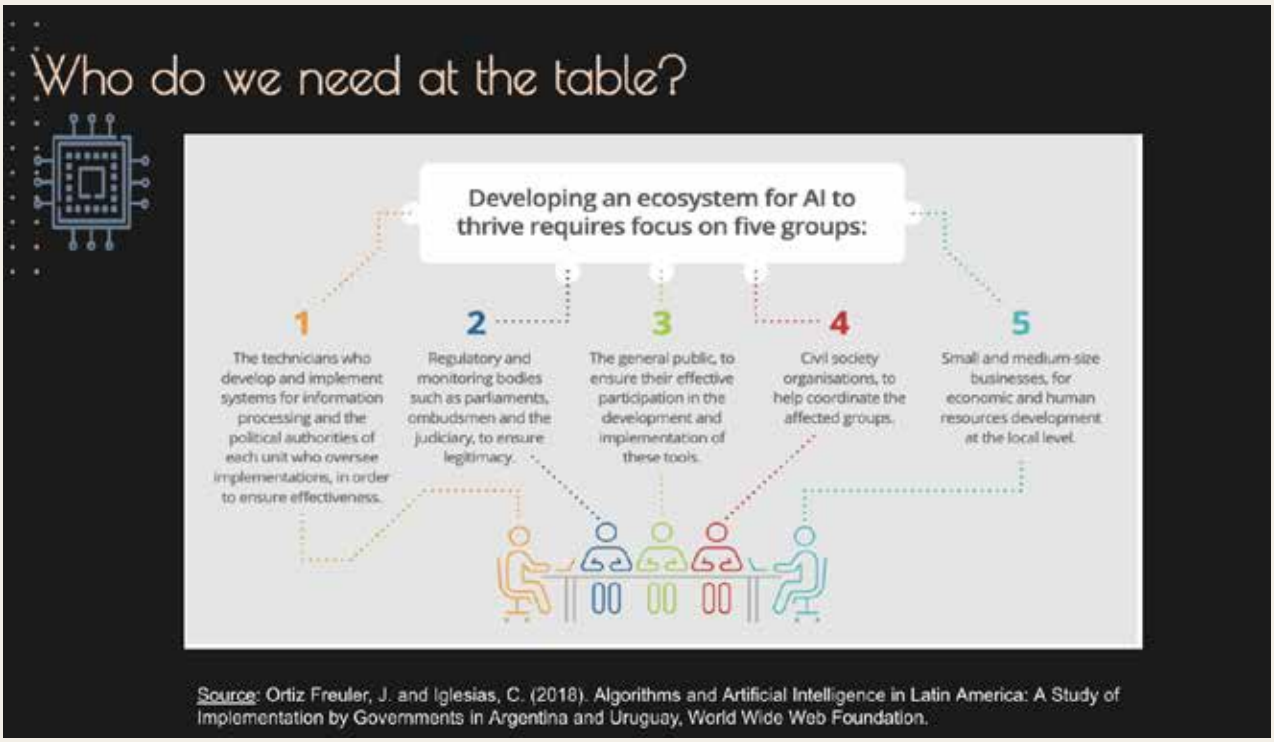
Slide Courtesy of Juan Ortiz Fueler

"In terms of the recommendations in the report, one was that governments start collecting data on these types of models that are being used and start sharing it across the region so that they can negotiate better with the company providers and they can design tenders that allow them to find the right fit…Many of these providers are global whereas the government officials are hyper local and so that there is an information asymmetry that might lead them to make not as good decisions as they could or should be making in terms of more specific recommendations."-*Juan Ortiz Fueler*



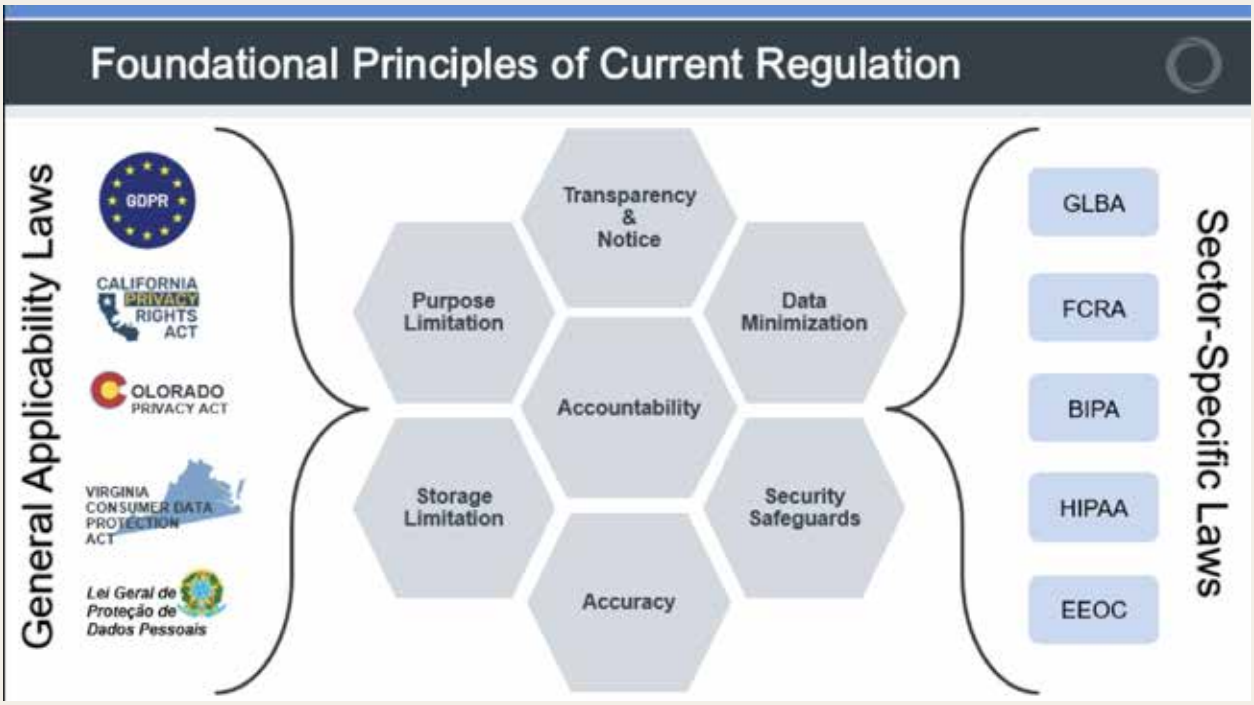
Slide Courtesy of Juan Ortiz Fueler

"We mentioned a couple of good practices that could be implemented in terms of transparency, public engagement, and accountability for each of these four phases that we mentioned. And so, for example, in terms of data collection and transparency, as I was mentioning that they should publish the metadata for each database including what is the purpose of the database that they have used? What are the potential use cases? What are the known reuse cases? Who's the point of contact? How is that data collected? So what is the methodology that they should start thinking from the first phase, what are the ethical and legal considerations that were taken into account? We're saying this because we understand that this si a system that has several points and that oftentimes the people who are operating on each of these steps are not necessarily connected. And so creating a system of accountability requires that they start sharing more information that they would otherwise share so that actors can look at that information that they would otherwise share so that other actors can look at that information."
- *Juan Ortiz Fueler*



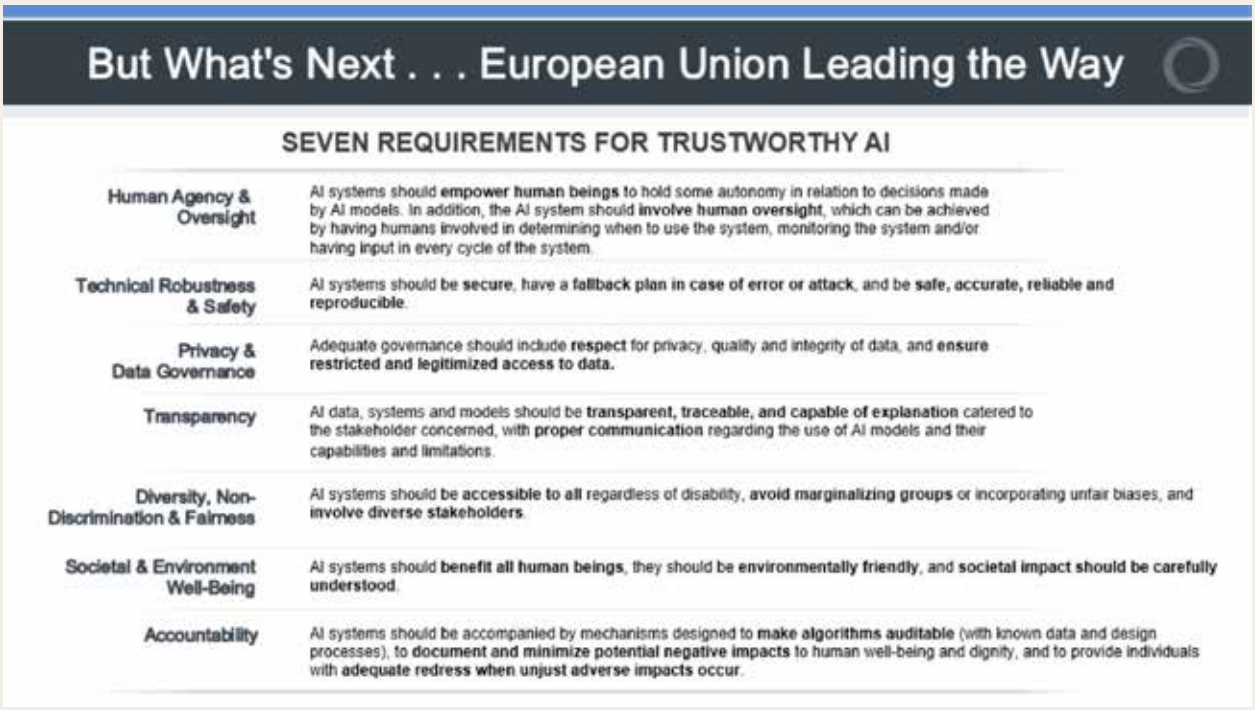
Slide Courtesy of Juan Ortiz Fueler

"We're telling governments that they should try to open up some of these discussions because we're at such an initial phase. And so they need to set up multi-stakeholder tables where they can discuss how these systems are going to be implemented and when and why they should be implemented. And what are the potential consequences in terms of the datafication report that I was mentioning, we're also concerned about how this might lead to a change in worldview in the coming years where the complete relationship with the gap between the government and the people between the bureaucrats and the technocrats. So this addresses kind of the first question of how the algorithms impact governance. We think it's a massive shift in the relationships take place across all of these stakeholders. And so in terms of how can policymakers assess whether or not to adopt and deploy an automated system? This is a values question. And so one of the big things that each team should be addressing and should disclose whenever they implement some of these systems is if they're focusing on the processes on the outcomes and what are the underlying frameworks that they're relying on? Lastly, what challenges might be distinct for a region like Latin America?" - *Juan Ortiz Fueler*



Slide Courtesy of Heather Egan Sussman and Ryan McKenney

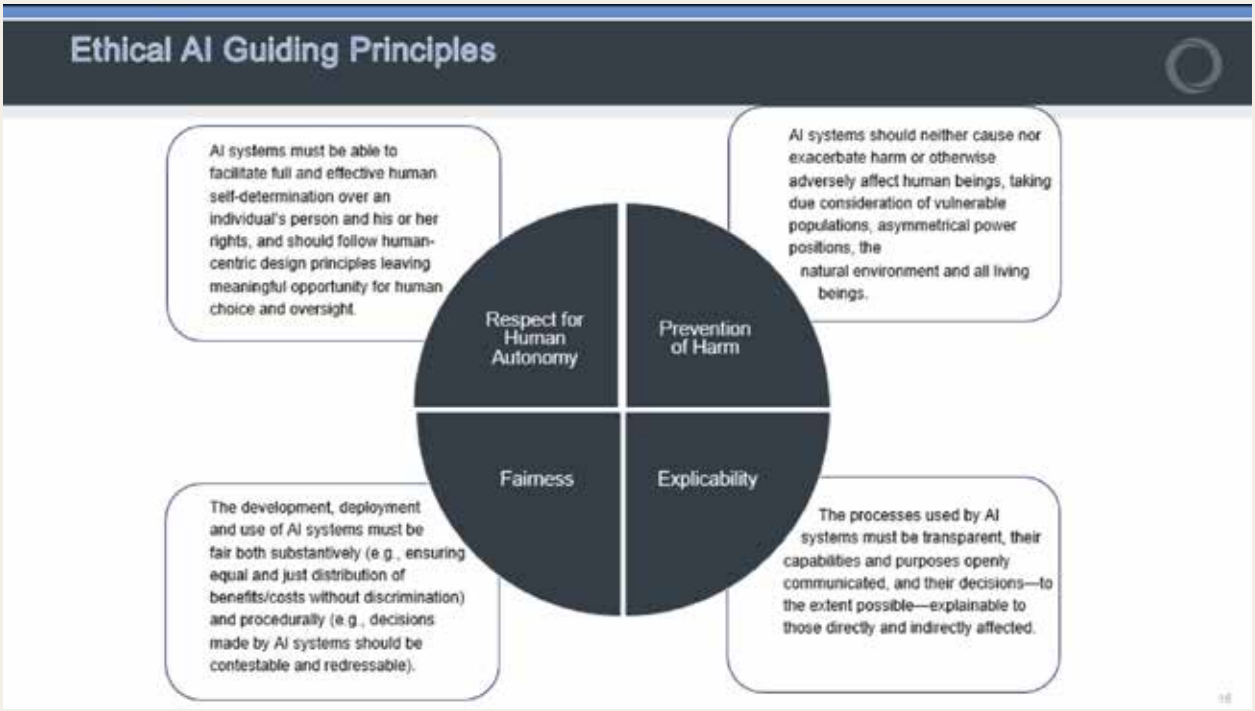
"We've raised the foundational principles of the current regulation that we have the slide on the left earlier that showed fair information practice principles, the one on the right. It talked about how to maximize AI. So just trying to ground this concept of regulation of AI in what we're currently seeing. We've got the potential for generally applicable laws like privacy law, right? These focus on the collection of personal information and personal data. So that necessarily isn't going to be protecting against these issues of interference with autonomy or intrusion upon seclusion and personal privacy. What you end up with is there are some challenges with the current way in which we're regulated. And yes, we have general consumer protection and general privacy laws. And yes we have this concept of fair information practice principles and we have sector specific laws. But if you can think about all the gaps in between all this space here at which I could potentially operate without regulation is something where we're going to see some of the future development of models." - *Heather Egan Sussman*



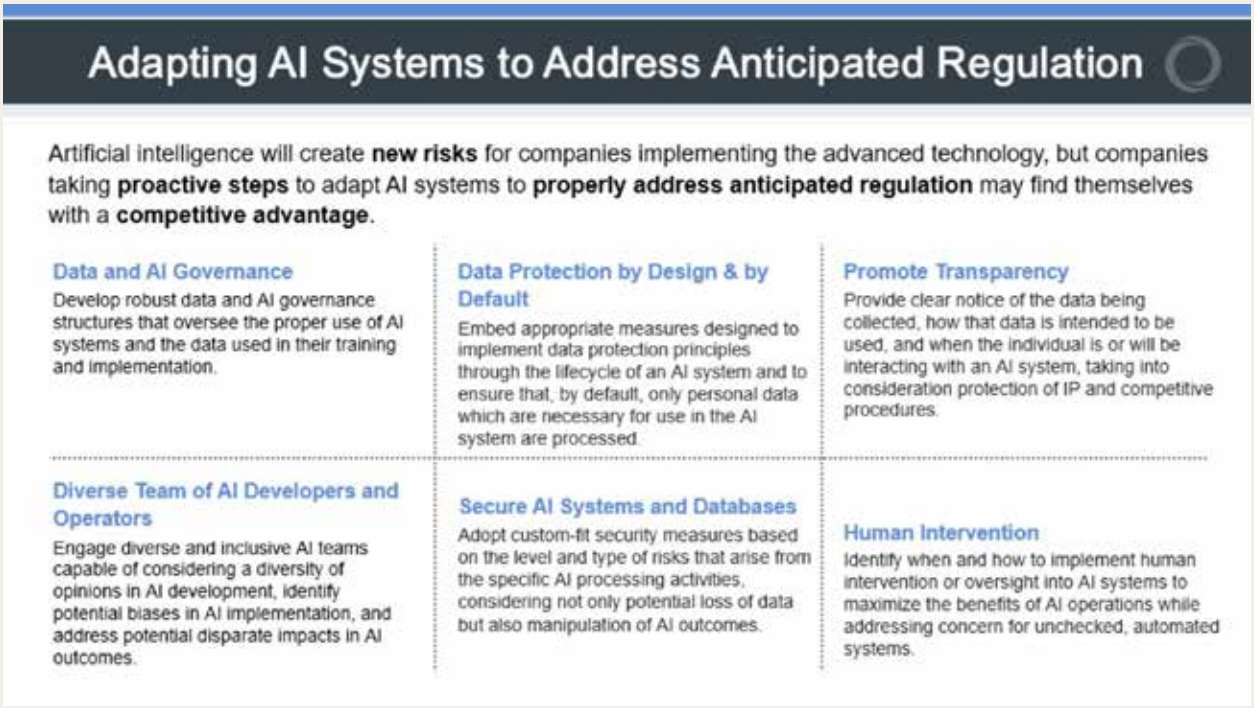
Slide Courtesy of Heather Egan Sussman and Ryan McKenney

"The European Union is really leading the way. We're the focus. Instead of this concept of notice, consent retention and use restriction is really more around trustworthiness. I write seven requirements for trustworthy AI and it's this concept of having human-oversight, technical robustness and safety, privacy and data governance transparency…You can see Europe really approaches it from a very different perspective. Of course, also focusing on diversity, nondiscrimination, fairness, societal and environmental well-being, and then ultimately, accountability and the concept of accountability just simply means that you need to be able to demonstrate your ability to comply with these other components. You need to have written policies and procedures and process by which complaints can be received, heard and ultimately addressed. When you find that something you've undertaken a business practice that may be in violation of one of the preceding components of this regulation…We've got another interesting concept under European law that not all AI is created equally and you've got low risk and higher risk and the types of AI that should just be outright prohibited."
- *Heather Egan Sussman*

"This concept of guiding principles with ethical AI, what I find really interesting is the practice of the work that we do is how many different companies have all different phases of the supply chain of tech in innovation and in the lifecycle of development of AI are still in a relatively massive stage of trying to implement any one of these components. And what does it look like? What experience? How do we determine fair? How are we measuring what we're going and then auditing, how are we making adjustments as we move forward? But this concept of respect for human autonomy, the prevention of harm, the ability, explainability and then fairness guiding ethical treatment and its throughout the lifecycle throughout." - *Heather Egan Sussman*

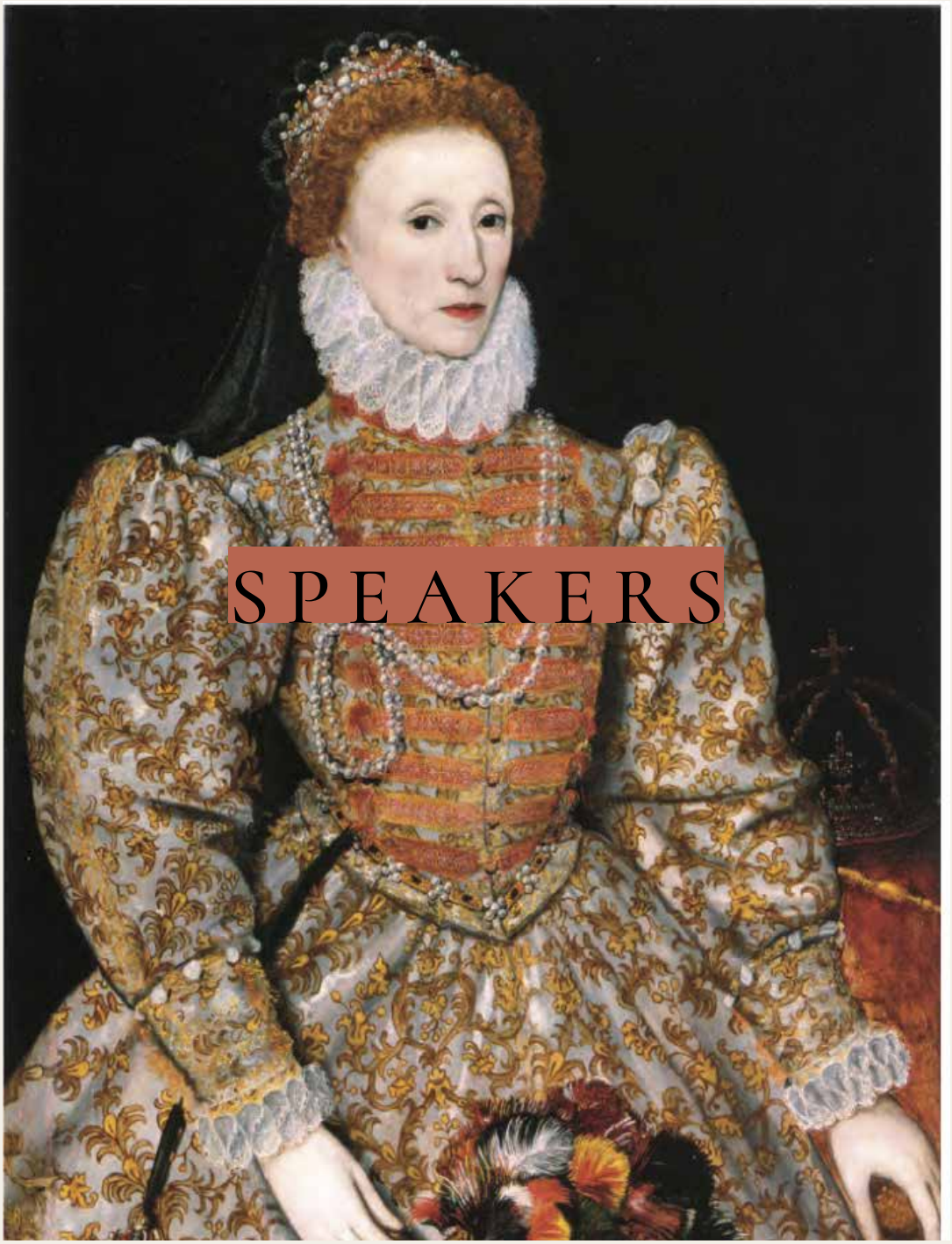


Slide Courtesy of Heather Egan Sussman and Ryan McKenney



Slide Courtesy of Heather Egan Sussman and Ryan McKenney

"This is a slide about adapting your AI systems to the current regulations as well as the anticipated future regulation that will likely follow that European model that Heather discussed earlier and is developing robust data. Obviously, data is the oil of the 21st century. Most companies businesses are built on data, but when you're building algorithms and using AI and ML, you've got to do it on consumer data and make sure that you're properly using data and that you're doing what you say you're doing with that data. Promoting transparency, ensuring that there's human oversight. Ensuring that there's lots of audits of your systems continuously and also that your compliance and your legal team…Compliance and your legal teams are not siloed that they're working with the engineers in the AI and ML folks who are building that model in a collaborative environment, working across borders, across teams and across disciplines to build better organizations that are thinking about data in a long term manner. They're not thinking about just what we need to use this model today. The thing about what are the risk from an AI perspective, from a reputational perspective, ethics, privacy law, all of these different aspects of law and regulation, thinking of how they come together and protecting companies from long term risk and also making sure that you're not taking advantage of customers, that there aren't unfair and deceptive practice in your business and that you're securing that data against the possibility of cyber attacks. That all goes to show you just have to continue to monitor, audit and ensure that there's human intervention over these animal models." - *Heather Egan Sussman*



SPEAKERS

Queen Elizabeth I (The Darnley Portrait), artist unknown, 1575.

Source: Data Portraits by Judith Donath

SPEAKER EXCERPTS

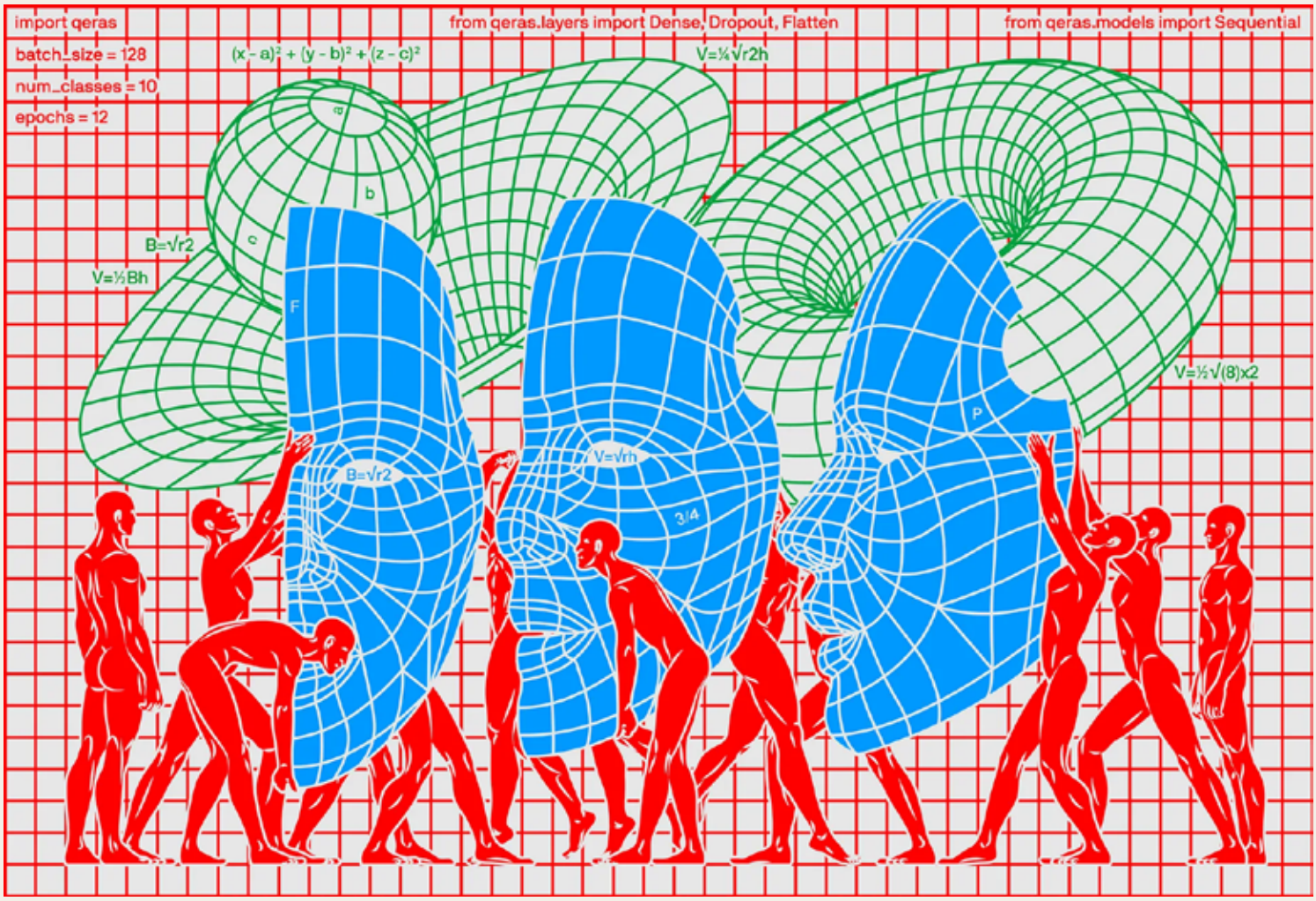


Illustration by Berke Yazicioglu

"The more open and available and accessible data is, the more we have the ability to address issues of bias."- Heather Sussman

"Fairness, as a non-scalable good is a feature, not a bug."- Professor Sandra Wachter

"There are six or twenty-one different types of fairness. Machine learning has made that very clear. Now, which is the best type? That's a really important question. It's one you have to face when you're dealing with machine learning, but it is not a question that fairness itself can answer. That is no answer in fairness. Considering fairness will not tell you which type of fairness is best in this situation, which means that fairness is not enough to ask, which is a very hard and difficult problem. That problem is not made by machine learning, but machine learning makes it really clear."- David Weinberger

"Bias comes into play the moment you try to boil down a social phenomenon–or really anything else to do with a human who has their own agency–into structured data."- Alexandra Pavliuc

"When people train AI systems, that data sets are often based on what they know, what their experience is, and this is one of the reasons, for example, when it comes to facial recognition, the training data sets sometimes overwhelmingly favor white faces. As a result, the facial recognition algorithms frequently fail when presented with faces that are not white." - Craig Newmark

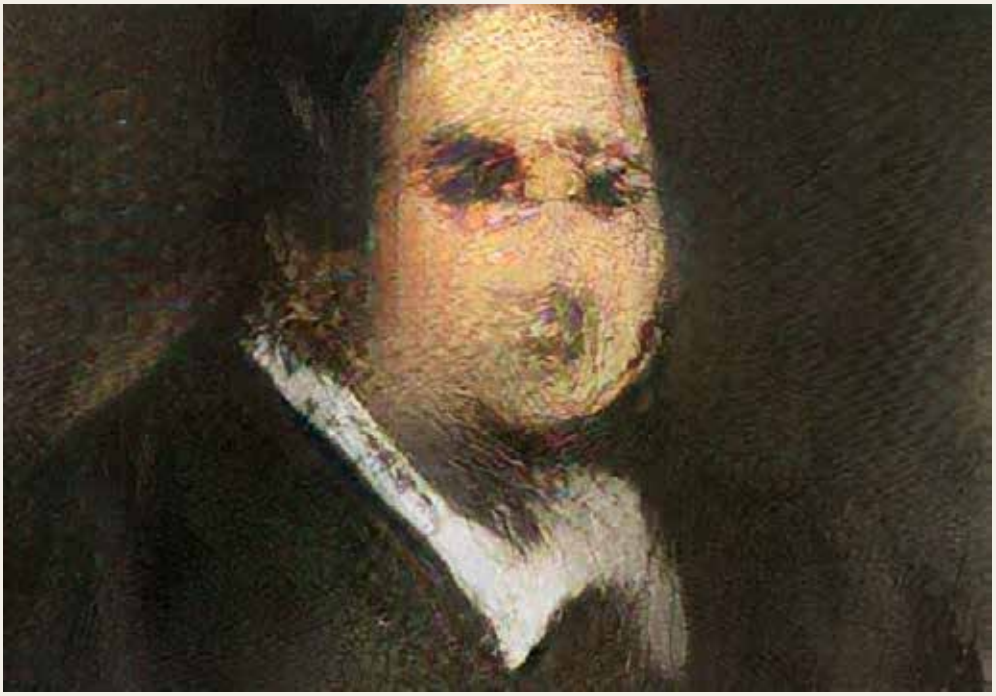
"If any of us who are diverse–if our kids went into these tech firms that I'm talking about right now, like AirBnB–the chance of promotion is 50% relative to a white woman–not even a white man."

- Rati Thanawala

"As women, we should embrace technology. We should really not show–men, they tend to say it's very complicated. It's really dangerous. It will set you off course. But it's because they feel threatened in their own leadership position. So the discussion around AI and bias is really fundamental, as it could contribute to a significant backlash in terms of diversity." - Ambassador Veronique Haller

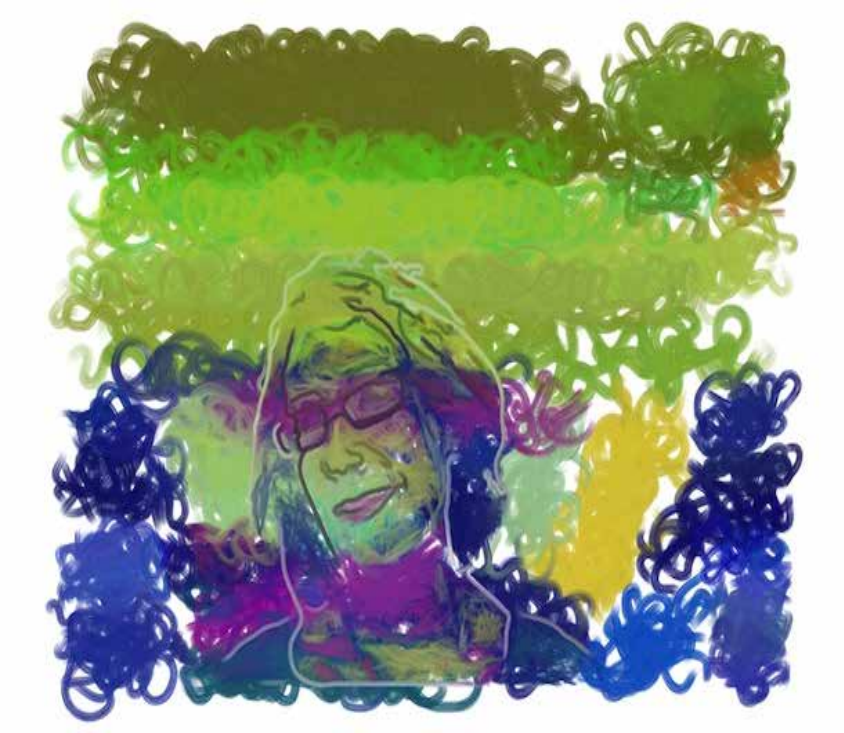
"There is a significant lack of what I'm going to call ethical capacity or a lack of an ethics ecosystem for AI."- John Basl

"In my case, even being able to connect with someone like Joy or Timnit as just immediate support or immediate allies has been really instrumental, like even as emotional support. It can even be peers that could provide emotional support as you're going through some of these situations and pushback. I found that to be sort of a super-critical component of surviving that kind of attack on your work or on your character or personhood in response to kind of doing this very vulnerable audit work." – Deborah Raji



Source: The first piece of AI-generated art to come to auction, Christie's. Portrait of Edmond Belamy, 2018, created by GAN (Generative Adversarial Network). Sold for $432,500 on October 25, 2018.

"We have huge areas of morality that are gray and we see the law flipping back and forth about what is and is not permitted. Even in this question about speech online, there is a very popular, in the US, line of morality and ethics that says the cure for bad speech is more speech. We should never be cutting off speech. We should be combating misinformation with correct information, combating hate speech with education about the need for diversity in our communities. But there's an equally strong line that says, well, look harm is harm and we really need to do something about that. We can't always wait for the marketplace of ideas to sort itself out." - Mason Kortz



By "The Painting Fool"

"These systems are much less accurate on darker-skinned individuals and, specifically, darker-skinned women."- Deborah Raji

"AI is just patterns and data. It is an assistance to decision-making. It makes us more organized. It gives us strengths, but only if we don't rely on it completely."- Ambassador Veronique Haller

"The idea of achieving substantive equality is about finding way to uncover hidden talent that you didn't know existed."- Professor Sandra Wachter

SPEAKER EXCERPTS

"Everyone wants to hire diverse talent, but, even in AI, they are saying we can't find enough diverse talent. That is not true…Tel Aviv has the largest diverse talent. One-third of the people in Israel in Tel Aviv for AI are women. Why don't you hire there? First of all, even though you're focused on hiring, you're not hiring, let's say, AI talent from urban areas outside of San Francisco. So I know where the 50 hubs are. So does everyone else."
- Rati Thanawala

Simon Colton

"This could be an opportunity for women to break silos and come up with their style of leadership, which is thus more adapted to what one needs in the digital transformations and when I look at what's happening in government, in international organization, they do really risk being left being, trying to impose and run after regulations, if they do remain rigid in their traditional structures. And I bring one more thought that is connected with what we've seen in the last two years with the COVID crisis, the transfer of meetings from a traditional table to Zoom…I believe this is a great opportunity for women to circumvent some of the bias that are connected with the authority that is given to the one sitting at the top of the table and on Zoom."
-Ambassador Veronique Haller

"The most important thing here is to make sure to identify the fairness related harms that are applicable to your AI…there are harms of allocation, quality of service, stereotyping, over or underrepresentation, denigration, et cetera, so that you can anticipate the most common causes of them and try to avoid them as much as possible." - Dr. Mehrnoosh Sameki

"It's important to distinguish between what is a best practice in the sense of being ethical or mitigating harm and what is legally required because those were not always the same thing." - Mason Kortz

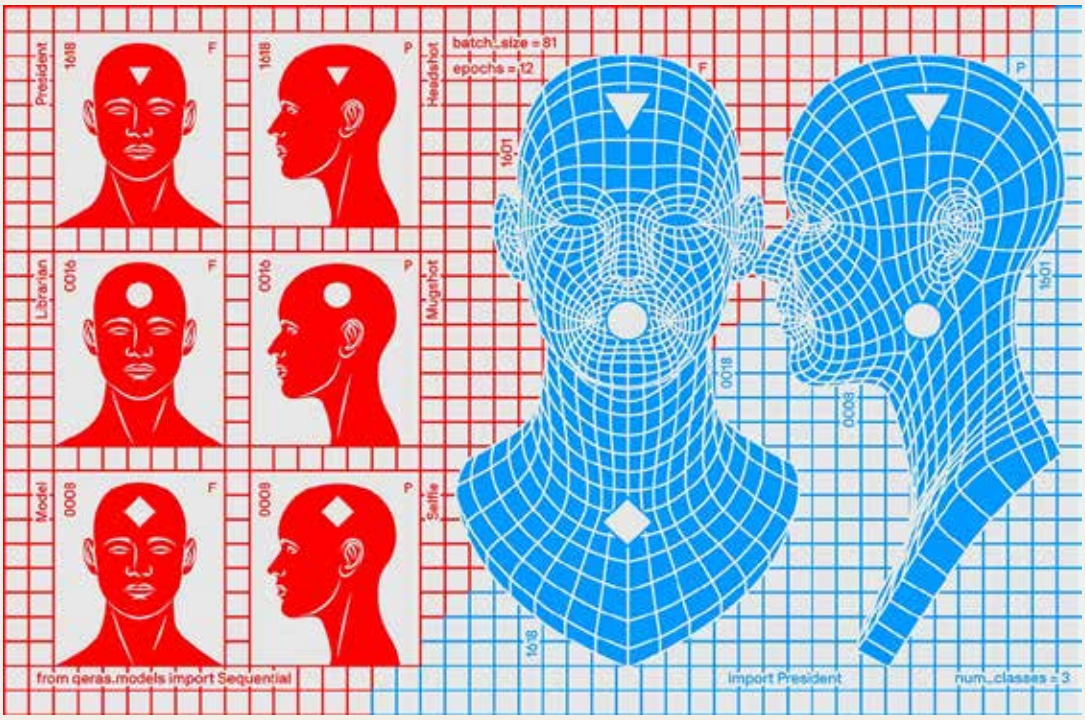


Illustration by Berke Yazicioglu

"Fairness in AI is about uncovering hidden talent."
– Professor Sandra Wachter

"I observed that on many, many meetings, that you have with other government on Zoom, is that the content of what you're saying matters often more than where you sit."
-Ambassdor Veronique Haller

"It's not a good idea to accept the fact that those [AI] systems are complete black boxes and embedded in our society and they make a lot of important decisions. The tool that we have developed give you a way of understanding something that is going on inside the black box that is meaningful for the individual."- Professor Sandra Wachter

"We can see how bias can be introduced, what effect that can have for populations for patients and we can also see how policymakers in countries try to address this issue and how they address it differently."
- Dr. Kerstin Noelle Vokinger

"We talk about transparency and bias, but a lot of the systems are just outright discriminatory in ways that are illegal. In the human brain, we know our social institutions aren't unbiased either when you're got humans running them, but now that we have software systems and algorithms, we're talking about bias or unintended bias instead of actual illegal discriminatory results." - Mitchell Baker

"We look to others to both validate our specific tastes and to inspire us with new tastes. Predominantly, we've looked at tastemakers to provide us with trusted recommendations. But part of the problem is that curation remains invisible labor on the internet." – Jad Esber

"It just goes without saying that the models really do reflect the data that's used to train them, and the data reflects real biased models likely to reflect the same bias."- Heather Sussman

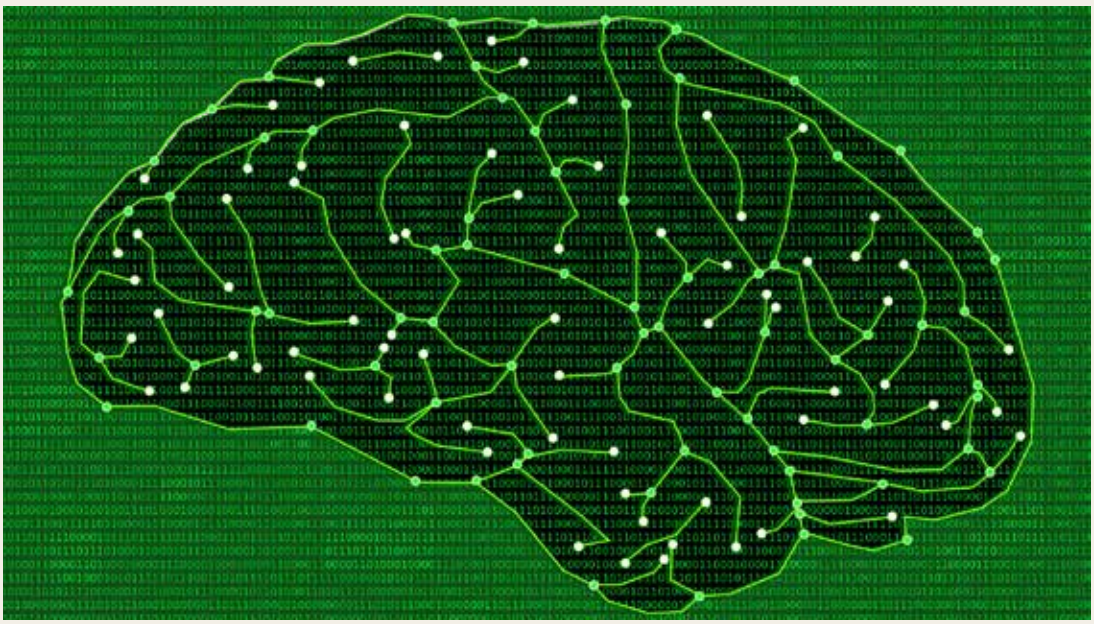
"I'm not in favor of compromising on fairness, but because fairness is so complex it often requires painful trade-offs."- David Weinberger

"Fairness and equality changes the whole lens in which you're looking at the problem." - Dr. Padmashree Gehl Sampath

Sougwen Chung, Putting the Art in Artificial Intelligence

"It's sort of a remarkable thing that machine learning is capable of doing, which is why there's such enormous interest in it, and why everyone in the tech industry is sort of looking for different ways to use AI. The trick is, often, we are using bad data and, often, there is no good data." - Ethan Zuckerman

"It has been more challenging being a parent in STEM. It's been extremely challenging because there's a way in which our society is just kind of set up expecting caregiving to happen somehow, right? Of course, I don't get taken care of, probably by your spouse. For example, the fact that our funding systems were doing this type of research involve extremely time intensive processes. And so when you have less time because you're spending time on caregiving as well as on your professional work, these things eat away." - Professor Finale Doshi Valez



"Women are half of the population in the world. We should also occupy at least half of the leadership roles in the world. Unfortunately, we don't." -Secretary Patricia Ellen

"Without a clear framework and an accountability framework, AI can worsen the historical biases and enhance inequality."- Dr. Padmashree Gehl Sampath

Medieval art was not concerned with creating an illusion of reality; the role of the image was to remind the viewer of the symbolic significance of the depicted people and objects, rather than their appearance. Spiritual relationship, not the demands of perspective or cohesive narrative, guided the placement of things (Dunning 1991). Bartolo di Fredi's panel (figure 8.1), for example, presents "the Adoration of the Christ Child" and "the Annunciation to the Shepherds"—events that took place at different times and places—in a single image; it depicts Mary as larger than the shepherds, showing her greater spiritual significance, not her physical size.

Although the concerns and beliefs of today's online world are vastly different from the mystical, nonscientific worldview of the Middle Ages, art from that distant time can provide inspiration for using the screen as a symbolic space in which to depict relationships, events, and status.

FIGURE 8.1
Bartolo di Fredi, *Nativity and the Adoration of the Shepherds* (14th c.).

Source: Data Portraits by Judith Donath

SPEAKER BIOS

MALAVIKA JAYARAM

- Executive Director of the Digital Asia Hub, Harvard University Berkman Klein Center
- Member of the Extended Programming Committee of Computer, Privacy and Data Protection Conference

JAD ESBER

- Co-founder of Koodos, web3 company.
- Researcher of digital identity, reputation systems, consumer marketplaces, curation and internet culture.

Deborah Raji

Mozilla Fellow. Computer scientist. Activist working on algorithmic bias, AI accountability and algorithmic auditing. Widely acclaimed author. Forbes 30 under 30.

John Basl

Assistant Professor of Philosophy, Northeastern University. Faculty Affiliate of the Berkman Klein Center. Primary Research areas: moral philosophy and applied ethics, especially the ethics of emerging technologies such as AI and synthetic biology.

RATI THANAWALA

- 2018-2020 Harvard Advanced Leadership Fellow
- Former Vice President, Bell Labs; Head of Bell Labs Network Planning, Performance and Economic Analysis

DAVID WEINBERGER

- Fellow, Senior Researcher, Member of Fellows Advisory Board, Harvard University Berkman Klein Center
- Explores the Philosophical Implications of AI

Katya Abazajian

Fellow, Georgetown Beeck Center of Social Impact and Innovation. Affiliate of the Berkman Klein Center. Open government advocate, researcher, and organizer with a multidisciplinary practice. Working to help cities and states use data for stronger local democracies.

Chinmayi Arun

Resident Fellow, Information Society Project, Yale Law School. Founder Director of the Centre for Communication Governance at National Law University Delhi. UN Human Rights Officer, and member of the UN Global Pulse Advisory Group on the Governance of Data and AI, and UNESCO India's Media Freedom Advisory Group.

MATTHEW BATTLES

Author, artist and associate director of metaLAB at Harvard University. Battles sees the institution of the library as more than just the building's contents. He headed a team which created a data visualization of the printing locations of books published in early-modern Europe, shown over time.

ETHAN ZUCKERMAN

Director of the Center for Civic Media at MIT. Associate Professor of the Practice at MIT Media Lab. Research focuses on the use of media as a tool for social change, role of technology in International Development, and use of new media tech by activists. Author, *Rewire: Digital Cosmopolitans in the Age of Connection* (2013).

PADMASHREE GEHL SAMPATH

Leading expert on technology, development and the global political economy. Visiting Professor at the South African Research Chair in Industrial Development (SARChI ID), University of Johannesburg. Professorial Fellow, United Nations -MERIT, and the Chair Person of the Technical Advisory Group of the World Health Organisation's COVID-19 Technology Access Pool (CTAP). Conducts empirical analyses in collaboration with academia, international agencies and the policy community to suggest reforms in the digital economy (AI, big data and internet of things), pharmaceuticals, life sciences and energy. Focuses on technology governance, development, and the global political economy.

KENDRA ALBERT

Public interest technology lawyer with a special interest in computer security law and freedom of expression. Clinical Instructor, Cyberlaw Clinic at Harvard Law School. Founder and Director of the Initiative for a Representative First Amendment.

MASON KORTZ

Clinical Instructor, Harvard Law School Cyberlaw Clinic, Berkman Klein Center. He is active in the emerging area of the law of artificial intelligence and algorithms, and has written and presented on the impact of algorithmic decision making on areas as diverse as intellectual property, products liability, and the criminal legal system.

RAM SHANKAR SIVA KUMAR

Data Cowboy in Azure Security at Microsoft, working on Security in Machine Learning. Primary focus is modeling security logs to surface malicious activity, with a secondary focus on using machine learning systems for offense. Broadly involved at the Berkman Klein Center on the following questions: How do we assess the safety of ML systems? What are the policy and legal ramifications of AI, in the context of security? Graduate from Carnegie Mellon University with a Masters in Electrical and Computer Engineering and a separate Masters in Innovation Management focusing on Telecom Policy.

CRAIG NEWMARK

Founder of Craig's List and Craig Newmark Foundation

FEDER COOPER

Cornell Engineering and Advocate for Transgender Rights in AI

JOSHUA BIXBY

Chief Executive Officer, Fastly, leading cloud computing service

SANJAY SARMA

Dean of Engineering and Vice President, MIT

RAJ AGARWAL

CEO and Founder Medocity, Telemedicine

STEVE CROWN

Vice President and Deputy General Counsel, Human Rights, Microsoft Corporation

SEAN WHITE

Vice President, Mozilla

KERSTIN NOËLLE VOKINGER

Faculty, University of Zurich. Affiliated Faculty, Harvard Medical School Interdisciplinary researcher on law, medicine and technology.

ALEXANDRA PAVLIUC

Oxford University, DPhil Social Data Science. Researcher of network analysis and disinformation focusing on visualization techniques to highlight computational propaganda. Montaigne, a French think tank.

FINALE DOSHI-VELEZ

Professor, Harvard Paulson School of Engineering and Applied Sciences. Focuses on the intersection of machine learning, healthcare, and interpretability. Head of the Data to Actionable Knowledge Group at Harvard Computer Science.

JUDITH DONATH

Writer, designer and artist whose work examines how new technologies transform the social world. Author of The Social Machine (MIT Press, 2014), she is known for her writings about identity, deception, privacy, online interaction and artificial minds. Formerly, she directed the Sociable Media Group at the MIT Media Lab, where she and her students created pioneering and influential social visualizations and interfaces that have been exhibited worldwide. Currently, she is writing a book about technology, trust and deception.

MEHRNOOSH SAMEKI

Technical program manager, Microsoft. Responsible for driving the product efforts on machine learning interpretability and fairness within the Azure Machine Learning platform. PhD Computer Science, Boston University.

MITCHELL BAKER

Chairwoman, Mozilla, Mozilla CEO, 1999-2008. Co-founded the Mozilla Project to support the open, innovative web and ensure it continues offering opportunities for everyone. Advocate for connecting technology to its impact on individuals and society. Mitchell has co-chaired the U.S. Department of Commerce Digital Economy Board of Advisors from its inception in March 2016 until August 2017, served on the United Nations High Level Panel on Women's Economic Empowerment, and the ICANN High Level Panel on Global Internet Cooperation and Governance Mechanisms.

VIRGILIO ALMEIDA

Former National Secretary for Information Technology Policies, Brazil. Faculty Associate, Berkman Klein Center. Virgilio received his PhD in Computer Science from Vanderbilt University. He has been awarded the Great Cross of the National Order of the Scientific Merit by the President of Brazil.

Article 9: Accessibility

I. To enable persons with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, to information and communications, including information and communications technologies and systems, and to other facilities and services open or provided to the public, both in urban and in rural areas. These measures, which shall include the identification and elimination of obstacles and barriers to accessibility, shall apply to, inter alia:

a) Buildings, roads, transportation and other indoor and outdoor facilities, including schools, housing, medical facilities and workplaces;

b) Information, communications and other services, including electronic services and emergency services.

II. States Parties shall also take appropriate measures:

a) To develop, promulgate and monitor the implementation of minimum standards and guidelines for the accessibility of facilities and services open or provided to the public;

b) To ensure that private entities that offer facilities and services which are open or provided to the public take into account all aspects of accessibility for persons with disabilities;

c) To provide training for stakeholders on accessibility issues facing persons with disabilities;

d) To provide in buildings and other facilities open to the public signage in Braille and in easy to read and understand forms;

e) To provide forms of live assistance and intermediaries, including guides, readers and professional sign language interpreters, to facilitate accessibility to buildings and other facilities open to the public;

f) To promote other appropriate forms of assistance and support to persons with disabilities to ensure their access to information;

g) To promote access for persons with disabilities to new information and communications technologies and systems, including the Internet;

h) To promote the design, development, production and distribution of accessible information and communications technologies and systems at an early stage, so that these technologies and systems become accessible at minimum cost.

Article 17: Human Rights Due Diligence

In order to identify, prevent, mitigate and account for how they address their adverse human rights impacts, business enterprises should carry out human rights due diligence. The process should include assessing actual and potential human rights impacts, integrating and acting upon the findings, tracking responses, and communicating how impacts are addressed. Human rights due diligence: (a) Should cover adverse human rights impacts that the business enterprise may cause or contribute to through its own activities, or which may be directly linked to its operations, products or services by its business relationships; 18 (b) Will vary in complexity with the size of the business enterprise, the risk of severe human rights impacts, and the nature and context of its operations; (c) Should be ongoing, recognizing that the human rights risks may change over time as the business enterprise's operations and operating context evolve.

FIGURE 8.7
Rebecca Xiong and Judith Donath, *People Garden (thriving)* (1999).

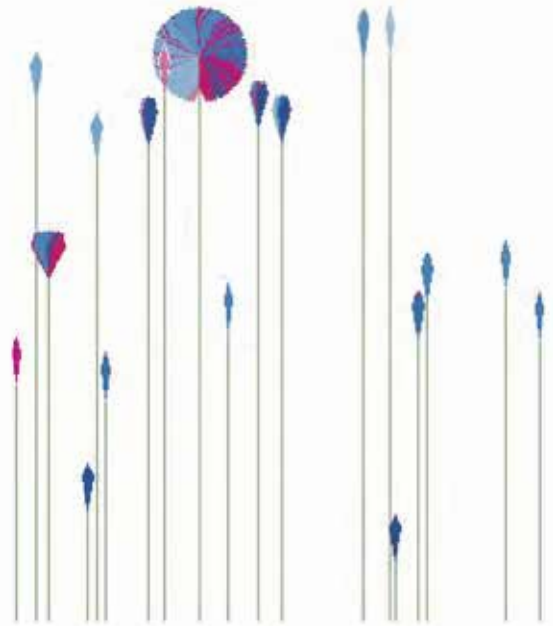


FIGURE 8.8
Rebecca Xiong and Judith Donath, *People Garden (abandoned)* (1999).

Source: Data Portraits by Judith Donath

Class Survey

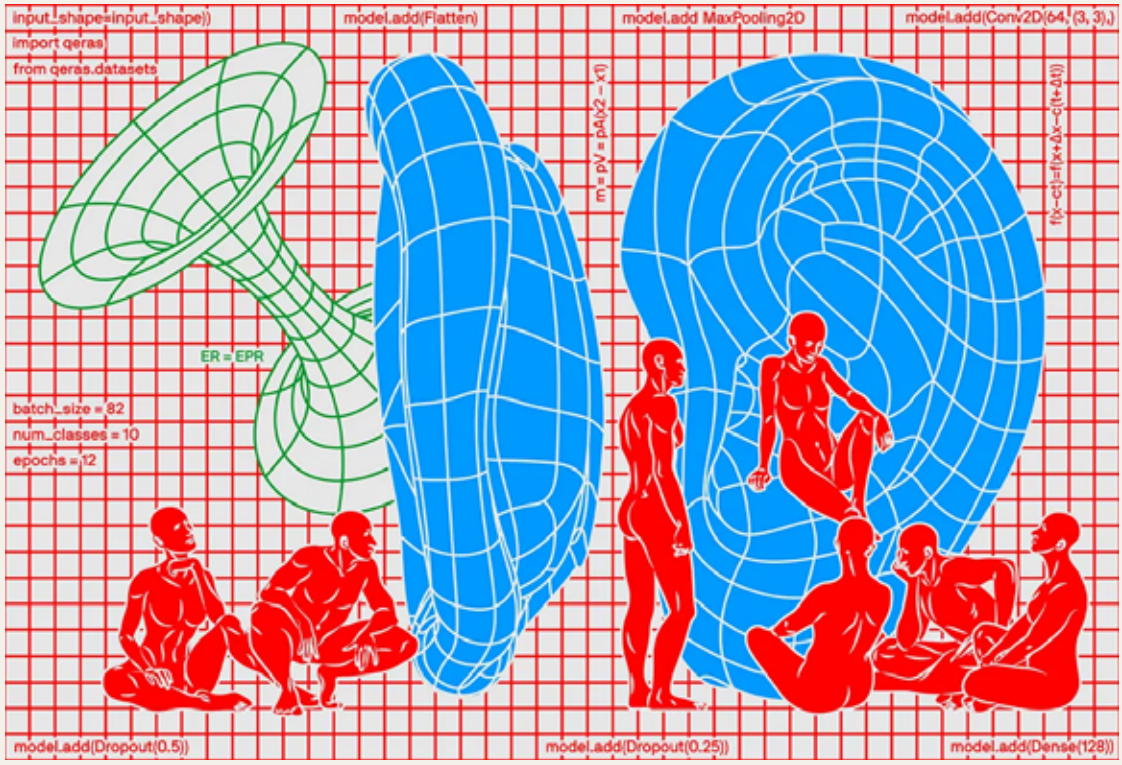


Illustration by Berke Yazicioglu

INTRODUCTION

BY BROOKE BERNSTEIN

CLASS SURVEY

BACKGROUND

A hallmark of the Policy Lab on AI and Implicit Bias each spring is an informal survey completed by the class and their colleagues on one aspect of AI and implicit bias to push boundaries, influence the focus of future research, and for others to build on. This semester, the survey that was developed, shared, and analyzed was focused on the use of AI on hiring platforms, like LinkedIn, and the implicit bias that can occur with their use.

RESEARCH QUESTIONS

We sought to answer multiple questions with this survey, including what are students', postgraduates', and young professionals' perceptions of the use of these systems, if the survey sample has lived experiences using these systems, and if there are differences in perceptions and lived experiences between the subgroups of the sample.

METHODOLOGY

The survey was developed using Qualtrics and was open to responses for two weeks starting February 28th, 2022, and was closed on March 14, 2022. Students in the Policy Lab were asked to complete the survey themselves and share the survey with five other colleagues that were also students, post-graduates, and/or young professionals ages 20 to 35. Respondents were asked to reflect on their hiring experiences over the last three years, since 2019, to reflect the changes that have occurred in the hiring process since the COVID-19 pandemic began. The survey received 201 responses, with 178 of those not being duplicates and having finished at least one substantive question after completing the demographics section.

STUDY VARIABLES AND ANALYSIS

The survey can be divided into four sub-sections. The first subsection collected the demographic information of the respondents. The second subsection collected the job related demographic information of the respondents. The third sub-section collected information on the experiences respondents had using job hiring platforms. The final, and fourth sub-section collected respondents' perspectives on the use of AI in job hiring platforms and the implicit bias that can occur with this use. The focus of the following analysis will be on the final subsection. Please see the appendix for the complete list of survey questions and brief analysis.

SAMPLE BREAKDOWN

DEMOGRAPHICS

- 53% of the sample were early career professionals ages 25-30.
- 37% identified as Asian, while 28% identified as White.
- 62% were born in the United States and are U.S. citizens, while 30% were not U.S. citizens*

  *Note: Those who are not U.S. citizens included Legal Permanent Residents.
- 65% identified as women, 33% identified as men.
- 22% identified as First Generation Professionals. 6% identified as LGBTQ+.
- 57% have a Bachelor's Degree. 24% also have a Master's Degree
- 21% of the sample's last degree was in Liberal Arts. 20% studied Law.
- 46% are students.44% are employed full-time.
- 33% are not currently working, while those who are working are in various fields like Information Technology (12%).

A majority of the sample of respondents were Asian and White early career professionals ages 25 to 30 that were born in the United States and are U.S. citizens. Additionally, a majority of the sample identified themselves as women. Most respondents in the sample have a Bachelor's degree and are either current students or employed full-time.

Age 25 « » Age 30

53%

65%
Women Identifying

33%
Men Identifying

57% At least Bachelor's Degree

46% Current Students

44% Employed Full Time

62%

Born in the United States and hold U.S. citizenship

37%
Identify as Asian

28%
Identify as White

*Note: Those who are not U.S. citizens includes permanent residents.

RESULTS

The final subsection of the survey can be divided into seven additional subsections of questions, which are illustrated below. This section received 178 complete responses.

Q29-31

Q29

A majority of the sample had not personally experienced discrimination or been treated unfairly by an employer in hiring, pay, or promotion. We were interested in seeing if those who had experienced discrimination were possibly born in the Global South in comparison to all other places of birth. Although, after considering this sub-group of respondents, it did not appear that comparing subgroups based on where they were born would be a robust metric.

Q30

A majority of the sample believe that online job platforms have increased their opportunities for the advancement of their careers, despite any implicit bias that might occur with their use.

Q31

There was also a distinct consensus among respondents that artificial intelligence systems will always reflect the biases of the people who designed them and the data the system was trained on.

QUALITATIVE RESPONSES

A respondent who had experienced discrimination from time to time shared this story: "I interviewed with a judge for a 1L summer internship. All seemed to be going well until he noticed that I was part of LALSA as well as the Jewish Law Students Association (JLSA). He asked how I could be both Jewish and Hispanic and made comments that I didn't "look Jewish" as though it were a compliment. He went on for maybe two more minutes about his surprise that I'm Jewish. He said I must be the only Argentinian Jew — "oh wait I know a guy who is Peruvian and Jewish and blonde, so he's got you beat" — until one of his law clerks asked him to stop. About a week later I did receive an offer, but based on the conversation I felt that he would create an uncomfortable work environment. I chose to work for a different judge." (Q29)

Another respondent who had experienced discrimination from time to time said, "Because I am out on my resume, I think that I have not received interviews for jobs that I was qualified for in more conservative geographies and more conservative offices." (Q29)

A respondent who had experienced discrimination regularly said that it was evident in the gender pay gap. (Q29)

Q33-35

Before answering the following set of questions, respondents were asked to consider the following scenario:

"In the past, when companies hired, they typically had someone read applicants' resumes and conduct personal interviews to choose the right person for the job. Today, a growing number of companies use artificial intelligence systems to provide a systematic review of each applicant without the need for human involvement. These systems often give each applicant a score based on the content of their resume, application, or standardized tests for skills such as problem-solving or personality type. Applicants are then ranked and hired based on those scores."

Q33

When asked if they had heard, read, or thought about AI and implicit bias before this survey, a majority of the respondents said that they had heard about it at least a little.

Q34

A majority of the sample only see a little benefit to society as a whole by using these systems to evaluate job candidates as a practice.

Q35

Although, a majority of the sample would still apply to a job that used this type of system to make a hiring decision. It is important to note that ~ 36% would not apply for such a job.

QUALITATIVE RESPONSES

Another respondent who said that they would not apply shared this perspective, "If resumes are not worded "correctly," these systems might miss some great applications." (Q35)

One respondent who said they would still apply said, "The system is somewhat irrelevant to my decision-making process, it just depends on if I want the job or not."
(Q35)

An additional respondent who said they would not apply said, "Working in a community is important to me and I don't think a company that uses AI is holistic as I would wish from my employer and its community."
(Q35)

Q37-39

Before answering the following two sets of questions, respondents were asked to consider the following scenario:

"In an effort to improve the hiring process, some companies are now using artificial intelligence systems to screen and evaluate resumes. The system outputs and assigns each candidate a score based on the content of their resume, and how it compares with the resumes of employees who have been successful in their position at the company. Only resumes that meet a certain score are sent to a hiring manager for further review."

Q37

Only a slim majority of respondents believe that this system is not fair to people applying for jobs in comparison to those who responded that this type of system is at some level fair to job applicants.

Q38

But, a majority of respondents believe that this type of system is at least somewhat effective at identifying successful job candidates.

Q39

Additionally, a slim majority of respondents believe that it is acceptable for companies to use this type of approach when hiring job candidates.

QUALITATIVE RESPONSES

Another respondent shared, "I understand that in a job market with more applicants than recruiters, this may be a necessary tool to through applicants. However, I'd also encourage recruiters to observe and be mindful of the AI model and how the data is trained. Ensuring that the model can differentiate between non-traditional paths into the industry or other crucial components of a resume that can be overlooked (e.g., DEI keywords, other unique work history & associated skills) will be crucial in determining the success of AI systems used in hiring." (Q39)

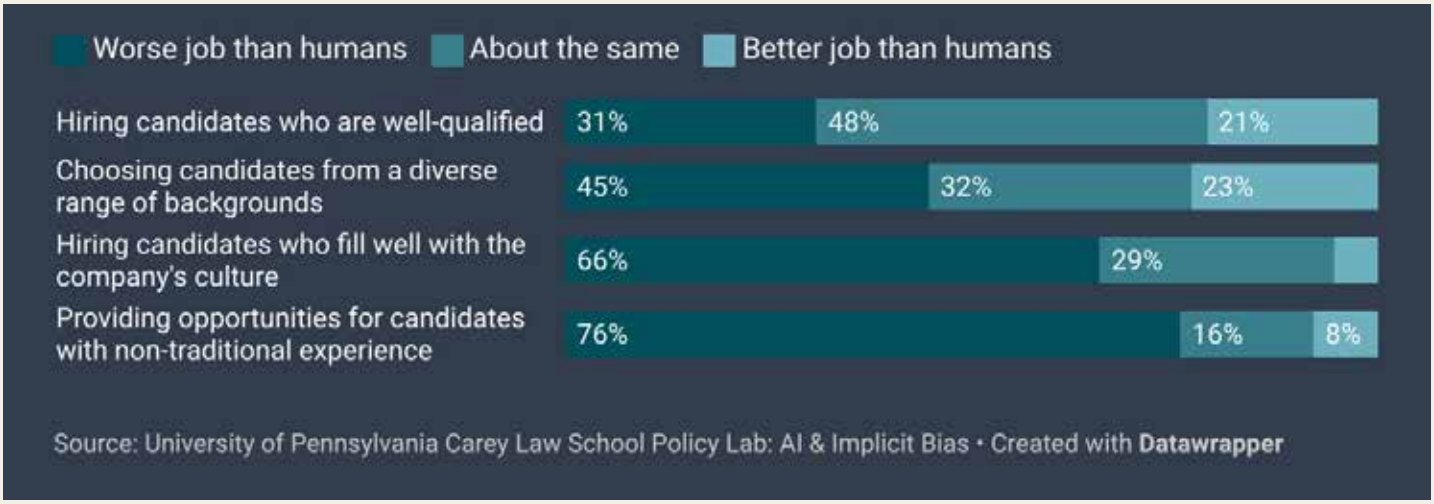
One respondent who said that it is acceptable for companies to use this type of approach shared this perspective, "Yes - but I'm worried that it will reproduce the demographics that already exist in the company. It could prevent underrepresented minorities from getting a foothold in the company." (Q39)

One respondent who believes that it is unacceptable said, "This type of system is great if you want to hire carbon copies of your existing employees. This sort of system would be biased against hiring individuals with different experiences or backgrounds (e.g., coming from different fields or less wealthy universities). By hiring through comparison to the existing team you build a monoculture and essentially create an automated "good old boys" network. Any employees that are hired would likely do a very good job but that ignores the greater societal impact of effectively building a machine that rejects people that are different."(Q39)

Q41-44

Q41-44

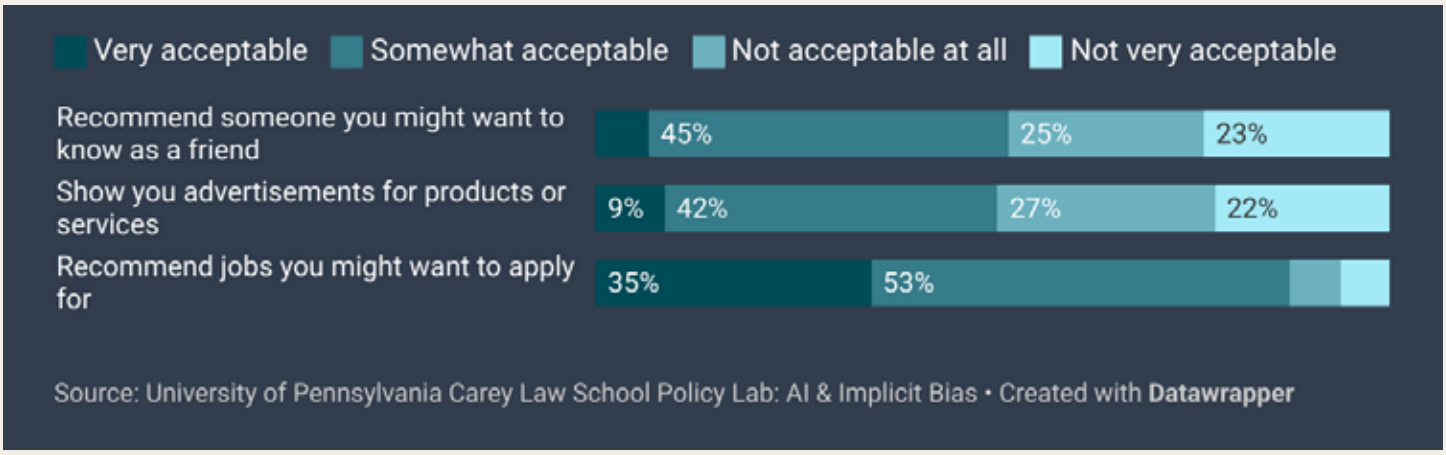
When asked to compare the use of these types of systems to hire job applicants to hiring done completely by humans, a significant majority said that systems would hire candidates who fit well with the company's culture and provide opportunities for candidates with non-traditional experiences worse than humans. But, there was almost a split between respondents who believe that systems hire well-qualified applicants and applicants from a diverse range of backgrounds about the same or better and respondents who believe that systems hire worse when compared to humans.



Q46–48

Q46-48

When asked how acceptable it was for these types of systems to recommend someone you might know as a friend or show you advertisements for products or services, there was almost a split between respondents who thought that these uses were acceptable or not. But, when asked if it was acceptable for these types of systems to recommend jobs you might want to apply for, a significant majority of the sample said that it was acceptable.



Q50-52

Before answering the following set of questions, respondents were asked to consider the following scenario:

“To improve the hiring process, some companies are now recording interviews with job candidates. Questions are often pre-recorded by the company and candidates are instructed to record their responses. These videos are analyzed by an artificial intelligence system, which compares the characteristics and behavior of candidates with traits shared by successful employees. Candidates are then given an automated score based on their predicted success that helps the firm decide whether or not they might be a good hire.”

Q50-52

A majority of respondents believe that this type of system would be unfair to job candidates, would not be effective for this particular scenario, and is not acceptable to use to hire.

QUALITATIVE RESPONSES

A respondent who believed that it is acceptable to use this type of system to hire said, “Yes with a caveat, as one would hope there are systems in place to ensure this AI can analyze different dialects and response types fairly equally.” (Q52)

Another respondent who believes that it is unacceptable to use this system to hire shared, “As I said before, I think AI systems are fundamentally biased. These videos are analyzed for specific characteristics and behaviors, and that is inherently biased. For example, many Americans like to make large hand gestures and talk animatedly. On the other hand, as an East Asian, I’ve grown up being taught that I should always behave in a calm demeanor to show maturity. Even within different cultures and different places where we grow up, there are different ways behaviors and characteristics are interpreted. If AI systems are looking for specific traits, then they would be inherently unfair.” (Q52)

Another respondent who also believed that it is unacceptable said, “These systems will very likely start to narrow in on things like facial structure and color rather than any actual meaningful characteristics. We don’t have the capabilities to parse the contents of an interview into a meaningful score with current technology.” (Q52)

Q54-56

For the last two questions, respondents were asked to consider if they would feel better, worse, or would not change their opinion of online job platforms and recruiters using artificial intelligence tools in the following scenarios:

1. “If these systems included public data about each job candidate, such as the material they post on social media, in making their evaluations?”
2. “If companies used these systems to conduct initial screenings of potential candidates but then traditionally interviewed those candidates?”

Q54

A majority of respondents believe that if these systems included public data about each job candidate, such as the material they post on social media, in making their evaluations, their opinions of the use of these systems would be worse.

Q55

But, when asked if companies used these systems to conduct initial screenings of potential candidates but then traditionally interviewed those candidates, a majority of respondents believe that their opinion of the use of these systems would be better or there would not be a difference.

QUALITATIVE RESPONSES

One respondent shared this final comment, “I am a researcher in AI and autonomy for the government. The lack of transparency regarding the hiring process (especially if AI is implemented) could have serious negative consequences for our society. Used as a general screening method would be tentatively acceptable, but really, humans are much more than their resumes and CVs, and AI systems aren’t built for evaluating someone’s appropriateness at a job.” (Q56)

Another respondent shared, “This was a very interesting question to pose! I think that AI has a place in reducing the friction between the very “anonymized” online job application process (i.e., applicants applying via a “faceless” job portal or via a job listing) and the actual candidate interview, but I don’t think it’s possible to fully automate job applications/hiring. I agree that much of the bias in models can come from the data and the people that train it, and until that kind of bias can be addressed, fully-automated hiring is not possible. And arguably, it may not be possible at all due to the social & human nature of certain occupations.”(Q56)

CONCLUSION

Regardless of the survey sample's knowledge and awareness of the implicit bias that can occur with the use of AI in job hiring platforms, a majority of the sample have normalized their use. This could potentially be the result of the demographics of the survey, which significantly were current students or alumni of the University of Pennsylvania and other esteemed universities. Students that graduate from these universities could possibly not be as concerned about the bias that can occur because they believe that an AI tool would not discriminate against their candidacy for a job. This is evident in a majority of the sample not having personally experienced discrimination or been treated unfairly by an employer in hiring, pay, or promotion before. This is particularly concerning, as if the use of these systems is normalized, there is potential for their use to be well-established during hiring, leading to biases being baked into the recruitment process, and eventually company culture.

It is important to note that there was a minority of respondents who do not support the use of AI on job hiring platforms, including not applying to companies which use these systems. As multiple respondents aptly said in the free response questions, if companies continue to use these systems, they could be discouraging qualified candidates from applying or rejecting qualified candidates before a human sees their resume.

There is an evident need for an information campaign to inform job candidates and companies of the risks, harms, and limitations of the use of these systems, as the knowledge that some already have about the issue seems to not be as robust as they might believe. Those in this demographic appear to possibly be more concerned about the use of AI in other industries, like marketing and social media, than in the hiring process.

Finally, it is important for this survey to be replicated to investigate this topic further, especially in a more formal manner and on a more diverse sample of respondents to compare the responses of those with different demographics and lived experiences.

BROOKE BERNSTEIN

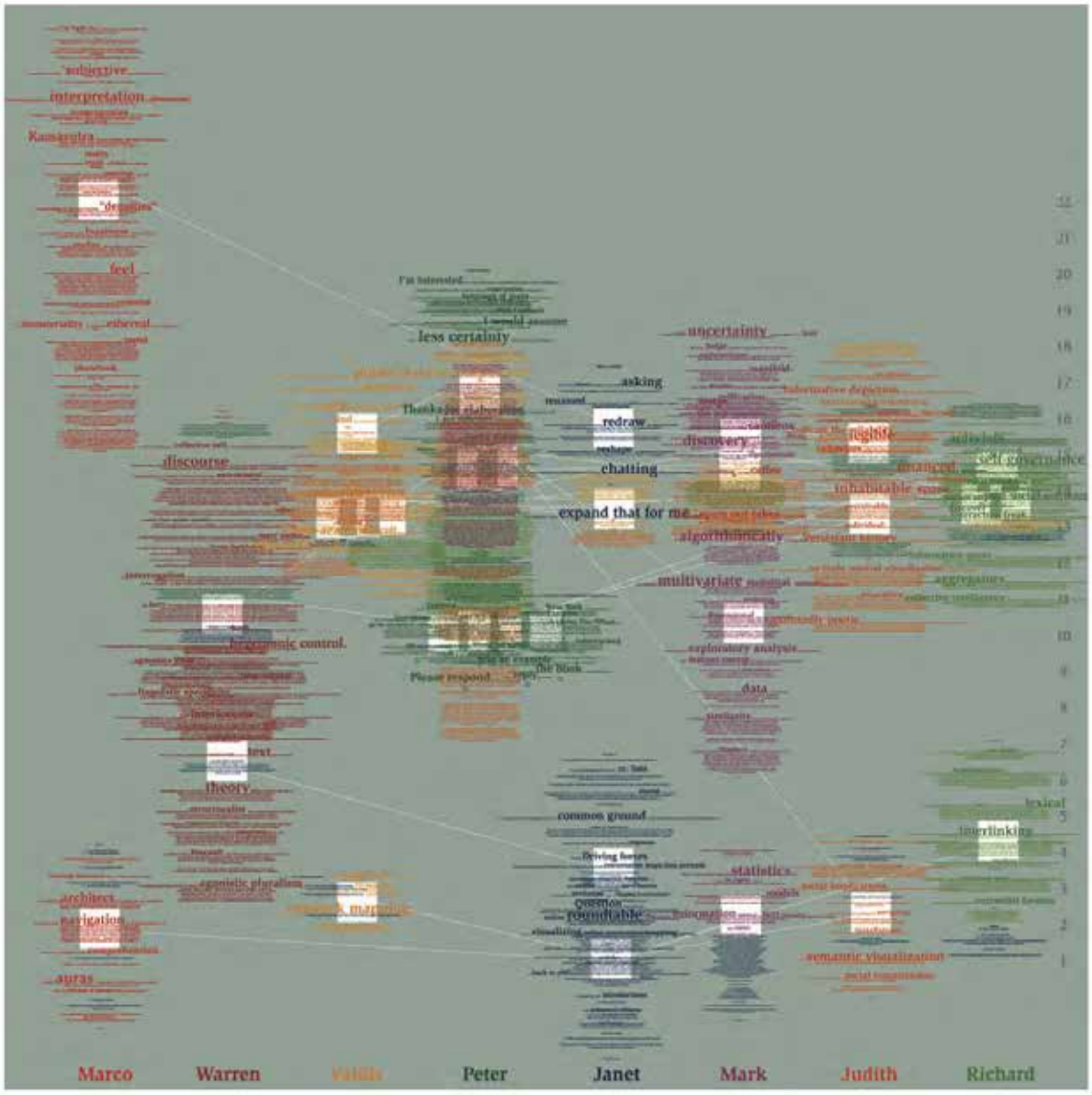


FIGURE 8.16
Judith Donath, *The Rhythm of Salience* (2006).

FIGURE 8.17
Aaron Zinman and Judith Donath, *Personas* (2008).

analyzes the resulting texts and attempts to characterize the person by fitting him or her into a set of categories of roles and interests (see figure 8.17). The result is sometimes surprisingly apt, but can also be very far off, given the computer's inability to distinguish different people with the same name and its errors in language comprehension. *Personas* is a reminder of the fallibility, social naiveté, and opacity of the computer as portrayer in an era when such computer analysis of people is increasingly prevalent.

Source: Data Portraits by Judith Donath

Appendix

AI Bias Policy Lab Spring 2022 Survey

Start of Block: Block 53

Q1 Thank you for participating in Penn Law's AI and Implicit Bias Policy Lab 2022 survey. The Policy Lab, conducted by Professor Rangita de Silva de Alwis, focuses on underlying biases in artificial intelligence systems and their impacts on various facets of society. With this current survey, the Policy Lab will examine the effects of AI in hiring for jobs and internships among current students, postgraduates and young professionals between the ages of 20-35. The results of the survey will communicate the impacts of AI on hiring and inform future policy proposals regarding development and improvement of AI.

For each survey question, please reflect on your job or internship experiences over the last three years (since 2019).

The survey can be completed in about 15 minutes. This form is anonymous. Your responses will be kept private and secure. Only the AI Bias Policy Lab professor and (two) teaching assistants will have access to your individual submissions. This information will not be used for any discriminatory purpose.

End of Block: Block 53

Start of Block: Block 1

Q2 What is your age?

- ○ 20-24
- ○ 25-30
- ○ 30-35

End of Block: Block 1

Start of Block: Block 2

Q3 How would you best describe your race and/or ethnicity?

- [ ] Asian
- [ ] Black/African-Descent
- [ ] White
- [ ] Hispanic/Latinx
- [ ] Native American
- [ ] Pacific Islander
- [ ] Jewish
- [ ] Prefer not to answer
- [ ] Other _____

End of Block: Block 2

Start of Block: Block 3

Q4 In what region were you born?

- North America
- Central America
- South America
- Caribbean
- Central Asia (includes Tajikistan, Uzbekistan, Kazakhstan, Turkmenistan, and Kyrgyzstan)
- East Asia (includes China, Mongolia, North Korea, South Korea, Japan, Hong Kong, Taiwan, and Macau)
- South Asia (includes Sri Lanka, Bangladesh, India, Afghanistan, Pakistan, Bhutan, Nepal, Iran, and the Maldives)
- Southeastern Asia (includes Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor Leste, and Vietnam)
- Europe (includes Central and Eastern Europe, Nordic countries, Southern Europe, Western Europe)
- Middle East and North Africa (includes Georgia, Armenia, Azerbaijan, Turkey, Cyprus, Syria, Lebanon, Israel, Palestine, Jordan, Iraq, Oman, Yemen, Kuwait, Bahrain, Qatar, Saudi Arabia)
- Pacific (includes Australia, New Zealand, and Pacific Islands)
- Sub-Saharan Africa (includes Central Africa, East Africa, Southern Africa, West Africa)
- Prefer not to answer

End of Block: Block 3

Start of Block: Block 4

Q5 What is your citizenship status?

- ○ Born in the U.S.
- ○ Born abroad to American parent(s)
- ○ U.S. citizen by naturalization
- ○ Not a U.S. citizen

End of Block: Block 4

Start of Block: Block 5

Q6 Which gender identity do you most identify with?

- ○ Woman
- ○ Man
- ○ Transgender woman
- ○ Transgender man
- ○ Gender variant/non-conforming
- ○ Prefer not to answer
- ○ Other _____

End of Block: Block 5

Start of Block: Block 6

Q7 **Do you identify with any of the following affiliations?**

- [ ] Disability
- [ ] LGBTQ+
- [ ] Veterans
- [ ] First Generation Professional

End of Block: Block 6

Start of Block: Block 7

Q8 What is the highest degree or level of school you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)

- ○ Less than a high school diploma
- ○ High school degree or equivalent (e.g. GED)
- ○ Some college, no degree
- ○ Associate degree (e.g. AA, AS)
- ○ Bachelor's degree (e.g. BA, BS)
- ○ Master's degree (e.g. MA, MS, MEd)
- ○ Professional degree (e.g. MD, DDS, DVM, JD)
- ○ Doctorate (e.g. PhD, EdD)

End of Block: Block 7

Start of Block: Block 8

Q9 Based on the previous question, what were your area(s) of study for the most recent degree you received? If not applicable (for high school), please select N/A. Select all the apply: the

following lists include fields of study within the categories but are not limited to fields within the list.

- ○ Business (includes business management, finance, marketing, accounting, human resources, and economics)
- ○ Science (includes life sciences, natural sciences, agriculture, medicine)
- ○ Engineering (includes computer and information
- ○ Liberal arts (including humanities/general studies, literature, history, art, journalism, philosophy/theology, music, architecture, foreign language)
- ○ Social science (including behavioral science, social work, anthropology)
- ○ Law
- ○ Education
- ○ Vocational (including technical, police and protective services, criminal justice)
- ○ N/A
- ○ Other _____

End of Block: Block 8

Start of Block: Block 9

Q10 What is your current employment status?

- ○ Employed full-time (40 or more hours per week)
- ○ Employed part-time (up to 39 hours per week)
- ○ Unemployed and currently looking for work
- ○ Unemployed and not currently looking for work
- ○ Student
- ○ Retired
- ○ Homemaker
- ○ Self-employed
- ○ Unable to work
- ○ Prefer not to answer

End of Block: Block 9

Start of Block: Block 10

Q11 What field do you currently work in?

- ○ Hospitality, services, arts, entertainment and recreation
- ○ Healthcare and social assistance
- ○ Manufacturing, mining, construction, agriculture, forestry, fishing and hunting
- ○ Retail, trade and transportation
- ○ Education
- ○ Banking, finance, accounting, real estate or insurance
- ○ Government, public administration or military
- ○ Information/technology
- ○ Professional, scientific and technical services
- ○ Not currently working
- ○ Other _____

End of Block: Block 10

Start of Block: Block 11

Q12 How closely related is your current job to the field, major, or area(s) of study for the most recent degree you received?

- ○ Somewhat closely
- ○ Not very closely
- ○ Not related at all
- ○ Not currently working
- ○ Don't know

End of Block: Block 11

Start of Block: Block 12

Q13 If your job is not very closely (or not at all) related to your most recent degree, what is the main reason you are not currently working in a job more closely related to the field you studied?

- ○ No job available in field/Bad economy
- ○ Wanted a change/Didn't like the work/wasn't for me
- ○ Happy with current job/Job security
- ○ Financial reasons/Better pay
- ○ Lacked experience/Not qualified
- ○ Circumstances/Moved/Had children/Flexible hours
- ○ Major was very general
- ○ Other (includes retired, still in school) _____

End of Block: Block 12

Start of Block: Block 13

Q14 Still thinking about your current job, would you say you have more qualifications than the job requires, the right amount of qualifications, or only some of the qualifications the job requires?

- ○ More
- ○ Right amount
- ○ Only some
- ○ Not currently working
- ○ Don't know

End of Block: Block 13

Start of Block: Block 14

Q15 Which of the following best describes your current job?

- ○ I am working in the same job as before the coronavirus pandemic
- ○ I have changed jobs since the coronavirus outbreak
- ○ I was not employed before the coronavirus outbreak
- ○ I am not currently working
- ○ Other _____

End of Block: Block 14

Start of Block: Block 15

Q16 Compared to before the coronavirus outbreak, are you...

- ○ More satisfied with you job
- ○ Less satisfied with your job
- ○ About the same
- ○ Not currently working
- ○ Other _____

End of Block: Block 15

Start of Block: Block 16

Q17 Here are two different ways of looking at your job: Some people get a sense of identity from their job. For other people, their job is JUST what they do for a living. Which best describes the way you usually feel about your job?

- ○ Sense of identity
- ○ Just what I do for a living
- ○ Don't know
- ○ Other _____

End of Block: Block 16

Start of Block: Block 17

Q18 Do you think of your current job as a career, a stepping stone to a career, or do you think of it as just a job to get you by?

- ○ A career
- ○ A stepping stone to a career
- ○ Just a job to get me by
- ○ Not Currently Working
- ○ Other _____

End of Block: Block 17

Start of Block: Block 18

Q19 For the most part, do you think someone without the level of education you currently have could develop the skills and knowledge needed to do your job on their own without formal education?

- ○ Yes
- ○ No
- ○ Not currently working

End of Block: Block 18

Start of Block: Block 19

Q20 Looking ahead, how important do you think it will be for you to receive training and develop new skills throughout your career to keep up with the changes in the workplace?

- ○ Essential
- ○ Important, but not essential
- ○ Not important
- ○ Retiring soon

End of Block: Block 19

Start of Block: Block 20

Q21 In general, do you feel you have the education and training necessary to get the kind of job you want, or do you need more education and training?

- ○ Have necessary education and training
- ○ Need more

End of Block: Block 20

Start of Block: Block 21

Q22 Do you have any experience using online job hiring programs like LinkedIn, Indeed, Monster, Glassdoor, Ziprecruiter, etc. to find or apply for job opportunities?

- ○ Yes
- ○ No

End of Block: Block 21

Start of Block: Block 22

Q23 Thinking about the jobs which appear as recommendations for you on these programs, to what degree do you feel like these recommendations capture your skills, expertise, and potential?

	Not at all 0	1	2	3	4	Somewhat 5	6	7	Very much 8	9	Have Not Used Programs 10
Degree of Capturing Skills, Expertise, & Potential											

End of Block: Block 22

Start of Block: Block 23

Q24 Have you every applied to a job through one of these platforms? (Select one or more)

- [ ] Glassdoor
- [ ] LinkedIn
- [ ] Ziprecruiter
- [ ] Monster
- [ ] FlexJobs
- [ ] Ladders
- [ ] AngelList
- [ ] Getwork
- [ ] Scouted
- [ ] Snagajob
- [ ] Have not applied
- [ ] Other _____

End of Block: Block 23

Start of Block: Block 24

Q25 Have you ever wanted to apply for a job you felt qualified for, but did not apply because it required more education or years of work experience than you had? If yes, please explain.

- ○ Yes _____
- ○ No

End of Block: Block 24

Start of Block: Block 25

Q26 How often do you receive an interview request when you submit an application through an online job hiring platform?

- ○ Always
- ○ Somewhat often
- ○ Not very often
- ○ Never
- ○ N/A

Q27 How often do you receive a rejection when you submit an application through an online job hiring platform?

- ○ Always
- ○ Somewhat often
- ○ Not very often
- ○ Never
- ○ N/A

Q28 How often do you receive no response when you submit an application through an online job hiring platform?

- Always
- Somewhat often
- Not very often
- Never
- N/A

End of Block: Block 25

Start of Block: Block 28

Q29 Have you personally experienced discrimination or been treated unfairly by an employer in hiring, pay or promotion? If yes, please explain.

- Yes, regularly _____
- Yes, from time to time _____
- No

End of Block: Block 28

Start of Block: Block 29

Q30 Thinking about opportunities for advancement in your career, do you feel that online job platforms have...

- Increased your opportunities
- Decreased your opportunities
- No impact either way

End of Block: Block 29

Start of Block: Block 42

Q31 Which of the following statements aligns closest with your opinion on these systems, even if neither is exactly the same?

- ○ It is possible for artificial intelligence systems to make decisions without human bias.
- ○ Artificial intelligence systems will always reflect the biases of the people who designed them and the data the system was trained on.

End of Block: Block 42

Start of Block: Block 30

Q32 **Please consider the following scenario…**

In the past, when companies hired, they typically had someone read applicants' resumes and conduct personal interviews to choose the right person for the job. Today, a growing number of companies use artificial intelligence systems computer programs to provide a systematic review of each applicant without the need for human involvement. These systems often give each applicant a score based on the content of their resume, application, or standardized tests for skills such as problem solving or personality type. Applicants are then ranked and hired based on those scores.

Q33 **How much have you heard, read, or thought about this idea before this survey?**

- ○ None at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ A great deal

Q34 Thinking about the possibility that job candidates might be evaluated using this sort of system, how much benefit, if any, do you feel that this practice adds to society as a whole?

- ○ A huge benefit
- ○ A little benefit
- ○ No benefit

Q35 Would you personally apply for a job that used this type of system to make hiring decisions? Please explain.

- ○ Yes _____
- ○ No _____

End of Block: Block 30

Start of Block: Block 35

Q36
Please consider the following scenario...
In an effort to improve the hiring process, some companies are now using artificial intelligence systems to screen and evaluate resumes. The system outputs and assigns each candidate a score based on the content of their resume, and how it compares with resumes of employees who have been successful in their position at the company. Only resumes that meet a certain score are sent to a hiring manager for further review.

Q37 **How FAIR do you think this type of program would be to people applying for jobs?**

- ○ Not fair at all
- ○ Not very fair
- ○ Somewhat fair
- ○ Very fair

Q38 How EFFECTIVE do you think this type of program would be at identifying successful job candidates?

- Not effective at all
- Not very effective
- Somewhat effective
- Very effective

Q39 Do you think it is acceptable for companies to use this type of approach when hiring job candidates? Please explain.

- Yes _____
- No _____

Q40 Thinking about the previous scenario,
Do you think these systems would do a better or worse job as humans when it comes to...

Q41 Hiring candidates who are well-qualified

- Better job than humans
- About the same
- Worse job than humans

Q42 Hiring candidates who fill well with the company's culture

- Better job than humans
- About the same
- Worse job than humans

Q43 Choosing candidates from a diverse range of backgrounds

- Better job than humans
- About the same
- Worse job than humans

Q44 Providing opportunities for candidates with non-traditional experience

- Better job than humans
- About the same
- Worse job than humans

End of Block: Block 35

Start of Block: Block 44

Q45 **How acceptable, if at all, do you think it is for job platforms to use data about you and your online activities to...**

Q46 Recommend someone you might want to know as a friend

- Very acceptable
- Somewhat acceptable
- Not very acceptable
- Not acceptable at all

Q47 Show you advertisements for products or services

- Very acceptable
- Somewhat acceptable
- Not very acceptable
- Not acceptable at all

Q48 Recommend jobs you might want to apply for

- Very acceptable
- Somewhat acceptable
- Not very acceptable
- Not acceptable at all

End of Block: Block 44

Start of Block: Block 47

Q49 **Please consider the following scenario...**

In an effort to improve the hiring process, some companies are now recording interviews with job candidates. Questions are often pre-recorded by the company and candidates are instructed to record their responses. These videos are analyzed by an artificial intelligence system, which compares the characteristics and behavior of candidates with traits shared by successful employees. Candidates are then given an automated score based on their predicted success that helps the firm decide whether or not they might be a good hire.

Q50 **How FAIR do you think this type of system would be to job candidates?**

- ○ Very fair
- ○ Somewhat fair
- ○ Not very fair
- ○ Not fair at all

Q51 How EFFECTIVE do you think this type of system would be at identifying successful job candidates?

- ○ Very effective
- ○ Somewhat effective
- ○ Not very effective
- ○ Not effective at all

Q52

Do you think it is acceptable for companies to use this type of system when hiring job candidates? Please explain.

- ○ Yes _____
- ○ No _____

End of Block: Block 47

Start of Block: Block 51

Q53 **Would you feel better, worse, or would your opinion of online job platforms and recruiters using artificial intelligence tools not change in each of the following scenarios…**

Q54
If these systems included public data about each job candidate, such as the material they post on social media, in making their evaluations

- ○ Better
- ○ Worse
- ○ No difference

Q55
If companies used these systems to conduct initial screenings of potential candidates but then interviewed those candidates in the traditional way

- ○ Better
- ○ Worse
- ○ No difference

End of Block: Block 51

Start of Block: Block 41

Q56 If you have any additional comments or relevant experiences to share, please add them here.

_____

End of Block: Block 41