An Equity Lens on Artificial Intelligence

INSTITUTE FOR gender AND THE economy

Rotman

An Equity Lens on Artificial Intelligence

Carmina Ravanera and Sarah Kaplan

Institute for Gender and the Economy,
Rotman School of Management, University of Toronto
August 15, 2021

Acknowledgments

The authors would like to acknowledge and thank Professors Joshua Gans, Avi Goldfarb, and Gillian Hadfield for their helpful feedback on early drafts of this project.

This report is available in French and English at:
www.gendereconomy.org/artificial-intelligence

"An Equity Lens on Artificial Intelligence" is co-funded by the Social Sciences and Humanities Research Council and the Government of Canada's Future Skills Program Grant #872-2020-0011.

« Une perspective d'équité en matière d'intelligence artificielle » est cofinancé par le Conseil de recherches en sciences humaines et une subvention dans le cadre du programme Compétences futures du gouvernement du Canada no 872 2020 0011.

Table of Contents

Executive summary

Artificial intelligence (AI) describes machines that can simulate some forms of human intelligence, such as identifying patterns and making predictions and decisions. Today, AI is used by organizations across many sectors for a variety of purposes, from hiring employees, to assessing risk, to making investment recommendations, to recommending criminal sentencing. However, it is well-known that social relations and contexts are reflected and reproduced in technology, and AI is no exception: it has the potential to reinforce underlying biases, discrimination, and inequities. Although AI can be used to benefit marginalized groups, a concerted focus on equity in AI by businesses and governments is necessary to mitigate possible harms. Here we provide a resource for scholars and practitioners for viewing AI through the lens of equity, with the objectives of synthesizing existing research and knowledge about the connection between AI and (in)equity and suggesting considerations for public and private sector leaders to be aware of when implementing AI.

The key insight: AI is a double-edged sword, with potential to both mitigate and reinforce bias:

- Because AI uses statistical prediction methods that can be audited, it has the potential to create outcomes that help groups facing marginalization in situations where human decisions may be clouded by cognitive biases.
- Despite this potential, because inequality and inequity are often reflected in technologies, some AI can and has reinforced marginalization of certain groups, such as women, gender minorities, and racialized and low-income communities. AI-powered products and services may use biased data sets that reproduce this bias; amplify stereotypes and marginalization, sometimes for profit; and/or widen asymmetries of power.
- The reinforcement of inequity and inequality has occurred because of embedded bias or significant omissions in datasets; the complexity and trade-offs involved in aligning AI with social values when profits are also at stake; a lack of transparency from those creating and implementing AI; a lack of accountability to the public or other users of AI; and limited participation by marginalized and diverse groups in the technology sector.
- There are also varied potential impacts of AI and automation on jobs and labour. It is possible that women, racialized, and low-income groups may be more susceptible to job loss or displacement due to automation across an increasing number of blue-, white- and pink-collar jobs.

These results suggest the following considerations for businesses and governments:

- Technology companies and governments can focus on initiatives for equitable representation in AI development
- Creators, researchers and implementors of AI can prioritize aligning AI with social values such as fairness, despite possible trade-offs for efficiency and profit
- Governments can create policies for AI that prioritize accountability and transparency, and require organizations to adhere to these principles
- Governments and companies can work towards economic security for workers who are being doubly impacted by new technologies and a global pandemic through attention on reskilling and/or upskilling programs
- Academic researchers can deepen knowledge on AI and inequity, such as by continuing cross-disciplinary work on the social, political and environmental impacts of AI and developing new and different alternatives that prioritize mitigation of harm.

An Equity Lens on Artificial Intelligence

Introduction

Artificial intelligence, or AI, describes machines that can simulate some forms of human intelligence. Some conceptualizations of AI refer to machines that act indistinguishably from humans, while others focus more on machine learning that can identify patterns, achieve an optimal outcome to a given problem, and/or make predictions and decisions based on prior information.[1] AI uses algorithms to achieve these outcomes, usually "learning" from large data sets and adjusting and improving based on new data. While AI is not a new or recent concept, today it is embedded in people's lives and is only becoming more pervasive.

AI is used by organizations across all sectors for a variety of purposes, such as hiring employees, performing surgeries, tutoring various subjects in schools, making decisions about criminal sentencing, automating driving, and predicting where crime will occur. It is used to make recommendations for what people watch on television or the music they listen to, to select which advertisements to show users on social media, and to display results on online search engines.[2] This is not an exhaustive list; in fact, it may be difficult to find a sector or field today where AI is not involved in some respect. AI has become ubiquitous to the point that some researchers have suggested that it has become a new type of infrastructure. Rather than being a physical product or an institution, such as roads or education systems, it is immaterial and often invisible, but it is nevertheless a "moderator of social relations, practices, and actions," including the distribution of power.[3] Social relations and values are reflected and reproduced in technology, and AI is no exception. This also means that enduring bias, discrimination, and inequality that are deeply rooted in society may also be deeply rooted in this technology.[4]

While such new technologies are altering the way the economy, organizations, and society function, communities across Canada continue to grapple with social, economic, and political inequality and inequities which may amplify or be amplified by the impacts of AI. In 2020 and 2021, the economic impacts of the COVID-19 pandemic have been felt most acutely by groups who were already marginalized, particularly women, racialized communities, and those experiencing low income. Researchers and policy analysts have suggested that recovery policies must be especially attuned to these groups to prevent rising inequality.[5]

Understanding the impacts of AI on the economy and society in Canada, especially in the context of the economic downturn caused by the COVID-19 pandemic, means understanding its impacts on marginalized groups. AI can potentially be used innovatively to lead to outcomes that benefit diverse communities. However, research has also shown that a focus on equitable AI for organizations and policymakers is necessary to mitigate harm.

Many disciplines and organizations are engaged in this discourse, with numerous perspectives coming from fields as diverse as law, computer science, and philosophy. By the time of publication, new research will likely have emerged. This is therefore not intended to cover all literature on this topic, but to provide a resource that highlights debates, questions and issues, explores the "why" and "how" of AI's impact on equity and equality, and conveys considerations for leaders, policymakers, and students on current and future AI use. Marginalization occurs based on numerous social locations and their intersections, such as gender, race, Indigeneity, socioeconomic status, immigration status, ability, and religion. However, the focus in the synthesis is primarily relating to gender and race, as a limited scope was necessary for its purpose. Nevertheless, the authors acknowledge that technology affects those with different social identities in varied ways, and that further research is necessary to understand these impacts more comprehensively.[6]

A double-edged sword: AI and (in)equity

The potential of AI

AI has the potential to result in improved outcomes for people across all sectors. Ideally it removes the possible impacts of human error by making accurate predictions and assisting humans with decision-making. For example, in workplaces, AI used in hiring could unbiasedly find the best candidate for a position; used in healthcare, it can help diagnose diseases such as dementia, and identify treatments; for financial institutions, it can predict likelihood of people defaulting on mortgages; for governments, it can assess refugee cases to result in more just outcomes for claimants.

The prediction power of AI is significant considering that humans' predictions and decisions are clouded by cognitive and other biases. People often do not fully understand why they make certain predictions, and their intuition can be impacted by their prior experiences or opinions. As researchers have noted, statistical prediction techniques as undertaken using AI tend to outperform prediction that is undertaken by humans with expertise and experience.[7] Further, human prediction and decision-making is often opaque – it is difficult to understand and probe the various factors that influence people. Human decision-making is also hard to audit. Thus, to the extent that algorithms can be audited and changed, AI could be a tool for mitigating discrimination, bias, and other forms of marginalization. That is, it is possible they could "turn into a powerful counterweight to human discrimination and a positive force for social good of multiple kinds." [8]

For example, researchers have shown that in the United States, judges commonly make errors about the likelihood that defendants released on bail will commit a crime, in part because of their biases and because of weighing a decision heavily on a current charge as opposed to other relevant information. Judges tend to treat high-risk defendants as low-risk if their current charge is minor and low-risk defendants as high-risk if their current charge is serious. This can lead to errors of two kinds, either unnecessarily high detention rates or high-risk offenders being released, increasing crime rates. Researchers have found that an algorithm created to statistically predict outcomes is more accurate than judges' assessments of these cases, which could result in less unnecessary detention and/or fewer crimes. Further, because algorithms can be adjusted to optimize specific outcomes, this type of AI could be implemented with the goal of decreasing detention that significantly impacts racialized communities. That is, it could be programmed to reduce racial disparities in detention rates while maintaining the same crime rate.[9] Another example is in an algorithmic tool used by Allegheny County's Office of Children, Youth and Families in Pennsylvania. It aims to predict children's risk of harm that call screeners may be unable to do quickly and as accurately, thus aiding by better directing resources to high-risk cases. The New York Times reported in 2018 that with the tool, high-risk calls are screened in more often and the percentage of low-risk cases needlessly investigated has dropped. [10]

Researchers have also suggested that an assistive AI system that has been used in refugee cases could reduce decisions that harm asylum seekers. Again, decisions about refugee cases are made by humans in what could be opaque, uncertain, and biased circumstances, often because of limited data about a claimant. This can result in asylum being denied where it should not have been. AI-based predictions can offer both a prediction and a probability that measures the degree of uncertainty of the prediction. Thus, its use in refugee processing would make explicit the uncertainties in the data informing decision-making for refugee claimants, for instance by demonstrating that there is not enough data to accurately conclude that a claimant will not be persecuted if they return to their home country. If legal structures shifted to resolve doubt in the claimant's favour rather than at their disfavour – which would require political will and change – such an AI system could help ensure that refugees are not denied necessary protections based on humans error.[11]

The overarching insight here is that, because algorithms and the information leading to their predictions can be programmed with specific decision parameters and later audited, they can assist in shifting circumstances of inequity and marginalization that human decision-makers may otherwise not address. However, researchers and advocates have shown that when AI is used in practice, it often does not act as the hoped-for counterweight to discrimination but instead can reinforce it. Societal inequality can and is replicated in AI, and mitigating this can be challenging. For instance, the risk assessment tool used for child services in Allegheny County has been critiqued for disproportionately impacting poor families: the algorithm uses poverty as an indicator of high risk for neglect and abuse, when this is in fact an unfair assumption. [12] Similarly, a study of algorithmic risk assessment used in assessing criminal defendants in Florida found that it incorrectly scored Black defendants as future criminals at twice the rate as white defendants, partly because race is closely correlated to factors deemed high risk, such as joblessness.[13] Further, people may not adhere to the assessments given by predictive tools. A study of a risk assessment tool in Kentucky's criminal justice system suggested that judges are more prone to overriding algorithmic recommendations for Black defendants than defendants of other races, leading to harsher bond conditions for Black defendants compared to similar white defendants.[14] These examples suggest that power relations and inequality embedded in society shape the data that are inputs to the algorithms, the algorithms themselves, and the way algorithms are used. This means that the transformative potential of AI comes with significant risks and challenges, many of which researchers and advocates are currently working to address.

AI and inequitable impacts

Technology and AI systems are not neutral or objective but exist in a social and historical context that can marginalize certain groups, including women, racialized and low-income communities. As such, there are many examples of the ways in which AI systems can reproduce existing biases and marginalization.[15] This may occur through biases or gaps in data used to train algorithms, as well as through the implementation of AI-powered products and services in ways that reinforce stereotypes, marginalization, and global power relations.

Biases and gaps in data

Because bias and inequality exist across all levels of society, it follows that the data on which some AI is built contains such biases, which AI may then reproduce. Attention to the reproduction of gender discrimination through AI is not new, yet it remains a persistent challenge. In the 1970s and 1980s, a medical school in the United Kingdom used a computer program to screen applicants. It ended up rejecting women and those with non-European-sounding names because the algorithm was based on prior data about successful applications where such candidates were poorly represented.[16] Similarly, in 2015, Amazon developed a now-defunct AI recruiting system that was found to have eliminated some women from candidacy, again based on previous hiring patterns in which men dominated.[17]

The same issues have occurred for racial gaps in data. In healthcare, an AI system used for detecting cancerous skin lesions was trained on a database containing mostly light-skinned populations, rendering it less likely to work on those with darker skin.[18] Racial and gender bias in data also intersect to replicate oppression of racialized women. Recently, researchers identified how AI facial recognition software from IBM, Microsoft, and Face++ is less accurate for darker-skinned subjects and especially darker-skinned women, leading to a higher likelihood of their misclassification when compared to white men. Again, this came about because the data on which they were trained did not have diverse racial and gender representation.[19] Depending on what facial recognition software is used for, this error could reinforce the surveillance and mistaken identification of racialized people and especially racialized women.

It is not only "big tech" companies that face this data issue. AI is also used in the public sector in areas such as policing. A recent study has shown that several police jurisdictions in the United States are using racially biased data for predictive policing systems, which in turn make biased predictions about who will commit crimes and where they will be committed. This could differentially focus on communities of racial minorities that are already over-policed.[20] This type of algorithmic policing is already being developed or used by several police forces across Canada as well as in airports, alongside surveillance technology that collects and monitors people's data online or from images.[21]

Reinforcement of stereotypes and marginalization through AI's implementation

Harms through AI not only come about through problematic datasets but also in the way companies and organizations have designed and used it to reinforce stereotypes, marginalization, and erasure of certain groups. For instance, AI-powered facial analysis software has been used to propagate the false idea that people with certain facial features are prone to criminality, and that the software can identify these people. This opens dangerous possibilities for racialized communities who are already stereotyped as being inherently criminal.[22] Researchers have further pointed to how common portrayals of AI, such as stock images and other representations of robots and robotics, tend to be racialized as white with Eurocentric appearances and voices. They suggest that this reproduces conceptions of intelligence, professionalism, and power as associated with whiteness. To the extent that AI and intelligent machines are often created to take over "dirty, dull, or dangerous jobs" that low-income, racialized people disproportionately take on, the whiteness of AI can also be seen as aiming to erase racialized people and their work.[23] Another example is that AI-powered digital assistants, such as Amazon's Alexa, Apple's Siri, and Microsoft's Cortana are named and gendered as women. Researchers have discussed how the gendering of this technology reaffirms the gender division of labour, where women are placed in caregiving and service roles to be commanded to fulfill household tasks.[24] These feminized digital assistants act as both assistants and companions, ensuring users' well-being in a friendly and empathizing manner, further entrenching stereotypes about women in subordination.[25]

In some cases, the reinforcement of stereotypes through AI is explicitly tied to gaining profits. For example, a recent independent audit of Facebook's algorithmic advertising delivery of job ads found that it perpetuates gendered job segregation based on current gender distributions in different job categories: e.g., a job ad for car sales associates was shown to more men than women, while the opposite was true for an ad for jewelry sales associates.[26] Developers could adjust the ad delivery algorithm to compensate for data biases or could remove algorithmic delivery from job ads altogether. However, this would come in conflict with the technology companies' short-term profit motives which are based on clicks on ads.[27] Thus, addressing these biases will require leadership commitment to making change. Research has also investigated how the algorithms behind Google searches, the results of which many consider to be both neutral and factual, reinforce racist and sexist narratives about Black women. In her book Algorithms of Oppression, researcher Safiya Noble details how her Google search of "Black girls" led to first-page results showing pornography and sexual objectification of Black girls and women, revealing how Google's algorithms reinforce oppression by pushing to the top results that will drive profit.[28]

Reinforcement of global power relations

There is also new and ongoing research on the impacts of AI through a lens of global power relations. In particular, researcher Kate Crawford's work discusses how AI is widening power asymmetries. She demonstrates that AI is an "extractive industry" which "depends on exploiting energy and mineral resources from the planet, cheap labor, and data at scale."[29] For instance, creating AI systems and software requires high consumption of energy and minerals, oil, and coal; and those creating its hardware may work in highly surveilled and dangerous conditions in factories. Such costs are borne by labourers in mining communities or outsourced manufacturers, while tech companies accumulate wealth and power.[30] In this way, AI can also be viewed as contributing to inequality on a global scale.

Also important to note is that AI is currently being used for decision-making on immigration and refugee claims by Canada's federal government, for example to evaluate applications and automate related activities. Even though there is the potential for it to create more accurate predictions for claimants, researchers have suggested that the complexity of refugee cases means that governments must assure that these decisions can be handled properly by AI. They have contended that AI used in such state apparatuses not only pose threats to privacy but could be interfering with rights to liberty and freedom from arbitrary detention, which in turn can have significant impacts on the lives of immigrants, migrants and refugees.[31]

Challenges in addressing inequity in AI

Research suggests different explanations for the mechanisms by which AI can perpetuate dynamics of inequality and inequity. To address the biases that may be introduced through data or assumptions used in developing algorithms, purposeful effort must be made to correct them. The questions are: which social values should be written into machines, who decides, how should it be done, and how can makers and users of AI be held accountable? A lack of transparency for the public and a lack of accountability of those developing and implementing AI can pose troubling scenarios for equity. Another influence is that AI is often created by companies that are not representative of marginalized groups and do not have their needs in mind.

Which values? Complexity and trade-offs

Teaching fairness and equity to algorithms is firstly challenging because these are highly complex concepts that must be articulated to a machine. That is, it is a difficult question as to how algorithms can be trained to adhere to social norms and values, many of which involve intricate structures such as law or culture. As researchers have asked, can we program AI such that taking certain wrongful actions would impose costs on it, just as humans avoid certain actions to avoid social penalties such as shame or guilt?[32] Others have suggested that supervising algorithms can act as "moral compasses" for algorithms, monitoring for bias and changing them accordingly.[33] But which values should be prioritized and in which cases? These questions are currently being tackled by those working on issues of ethical AI.

Secondly, there may be trade-offs with accuracy when programming such values into AI.[34] For instance, an algorithm making predictions about who will default on credit loans would have to be explicitly programmed to reduce racial disproportionality, but this could result in less accurate predictions. Yet, not doing so would reinforce inequity, considering histories of disenfranchisement and oppression that have led to increased rates of poverty and financial insecurity for racialized communities. This then brings about the question as to whether AI use should be limited and where, as well as whether humans making the same decisions and predictions are more or less likely to perpetuate bias. As researchers have noted, "In the era of data and machine learning, society will have to accept, and make decisions about, trade-offs between how fair models are and how accurate they are…In fact, such trade-offs have always been implicitly present in human decision-making; the data-centric, algorithmic era has just brought them to the fore."[35] Engaging with these problems and connecting AI to social and historical contexts is essential as AI becomes ever more ubiquitous.

Who decides? Diversity in AI teams

Increased representation of those in marginalized groups in AI development could lead to more equitable outcomes for AI. While there are few studies on the impacts of diverse teams on creating more equitable products specifically,[36] across a variety of sectors there are examples of teams led by women, racialized peoples, and other marginalized groups who have created products and services that are purposely inclusive. For instance, Fenty Beauty, a cosmetics company founded by Barbadian pop star Rihanna, creates makeup shades for those with darker skin tones who have often been excluded from cosmetic lines; AccessNow, an app made by a founder with a disability, indicates to users the accessibility of different locations in a central information source.[37] In the case of AI, products that have been tested and created by a homogenous group logically may not take others' needs or perspectives into account.[38] If a racially diverse team was working on facial recognition software, one can imagine they would have been likely to notice the potential for race-based misclassification.

Around the world, women are under-represented in computer science and computer engineering fields. Globally, only 22% of AI professionals are women. In Canada, despite a relatively high concentration of AI professionals compared to other countries, just 24% are women.[39] Women comprise just 15% of AI research staff at Facebook and only 10% at Google.[40] Data also show that visible minority STEM (science, technology, engineering, and math) graduates in Canada are significantly less likely than non-visible minority STEM graduates to work in a STEM occupation.[41] Further, only 2.5% of Google's full-time workers are Black, as are 4% of Microsoft's.[42]

Research has shown how gender and racial segregation of occupations relegates women and racialized people away from influential jobs in technology. Women and racialized groups in STEM professions have reported feeling constantly excluded in organizational culture and that they experience workplaces that are not flexible to caregiving needs.[43] Technology workplaces may also relegate women to people-focused rather than technical roles, steering them away from jobs that would affect how technology is developed.[44] Women and racialized groups have further reported facing blatant bias, erasure, and marginalization at work.[45] In December 2020, Dr. Timnit Gebru, who was a leading AI research scientist at Google, made media headlines when she was fired for a paper she wrote on risks and harms of language models (i.e., AI trained on text data). Some have shown that her firing revealed abusive tactics, including gaslighting, dismissal, and discrediting—tactics that are commonly used against Black women who aim to advance justice, not only in technology but across society.[46]

How can we address bias in data?

AI functions by "learning" from data sets: algorithms are created to mine data, analyze it, identify patterns and make predictions. Datasets may come from any number of sources including books, photos, health data, government agency data, or Facebook profiles. Societal biases and inequality are often embedded in such data, and AI will not promulgate social values such as fairness unless directly programmed to do so. Thus, if an AI hiring system is based on previous hiring data where few women were hired historically, the algorithm will perpetuate this pattern.[47] Similarly, racialized and low-income groups are more likely to be subject to surveillance, for example when their neighbourhoods are heavily policed. As a result, AI used for predictive policing will be more likely to predict crime in an area which has been policed in the past more than others. As researchers note, "data acts to reinforce [people's] marginality when it is used to target them for suspicion and extra scrutiny. Those groups seen as undeserving are singled out for punitive public policy and more intense surveillance, and the cycle begins again," creating, in their words, a "feedback loop of injustice."[48]

On the other hand, data may also be biased due to omissions. For instance, data that is used for training AI systems on language may come from free public texts, such as books which only enter the public domain when authors died more seventy years previously. Since that literary canon is based on books written mainly by white, Western men, certain vocabulary and perspectives are omitted.[49] Another example is that AI systems trained to recognize gender tend to have little or no data on transgender and non-binary people, potentially leading to misgendering.[50] Datasets may also omit entire populations who do not have internet histories or social media presence, credit card histories, or electronic health records, leading to skewed results. Those omitted are often racialized communities, people with low socioeconomic status, and others on the margins.[51]

Ensuring fairness in the data used for AI is a complex problem considering how inequality and inequity influence people's lives in complex and overlapping ways. It is not effective to simply remove variables such as gender and race to avoid discrimination by algorithms, because proxy variables may end up creating the same impacts.[52] Recently, Apple's credit card was in the news because its algorithm appeared to give smaller lines of credit to women than men, even to those who were married and sharing assets. Initially, Apple and its banking partners said the results could not be biased because gender was not a variable in the algorithm, and that the credit scoring was gender blind. Ultimately, while an investigation into the Apple Card concluded it did not discriminate against women, experts noted that creating a gender-blind algorithm would not prevent gender discrimination from happening inadvertently.[53] As another example, a tool that a family services department uses to analyze children's risk of harm may not take race into account, but other variables that are included such as poverty levels correlate strongly with race, so racial bias may not be removed even if race is not explicitly measured. [54] [55]

How can we assure accountability?

Research suggests that, in general, there is a lack of accountability to people who are being harmed by AI systems. That is, the scope of AI's impacts as well as who is responsible for creating and mitigating them is often unclear. This suggests the need for more assessments and audits on what AI-driven products and services mean for people, including evaluations on how fair they are.[56]

The first challenge for accountability is transparency. There is often a lack of transparency around AI systems' purposes, their algorithms, and the data they use. This is sometimes called the "black box" problem, where the inscrutability of these systems can prevent public understanding of risks and impacts.[57] If people are not aware of how algorithms are being used on them, then it is not possible to question or change their predictions and decisions.[58] Even when AI is used by the public sector for processes as wide-ranging as surveillance and immigration decisions, the public may not have knowledge or ownership of it. As such, some researchers have proposed complete transparency of AI, where algorithms and/or data as well as the results they are aiming to achieve are available for public scrutiny. Helping the public understand how algorithms and AI are influencing their lives can be a step towards mitigating potentially harmful outcomes.[59]

At the same time, there are debates around how transparent AI systems feasibly can be. Some researchers suggest that requiring such transparency would stifle the ability for companies to innovate because intellectual property would not be protected. Further, since algorithmic code is generally inscrutable for the average person, transparency may not necessarily increase people's trust of AI nor decrease its harms.[60] Software is also proprietary, and transparency may not be possible for security, safety, or legal reasons.[61] Finally, governments or companies using algorithms may not want to share them for fear that people will figure out how to circumvent them or manipulate outcomes.[62] Thus, transparency to benefit the public will need to be balanced with the benefits of intellectual property and innovation.

The second challenge for accountability is a lack of appropriate governance structures. Researchers are currently working on governance structures and auditing procedures that can be put in place within technology companies that explicitly evaluate an AI system in terms of social benefits and values.[63] Even though many companies may already conduct audits on their AI, these are unregulated and not standardized, thus making it hard for users to assure any results of the audits are used to change the algorithms.[64] Further, third-party researchers conducting audits tend to face many challenges: companies such as Google and Facebook create barriers to outside audits by prohibiting creating fake profiles for research purposes [65] and often do not make necessary data available. Providing such access involves a balance of privacy and auditability. Third-party auditing is also costly and involves substantial time and effort.[66]

Some researchers have proposed that there should be regulatory mechanisms ensuring companies and governments are held accountable for unfair and unjust impacts. If a law or policy were in place such that those who created and owned algorithms were held directly responsible for its outcomes, this might help ensure that AI is developed with ex ante considerations of its social impacts, rather than through ex post efforts to address harms after they have occurred.[67] Policies that allow the public and civil society to intervene into AI development and its use in the public sector, such as through consultations that are inclusive of people outside of the technology sector, may also allow for more accountability.[68]

AI, automation and employment

In addition to the conversations about bias in AI, there is a parallel discussion about the potentially inequitable effects that AI and automation more broadly will have on employment. Note that AI and automation are not identical: while automation tends to refer to any tasks that are done by machines, particularly mundane and repetitive ones, AI specifically refers to work done by machines that imitates human intelligence, making predictions or decisions. However, AI can be a form of or be directly involved in automation, so the evidence on automation and socioeconomic inequity is important here.

Research has suggested that automation will restructure or displace many jobs. One study has estimated that around 47% of employment in the US is at risk of automation[69] , another estimates 42% of Canadian employment.[70] Researchers have also estimated that in the United States, each additional "robot" (fully autonomous, programmable machine) per thousand workers reduces the employment-to-population ratio by 0.2 percent and wages by 0.42 percent.[71] Further, analyses suggest that over the last thirty years, job displacement in the US as a result of automation was 16 percent, while there was only a 10 percent increase in new opportunities. These new opportunities often benefitted high-skilled workers, leading to job loss and stagnating wages for lower-skilled workers.[72] Workers whose jobs tend to be standardized and involve routine tasks, such as those in factories, retail, or some office employment, are more likely to be displaced by technology than jobs requiring manual dexterity, technological skills, or creative or emotional labour.[73]

Some studies have therefore suggested that women and racialized groups will be relatively more impacted, due to their concentration in specific jobs and industries. A recent study from the United Kingdom found that women hold 70% of jobs at high risk of becoming automated.[74] Women are overrepresented as cashiers, secretaries, bookkeeping clerks, receptionists, and accountants, among other occupations which are predicted to be at a high risk. Further, women are less likely than men to fill the high-paid jobs that increases in automation will require, such as computer scientists. Race and Indigeneity may also have an impact. In Canada, around 250,000 jobs that are held by Indigenous peoples are at risk of automation. Indigenous employment is more concentrated in industries such as accommodation and food services; retail; construction; and transportation, relative to those who are not Indigenous.[75] Similarly, a recent study from the United States found that 31% of Latino workers and 27% of Black workers are concentrated in 30 occupations that are at high risk of automation, compared to 24% of white workers.[76] Automation of labour may therefore particularly disadvantage groups that already face poverty and marginalization at disproportionate rates, and existing economic and social inequity may be exacerbated.[77]

On the other hand, automation can cause employment growth or complement labour in various ways. In the past, automation in textile, steel, and automotive industries led to an increase in jobs in these sectors by reducing costs and thus increasing demand.[78] However, jobs do not normally grow proportionally as fast as revenues. Across sectors, people are already working with automated machines, such as in healthcare, education, and the legal sector. [79] The question about whether or not jobs will be replaced by automation is somewhat difficult to predict: whereas increased efficiency created by technology may reduce the number of jobs, sales growth driven by the increased efficiency may increase the demand for jobs. This can be seen in the case of Amazon which is increasingly using robots in its fulfillment centers to increase the efficiency of staff, but which is still hiring thousands of new staff to keep up with increased demand. Researchers have also shown that AI-powered digital platforms can increase the pool of employers and workers by removing barriers for employers to find workers, reducing transaction costs, and improving matching between employers and employees.[80]

What is new with predictive AI is that it increasingly can substitute for many white-collar jobs that involve judgment and prediction, such as those in real estate, investment advising, legal professionals, and software developers. This is because skills such as document review, reading, writing, coding, and even teaching are becoming more commonly automated.[81] [82] Taxis and other driving services – a sector dominated by men – have also been "deconstructed" and transformed by technology companies such as Uber and Lyft. The impact is ambiguous as this substitution of tasks that involve prediction may also increase the need for labour for complementary tasks that are upstream or downstream from the automated task. One example is that using automation in radiology to interpret image results speeds up this task, but it also may result in an increased need for labour to communicate results and to decide on the actions to take following the prediction that the AI has made.[83]

There is also research discussing how employers are using algorithms to control and direct workers. This has implications for how work is done as well as for new types of labour. Employers have been using algorithms to record worker behaviour in real time, e.g., calculating how long workers take to do tasks or monitoring how employees are communicating. They can also be used to measure productivity, predict and rate workers' performances, make recommendations for employees when doing their tasks, and more. A familiar example here is Uber, which uses algorithms to suggest to drivers to rest if their driving seems erratic, or to remove them from the platform if they receive many low ratings. Such algorithmic use may lead to new occupations, such as analysts for the new data gathered through algorithms and engineers, developers, and technology support staff who design and manage algorithms. Simultaneously, this algorithmic use may have negative impacts on workers, including privacy invasion, increased stress and frustration, and lack of awareness of reasons for being fired or losing wages.[84]

Thus, it will be important to study how these new technologies are affecting various professions, as well as to ensure worker adaptability to new norms and people's ability to access new skills and training.[85] That is, the workforce must have the skills required to take on new and different labour created by automation, as those who do not have skillsets to adjust and respond to technological change may face the most impact.[86]

These changes in AI and automation are now intersecting with the impact of the COVID-19 pandemic. Women, racialized, and low-income groups have borne the brunt of job losses and economic instability from the COVID-19 pandemic's fallout in Canada.[87] [88] Caregiving responsibilities have further hindered women's participation in the workforce since the pandemic, and may hinder their ability to reskill or upskill to adjust to new technologies.[89] Policymakers and organizations therefore have an opportunity to focus on strategies and policies that will ensure gainful, decent employment is attainable for all, especially in the context of a rapidly changing digital economy.[90]

Implications for research, policy and practice

Given these considerations, research, policy and practice can be mobilized to understand and address the impacts of AI and ensure that it is implemented in an equitable and just way.

Regulation and policy

It is widely recognized that governments have some catching up to do to ensure AI is developed and used in a way that reduces harm for marginalized groups. More regulation is not a catch-all solution because it can often fall behind the fast pace of AI development. Thus, some researchers have suggested that regulation will be effective only if developers and others in the technology industry are making concerted efforts to reduce unfairness and injustice in the algorithms they create.[91] Nevertheless, new laws or policies regulating AI could help protect people and craft a future that is more inclusive, where AI can "capitalize on human strengths" and complement humans rather than aiming to replace them.[92]

To establish greater accountability, new policies or laws could ensure that it is clear who created, owns, and controls AI, thus attributing responsibility where there currently is little.[93] Others have suggested that audits and impact assessments of AI should be mandatory and undertaken before and during AI implementation.[94] Further, although there are debates around transparency, standard processes of "explainability" can still be put in place so that organizations provide justifications for decisions made about and by AI (including its purpose, design, and datasets) and disclose risks, such as through public records and published reports.[95] [96] Together these ideas suggest the need for stronger oversight of AI, where governments create regulatory bodies and/or frameworks that account for social and historical contexts in AI and comprehensively regulate practices.[97]

In many countries, and through the work of AI ethics organizations, such frameworks and initiatives are already being implemented. In the United States, the Algorithmic Accountability Act was introduced in 2019, proposing that large companies must evaluate their algorithms for bias and the risks they pose to users.[98] In 2021, the EU created a proposal for an Artificial Intelligence Act, "the first ever legal framework on AI."[99] Further, some researchers and advocates have recommended a comprehensive global framework that will broadly govern AI use, similar to those of universal human rights from the United Nations.[100] The Toronto Declaration of 2018, led by Amnesty International and Access Now, is one example of a comprehensive statement of calls to action to uphold human rights in AI.[101]

Canada is in the process of developing its own policies and frameworks. Following a $125 million investment in a Pan-Canadian Artificial Intelligence Strategy in 2017, the federal government developed a Directive on Automated Decision-Making and a public Algorithmic Impact Assessment.[102] It further created an Advisory Council on Artificial Intelligence in 2019, although this council has notably been critiqued for lack of representation of racialized and other marginalized groups.[103] The Office of the Privacy Commissioner of Canada has also recently made recommendations for updating the Personal Information Protection and Electronic Documents Act to better regulate AI,[104] and in Ontario at the time of writing this report, public consultations are underway to create a provincial Trustworthy Artificial Intelligence Framework.[105] The effects of these policy efforts for fairness, transparency and accountability are yet to be seen. Note, however, that as of June 2021, there is no legislation pending on AI governance in Canada.[106]

Industry standards

Beyond regulation, researchers and advocates are working towards AI that centres social considerations instead of discovering and addressing problems after the fact. Indeed, AI can be developed and created to align with the goals of reducing systemic inequality and inequity, but as mentioned earlier it is not an easy task to program the complex norms and values into AI that humans understand when they are making predictions and decisions.[107] Another question arises from this challenge: if such AI has not yet been robustly developed, what are the circumstances in which AI should not be used, and what are the best alternatives?

This also becomes a moral question involving trade-offs and values. Purposely aligning AI with social values means organizations may have to prioritize equity and other social considerations over profit or efficiency. Such a shift may require significant time and money, such as the costs of conducting research on social impacts or the potential revenue losses from not implementing new AI due to ethical reasons. Thus, there is a need for industry cooperation and collective action involving the establishment of standards, so that safe and responsible AI becomes accepted as a norm.[108]

Representation

If a lack of representation of marginalized communities in the development of technology hinders inclusive and fair AI, this problem could be mitigated through more equitable hiring and promotions. There have been many studies on solutions for making workplaces, including technology companies, more inclusive to women and racialized groups. These include being more flexible to workers who need to prioritize caregiving (usually women); transforming non-inclusive hiring and recruiting practices that favour certain candidates (such as young men or people who have trained in elite schools); and working towards anti-racist and anti-sexist policies and culture.[109] Schools teaching STEM also can usefully transform their cultures, as studies have shown that young women may be treated poorly by their men classmates in engineering and made to feel like they do not belong there.[110]

Another possibility that could result in teams being better equipped to address issues of equity would be by ensuring the involvement of different disciplines in AI development. Researchers have suggested that hiring humanities and social sciences scholars who have a comprehensive understanding of socioeconomic inequality, history, and critical theory involving gender, race, and other social identities could help organizations identify and solve ethical problems related to implementing AI.[111]

Reskilling and upskilling

The COVID-19 pandemic has resulted in the decimation of high-contact industries such as tourism, hospitality, food services, and retail. These are also industries that disproportionately employ women. Concurrently, some research suggests that the high-contact jobs that are more likely to be replaced by automation tend to be done by women as well as racialized, Indigenous, and low-income groups. A focus on skills development and reskilling from both companies and governments may address these issues by helping ensure that some groups are not left behind.[112] These initiatives could include designing new training opportunities; partnerships between government, companies, and post-secondary schools to ensure equitable access to reskilling or upskilling programs; and government and corporate support for initiatives and organizations that work to involve equity-seeking groups in STEM.[113] It could also involve robustly funding the care sector, i.e. childcare and eldercare. This could provide hundreds of thousands of new jobs that are not susceptible to automation, especially for women, and would facilitate services for which the pandemic has revealed Canadian families are urgently in need.[114]

Towards equitable AI

Current research on AI suggests that this is an important moment for leaders, policymakers, and researchers to prevent the reinforcement of inequality and inequity through technology. Much work is already being done by organizations across the country and around the world to advocate for equitable AI. In the academic landscape, there are several avenues for research, some of which are listed below:

- Conducting cross-disciplinary and interdisciplinary work on AI: Technical AI researchers can collaborate with researchers in social sciences and the humanities for a better understanding of the impacts of AI on groups facing marginalization, including on a global scale, as well as how these can be changed. Stronger connections can also be forged between marginalized groups and those creating the technology that impacts them. [115] [116]
- Creating new AI that aligns with social values: Innovative thinking and research is ongoing to understand how to implement algorithms and AI that align with values such as fairness and correct for biases and other harms. Researchers have explored how a focus on profit in AI is linked to reinforcement of marginalization. Future research could further develop alternatives focusing on other considerations.[117]
- Investigating how regulation and policy can better mitigate harmful impacts: Research can continue to be undertaken into the optimal methods of regulating AI (i.e., laws and policies) and ensuring such regulations manage the potential tension between mitigating marginalization while also not stifling innovation.

AI has changed the economy and society. While it has the potential to better many lives, it can lead to significant harms. Since technology is created within contexts and histories of inequality and power, these are easily, though perhaps unintentionally, reproduced through AI systems. Numerous examples have shown that AI that is applied without attention to such contexts may reinforce discrimination and bias against women, racialized communities, and others experiencing inequity and inequality. AI also has the potential to place marginalized communities at further socioeconomic risk by replacing or restructuring jobs, especially considering that the economic stability of many livelihoods has already been damaged by COVID-19. These findings suggest the need for ongoing work in the following areas for businesses, policymakers and researchers:

- Technology companies and governments can focus on initiatives for equitable representation, especially in AI development,
- Creators, researchers and implementors of AI can prioritize aligning AI with social values such as fairness and equity, despite trade-offs for efficiency and profit,
- Governments can create policies and laws for AI that prioritize accountability and transparency, and require tech organizations to adhere to these principles,
- Governments and companies can work towards economic security for workers who are being doubly impacted by new technologies and a global pandemic through attention on reskilling and/or upskilling programs,
- Academic researchers can deepen knowledge on AI and inequity, such as by continuing interdisciplinary work on the social, political and environmental impacts of AI and developing new and different alternatives that prioritize mitigation of harm.

Research suggests that preventing the reinforcement of inequity through AI requires cross-sectoral and interdisciplinary work, from governments to academia to companies. That is, solutions will involve a combination of regulation and policy, new research and development towards fairer AI, shifting norms around who develops and makes decisions about AI, and ensuring accountability towards those who are most impacted. Without concerted efforts, the reinforcement of systemic bias and discrimination will continue to perpetuate through these technology systems that are becoming ubiquitous.

Endnotes

1. Russell, S.J. and Norvig, P. (2013). Artificial Intelligence: A Modern Approach. Pearson.

2. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M. and Teller, A. (2016). Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence, Stanford University. Retrieved on March 19, 2020 from https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf

3. Jaume-Palasi, L. (2019). Why We Are Failing to Understand the Societal Impact of Artificial Intelligence. Social Research 86(2), 477-498.

4. Birhane, A. and van Dijk, J. (2020). Robot Rights?: Let's Talk about Human Welfare Instead. AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics and Society.

5. Kabeer, N., Razavi, S. and van der Meulen Rodgers, Y. (2021). Feminist Economic Perspectives on the COVID-19 Pandemic. Feminist Economics. Retrieved on March 2, 2021 from https://doi.org/10.1080/13545701.2021.1876906

6. This research overview was based on a review of the existing scholarly research and other reports examining questions of: What are the impacts of AI on groups facing marginalization and inequity, and why do these occur? What is the future of Canada's workforce in the context of AI and automation? How can our society effectively create an inclusive and equitable economy, in concert with the rise of AI? What research gaps exist and how can these be addressed to create more equitable AI? The synthesis is meant to investigate these questions in a way that is generative and informative rather than comprehensive. Considering there are many groups outside of academia doing work on the ethical and societal impacts of AI, the authors went beyond scholarly publications to include policy reports and other articles produced by researchers and advocates for equitable AI. To identify the literature to review and synthesize, searches for research and documents were conducted on databases such as JStor, EBSCOHost, Google Scholar, and Google search, and citations were examined to find further sources. Search terms included the various terms used to describe AI and its processes, such as "artificial intelligence", "machine learning", and "algorithms", as well as terms that would speak to its impacts on equality and equity, such as "bias", "gender", "race", "stereotypes", "inequality" and "employment". As this is a quickly developing field of research, searches were limited to the previous five years. We also consulted a number of academic experts in the social impacts of AI to identify additional research sources and to obtain feedback on earlier drafts.

7. Sunstein, C.R. (2019). Algorithms, Correcting Biases. Social Research 86(2): 499-511.

8. Kleinberg, J., Ludwig, J., Mullainathan, S. and Sunstein, C.R. (2018). Discrimination in the Age of Algorithms. Journal of Legal Analysis 10(2018), 113-174.

9. Sunstein, C.R. (2019). Algorithms, Correcting Biases. Social Research 86(2): 499-511.

10. Hurley, D. (2018). Can an Algorithm Tell When Kids Are in Danger? New York Times Magazine. Accessed April 28, 2021 from https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html

11. Cameron, H.E., Goldfarb, A. and Morris, L. (2021). Artificial intelligence for a reduction of false denials in refugee claims. Working paper.

12. Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

13. Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine Bias. ProPublica. Retrieved on April 28, 2021 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

14. Albright, A. (2019). If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions. https://thelittledataset.com/about_files/albright_judge_score.pdf

15. Birhane, A. and van Dijk, J. (2020). Robot Rights?: Let's Talk about Human Welfare Instead. AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics and Society.

16. Garcia, M. (2016-2017). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. World Policy Journal.

17. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved on March 23, 2020 from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

18. Noor, P. (2020). Can we trust AI not to further embed racial bias and prejudice? BMJ 368.

19. Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81(1-15).

20. Richardson, R., Schultz, J.M. and Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. New York University Law Review Online 94(15).

21. Roberson, K., Khoo, C. and Song, Y. (2020). To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada. The Citizen Lab, University of Toronto. Retrieved on February 12, 2021 from https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf

22. Agüera y Arcas, B., Mitchell, M. and Todorov, A. (2017). Physiognomy's New Clothes. Retrieved on Feb. 18, 2021 from https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a

23. Cave, S. and Dihal, K. (2020). The Whiteness of AI. Philosophy & Technology 33(685-703).

24. Schiller, A. and McMahon, J. (2019). Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence. New Political Science 41(2).

25. Costa, P. (2018). Conversing with Personal Digital Assistants: on Gender and Artificial Intelligence. Journal of Science and Technology of the Arts 10(3), 59-72.

26. Imana, B., Korlova, A. and Heidemann, J. (2021). Auditing for discrimination in Algorithms Delivering Job Ads. Proceedings of the web Conference 2021, April 19-23, 2021. Retrieved on May 5, 2021 from https://ant.isi.edu/datasets/addelivery/Discrimination-Job-Ad-Delivery.pdf

27. Hao, K. (2021) Facebook's ad algorithms are still excluding women from seeing jobs. MIT Technology Review. Retrieved on May 1, 2021 from https://www.technologyreview.com/2021/04/09/1022217/facebook-ad-algorithm-sex-discrimination/

28. Noble, S.U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

29. Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

30. Crawford, K. and Joler, V. (2018). Anatomy of an AI System. Retrieved on May 1, 2021 from https://anatomyof.ai/

31. Molnar, P. and Gill, L. (2018). Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System. The Citizen Lab, University of Toronto. Retrieved on April 4, 2020 from https://ihrp.law.utoronto.ca/sites/default/files/media/IHRP-Automated-Systems-Report-Web.pdf

32. Hadfield-Menell, D.H. and Hadfield, G.K. (2019). Incomplete Contracting and AI Alignment. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Retrieved on March 20, 2021 from https://dl.acm.org/doi/10.1145/3306618.3314250

33. Wellner, G. and Rothman, T. (2020). Feminist AI: Can We Expect Our AI Systems to Become Feminist? Philosophy and Technology 33, 191-205.

34. Kearns, M. and Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.

35. Kearns, M. and Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.

36. The research on the effects of diverse work teams on their products and processes suggests they depend on circumstances, such as the types of tasks teams are undertaking. Some studies suggest culturally or racially diverse teams result in higher levels of product innovation as well as more creative performance and thinking, while others have found that it may lead to communication costs or conflict that reduces teams' ability to innovate.

37. See "Creating a More Accessible World: Q&A with Maayan Ziv, Founder and CEO AccessNow (2018), https://news.microsoft.com/en-ca/2018/12/03/creating-a-more-accessible-world-qa-with-maayan-ziv-founder-and-ceo-accessnow/"; "Highlighting Fenty Beauty's Diversity-Based Business Model (2020), https://www.gendereconomy.org/fenty-beauty-diversity-based-business-model/"; and "Backstage Capital: Investing in Black women founders (2020), https://www.gendereconomy.org/backstage-capital-investing-in-black-women-founders/" for some examples.

38. Gebru, T. (2020). Race and Gender. The Oxford Handbook of Ethics of AI. Eds. Dubber, M.D., Pasquale, F. and Das, S. Retrieved from https://arxiv.org/abs/1908.06165

39. World Economic Forum (2018). The Global Gender Gap Report 2018. Retrieved on March 4, 2021 from http://www3.weforum.org/docs/WEF_GGGR_2018.pdf

40. West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved on March 24, 2020 from https://ainowinstitute.org/discriminatingsystems.pdf

41. Frank, K. (2019). A Gender Analysis of the Occupational Pathways of STEM Graduates in Canada. Statistics Canada. Retrieved on March 19, 2020 from https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2019017-eng.htm#a4_2

42. West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved on March 24, 2020 from https://ainowinstitute.org/discriminatingsystems.pdf

43. Seron, C., Silbey, S., Cech, E. and Rubineau, B. (2018). I am Not a Feminist, but…: Hegemony of a Meritocratic Ideology and the Limits of Critique Among Women in Engineering. Work and Occupations 45(2).

44. Alegria, S. (2019). Escalator or Step Stool? Gendered Labor and Token Processes in Tech Work. Gender & Society 33(5), 722-745.

45. Seron, C., Silbey, S., Cech, E. and Rubineau, B. (2016). Persistence is Cultural: Professional Socialization and the Reproduction of Sex Segregation. Work and Occupations 43(2).

46. Turner, K., Wood, D., and D'Ignazio, C. (2021). The Abuse and Misogynoir Playbook. The State of AI Ethics, Montreal AI Ethics Institute. Retrieved on February 2, 2021, from https://montrealethics.ai/wp-content/uploads/2021/01/The-State-of-AI-Ethics-Report-January-2021.pdf

47. Williams, B.A., Brooks, C.F. and Shmargad, Y. (2018). How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications. Journal of Information Policy 8, 78-115.

48. Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

49. Levendowski, A. (2018). How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem. Washington Law Review 579.

50. Gebru, T. (2020). Race and Gender. The Oxford Handbook of Ethics of AI. Eds. Dubber, M.D., Pasquale, F. and Das, S. Retrieved from https://arxiv.org/abs/1908.06165

51. Katyal, S.K. (2019). Artificial Intelligence, Advertising, and Disinformation. Advertising and Society Quarterly 20(4).

52. Wellner, G. and Rothman, T. (2020). Feminist AI: Can We Expect Our AI Systems to Become Feminist? Philosophy & Technology 33: 191-205.

53. Knight, W. (19 November, 2019). The Apple Card Didn't 'See' Gender—and That's the Problem. Wired. Retrieved on June 29, 2021 from https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/

54. Hirschman, D. and Bosk, E.A. (2019). Standardizing Biases: Selection Devices and the Quantification of Race. Sociology of Race and Ethnicity 115(6):974-1001.

55. Kearns, M. and Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.

56. Donahoe, E. and MacDuffee Metzger, M. (2019). Artificial Intelligence and Human Rights. Journal of Democracy 30(2).

57. Coglianese, C. and Lehr, D. (2019). Transparency and Algorithmic Governance. Faculty Scholarship at Penn Law 2123.

58. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

59. Wellner, G. and Rothman, T. (2020). Feminist AI: Can We Expect Our AI Systems to Become Feminist? Philosophy & Technology 33: 191-205.

60. De Fine Licht, K. and de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. AI & Society.

61. Askell, A., Brundage, M., and Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. Working paper. Retrieved on March 24, 2021 from https://arxiv.org/abs/1907.04534

62. Brauneis, R. and Goodman, E.P. (2018). Algorithmic Transparency for the Smart City. Yale J.L. & Tech.

63. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020). Closing the AI Accountability Gap. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44.

64. Ng, A. (2021). Can Auditing Eliminate Bias from Algorithms? The Markup. Retrieved on March 2, 2021, from https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms

65. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

66. Imana, B., Korlova, A. and Heidemann, J. (2021). Auditing for discrimination in Algorithms Delivering Job Ads. Proceedings of the web Conference 2021, April 19-23, 2021. Retrieved on May 5, 2021 from https://ant.isi.edu/datasets/addelivery/Discrimination-Job-Ad-Delivery.pdf

67. Pasquale, F. (2020). New Laws of Robotics: Defending human expertise in the age of AI. Belknap Press.

68. McKelvey, F. and MacDonald, M. (2019). Artificial Intelligence Policy Innovations at the Canadian Federal Government. Canadian Journal of Communication Policy Portal 44, 43-50.

69. Frey, C.B. and Osborne, M.A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change 114, 254-280.

70. Lamb, C. (2016). The Talented Mr. Robot: The impact of automation on Canada's workforce. Brookfield Institute for Innovation + Entrepreneurship. Retrieved on March 23, 2021 from https://brookfieldinstitute.ca/wp-content/uploads/TalentedMrRobot_BIIE-1.pdf

71. Acemoglu, D. and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. Journal of Political Economy 128(6), 2188-2244.

72. Acemoglu, D. and Restrepo, P. (2020). Unpacking Skill Bias: Automation and New Tasks. AEA Papers and Proceedings 110, 356-361.

73. Dahlin, E. (2019). Are Robots Stealing Our Jobs? Socius: Sociological Research for a Dynamic World 5(1-14).

74. Roberts, C., Parkes, H., Statham, R. and Rankin, L. (2019). The Future is Ours: Women, Automation and Equality in the Digital Age. The Centre for Economic Justice. Retrieved on March 25, 2020 from https://www.ippr.org/files/2019-07/the-future-is-ours-women-automation-equality-july19.pdf

75. Digital Differences: The impact of automation on the Indigenous economy in Canada. Diversity Institute, Canadian Council on Aboriginal Business, and Future Skills Centre. Retrieved on Feb. 8, 2021 from https://fsc-ccf.ca/research/digital-differences-the-impact-of-automation-on-the-indigenous-economy-in-canada/

76. Broady, K. (2017). Race and Jobs at High Risk to Automation. Joint Center for Political and Economic Studies. Retrieved on Feb. 8, 2021 from https://jointcenter.org/wp-content/uploads/2017/12/Race-and-Jobs-at-High-Risk-to-Automation-12-18-17-11_30-am.docx-2_0.pdf

77. Hegewisch, A., Childers, C. and Hartmann, H. (2019). Women, Automation, and the Future of Work. Institute for Women's Policy Research. Retrieved on March 19, 2020 from http://iwpr.org/wp-content/uploads/2019/03/C476_Automation-and-Future-of-Work.pdf

78. Bessen, J. (2019). Automation and Jobs: When Technology Boosts Employment. Boston University School of Law, Law & Economics Paper No. 17-09.

79. Pasquale, F. (2020). New Laws of Robotics: Defending human expertise in the age of AI. Belknap Press.

80. Roos, G. and Shroff, Z. (2017). What will happen to the jobs? Technology-enabled productivity improvement – good for some, bad for others. Labour & Industry 3, 165-192.

81. Baldwin, R.E. (2019). The Globotics Upheaval: Globalization, Robotics and the Future of Work. Oxford University Press.

82. Schlesinger, J. (June 30, 2021). Why everyone from students to NASA is using this Canadian invention. The Globe and Mail. Retrieved on July 4, 2021 from https://www.theglobeandmail.com/business/technology/article-why-everyone-from-students-to-nasa-is-using-this-canadian-invention/

83. Agrawal, A., Gans, J.S. and Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. Journal of Economic Perspectives 33(2), 31-50.

84. Kellogg, K.C., Valentine, M.A. and Christin, A. (2020) Algorithms At Work: The New Contested Terrain of Control. Academy of Management Annals 14(1), 366-410.

85. Latham, S. and Humberd, B. (2018). Four Ways Jobs Will Respond to Automation. MIT Sloan Management Review. Retrieved on May 3, 2021 from https://sloanreview.mit.edu/article/four-ways-jobs-will-respond-to-automation/

86. Acemoglu, D. and Restrepo, P. (2018). Artificial Intelligence, Automation and Work. NBER Working Paper Series.

87. Lim, J. (2020). Survey of labour force shows racialized unemployment gap due to COVID-19. Retrieved on March 22, 2021 from https://ipolitics.ca/2020/08/07/july-labour-force-survey-shows-racialized-gap-in-unemployment-due-to-covid-19/

88. Desjardins, D. and Freestone, C. (2021). COVID Further Clouded the Outlook for Canadian Women at Risk of Disruption. RBC Economics. Retrieved on March 10, 2021 from https://thoughtleadership.rbc.com/covid-further-clouded-the-outlook-for-canadian-women-at-risk-of-disruption/

89. Jivani, A.I. (2020). Gender Lens to the Future of Work. Daedalus 149 (1).

90. Roos, G. and Shroff, Z. (2017). What will happen to the jobs? Technology-enabled productivity improvement – god for some, bad for others. Labour & Industry 3, 165-192.

91. Kearns, M. and Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.

92. Pasquale, F. (2020). New Laws of Robotics: Defending human expertise in the age of AI. Belknap Press.

93. Pasquale, F. (2020). New Laws of Robotics: Defending human expertise in the age of AI. Belknap Press.

94. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

95. De Fine Licht, K. and de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. AI & Society.

96. Coglianese, C. and Lehr, D. (2019). Transparency and Algorithmic Governance. Faculty Scholarship at Penn Law. 2123.

97. Collett, C. and Dillon, S. (2019). AI and Gender: Four Proposals for Future Research. Cambridge: The Leverhulme Centre for the Future of Intelligence. Retrieved on Feb. 20, 2021, from http://lcfi.ac.uk/media/uploads/files/AI_and_Gender___4_Proposals_for_Future_Research_210619_p8qAu8L.pdf

98. H.R.2231 - Algorithmic Accountability Act of 2019 (2019). Retrieved on March 24, 2021 from https://www.congress.gov/bill/116th-congress/house-bill/2231

99. European Commission (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Retrieved on May 12, 2021 from https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

100. Donahoe, E. and MacDuffee Metzger, M. (2019). Artificial Intelligence and Human Rights. Journal of Democracy 30(2).

101. The Toronto Declaration. Retrieved on March 2, 2021 from https://www.torontodeclaration.org/declaration-text/english/

102. Government of Canada (2019). Directive on Automated Decision-Making. Retrieved on March 3, 2021 from https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

103. Poetranto, I., Heath, V. and Molnar, P. (2019). Canada's Advisory Council on AI lacks diversity. Retrieved on March 1, 2021 from https://www.thestar.com/opinion/contributors/2019/05/29/canadas-advisory-council-on-ai-lacks-diversity.html

104. Office of the Privacy Commissioner of Canada (2020). A Regulatory Framework for AI: Recommendations for PIPEDA Reform. Retrieved on March 1, 2021 from https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/

105. Government of Ontario (2021). Consultation: Ontario's Trustworthy Artificial Intelligence (AI) Framework. Retrieved on May 1, 2021 from https://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework-consultations

106. McKelvey, F. and MacDonald, M. (2019). Artificial Intelligence Policy Innovations at the Canadian Federal Government. Canadian Journal of Communication Policy Portal 44, 43-50.

107. Jaume-Palasi, L. (2019). Why We Are Failing to Understand the Societal Impact of AI. Social Research: An International Quarterly 86(2).

108. Askell, A., Brundage, M., and Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. Working paper. Retrieved on March 24, 2021 from https://arxiv.org/abs/1907.04534

109. Wynn, A.T. and Correll, S.J. (2018). Puncturing the pipeline: Do technology companies alienate women in recruiting sessions? Social Studies of Science 48(1).

110. Seron, C., Silbey, S., Cech, E. and Rubineau, B. (2016). Persistence is Cultural: Professional Socialization and the Reproduction of Sex Segregation. Work and Occupations 43(2).

111. Noble, S.U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

112. Illanes, P., Lund, S., Mourshed, M., Rutherford, S. and Tyreman, M. (2018). Retraining and reskilling workers in the age of automation. McKinsey Global Institute. Retrieved on April 8, 2020 from https://www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation

113. Dessanti, C. (2020). The She-Covery Project: Confronting the Gendered Economic Impacts of COVID-19 in Ontario. Ontario Chamber of Commerce. Retrieved on February 12, 2021 from https://occ.ca/wp-content/uploads/OCC-shecovery-final.pdf

114. De Henau, J. and Himmelweit, S. (2021). A Care-Led Recovery From Covid-19: Investing in High-Quality Care to Stimulate and Rebalance the Economy. Feminist Economics. Retrieved on March 20, 2021 from https://doi.org/10.1080/13545701.2020.1845390

115. Collett, C. and Dillon, S. (2019). AI and Gender: Four Proposals for Future Research. Cambridge: The Leverhulme Centre for the Future of Intelligence. Retrieved on February 20, 2021 from http://lcfi.ac.uk/media/uploads/files/AI_and_Gender___4_Proposals_for_Future_Research_210619_p8qAu8L.pdf

116. Noble, S.U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

117. Noble, S.U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

An Equity Lens on Artificial Intelligence

Carmina Ravanera and Sarah Kaplan

Institute for Gender and the Economy,
Rotman School of Management,
University of Toronto
August 15, 2021

Rotman School of Management
UNIVERSITY OF TORONTO

INSTITUTE FOR gender AND THE economy

www.gendereconomy.org/artificial-intelligence